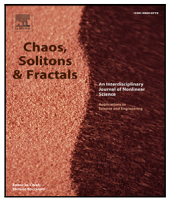
Chaos, Solitons and Fractals 185 (2024) 115175

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

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



A visit generation process for human mobility random graphs with location-specific latent-variables: From land use to travel demand

Fabio Vanni *

Department of Economics, University of Insubria, Via Monte Generoso 71, Varese, 21100, Italy
Center for nonlinear Science, Department of Physics, University of North Texas, Union Circle, Denton, 1155, TX, USA
GREDEG-CNRS, Université Côte d'Azur, 250 rue Albert Einstein, Sophia Antipolis, Nice, 06560, France

ARTICLE INFO

Keywords:
Complex mobility networks
Inhomogeneous random graph
Stochastic visitation process
Origin–destination transport flows
Income elasticity of travel demand

ABSTRACT

This research introduces a novel mathematical framework for understanding collective human mobility patterns, integrating mathematical modeling and data analysis. It focuses on latent-variable networks to investigate the dynamics of human mobility using stochastic models. By analyzing origin–destination data, the study uncovers scaling relations and explores the economic implications of mobility patterns, particularly regarding the income elasticity of travel demand. The mathematical analysis begins with the development of a stochastic model based on inhomogeneous random graphs, constructing a visitation model with multipurpose drivers for travel demand. Through this model, the study gains insights into the structural properties and dynamic correlations of human mobility networks, deriving analytical solutions for key network metrics: visit distribution, assortativity behavior and clustering coefficient. Empirically, the study validates the model's assumptions and reveals scaling behaviors in origin–destination flows within a region, reproducing statistical regularities observed in real-world cases. Notably, the model's application to estimating income elasticity of travel demand provides significant implications for urban and transport economics. Overall, this research contributes to a deeper understanding of the interplay between human mobility and regional demographics and economics. It sheds light on critical scaling relations across various aspects of collective human mobility and underscores the importance of incorporating latent-variable structures into mobility modeling for accurate economic analysis and decision-making in urban and transportation planning.

1. Introduction

How people move from one place to another represents a crucial key to depicting human relations and social interactions in complex societies [1–4]. Human mobility encompasses a wide range of spatial and temporal scales, from daily commuting within a city to long-term migration across countries. This mobility is influenced by factors such as economic development, technological advancements, political stability, and geographical features. Researchers have explored human mobility and transportation dynamics using various mathematical approaches, ranging from individual path-based models to collective population-based analyses [5–7]. In particular, data-driven studies on human mobility are valuable for identifying the factors driving the movement of individuals and goods, and how they affect economic outcomes such as labor market dynamics, economic growth, transportation planning, and consumer behavior [8,9]. This paper addresses the modeling and explanation of the statistical properties of collective human mobility using a data-driven approach. It aims to provide a stochastic model for complex networks of human mobility, rooted in an origin–destination structure that represents the foundational flows of mobility. In particular, the point locations are characterized by intrinsic features, considered latent variables, that define their ability to generate new trips and attract new visits. These features are drivers of travel demand dynamics and will be estimated from the land-use structure of a region and the resident population.

The research purpose of this manuscript is to investigate whether there exists a time-invariant and regionally consistent relationship law between key components of the mobility network. This approach enables us to tackle the issue of how various statistical configurations of latent variables can yield different structures within the mobility network, leading to the emergence of scaling properties in the visit distribution. Specifically, we seek to explore the potential relation between the number of visits to locations, the distance traveled (or other "trip costs") to reach those locations, and the inherent attractiveness of destinations (inferred by land-use data). Through a network approach,

* Correspondence to: Department of Economics, University of Insubria, Via Monte Generoso 71, Varese, 21100, Italy.
E-mail address: fabio.vanni@uninsubria.it.

https://doi.org/10.1016/j.chaos.2024.115175
Received 27 July 2023; Received in revised form 3 June 2024; Accepted 14 June 2024
Available online 25 June 2024

it becomes feasible to uncover higher-order relationships between origin and destination locations. This is achieved by examining assortative behavior of the mobility graph and clustering coefficient, which serve as measures of transitivity. Finally, from an economic standpoint, a crucial aspect of this research is to comprehend the significance of a latent variable (land-use) in elucidating income elasticity within multi-purpose travel demand across transportation modes.

More in details, as research methodology, the paper provides a mathematical description based on stochastic processes for mobility network formation. The objective, from a mathematical standpoint, is to integrate a stochastic model of origin–destination visitation processes for collective human mobility. Specifically, complex network theory will be utilized for both mathematical description and analysis of mobility patterns. The model is rooted in the class of inhomogeneous random graphs [10,11], also known as latent-variable networks [12–16]. In this framework, the nodes represent locations, and the links are weighted since they denote trips between origins and destinations made by travelers who makes a certain effort for undertaking those trips. So the trip cost (as distance traveled) is considered as the link weight between to nodes. Consequently, the network takes the form of a directed and weighted multigraph, serving as a representation of collective mobility. Furthermore, the graph is labeled, as each node is characterized as both an origin and a destination, with unique features, taken as latent variables, that define the travel demand towards different locations. The process describing the number of visits to destinations has been formalized using a master equation for the probability density function, following the literature on continuous-time Markov processes [17,18]. Specifically, the evolution of the visit distribution and graph correlations (as assortativity and clustering coefficient) over time is described through an integro-differential equation with an explicit asymptotic solution. This solution highlights the relationship between visiting generation patterns and the intrinsic and environmental features of locations, as specified by latent-variable properties.

On the data analysis side, the study utilizes the SafeGraph dataset for the New York metropolitan area [19]. The analysis of the Origin–Destination (O-D) network is centered on the metropolitan area of New York City, comprising 1076 Census Block Groups (CBGs). However, arrivals to destinations from all CBGs across the US are considered, while outgoing distributions are inaccessible due to data limitations beyond the New York City region. The analysis is conducted on an hourly basis, encompassing each day throughout November 2029 and partially in June 2020. Examination of this data unveils clear scaling relations in commuting patterns, consistent with findings from other empirical studies [20–23]. I utilize monthly data to establish scale-free distributions of arrival and visit frequency. In this context, visits represent individuals arriving at destinations with specific trip weights, reflecting travel effort in terms of distance traveled or travel time. The presence of these scale-free distributions is attributed to the statistical properties of the attractiveness latent variable, identified as land-use, and the probabilistic mechanism of destination selection. Additionally, at this particular time and spatial scale, the analysis reveals the absence of network correlations in the Origin–Destination (OD) matrices. This insight aids in identifying potential probabilistic laws driving OD trips and their costs relative to destination attractiveness.

The paper's novel contribution is twofold: analytical and empirical. The analytical aspect addresses the problem of how different configurations of latent-variable patterns determine the structure of the mobility network, encompassing degree and strength distributions as well as graph correlations such as assortativity and transitivity properties. This sheds light on how land uses or attractiveness proxies can shape individuals' preferences regarding where, when, and how frequently they move and visit places of interest. Additionally, the current model serves as a generalized framework capable of encompassing popular models describing collective human mobility dynamics, including gravity [24,25], radiation [26–28], and opportunity models [29,30]. These models can all be, in fact, explained with the mobility network framework, provided proper selection and specification of parameters and relationships between latent variables are made.

Empirically, the study aims to validate that land-use latent variables effectively represent destination attractiveness and demonstrate the impact of trip weights on the visitation process. Eventually, based on the model's assumptions validated by empirical analysis, we can observe inverse power law shapes in both the degree arrival distribution and the strength visit distribution and we can observe a topological relationship between node degree and strength, which will be expressed through a scaling relation. An important finding will be the connection between the topological factor and the visitation process mechanism built on latent variable structure, shedding light on the underlying dynamics. Essentially, meanwhile the theoretical framework establishes the existence of critical scaling relations across multiple aspects of collective human mobility, the empirical analysis confirms that specific scaling coefficients are unique to the analyzed time period and region, yet the formulas governing mobility scaling variables remain universally applicable.

The emergence of power law visit distributions is shaped by the statistical characteristics of the latent variable across regions and its connection to the mechanism of generating trips between locations. This connection dictates whether the mobility graph displays a neutral, scale-free distribution of visits, as in the case our study, or introduces correlations and alternative visit patterns. Through data analysis, we engage in a model selection process to discern among various latent variable characteristics.

Finally, the study underscores a reciprocal interplay between human mobility and the demographic and economic structure of a region, a topic extensively discussed in the literature [31–36]. Using a bottom-up view of the research, attractiveness is the basic impulse that draws people to a location. Although attractiveness is a latent variable that cannot be directly measured, it can be inferred through proxies or indexes. In this article, attractiveness is identified with land use. "Land use" describes how humans utilize land and the landscape, encompassing the economic and cultural activities at a given location. Travel demand is directly determined by this land-use variable, which includes factors such as zoning divisions (tax lots), and is influenced by socio-economic status, including the average income of travelers, which reflects individuals' willingness to pay for traveling from one place to another.

As an economic application, the study will illustrate the relationship between the scaling exponents of attractiveness and the income elasticity value, leveraging information on the allocation and utilization of land resources for various economic and social activities. The estimated income elasticity of demand is often utilized to forecast future changes in consumption in response to income fluctuations [37–40].

An Appendix A section and supplementary material are provided to enhance the discussion on the stochastic interpretation of the visit process, further statistical analysis of the data, and a detailed description of data and interpretation of the latent variables in travel demand modeling.

2. Model

A trip mobility network is built upon an origin–destination rationale and it will be represented as a directed and weighted labeled multigraph where the nodes are administrative units of an urban or regional area (a city, a county, a state etc...). The graph is directed (trip direction) and it allows self-edges and multiple edges (many travelers from and to the same origin and destination). Furthermore, the network is a labeled graph since the nodes will be specified by intrinsic location attributes. The modeling argument is based upon some fundamental assumptions common in literature of collective human mobility [5,22,42]. Firstly, the fluxes directed towards a specific destination represent a observable manifestation of individuals' inclination to visit that

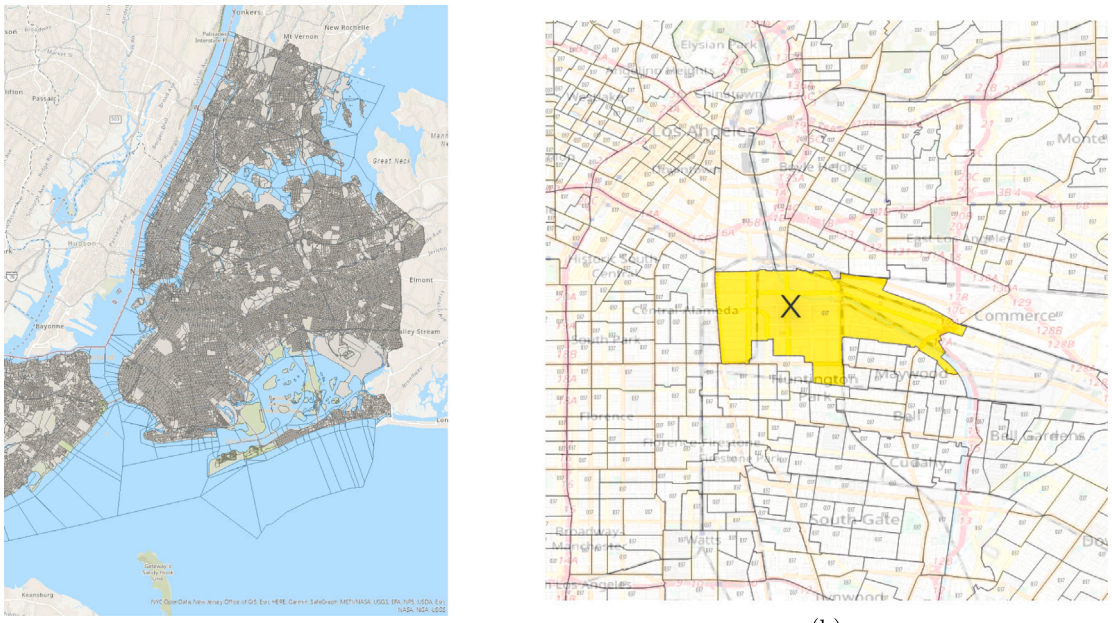


Fig. 1. Example of a tessellation of a large urban area (i.e. New York City) in (Census) blocks (a) as in [41]. In particular, a block is characterized by an intrinsic attractiveness x as a latent variable depending on many features of location (b).

location. Secondly, the motivation to visit is indicative of the effort required, and it can be articulated in terms of the energy individuals are willing to expend to reach said destination. Lastly, the aggregated willingness to visit, across numerous individuals, depends only on the attractive characteristics of the destination.

2.1. Mobility graph

A geographical region of interest, R, is specified as the portion of territory for which we are interested in generating the flows. Over the region of interest, a set of geographical tiles called tessellation, $\mathcal{T}$, made up of locations l_i so that $T = \{l_i : i = 1 \dots n\}$ so that the locations are non-overlapping, $l_i \cap l_j = \emptyset, \forall i \neq j$, and the union of all locations completely covers the region of interest, $\cup_{i=1}^{n} l_i = R$. The tessellation for real geographical regions can be obtained in many ways according to the scope. In the case of interest, location tiles are the census areas defined by national authorities for administrative and demographic purposes. In particular e census blocks are the smallest geographic unit used by the United States Census Bureau for tabulation of data collected from all houses. An example of tessellation is given, as an example, in Fig. 1 for New York city, where tiles are the census block division of the urban area. Similarly, the locations, other than being destinations, are, at the same time, labeled as origins. The correspondent intrinsic feature variable will represent the resident active population which stays in the area from which a trip originates.[1]

At this point, it is possible to frame the network structure according to the following definition:

Definition 1 (*Trip Mobility Graph*)**.** We denote with $\mathcal{G} = (T, E)$ a directed and weighted multigraph, where T is the set of n locations (nodes) and E is the set of trips (edges).

(T1) Each location $\ell_i \in T$ is assigned a destination node-type label denoted by the latent variable x_i, which takes values in the space $\Omega_x \subseteq \mathbb{R}$. The latent variable represents the attractiveness of the destination where a trip ends, serving as the driving factor for travel demand to each destination.

(T2) Conversely, each location $\ell_i \in T$ is also labeled with another origin node-type variable denoted as y_i, which takes values in the space $\Omega_y \subseteq \mathbb{R}$. The latent variable y_i represents the productivity of origin locations where trips start, serving as the driving force for travel supply and demand.

(E1) Each edge $e_{ij} \in E$ represents a journey originating from a location labeled y_i and heading towards a destination characterized by its attractiveness x_i.

(E2) Additionally, each edge is assigned a weight representing a specific aspect of the trip. More precisely, the arrival at a destination reached from a departure point is comprised of a trip undertaken at a cost τ_{ij}.

The latent-variable probability measure is defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra generated from the real line and the probability will be assumed absolutely-continuous respect to the Lebesgue measure, so that the probability density function can be defined as $\rho(x) = F'(x)$.

Let us consider that the latent variables can be viewed as inherent characteristics of the region encapsulating various factors. Specifically, each node is defined by its intrinsic attractiveness denoted by x, reflecting the travel demand for a destination area. This attribute is influenced by several location-specific properties such as job opportunities, the presence of retail stores, geographic features, infrastructure, amenities, and school districts. Similarly, each node is also characterized by another property, denoted as y, which represents the location's role as an origin of trips. This attribute identifies the potential users who may depart from the area, thereby indicating the area's capacity to generate trips. For a more detailed discussion, please refer to the Supplementary Materials. In this context, these latent variables, namely attractiveness and productiveness, can be viewed as drivers of travel demand from an economic standpoint.

From a dynamical perspective, the model is presented as a stochastic process that describes a random graph evolving in time [43,44]. The model presented describes the occurrence dynamics of trips that occur between origins with a given population y and destination blocks

[1] Despite of not being a proper hidden variable, the people which can actively considered travelers undergoes to many factors (such as residence, age, employment status, and many others).

with a given attractiveness label x. In this paper, I will, primarily, focus on the case of exchangeable origin–destination blocks, so that the trip generating process can be represented by two processes: the trip production and the trip arrival process. The first accounts for the number of trips (departures) originating from a block and the latter accounts for the number of trips (visits) ending in each destination block.

A visit generation process consists on a trip mobility graph evolving over time $\mathcal{G}_t$ which is a temporal inhomogeneous random graph model under the assumption of an infinitely large graph so that one can consider a continuous approximation on the variables [10,11,45]. Consequently, the evolution of the mobility graph can be considered as a process with the following properties:

Definition 2 (*Trip Arrival Process*)**.** The graph processes for directed multi-edges can be defined as random graph $\mathcal{G}_t$ evolving so that $\forall t$ a new edge is added $\mathcal{G}_{t+1} = \mathcal{G}_t \cup \ell$ where $\ell \in 2\binom{T}{2} \setminus E(\mathcal{G}_t)$ is chosen randomly with replacement with probability proportional to a kernel function $\mathcal{K}(x, y)$ defined in the latent-variables space $\Omega_x \times \Omega_y \to [0, \infty)$.

- Let $(x_i)_{i\in[n]}$ be a sequence of attractiveness latent variables such that their empirical distribution approximates a probability measure μ_x as $n \to \infty$. Similarly, the productiveness variables $(y_i)_{i\in[n]}$ are random variables with an associated probability measure μ_y.
- The kernel function $\mathcal{K}(x, y)$ rules out the chances that a location of attractiveness x will be a destination of a visit whose trip originated in a location of feature y. Consequently, the stochastic process of trip arrivals is defined through the infinitesimal arrival intensity $d\mathcal{V}_x = \mathcal{K}(x, y)\mu_y(dy)$ where $\mu_y(dy)$ represents the measure assigned to the infinitesimal interval of productiveness for origin locations.

Essentially, the latent variables x and y are considered to be independently and identically distributed, with their empirical distribution converging almost surely to the cumulative distribution function $F(x)$ and $F(y)$ respectively. The graph evolves when a new visit from an origin to a destination is completed (a new link in the network) according to the attraction and the production rates in the degree-space of the mobility graph. The trip arrival process is driven by the rate $d\mathcal{V}_x$ which is interpreted as the propensity (or intensity) of new travels landing in destinations of attractiveness x conditional to trips that departed from locations of infinitesimal productiveness y. Similarly, as dual problem, a trip departure process can be defined by the infinitesimal departure intensity $d\mathcal{V}_y = \mathcal{K}(x, y)\mu_y(dy)$.

From a computational point of view, the processes studied here will have a fixed vertex set, and they will start without any edges and grow by adding edges according to linking rule, without deleting any, since the total number of visits up to a certain time is studied. The evolving mobility graph $\{\mathcal{G}^{(t)}\}$ is fully described by means of a time-varying adjacency matrix $A^{(t)}$ which represents the origin–destination table of the mobility problem at time t and the sum along the columns represent the in-degree of the nodes or equivalently the number k of visits received up to time t.

2.2. Visitation model

At this juncture, the visitation model is introduced by embedding trip weights into the arrival process, following the framework outlined in Definition 1, in line with the interpretation of the mobility graph outlined in Definition 2. Here, each edge of the mobility graph is assigned weight τ representing a trip cost random variable with probability density function $\varrho_x(\tau)$.

Those weights have the meaning of trip 'size' as for example the distance traveled from an origin to the selected destination, or the emission impact of each trip, or any type of travel cost or visit benefit. Consequently, the mobility graph process $\tilde{\mathcal{G}}_t$ can be studied in terms of a weighted and labeled directed multigraph adjacency matrix $\tilde{A}^{(t)} = C^{(t)} \circ A^{(t)}$, where $\circ$ denotes element-wise matrix multiplication, and C is regarded as either a coefficient matrix or a random one, depending on the real-world phenomena under study and the available data. Let us notice that the in-strength (i.e. weighted in-degree) of a destination node κ is defined as the sum of weights of all the visits received. In the case that weights are all equal to 1 the strength κ is equivalent to the degree k. A general framework which describes ensembles of dynamic networks can be assessed by making a Markov assumption on the evolution of the network by studying the probability of realization of a member of configuration ensemble graph over time [46,47]. Specifically, the temporal evolution of finding the trip mobility network in the configuration $\mathcal{A}$ at time t after L steps (trips) can be written as:

$$\mathbb{P}(\mathcal{A}, t) = \prod_{l=0}^{L} \mathbb{P}(\tilde{A}^{(t_l)} | \tilde{A}^{(t_{l-1})}, \boldsymbol{\Lambda})$$

where $\boldsymbol{\Lambda}$ is any transition rule built on latent variables, which describes the creation, at each time t_l, of a new trip from an origin towards a destination. From a modeling perspective, the complexity of knowing the configurational distribution of the origin–destination mobility network can be reduced trough the study of the degree (or strength) distribution and its higher order statistics.

The first statistic is the in-strength distribution, that results to be the visit distribution defined as:

$$P(\kappa, t) = \sum_{\{\mathcal{A}\}} \sum_{i} \delta(\kappa - \kappa_i)\mathbb{P}(\mathcal{A}, t)$$

where δ is the delta function, and $\mathbb{P}(\mathcal{A}, t)$ is the probability to find our trip mobility network in the configuration $\mathcal{A}$ at time t, where each node has a strength degree κ_i. If the in-degree of a node is simply the number of arrivals of new travelers, the in-strength of a location node is defined as the number of arrivals of new travelers who have faced a cost. Such weighted version of arrivals is named visits where each trip has an intensity. In the case of this work, the derivation of the visit distribution is assessed on the basis of the latent variable framework. The model is developed upon the asymptotic regime assumption of an infinite network where the number of locations in a tessellation is extremely large, where each location can generate and attract an unlimited number of edges (trips). Consequently, the continuous mean-field approach is used so that single locations can be studied as uncorrelated nodes in the same class of locations with the same attractiveness [48–50]. So rather than studying the single location ℓ_i, one analyzes those locations which belong to the same class of attractiveness level, i.e. ℓ_x seen as a absolutely continuous random variable so that it admits a probability density function $\rho(x) = F'(x)$. Let us call conditional visit distribution $p(\kappa, t|x)$ the strength distribution conditional to destinations of the attractiveness type x, and each class evolves independently one from any another, so that a superposition of co-evolving conditional degree distributions is possible. The overall visit distribution can be derived according to the following proposition:

Proposition 1 (*Visit Distribution*)**.** *The visit distribution of the trip-mobility network is fully characterized by the attraction rate ν_x that defines the transition probability, per unit of time, that a destination of attractiveness x increases its number of visits by one from any destination. The attraction rate is defined as the mean intensity of the trip arrival process $\nu_x = \int_{\Omega_y} d\mathcal{V}_x$.*

- *The evolution of the conditional visit distribution can be described by a master equation for destinations with attractiveness x as:*

$$\frac{\partial}{\partial t} p_x(\kappa, t) = \int_0^{\kappa} \nu_x \varrho_x(\tau) [p_x(\kappa - \tau, t) - p_x(\kappa, t)] d\tau \tag{1}$$

with the initial condition $p_x(\kappa,0)=\delta(\kappa)$. In the asymptotic regime, the conditional probability $p(\kappa,t|x)$ can be written as:

$$p(\kappa,t|x) \sim \epsilon_t \frac{1}{\sqrt{2\pi\nu_x\langle\tau^2\rangle_x t}} e^{-\frac{(\kappa-\nu_x\langle\tau\rangle_x t)^2}{2\nu_x\langle\tau^2\rangle_x t}} \quad , \text{ where } \epsilon_t = \frac{2}{1+erf\left[\frac{\nu_x\langle\tau\rangle_x t}{\sqrt{2\nu_x\langle\tau^2\rangle_x t}}\right]} \tag{2}$$

where the correction factor $\epsilon_x \to 1$ in the asymptotic limit $t\to\infty$, and $\langle\tau\rangle_x$ and $\langle\tau^2\rangle_x$ are the first and the second moment of the trip-weight distribution $\varrho_x(\tau)$.

- *Finally, the temporal asymptotic expression of the visit distribution is a mixture distribution as:*

$$P(\kappa,t) = \int_{\Omega_x} p(\kappa,t|x)\rho(x)dx \sim \sum_{i=1}^{m} \left|\frac{\partial z(x)}{\partial x}\right|^{-1}_{x_{0,i}} \rho\left(x_{0,i}\right) \tag{3}$$

where $x_{0,i}=x_{0,i}(\kappa,t)$ are the zeros of the expression $z(x)=\kappa-\nu_x\langle\tau\rangle_x t$, and $P(\kappa,t)$ the in-strength distribution of the overall trip mobility network.

The functional form of $\langle\tau\rangle_x$ in Eq. (3) is crucially important to determine the effect of trip-cost on the topology of visitation network model and its related visit distribution. Let us observe that if, for example, the trip weight distribution $\varrho_x(\tau)$ is a dirac delta, then the strength distribution is equivalent to the degree distribution, this is a *degenerated case* where trip-costs are not considered. Another particular case is when the trip weight distribution is identical over the attractiveness variable so that $\varrho_x(\tau)=\varrho(\tau)$ then the strength is proportional to the degree $\kappa=\langle\tau\rangle k$, which I call the *constant distance condition*. A more general assumption in a macroscopic perspective is to assume an *invariant trip condition*, when considering origin–destination locations aggregated by their latent variables. It means that the center of resident population of type y is the same of any other resident population of different type.[2]

In particular, there is a model selection issue, since different choices of attractiveness features and trip-cost functional shapes can generate the same effect on the strength-distribution, one could clarify the ambiguity investigating higher order characterization of the degree distribution of the mobility network.

Let us observe that the type of processes described in Proposition 1 can be reinterpreted in terms of a mixture of compound Poisson processes as described in the Appendix which represents to be very popular in financial mathematics to model stock prices, insurance claims, and other financial phenomena [52,53]. The visitation model of the mobility network can also be interpreted in combinatorial terms by using urn processes for solving balls in bins problems and finding the occupancy distributions as sketched in Supplementary Information. Such approach is used in world trade literature as in [54–56]. Despite different approaches, the one introduced in the paper has the advantage to provide direct, though asymptotic, solutions to the scaling relations in the mobility networks.

A more detailed characterization concerns the exploration of the connectivity correlations in the origin–destination correspondences of trip mobility network. Higher order statistics of a network in the degree space, can be obtained trough by the conditional probability $P(\kappa^{(1)},\kappa^{(2)},\dots,\kappa^{(k)}|\kappa',t)$ that a node with strength κ' connects to nodes with strength $\kappa^{(1)},\kappa^{(2)},\dots,\kappa^{(k)}$ at time t. The simplest of these degree correlations is the two-point correlation being described by the conditional probability $P(\kappa|\kappa',t)$ as the probability that a trip departing from an origin location of out-strength (departure) κ' reaches a destination node of in-strength (visit) κ. The correlations between degrees of the nearest-neighbor vertices are described by the probability distribution:

$$P(\kappa,\kappa',t) = \sum_{\{\mathcal{A}\}}\sum_{ij} \delta(\kappa-\kappa_i)\mathbb{P}(\mathcal{A},t)\delta(\kappa_j-\kappa')$$

However, the empirical evaluation of such conditional probability in real networks is cumbersome, so the weighted degree-degree correlations are commonly accounted by average-nearest-neighbor's strength function $k_{nn}(\kappa,t)$ which makes use of a smoothed conditional probability [57] often used as a measure of degree homophily of the nodes. In the latent variable framework, as shown in Fig. 2(a), the conditional assortativity $k_{nn}(x)$ measures how much a location with attractiveness x tend to be a destination of an origin location of population y as defined in [48,58,59]. In a similar way, the three-point correlations can be studied in terms of the clustering coefficient spectrum $c(\kappa,t)$ which indicates the probability that two neighbors of strength-κ node are neighbors themselves. In the case of weighted and directed networks there many different ways to define the cluster coefficient [60,61].

At the latent variable level, the conditional clustering coefficient of a destination with attractiveness x can be interpreted as the probability that two randomly chosen locations with trips towards a destination with attractiveness x are neighbors. Consequently, the Markovian property at the latent variable level [58,62,63] allows to calculate analytical expressions for the assortativity $k_{nn}(\kappa)$, quantifying two vertices correlations, and clustering coefficient spectrum $c(\kappa)$, as a measure of three vertices correlation. A very important result is that the degree correlations of trip-visit distributions are completely determined by the attraction (and production) rate and by the origin–destination conditional probability $\chi(y|x)$.

Proposition 2 (*Visit Correlations*)**.** *In the visiting mobility network under the latent variable assumption, the origin–destination correlation is defined as the conditional probability that a visit in the destination of attractiveness x has originated from a location of population y, and it is written as:*

$$\chi(y|x) = \frac{\partial}{\partial y}\log\mathcal{V}_x \tag{4}$$

As consequence, the following estimates of the two-point and three-point correlations hold:

- *the average out-strength of origin neighbors of destinations with in-strength κ, can be written as:*

$$k_{nn}(\kappa,t) \sim \frac{t}{P(\kappa,t)} \iint \nu_y p(\kappa,t|x)\chi(y|x)\rho(x)dydx \tag{5}$$

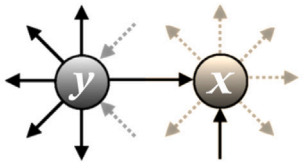
If destinations and origins are independent $\chi(y|x)=\chi(y)$ and $\langle k_{nn}\rangle(\kappa,t)=const.$

- *the clustering coefficient for destinations of in-strength κ is:*

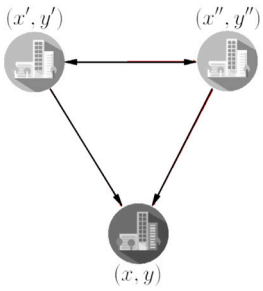
$$c(\kappa,t) \sim \frac{1}{2\nu_0 P(\kappa,t)} \iiint p(\kappa,t|x)\rho(x)\left(\nu_{y'}+\nu_{y''}\right)\chi(y'|x)\chi(y''|x)dy'dy''dx \tag{6}$$

The Markovian nature of this class of networks implies that all higher-order correlations can be expressed as a function of the attraction and production rates ν_x,ν_y and the conditional origin–destination probability $\chi(y|x)$, allowing an exact treatment of mobility models at the mean-field level. Under the hypothesis that origins and destinations are independent, that is $\chi(y|x)=\chi(y)$, so that visit production process and visit attraction process are independent. As a consequence, the average-nearest-neighbor's strength function and the clustering coefficient are constant along κ as discussed later in the text.

[2] The center of population is a concept related definition at [51] defined as the point at which an imaginary surface representation of a region would balance if weights of identical size were placed on it so that each weight represented the location of one person. The assumption of invariant trips implies that all origin locations are exchangeable with respect to the node type y, and movements are isotropic, meaning that all directions towards a destination are equal. So under a mean-field perspective, for a given destination of attractiveness x, we can consider a representative mean distance from all the origins of type y condensed into their population center.



(a) destination assortativity for the node x



(b) destination clustering for the node (x, y)

Fig. 2. The two point and three point correlations of the trip mobility network can be calculated in terms of the hidden variables x and y. In particular the conditional average-nearest-neighbor's strength at the latent variable level (a) is the in-out strength (origin–destination) assortativity coefficient, and the cluster coefficient at the latent variable level (b) is the "In" clustering (or destination clustering) coefficient for weighted and directed networks as [60,61], i.e. a triangle such that there are two trips coming into of the destination node $(x \leftarrow y', x \leftarrow y'', x' \leftarrow y'' \vee x'' \leftarrow y')$.

2.3. Numerical validation & Monte Carlo simulations

For computational purposes, we now consider only the constant distance case where the trip-costs have no effect on the visit distribution since $\kappa = \langle \iota \rangle k$ so that it is equivalent to study the degree distribution of the arrival process. As already assessed, for every couple of origin–destination locations, trips can be realized according to a kernel function $\mathcal{K}(x, y)$ [16,64–66]. As a crucial case in the context of scale-free networks, one can consider the attraction rate to be proportional to some power of the destinations' attractiveness:

Remark 1. Le us assume that the attraction rate is of the form $\nu_x = \nu_0 x^{\alpha_0}$, with $\alpha_0 > 0$, and homogeneous trip-weight distribution $\varrho_x(\cdot) = \varrho(\cdot)$, the asymptotic trip-visit distribution can be written as:

$$P(\kappa, t) \sim t^{-\frac{1}{\alpha_0}} \kappa^{\frac{1}{\alpha_0}-1} \rho(x_0) \tag{7}$$

where $x_0 = x_0(\kappa, t) = \left(\frac{\kappa}{\nu_0 t}\right)^{\frac{1}{\alpha_0}}$, and where ρ is the attractiveness probability density function. For $\alpha_0 = 0$ the Erdos–Renyi random graph is recovered.

In the particular case that the attractiveness distribution is $\rho(x) \sim \rho_0 x^{-\eta}$, the visiting in-degree distribution has the following asymptotic tail distribution

$$P(\kappa, t) \sim t^{\frac{\eta-1}{\alpha_0}} \kappa^{-(1+\frac{\eta-1}{\alpha_0})} \tag{8}$$

which shows the typical scale-free structure of an inverse power law distribution for the visiting degree of the mobility network.

The analytical results in the previous remark is confirmed by numerical integration of the compound distribution Eq. (3) by using the truncated normal conditional probability Eq. (2). Moreover, a graph process is performed via montecarlo (MC) simulation of the network where occupation probability is expressed via a separable linking function $\mathcal{K}(x, y) = g(x)h(y)$ for the evolution of the adjacency matrix. Such kernel gives arise to an attraction rate as $\nu_x = \nu_0 g(x)$, where ν_0 is a normalization constant as shown in details in the supplementary materials. So by choosing $g(x) = x^{\alpha_0}$ we are in the case as specified in Remark 1. At this point, it is possible to compare the three approaches and confirm the consistency of results obtained. In Fig. 3 it can be observed how the three approaches provide the same visit distribution for a particular choice of the parameters as discussed in the caption. Moreover, in this circumstances, the mobility graph has neutral correlations in terms of assortativity and transitivity as confirmed by simulations Figs. 3(c) 3(d).

As expressed in different research works [15,66–69], different combinations of the attraction rate and attractiveness distribution can generate the same trip-visit distribution. For example, a scale-free visitation distribution can be obtained with the following relations:

Remark 2. Le us assume that the attraction rate is of the exponential form $\nu_x = \nu_0 e^{\gamma_0 x}$, and homogeneous trip-weight distribution $\varrho_x(\cdot) = \varrho(\cdot)$, the asymptotic trip-visit distribution can be written as:

$$P(\kappa, t) \sim \frac{\kappa^{-1}}{\gamma_0 \nu_0} \rho(x_0) \tag{9}$$

where $x_0 = x_0(\kappa, t) = \frac{1}{\gamma_0} \log \frac{\kappa}{\nu_0 t}$, and where ρ is the attractiveness probability density function.

In the case that the attractiveness distribution is exponential $\rho(x) \sim \rho_0 e^{-\lambda x}$, the visiting in-degree distribution has the following asymptotic tail distribution

$$P(\kappa, t) \sim t^{\frac{\lambda}{\gamma}} \kappa^{-(1+\frac{\lambda}{\gamma_0})} \tag{10}$$

which shows another way to obtain a scale-free structure with an inverse power law distribution for the visiting degree of the mobility network.

It is very common in literature to assume, at a collective level for wide regions, that the visit production and the visit attraction process are assumed to be independent. As consequence of Proposition 2, for neutral networks the estimate of assortativity and clustering coefficient is:

Remark 3. Under the hypothesis that visit production process and visit attraction process are independent the average in-strength of nearest neighbor function is constant, and in the asymptotic limit:

$$k_{nn}(\kappa, t) \sim \frac{t\mathbb{E}[h^2(y)]}{N\mathbb{E}[h(y)]^2} = \frac{\langle \kappa_{out}^2 \rangle}{\langle \kappa_{out} \rangle} \tag{11}$$

As regard with the clustering coefficient spectrum, under the same hypothesis:

$$c(\kappa) \sim \frac{\mathbb{E}[g(x)]\mathbb{E}[h^2(y)]}{\mathbb{E}[h(y)]} = \frac{\langle \kappa_{in} \rangle}{tN} \left(\frac{\langle \kappa_{out}^2 \rangle - \langle \kappa_{out} \rangle}{\langle \kappa_{out} \rangle^2} \right)^2 \tag{12}$$

Proof. The proof is given in Appendix A.3 □

Let us notice that the clustering coefficient is normalized respect to time t since $c(\kappa) \in [0, 1]$ and so it results to be the generalization for directed multigraphs without correlations as for simple graphs in [48,66,70]. As a result, we can notice that in the case of neutral networks, the two and three point correlations can be obtained by using three different approaches providing the same estimate: latent variables ('latent estimate'), the adjacency matrix ('expected value') and, finally, the algorithm computation of assortativity and clustering coefficients for directed and weighted networks ('simulation approach').

Simulations for such results are shown in Fig. 4 where several computational simulations of a graph for different time length t is

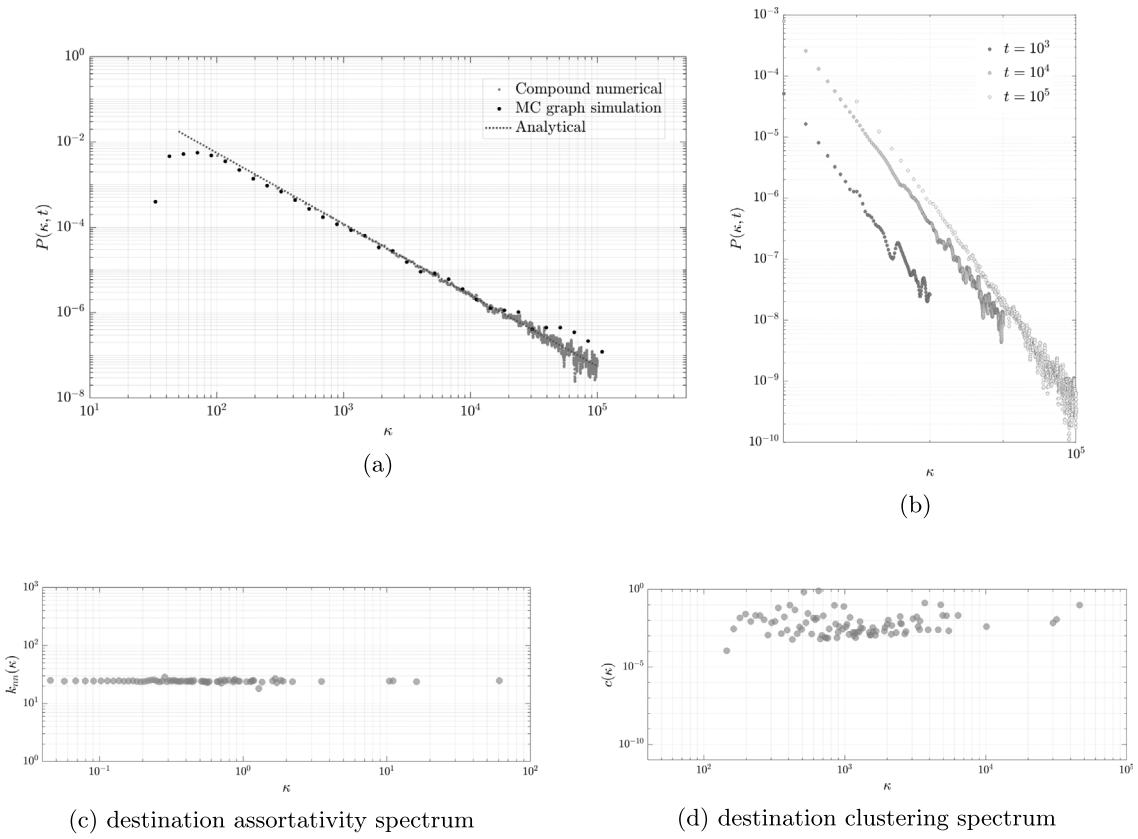


Fig. 3. Visit distribution computed in the case of constant trip weights, and the attractiveness distribution is $\rho(x) \sim \rho_0 x^{-2}$ and kernel function $\mathcal{K}(x, y) \propto x^{1.5} h(y)$ so that $\nu_x = \nu_0 x^{1.5}$. In (a), the probability density $P(\kappa)$ has been estimate with three different approaches: (1) it is evaluated through the numerical integration of the compound probability as in Eq. (3). (2) It is evaluated through montecarlo (MC) graph simulation of sequential adjacency matrices with $N = 300$ locations and with a simulation time of $t = 10^5$ time steps. The three approaches provide the same scale-free behavior of the visit distribution as $P(\kappa, t) \sim t^{\frac{2}{5}} \kappa^{-\frac{5}{5}}$ as expected by the analytical asymptotic estimate Eq. (8). In (b) the compound distribution approach is calculated at three different time snapshots. With the additional specification of $\phi(y) \sim \phi_0 y^{-2}$ in (c) the average-nearest-neighbor's strength function shows a neutral assortativity in the network, meanwhile the dashed line represents the assortativity mean value. In (d) the local clustering coefficient for different values of κ shows a constant transitivity, and the dashed line represent the average global cluster coefficient.

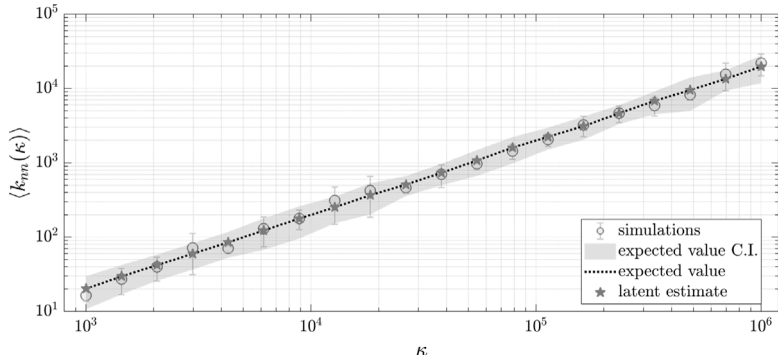
presented alongside the prediction results for uncorrelated networks for assortativity and clustering coefficient. It is worth noticing that the analytical prediction are asymptotically valid so that no isolated nodes or leafs exists since local neighborhood clustering is typically not defined if a node has one or no neighbor and such situation influences the estimation of the global clustering in sparse networks [71]. In the present work, the clustering coefficient algorithm removes all the local clustering of all the nodes with less than 2 neighbors, so the global clustering coefficient is over-estimated.[3] Many other examples of visit distributions and degree correlations are presented in Appendix C and more deeply discussed in the Supplementary Material. The emergence of power laws in our analysis is driven by the statistical properties of the latent variable and its influence on the kernel probability and attraction rate. When land-use distribution follows a power-law pattern and destination selection correlates with a power of land use, the mobility graph exhibits neutrality, resulting in a scale-free distribution of visits. However, different configurations of latent-variable structures can also yield similar scale-free distributions. Conversely, if the kernel probability lacks multiplicative factorization, diverse visit distributions may arise, including power laws with exponential cut-offs or non-power-law distributions.

In the recent decades, multiple models have emerged to elucidate the universal principles of human mobility including the gravity model, radiation model, opportunities model, and their related extensions [5, 25–27,72,73]. Let us notice that the visitation network model incorporates many of the main popular collective human mobility models under the lens of a latent variable perspective as shown in Appendix D. For example, gravity models is equivalent to a neutral mobility network with a visitation model driven by an attraction rate of the type $\nu_x \sim \langle \tau \rangle_x x^a$ with $\langle \tau \rangle_x = 1/\langle d \rangle_x$ which indicates that the trip weight is inversely proportional to the travel distance to reach a destination of attractiveness x. Alternatively, an intervening opportunity model is equivalent to a neutral visitation model with an exponential-like attraction rate. Finally, a radiation model rational can be interpreted in terms of a network visitation model with neutral correlations and an attraction rates similar to the gravity model one but with a sort of cut-off for extremely large values of attractiveness. Consequently, it is evident the importance of knowing the relation between attractiveness and trip weights. We will assess such issue in the next sections where empirical evidences will drive the research on the estimate of scaling relations.

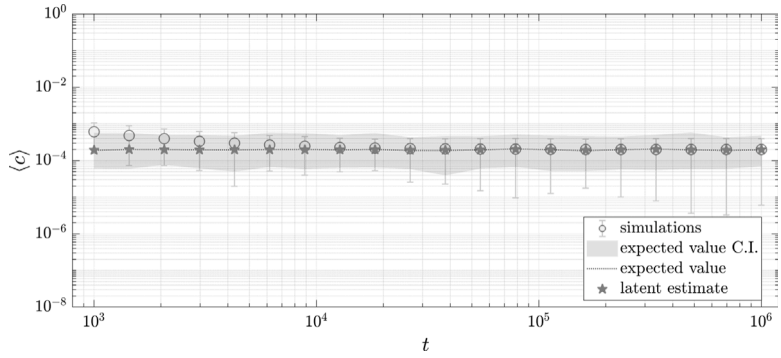
3. Data description

In the present section a network analysis of the main graph measures and topology will be conveyed in the particular case study of New York metropolitan area by using Safegraph Mobility Dataset [19] for the year 2019. Origin–destination (OD) data represent movement flows through geographic space, from an origin (O) to a destination (D). OD datasets represent information on trips between two geographic areas often represented by the geographical centroids of the

[3] Another choice would be to set to zero the local clustering coefficient for all nodes with less than two neighbors. Is such case the global clustering coefficient would be under-estimated.



(a) mean in-assortativity for different simulation times



(b) mean in-clustering coefficient for different simulation times

Fig. 4. Two points and three point network correlations using three different approaches, in the case of uncorrelated graph as reported in Remark 3. The overall average strength of nearest neighbors $\langle k_{nn}(\kappa)\rangle_\kappa$ is replicated for each t so to obtain a mean global value $\langle k_{nn}\rangle$ over a ensemble of $S = 50$ replications as in (a). Similarly one can obtain the global mean in-clustering coefficient $\langle c\rangle$ in (b).

areas. Typically encoded with a square symmetric matrix, OD flow data contain numerical data on the aggregate quantity of individuals traveling from one geographic area to another over a specific time period. Mostly used in transportation planning, OD flows are an invaluable source of data for understanding spatial and temporal patterns of urban mobility and dynamics [2,9,74,75]. Visit flows can be in practice estimated in various ways in real world data. In particular, mobile phone location data are provided by SafeGraph trough dynamic population Origin–Destination flow matrices with hourly temporal resolution and aggregated by census block groups (CBG) in the USA as discussed in [76]. In the daily CBG to CBG visitor flows metric, each row contains an origin CBG and a destination CBG, as well as the number of mobile phone-based visitor flows from the origin CBG to the destination CBG. Every day, the number of unique mobile phone users who live in the origin CBG and visits to the destination CBG are recorded. As regarding with visit production model, the population in each block is the key information to obtain from data in order to define the variable y and its respective distribution. However the population data is susceptible to the way data are collected and sampled by the provider. In fact the demographic sampling depends on many factors as the geographical boundaries which define a block, a tract, or any administrative tessellation. Moreover in the statistical sampling methodology the individual measurements in each block go through a few transformations and aggregations which impacts the final measurement [77]. Safegraph Mobility Data is one of the data sources that allows users to track and analyze visits places and points of interest. This data is collected by smartphone devices, and then anonymized and aggregated by Safegraph [19]. In the present work I have made use of SafeGraph's Neighborhood Patterns dataset that contains footfall data aggregated by census block group (CBG) in the United States of America. This data includes the number of visitors to a given Census Block Group parsed by the timing of their visit (hour of the day) and the visitors' primary daytime/nighttime Census Block Group. Each row of the database represents a single area (a census block group ID) as destination location and the columns add context to the specified area. By analyzing millions of anonymous mobile phone users' visit trajectories to various places provided by SafeGraph, the daily dynamic origin-to-destination (O-D) population flows are computed, aggregated, and inferred at the census block group geographic scales. In order to build the O-D table, the column used in the work is given by the number of devices that stopped in this area by home origin area. I will associate to each row of the O-D matrix all the visits received in one month by a destination that originated from devices located in origin areas specified in the SafeGraph as shown in Table 1. The destination CBG is described in the first column through a geographic identifier FIPS code of 12 digits which concatenates County FIPS + Tract Code + Block Group, see [78]. The second column of the table is a string reporting a list of origin home area IDs and the correspondent numbers of visits received from devices from those home locations. The second group of data in Table 1 used in the research are the number of stops in destination CBG area each day (local time) over the covered time period (month).[4] As fully described in [76,79], a home place of a user refers to his/her most common nighttime location during

[4] In SafeGRaph data there is a difference between number of devices in a location and number of stops in a location. The first represents the number of stops by devices in our panel to this area during the date range. The count includes stops by devices whose home area is the same as this area. The second represents the number of unique devices in our panel that stopped in this area during the date range. This includes devices whose home area is the same as this area.

Table 1
Example of the SafeGraph Neighbor Patterns data with columns used to build an O-D adjacency matrix and the number of hourly visits. Data for a selected month.

CBG (*destination*)	Device home areas in a month (*origins*)
"360610144023"	{"360610144023":32 ; "360610144024":27 ; "360610152006":13 ; ...}
"360610041001"	{"360610041001":30; "360610041006":16; ... }
⋮	{...}
CBG	**Stops by each hour in a month**
"360610144023"	[5;3;1;1;0;1;1;3;2;0;2;1;1;1;3;6;4;3;1;2;12;6;9;...]
"360610041001"	[18;4;2;2;1;1;0;3;8;10;13;7;17;8;1;7;17;15;11;5;...]
⋮	[...]

the last six weeks. For each day, GPS pings of each device are clustered and only those clusters during nighttime hours (6 pm–7 am local time) are kept. The CBG with the most clusters in that day is recorded. Based on this, the most frequent CBG over the last six weeks that reflects the primary night-time location is used as the "home location" for each user. The daily CBG to CBG visitor flows metric, every day, the number of unique mobile phone users who live in the origin CBG and visits to the destination CBG are recorded. More specifically, GPS pings of each user are clustered first. Only those clusters (and not single trajectory points) with a duration of at least one minute are counted as a "visit". By doing so, the daily mobile phone-based visitor flows between CBG and CBG are grouped and summed up. A comprehensive analysis of biases in mobile location SafeGraph data across spatial scales and over time is deeply discussed in [80].[5] The data utilized was obtained as a free sample for academic institutions, acquired as a single file directly downloaded from the SafeGraph website. This dataset encompasses information for the year 2019 in the month of November and for the year 2020 in the month of June.

From SafeGraph data it is possible to build a matrix of trip flows between locations in a day for arbitrary large region R of the US. In particular, a county is considered with a tessellation at the resolution of census block group level, then it is possible to reconstruct the adjacency matrix of directed trips from an origin location towards a destination as stops of devices as described in [19]. Fig. 5 plots a sequence of visitation counts in different census block group areas during different time windows of a day for New York city.

The data has been re-organized as shown in Table 2 where the $(N+1)\times(N+1)$ matrix $\boldsymbol{A}$ indicates the global origin–destination table visiting flow between the N blocks of the region R plus one external node which represents the resto of the world available in the data outseide the region of interes R. In particular, a_{ij} is the number

Table 2
Data matrix format. The vector $\boldsymbol{D}$ represents the array of locations as destination units in the tessellation region. Similarly $\boldsymbol{O}$ represents the same array locations but as origins of trips inside the tessellation region. The array W_D is the set of destinations located outside the region and W_O is the set of all the origin locations outside the region. So $\boldsymbol{A}_0$ is the open O-D table, and $\boldsymbol{A}$ is the close O-D table.

$$\boldsymbol{A}=\left[\begin{array}{c|cccc|c} & \multicolumn{4}{c|}{\boldsymbol{O}} & W_O \\ \hline \boldsymbol{D} & a_{11} & a_{12} & \dots & a_{1N} & w_1 \\ & \vdots & \vdots & \ddots & \vdots & \vdots \\ & a_{N1} & a_{N2} & \dots & a_{NN} & w_N \\ \hline W_D & w^1 & w^2 & \dots & w^N & v_W \end{array}\right]=\left(\begin{array}{cc|c} \multicolumn{2}{c|}{\multirow{2}{*}{$\boldsymbol{A}_0$}} & w_1 \\ & & \vdots \\ & & w_N \\ \hline \multicolumn{2}{c|}{w^1\cdots w^N} & v_w \end{array}\right)$$

of visits registered as stop in the destination j in R that originated in the location i in R. For locations outside the selected region, the entry w_i counts the visits in the destination i of the region R that originated from a location outside the region R. The in-degree for the destination location i is the sum along the columns of the row i from the matrix $\boldsymbol{A}$ from which the empirical visit distribution is evaluated. In addition, the trip-visit distribution, aka the in-strength distribution, is evaluated by associating a weight to each visit. As a typical choice, the weight is taken to be the distance between the origin block and the destination one in kilometers as estimate of the distance from home traveled by devices. Such information is recovered by the census bureau geographical data using Census Block Group geometries with longitude and latitude coordinates of the block centroid [19,83], calculated as the haversine distance between the visitor's home geohash-7 and the destination location geohash-7 for each visit. The median distance of all the travelers who visit a given destination, is also reported in SafeGraph column. A more detailed estimate would be the effective distance traveled by each visitor in the trip between its main location to the selected destination. Such information is not reported in SafeGraph at the moment neither in other data sources consistent with the data structure in the study. However, the radial movement approximation is motivated by the fact that travelers typically seek the shortest route [2,22].

Additionally, the SafeGraph dataset provides the count of visits received by each CBG at every hour of the day throughout an entire month, albeit lacking information regarding the origin. Understanding how the distribution of visits and mobility flow evolves throughout each hour of the day can be valuable.

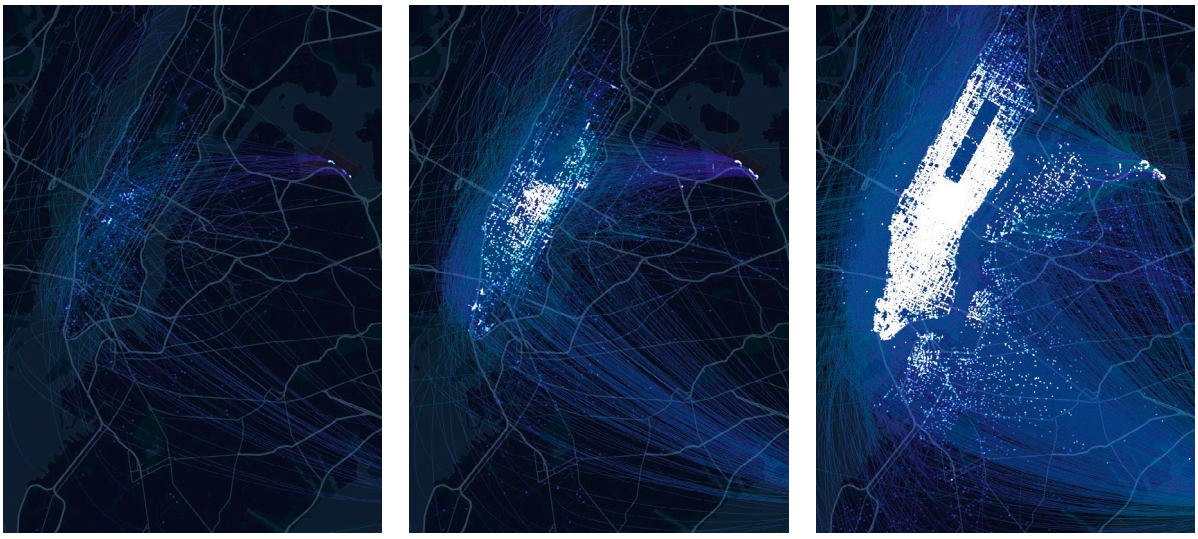
4. Mobility network analysis

The analysis of the Origin–Destination (O-D) network focuses on a spatial scale corresponding to the metropolitan area of New York City, comprising 1076 Census Block Groups (CBGs). However, in tallying arrivals to destinations, all other CBGs outside the considered region are also included, allowing for a comprehensive view of the visitation process and accurate estimation of the in-degree and in-strength distribution of the mobility network. Unfortunately, equivalent information for outgoing distributions is unavailable, as it would necessitate access to all other locations in the US, which is currently unavailable at this stage of the research.

[5] The SafeGraph Patterns dataset demonstrated an average sampling rate of approximately 7.5%, a considerable figure considering the US population's size. Additionally, this sampling rate remained relatively consistent across counties, with the number of sampled devices strongly correlating with census population figures. Consequently, the sampling bias correction can be performed by utilizing the official ACS population data [81], from Census Bureau in the same year, with mobile phone visitor patterns, so the population flows are inferred using the following equation:

$$pop_{flows}(o,d)=visitor_{flows}(o,d)\cdot\frac{pop(o)}{pop_{tot}}\cdot\frac{devices(o)}{devices_{tot}}$$

where $pop_{flows}(o,d)$ is the estimated population flows from geographic origin area o to geographic destination d, $visitor_{flows}(o,d)$ is the computed mobile phone-based visitor flow from o to d, $pop(o)$ indicates the population at the geographic origin area o extracted from the ACS, $devices(o)$ refers to the number of unique mobile devices residing in o, pop_{tot} is the total population in the considered region and $devices_{tot}$ is the total number of unique devices registered in the same total region. While sampling bias was generally minimal across demographics like gender, age, and moderate income, minority groups such as Hispanic populations, low-income households, and individuals with lower education levels tended to experience higher levels of underrepresentation bias. This bias varied across space, time, urbanization levels, and spatial scales.



(a) 5am-6am (b) 5am-9am (c) 5am-12pm

Fig. 5. Number of cumulative visits in New York city [82] for different time windows.

Let us start with the estimate of the distribution of visits among different destinations, which namely represent the in-strength distribution so that the in-strength of destinations is $\kappa_i = \sum_j C_{ij} A_{ij}$ where A_{ij} is the entry of the adjacency matrix indicating the number of arrivals in destination i of a trip originated in the location j. Whereas C_{ij} is the visit "size" of the traveler who departed from origin j and has arrived to destination i. Such value is taken from weight matrix C that represents, in this particular case, the distances between origin–destination pairs. In the following sections the analysis is referred to monthly data in November 2019.

4.1. Visitation distribution and graph correlations

In this section I show that the O-D SafeGraph mobility network is consistent with the hypothesis of uncorrelated graph with scale-free visit distribution at a macroscopic scale as also discussed in [9]. In Fig. 6(a) the empirical complementary cumulative distribution function is plotted for the case of New York metropolitan area in November 2019. The inspection of in-strength distribution shows that the visit distribution has a scale-free asymptotic behavior as $P(\kappa) \sim \kappa^{-\mu}$ with power law coefficient of $\mu \approx 1.8$. Let us now discuss the degree correlations such as assortativity and clustering just in the case of the origin–destination matrix. As already discussed, the average out-strength of neighbors of destinations of in-strength κ measures the tendency of having a directed trips from an origin location to a destination, defined as in [84,85], and from here the assortativity spectrum can be built. As plotted in Fig. 6(b), the average nearest neighbor in-strength function is flat, and this shows a neural assortativity behavior with a mean value of $\langle k_{nn}(\kappa)\rangle_\kappa \approx 15.4$. Such estimate is in agreement with the analytical prediction of the expected average in-strength of the nearest neighbor for uncorrelated networks $\mathbb{E}[k_{nn}^{(u)}] = \langle \kappa_{out}^2 \rangle / \langle \kappa \rangle = 15.7$ as proposed in Remark 3 . Under the same conditions, the average in-clustering coefficient can be computed accordingly to the definition [86,87] adapted to the in-clustering coefficient defined in the present work. The clustering spectrum for the data is shown in Fig. 6(c) where the clustering spectrum is flat as for uncorrelated-graphs, and the global clustering coefficient is given by $\langle c_{in}(\kappa)\rangle \approx 0.02$ consistently with the analytical prediction of $\mathbb{E}[c^{(u)}]$ for uncorrelated networks as reported in Remark 3. The absence of degree-correlations allows to consider the origin–destination conditional probability to be $\chi(y|x) = \chi(y)$. This means that a destination receives a visit from a randomly chosen origin location, without any particular choice of the origin but the number of resident populations in it. Consequently, the correlations between origins and destinations are entirely due to trip costs represented in the weight matrix is a well-mixed locations at level of attractiveness property as specified in the model assumptions.

The scale free behavior of the visit distribution together with the neutral tendency of degree correlations, allows us to narrow the type of kernel function to be considered in the model. Despite that, we do not have enough information to uniquely determine the attraction rate and the latent variable distribution. We will face such issue in the nex section when we will discuss possible proxies of attractiveness latent-variable.

4.2. Degree-strength scaling network topology

In this section, I demonstrate that the inclusion of trip weight information, representing the distance traveled, influences the scaling relationship between weighted visits and degree (arrivals). This leads to a notable alteration in the shape of the visit distribution, which remains scale-free but exhibits a distinct power-law coefficient.

As consequence, a very important analysis of the mobility network, is to study how the trip costs affect the topology of network. The number of trip arrivals at the ith destination can be written as the in-degree $k_i = \sum_j A_{ij}$, meanwhile the visit strength of the ith destination can be written as $\kappa_i = \sum_j C_{ij} A_{ij}$ where C_{ij} is the entry of the Origin–Destination distance matrix C. The in-degree and in-strength distribution have been plotted in Figs. 7(a) and 7(b) respectively, with a clear scale-free asymptotic behavior but with different power law coefficients. Such evidence suggests a very interesting aspect of visiting patterns where trip cost weights have a significant effect of on the mobility network structure. In particular, the relation between strength and degree of a location node can be written the average strength of destinations with degree k changes as:

$$\kappa(k) \sim k^{1+\delta} \tag{13}$$

where the exponent δ represents the rescaling factor and $\delta = 0$ occurs in the absence of correlations between the weight of links and the degree of nodes [85] so that the strength of a node is simply proportional to its degree and the two quantities provide therefore the same information on the system. The action of some correlation in the weight can bring cases where $\delta \neq 0$. In such situation such relation induces a change in the scaling of the degree distribution (i.e. visit distribution) $P(k) \sim k^{-\mu_0}$

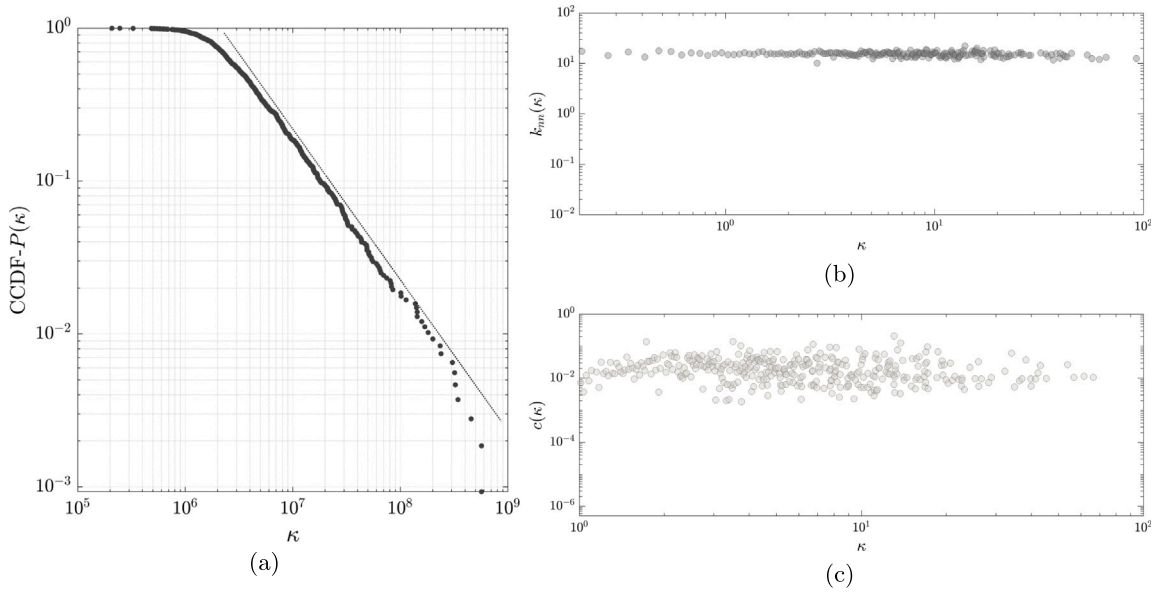


Fig. 6. Visit distribution and correlations of the Safegraph Neighbor Patterns mobility network for New York city in the month of November 2019. The complementary cumulative in-strength distribution function is plotted in (a) of the trip-visit probability density function with a clear scale-free behavior. As for the trip-visit correlations, the network shows a neutral assortativity (b) and a neutral clustering (c) computed on the normalized version of the OD adjacency matrix.

and the strength distribution (i.e. trip-visit distribution) $P(\kappa) \sim \kappa^{-\mu}$ according to the relation:

$$\mu = \frac{\mu_0 + \delta}{1 + \delta} \tag{14}$$

so that the visiting distribution can be written as:

$$P(\kappa) \sim \kappa^{-\frac{\mu_0+\delta}{1+\delta}} \tag{15}$$

Let us notice that in the case of $\delta = -1$, $P(\kappa)$ is a Delta distribution of constant strength. In the case of the data under study, there is a clear linear relation between strength and degree as in Fig. 7(c), consequently the strength distribution shows a scale free coefficient $P(\kappa) \sim \kappa^{-\mu}$ different from the one in the degree distribution $P(k) \sim k^{-\mu_0}$ consistently to the trasformation in Eq. (14).

The value of the rescaling exponent δ has been estimated through a regression analysis which also allows to perform the significance level for a linear relation between degree and distance-strength. As regarding with the data used in this study, the weight of a links correspond to the physical distance between the origin and the destination of the trip which defines the trip-visit distribution as indicated in Fig. 7. As reported in the caption, there is a neat change of the slope in the scale-free distribution, the visit-distribution based on the degree Fig. 7(a) shows a slower power law coefficient respect to trip-visit distribution based on node strengths Fig. 7(b). This is due to the linear relation between the degree and its weighted version through the origin–destination distances. Let us notice that the weights have an impact on the degree which is significant in determining a change in the scaling relation of the distribution. Such result of $\delta > 0$ suggests that the strength of nodes grows faster than their degree, in other words, the trip distances associated to highly visited locations have higher values than those expected if the trip distances were assigned at random. Such tendency denotes a strong correlation between the trip weight and the topological properties in the mobility network, where the higher the number of visits in a location, the more traffic the location can handle. As a conclusion, a general scheme arises where the number of visitors to any destinations decreases as the inverse power law of the product of their visiting frequency and travel distance, as already suggested by [22]. On the contrary, random weights would have produced a δ close to zero, which occurs in the case when link weights are independent from the network topology, so that the strength distribution would carry no information than the degree distribution. For a more detailed estimate of the rescaling exponent δ, a regression analysis is reported in the first row of Table 3, for different types of trip weights: distance between origin–destination pairs and travel time to reach a destination, and the income of travelers. As a additional study to the linear regression statistics, performing a residual analysis makes it possible to test the assumption of a linear regression model such as the errors are independent and normally distributed, as shown in the Supplementary Materials (SM).

Table 3
In-degree vs. in-strength regression analysis for different trip-weights in November 2019. The slope of linear fit reported is the coefficient $1 + \delta$. Let us notice in the scaling relation between distance traveled and attractiveness as land use is $\langle z \rangle_x \sim x^{\theta}$, where the relation $\theta = \alpha - \alpha_0$ holds.

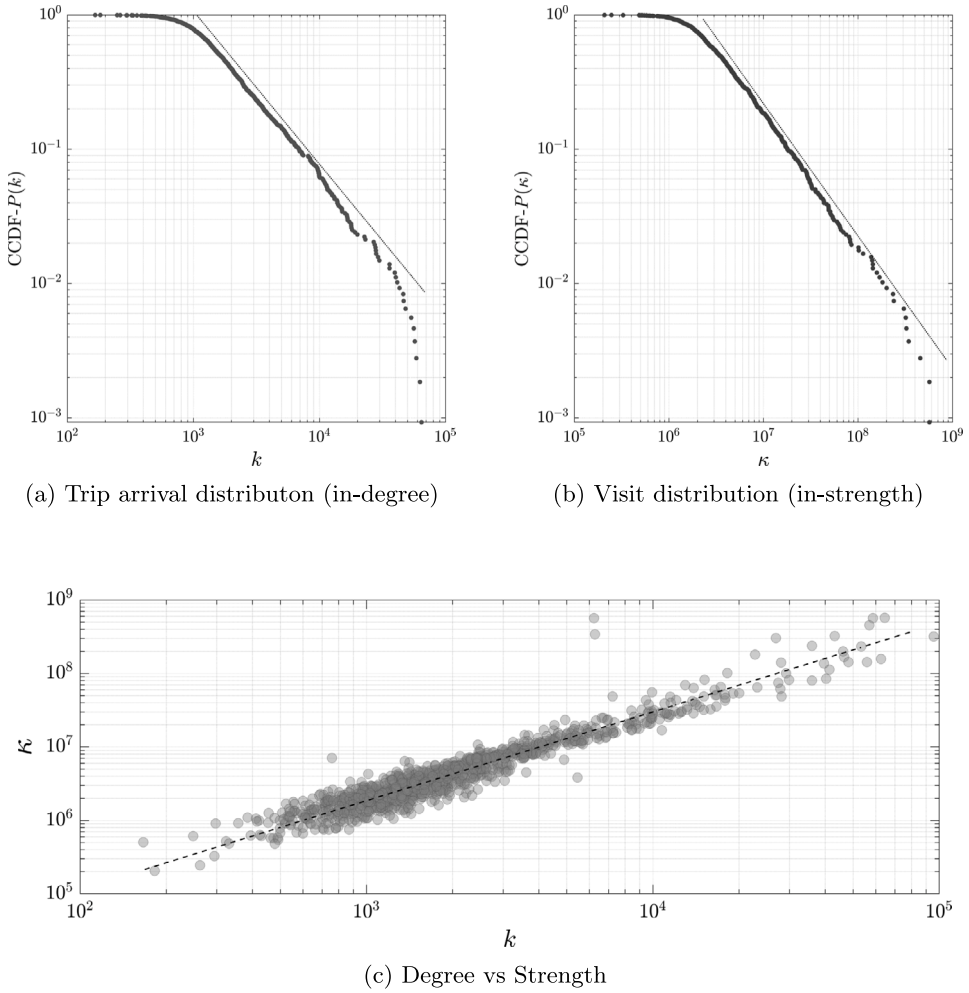
	Distance	Travel time	Income
δ	0.531* [0.467, 0.595]	0.033* [0.025, 0.042]	0.017* [0.010, 0.024]
α_0	0.842* [0.790, 0.894]	0.842* [0.790, 0.894]	0.842* [0.790, 0.894]
α	1.322* [1.223, 1.421]	0.874* [0.822, 0.932]	0.858* [0.805, 0.912]

Note for the linear fit the p-value of the linear regression: *p < 0.01. Numbers in square parenthesis indicates the confidence intervals at a significance level of 99%.

4.3. Attractiveness variable from land-use data

In this section, ultimately, I demonstrate how to infer the scaling properties mentioned earlier, by directly leveraging the data concerning the land-use of CBGs as a proxy for the attractiveness of destinations. Consequently, we can express the visit distribution in terms of land-use, thereby establishing scaling relations between latent variables and trip weights.

Regional movement patterns, and consequently travelers' distribution, can be explained from land use, since purposes of people's trips are strongly correlated with the land use of the trip's origin and destination [35,36,75,88]. In this section, the investigation on possible properties of latent variables will be useful to select which kernel function is suitable to meet the trip mobility network characteristics. For example, the latent-variable framework can provide an interesting



(a) Trip arrival distributon (in-degree)

(b) Visit distribution (in-strength)

(c) Degree vs Strength

Fig. 7. Visit patterns for in-strength defined in terms of distance traveled in trips as visits in one month (November 2019). The log–log linear fitting of degree-strength scatter plot provides a significant scaling coefficient as $\kappa \sim k^{1+0.53}$ in (c), so that the visit distribution asymptotic behavior is $P(k) \sim k^{-2.2}$, in (a), meanwhile the trip-visit distribution by distance weights is $P(\kappa) \sim \kappa^{-1.8}$, in (b).

interpretation of the effect of trip distances on the visit distribution, which can be formalized in terms of the attractiveness latent variable model. It is out of scope of this work to detect the best combination of factors which define the attractiveness of locations, however, how supported by some studies [34,89,90], let us take the non-residential land-use of census blocks to be the primary cause of travel demand and so it could be consider a proxy of destination attractiveness x in a multi-purpose travel model.

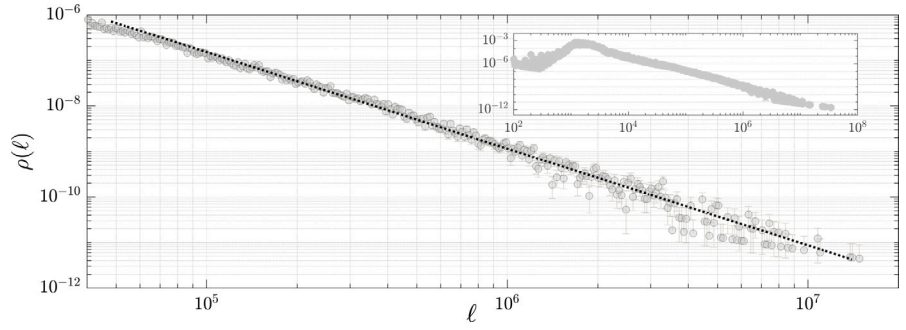
At this point, it is possible to formulate the observed mobility network in terms of the latent-variable model so that the scale-free distribution shown in the real-world trip mobility network can be stated in terms of the latent variable statistical attributes. In such perspective, the data from the New York Open Data [91] has been used where the land use zones is reported as the square feet occupied by building, parks and areas with a given use of destination (except residential), see the Supplementary Materials for more details on data used and correspondent interpretations. Administrative regions are classified by their physical features and land use which refers to the way in which land is utilized, developed, and transformed for different purposes such as residential, commercial, industrial, and agricultural purposes.Land-use can be seen as a "ceteris paribus" candidate for attractiveness of locations and in Fig. 8(a), the plot shows the probability density function of square feet of land-use lots and it shows a scale-free behavior with a power law exponent of 2 as plotted in Fig. 8(a). The same analysis is performed after aggregating tax lots into census block groups, the land-use areas for non-residential purposes keep the same asymptotic fat-tail distribution with an inverse power law probability density function $\rho(x) \sim x^{-\eta}$ with $\eta \approx 2$ as plotted in Fig. 8(b).

At this point it is possible to investigate the relation between mobility and land-use data (as attractiveness indicator). First, it is possible to analyze the relation between the number of arrivals k in each destination with the non-residential land-use of that area x. Similarly the relation of land-use versus the visits κ as weighted arrivals is analyzed as well. So, let us investigate a log–log linear regression analysis of such variables. The linear fit analysis of the relation $k_x \sim x^{\alpha_0}$ reveals an estimated value $\alpha_0 \approx 0.842$ as shown in Fig. 9(a) and linear fit analysis of the relation $\kappa_x \sim x^{\alpha}$ reveals an estimated value $\alpha \approx 1.322$ as shown in Fig. 9(b). In Table 3 such estimations for α_0 and α are reported for different types of trip weights. By knowing that $\kappa_x = \mathbb{E}[\kappa|x] \propto \langle \iota \rangle_x k_x \sim x^{\theta+\alpha_0}$, it is possible estimate the scaling of the trip size goes as $\langle \iota \rangle_x \sim x^{\theta}$, as confirmed by a direct linear fit analysis where $\theta \approx 0.48$.

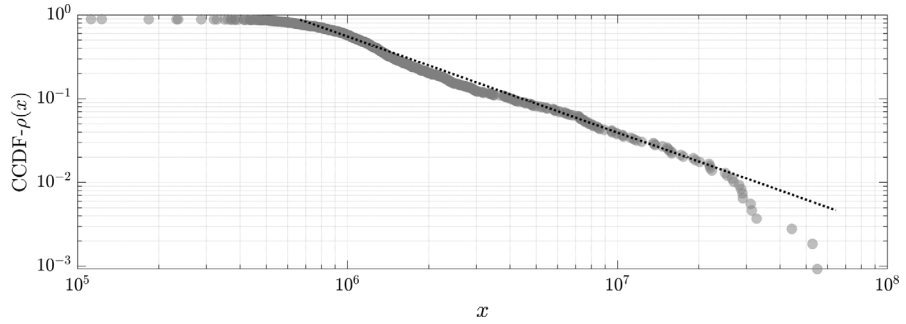
Finally, by using the latent variable framework one can recover the visit distribution as

$$P(\kappa) \sim \kappa^{-(1+\frac{\eta-1}{\alpha_0+\theta})} \tag{16}$$

which is the same scale-free distribution directly observed during the analysis of the origin–destination network. It is worth noticing the relation between the scaling exponents θ and δ. It can be easily verified that the visit distribution in Eq. (15), obtained by distance traveled, and the visit distribution Eq. (16), obtained by land use hidden variable,

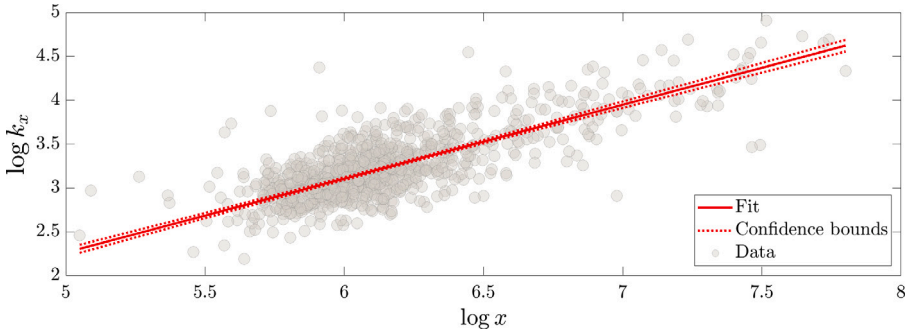


(a) Tax lot square feet for non-residential land use

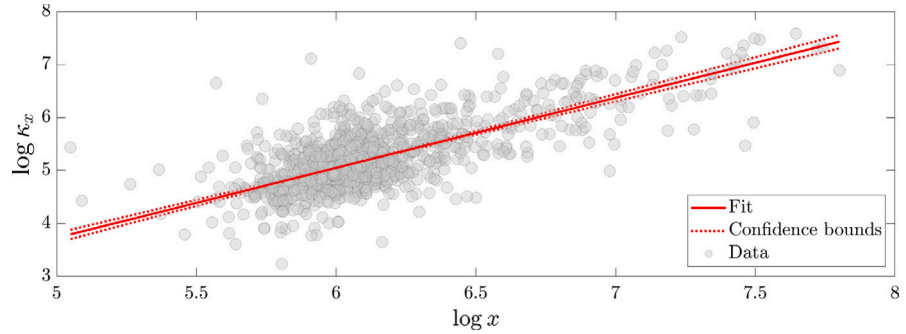


(b) Tax lot square feet aggregated in census block groups

Fig. 8. Log-binning procedure for the probability density function of square feet for land-use zones as hypothetical index for destination attractiveness. It shows an inverse power law fat-tail distribution so that the asymptotic behavior the probability density function can be written as $\rho(x) \sim x^{-2}$. In the inset the whole distribution is plotted where the data is fully represented even at a low spatial scale. In (b) the complementary cumulative density function of the land-use square feet aggregated for census block from tax lot data.



(a) Scatter log-log plot of the destination with land-use x versus the number of arrivals k_x to reach that destination. The slope is $\alpha_0 \approx 0.84$.



(b) Scatter log-log plot of the destination with land-use x versus the number of visits κ_x (i.e. number of arrivals weighted by distance traveled) to reach that destination. The slope is $\alpha \approx 1.32$.

Fig. 9. Regression fit between the land-use of census block group and (a) arrivals and (b) visits.

are the same in-strength pdf so the power law coefficient be $\mu = \frac{\mu_0+\delta}{1+\delta} = 1 + \frac{\eta-1}{\alpha}$, where $\mu_0 = 1 + \frac{\eta-1}{\alpha_0}$ and $\alpha = \alpha_0 + \theta$. Solving, we find that, the following crucial relation holds:

$$\delta = \frac{\theta}{\alpha_0} = \frac{\alpha}{\alpha_0} - 1 \tag{17}$$

which can be checked by comparing the values reported in the Table 3 under their relative error margins. In conclusion, the attraction rate of a location is higher than another destination in the sense that that destination is able to attract more distant travelers respect to the second. Such effect reveals the action of human travel demand on mobility, as formulated in the present paper, in terms of latent variable network model. The scaling relation between land use versus both travel distances and visits are power laws, so that the attraction rate must be of a power law type and the distribution of land use (i.e.the attractiveness latent variable) has a Pareto-type distribution. So no other attraction rate is compatible with the observation. Moreover, since degree correlation are absent, it is plausible that $\chi(y|x) = \chi(y)$ so that the kernel function $\mathcal{K}(x, y)$ is a multiplicative separable function.

As a result, we can interpret the findings by stating that the trip weight decreases with increasing attractiveness, indicating that the perceived cost of traveling towards more attractive regions diminishes when the destination itself is more attractive.

4.4. Mobility network evolution

In the previous analysis temporal granularity was at a monthly scale, i.e. the total number of trips traveled in one month. Moreover, it is possible to investigate a more granular temporal resolution, and observing how the counting of new visits changes each hour in a day. I then compare it with the analytical prescription through the visitation network model. In particular, I show how the average number of total visits changes over time during a day, highlighting the contribution of attractiveness in hourly commuting.

At this purpose, the average-degree of mobility network sheds some light on the change of aggregated attractiveness in 24 h. I show how the analytical description of the degree evolution can explain the trend of total number of arrivals in all the destinations.

In the assumption of infinitely large non-sparse networks, we can approximate the evolution of the average degree as an Ito stochastic differential equation related to the master equation Eq. (1) in the limit of $\epsilon \to 1$. So, given the density probability function p_x, the strength of stochastic process is $Y_t := \mathbb{E}_t[k|x]$ with $Y_t \sim p(Y, t)$ [92,93]. Consequently, the visiting process can be described by an Ito process in terms of the counting of new visits in location of attractiveness x as: $d\mathbb{E}_t[k|x] = \nu_x dt + \sigma_x dW_t$ where W_t is a Wiener process and a standard deviation $\sigma_x = \sqrt{\nu_x}$. Finally, the total amount of new visits in the network is obtained by $\bar{k}_t = \int_{\Omega_x} \mathbb{E}_t[k|x]\rho(x)dx$, so that the Ito daily process for the network degree is described by the following stochastic equation:

$$\begin{aligned} d\bar{k}_t &= \left(\int_{\Omega_x} \nu_x \rho(x)dx\right) dt + \left(\int_{\Omega_x} \sigma_x \rho(x)dx\right) dW_t \\ &= \bar{\nu}_t dt + \bar{\sigma}_t dW_t \end{aligned} \tag{18}$$

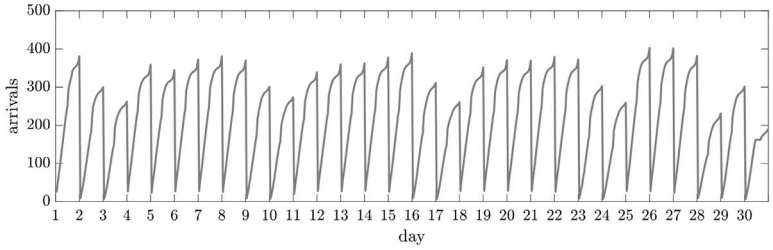
where $\bar{\nu}_t$ is the mean attraction rate and $\bar{\sigma}_t$ is the mean standard deviation, and the can change over time during a day. The Eq. (18) describes the continuous sample path of the transport diffusion degree process in the visiting network model. In Fig. 10, we observe the typical progression of total arrivals (cumulative new visits) throughout a day. Subsequently in Fig. 10(b), I employ the stochastic equation Eq. (18) to generate simulated real-world data, estimating the mean attraction rate $\bar{\nu}_t$ to match the observed trend precisely. Notably, there is a discernible shift in trend around 5–6 pm, coinciding with the time when typical workflow undergoes inversion, thereby redefining the attractiveness of destination points locations. Although attractiveness has been linked to a latent variable representing land use, it remains accurate on a broad temporal scale. However, within a single day, the land use must be adjusted by a factor corresponding to the actual perception of willingness to travel towards those destinations. In this specific scenario, the mean attractiveness of destination locations starts with a value of $\bar{\nu}_t \approx 20$ (daytime period) and then it decreases by approximately 80% after late afternoon. Nonetheless, the mean attractiveness averaged over a month aligns with what has been inferred in the preceding sections. Finally, in Fig. 10(c) the evolution of the empirical distribution of total arrivals obtained from mobile phone data is well approximated by a truncated power-law distribution.[6] The power law part replicates the analytical results in Eq. (8) with a scale free parameter of $\mu_0 \approx 2.2$ as expected. At this stage of the research, I lack access to information regarding the distance traveled at an hourly level since the dataset provides data at a monthly resolution as for origin–destination pairs, meanwhile it offers the numbers of stops (arrivals) each hour of the month. So it is not possible to estimate the strength distribution at a daily scale but only at monthly scale as explored in the previous sections.

Another crucial observation must be made here. The specific values of power-law coefficients (μ and μ_0) and scaling factors (δ, α_0 and θ) are characteristic only of the analyzed time period. What remains universal are the formulas governing the scaling variables in mobility. Indeed, a comparison of the same dataset from June 2020 reveals different power-law coefficient values for degree and strength distributions, yet the scaling relations persist. The same holds true for the relationship between land-use variables and strength degrees; despite varying values, the scaling relations remain intact. The Fig. 11 shows those results, which can be interpreted under the light of period analyzed, in fact the Eq. (14) confirms the scaling between strength and degree and the role of trip-cost (i.e. distance traveled) in the network topology. However, the values of power law coefficients of degree and strength distributions are not the same as for the month November 2019, even the aggregated attractiveness is about 50% respect toe the estimated value in November 2019. During June 2020 NY metropolitan area was under one of the more stringent periods of social distancing restrictions, as in [94], that was put in action as mitigation policies for the containment of Covid19 disease spreading. In the periods of the first part of the year 2020, containment measures and isolation responses can be seen as a shock to human mobility [95–97], the results show how the values of coefficients and variables change, but the scaling relation remains unvaried.

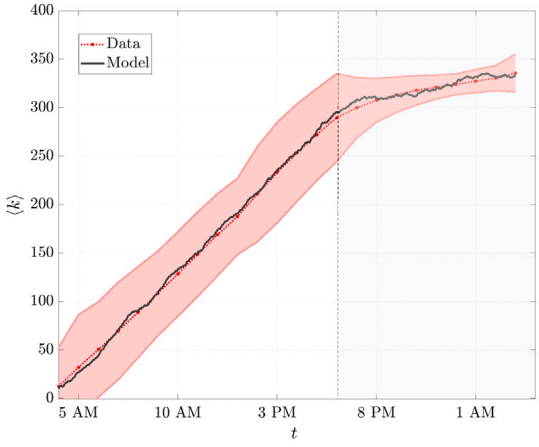
In the data analysis section of the paper, empirical evidence guides the selection of ν_x towards a power-law form for attraction rates and trip weights. This choice holds true within the temporal and spatial scale of our analysis. However, finer spatial resolutions, more detailed traveler preferences, and shorter temporal scales may necessitate adjustments to the model, potentially resulting in different shapes of visit distributions or the presence of assortativity and transitivity, prompting consideration of network correlations. Power-law emergence is anticipated at larger spatial scales due to the multiscale nature of transport networks, meanwhile, other distributions can emerge as for example at urban scales where a single transportation mode predominates [98,99].

It is noteworthy that the emergence of power laws in our analysis is attributed to the statistical properties of the latent variable, particularly its influence on the kernel probability and corresponding attraction rate. When the land-use distribution follows a power-law pattern and the probability of selecting destinations is proportional to a power

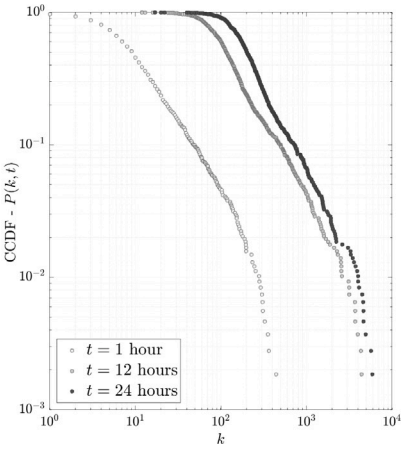
[6] The exponential cut-off is typically due to a lack of statistics or small samples. As an example, in the case of small population size, we can modify the Remark 1 replacing the attractiveness distribution to be power law with exponential cutoff like $\rho(x) \sim \rho_0 x^{-\eta} e^{-\lambda x}$, it is possible to prove, see [84], that the degree distribution is $P(k) \sim k^{\frac{\eta-1}{\alpha_0}-1} e^{-c\lambda k^{1/\alpha_0}}$ where c is a constant. So the arrivals distribution is of a Weibull-type, with (stretched) exponential tail. As similar discussion can be found in the Supplementary Materials.



(a) Total number of arrivals each hour in a day, over different days of a month.



(b) Cumulative number of trip arrivals.



(c) Trip arrival distributions.

Fig. 10. The analysis entails examining the total number of arrivals per hour for each day throughout November 2019. Cumulative counting occurs within each day, beginning at 4 a.m. and concluding at 3 a.m. the following day, with the counting process resetting thereafter. Figure (a) illustrates this process for the entire month. In Figure (b), the average cumulative visit count in a typical day is depicted alongside confidence intervals, followed by the presentation of a stochastic trajectory through the fit of Eq. (18), with a daytime aggregated attractiveness of estimated as $\bar{v}_x \approx 20$, which then drops by 80% after late afternoon. Lastly, Figure (c) displays the arrival (i.e., degree) distribution across three distinct times of the day.

of the land-use, the mobility graph remains neutral, resulting in a scale-free visit distribution. However, various combinations of latent-variable structures can also lead to similar scale-free distributions. Alternatively as listed in Appendix C, if the kernel probability $\mathcal{K}(x, y)$ lacks multiplicative factorization, we may observe both graph correlations and diverse forms of visit distribution, ranging from power laws with exponential cut-offs to distributions that deviate from power laws entirely.

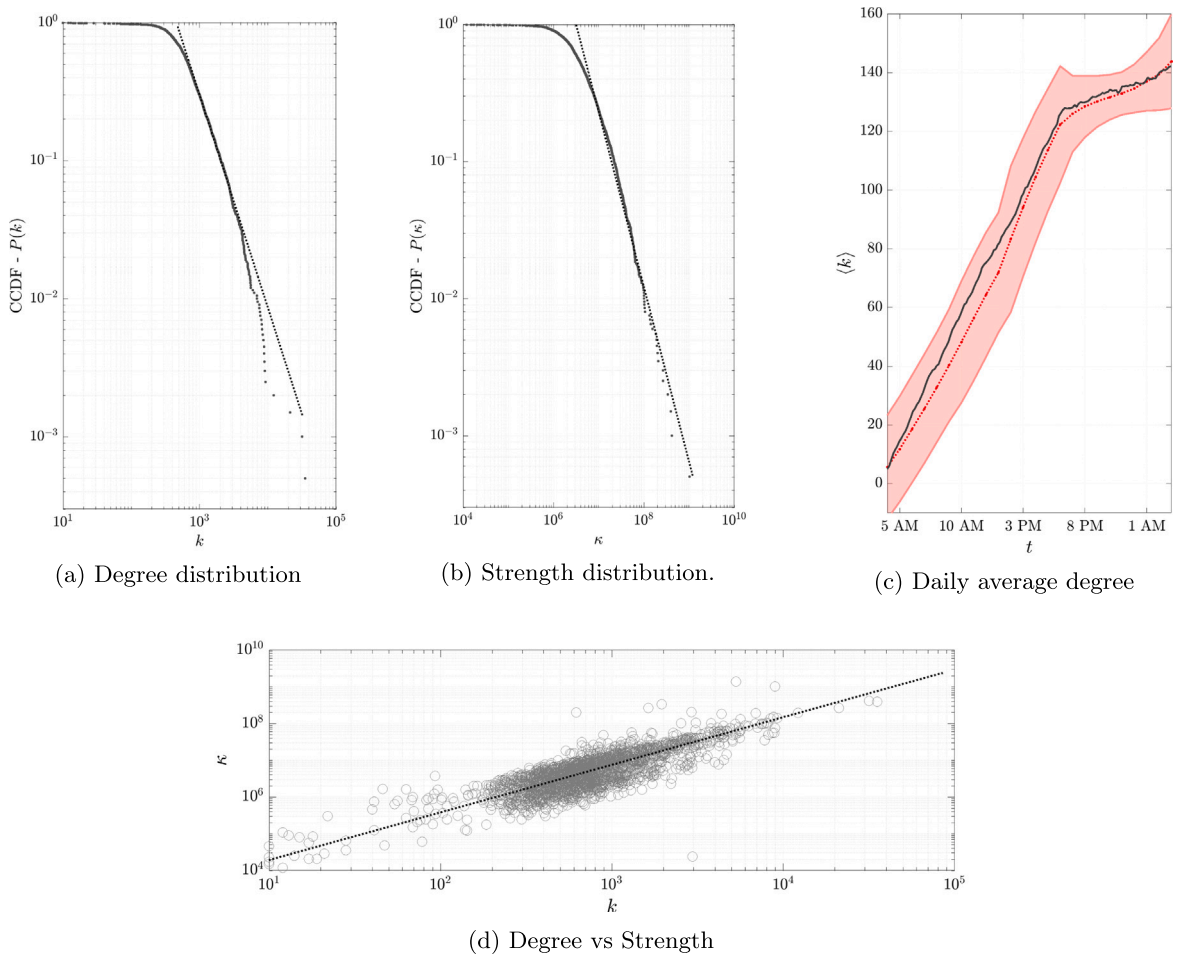
5. Economic analysis: mobility network and income elasticity

In this section, I demonstrate the significance of the latent variable as a key factor justifying the income elasticity of multi-purpose travel demand across various transportation modes. The visitation model approach can provide economical interpretations of some empirical urban scaling evidences [9], where scaling laws are also present in economical values of each location respect to its attractiveness. A crucial attention will be focused on the income elasticity of travel demand of each location can be estimated through land use and the number of visit.

In a economic modeling framework, those latent variables allows the characterization of travel demand for different areas in the region, and they can be inferred through the analysis of important demographic, geographic and economic indicators. As discussed in literature [34,35,100], many different factors are involved for a place to be considered as attractive, such as trip purposes, job or leisure opportunities, infrastructure facilities, geographical characteristics, urban zoning planing. Generally, all those physical and human factors can be captured by land use and travel behavior analysis which are at the basis of two primary approaches in transportation economics, urban science and engineering literature. In such literature, in fact, travel is considered to be a derived demand, that it is generated in response to people satisfying personal needs and desires [4,90,101]. Let us, first, define the "benefit" that travelers received by visiting a location, and in particular, the variable I_x indicates the income level associated to the visitors who have traveled towards a given location with attractiveness x for a job purposes. Census data from Census Bureau surveys [81] have been used; I extracted the median aggregated household income of a specific Census Block Group (CBG), assuming that a traveler originating from that location shares the reported income level. In this way the benefit can be determined through the strength-by-income variable κ_i which converts a visit into potential economic output through a conversion factor i_0 that is the traveler's income per unit of visit time.[7] On the other side, let us define the "cost" Q_x faced by travelers to reach a location as the strength-by-distance variable κ_q taking into consideration that commuting costs are proportional to distance or time traveled, and the proportionality conversion factor c_0 is the cost of transportation per unit of quantity traveled.[8] Finally, the relation between the two variables can be written in terms of the attractiveness

[7] The factor i_0 is an average conversion which attributes a benefit-value to each trip, but this value depends on the distribution among possible trip purposes (as business, consumption, leisure, ...) for each traveler by their origin locations. If origins and destinations are, again, randomly independent i_0 can be considered a mean-field constant correction factor. Further it can change over the observation time during which data have been collected. Considering i_0 to be location independent it can be considered a proportionality constant.

[8] Quantity can be measures in distance or travel time. A lower c_0 correspond to a better transportation system. Let us observed that in the case of cities the transportation costs and distance are well approximated by a linear relation. However for larger areas, empirical studies [102,103] show that transport cost



(a) Degree distribution (b) Strength distribution. (c) Daily average degree

(d) Degree vs Strength

Fig. 11. The scaling analysis entails examining the principal estimates in June 2020. In (a), the empirical survival distribution of arrivals in June 2020 shows a probability function compatible with $P(k) \sim k^{\mu_0}$ with $\mu_0 \approx 2.50$. In (b) the empirical survival distribution of visits (arrivals weighted by distance traveled) in June 2020 shows a probability function compatible with $P(\kappa) \sim \kappa^{\mu}$ with $\mu \approx 2.25$. In (c), the average cumulative visit hourly count in a typical day is depicted alongside confidence intervals, followed by the presentation of a stochastic trajectory resenting the average degree evolution. The daytime aggregated attractiveness has been estimated as $\bar{v}_t \approx 10$, which then drops by 80% after late afternoon. Finally, in (c), the log–log linear fitting of degree-strength scatter plot provides a significant scaling coefficient as $\kappa \sim k^{1+\delta}$ with $\delta \approx 0.22$. Those estimations confirm the scaling relation Eq. (14).

variable as:

$$\frac{\text{travelers' income } (I_x)}{\text{travel quantity } (Q_x)} = \frac{i_0 \mathbb{E}[\kappa_i|x]}{c_0 \mathbb{E}[\kappa_q|x]} \propto \frac{i_0\, x^{\alpha_0+\theta_I}}{c_0\, x^{\alpha_0+\theta_Q}} \tag{19}$$

where θ_Q is the scaling exponent derived from regression slope by income in Table 3 and the exponent θ from regression slope by amount of travel given by distance traveled in the same analysis. The relation between three variables is represented graphically in Fig. 12(a). At this point it is possible to write the income elasticity of travel demand which has the meaning of how sensitive the demand for traveling a certain distance is to changes in income levels, the direct relation between the income reward and the quantity of travel demanded for visiting locations can be derived from Eq. (19) as:

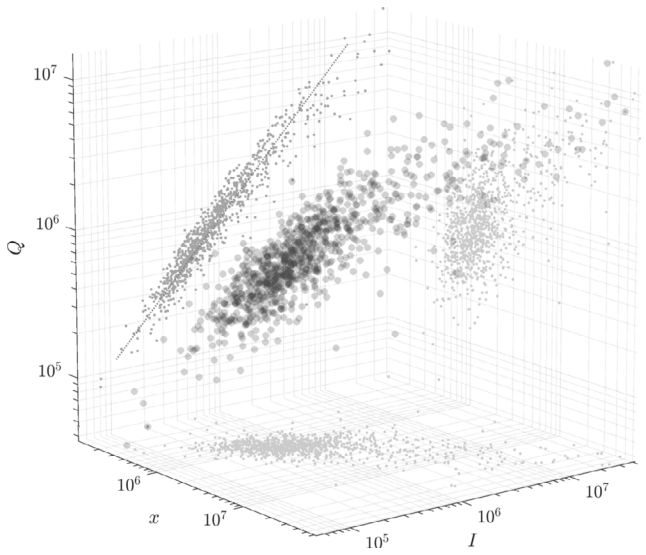
$$Q_x \sim I_x^{\varepsilon} \quad \text{, with } \varepsilon = \frac{\partial Q_x}{Q_x}\frac{I_x}{\partial I_x} = \frac{\alpha_Q}{\alpha_I} = \frac{1+\delta_Q}{1+\delta_I} \quad ,\forall x \in \Omega_x \tag{20}$$

where the exponent ε is the income elasticity as estimated in Fig. 12, and in the case under study $\varepsilon < 1$, indicating that distance traveled is a necessity good, i.e., as income increases, people spend proportionally less on traveling when the income levels of travelers increase for any transportation mode.[9] Such prediction of an income elasticity of about $\varepsilon \approx 0.65$ can be compared to results presented in literature reviews of empirical evidences and meta-analysis studies [39,104–106] and also discussed with a theoretical interpretations [40,103], where average income elasticity for aggregated travel demand is estimated to be in a range values compatible with the elasticity ε estimated here.[10] It is important to notice that the income elasticity of travel demand can be influenced by many factors, such as the availability of transportation options, the price of transportation, and individual preferences. From a network analysis perspective, it would be convenient to compare the number of trip arrivals at destinations against the travelers' income.

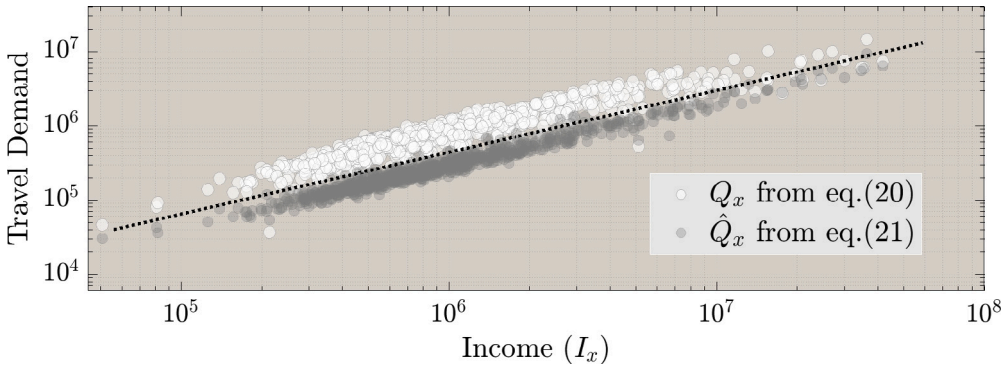
is an increasing and concave function of distance so that $c(\tau) = c_0 \tau^{\upsilon}$ with $0 < \upsilon < 1$, since travelers switch to faster transport modes for longer trips.

[9] For any mode and any purpose condition, undistinguished transportation mode is considered here, so all the possible modes are combined and only the distance necessary to reach the destination is taken into account.

[10] Let us notice that income elasticity shows a large variability in the empirical evidences since distance traveled is not homogeneous across different sources of income, type of jobs and age [38]. Moreover, travel demand is reported in different units (individual or aggregate distance km/day, travel time, fuel consumption) and in different behaviors (commuting vs non-commuting, essential vs non-essential) or travel purposes (business, job, shopping, leisure). Travel can also vary in terms of different traveling modes according to transportation infrastructure. Moreover the estimations reported can even change over the years.



(a) Three dimensional scatter plot for the variable $\{x, I_x, Q_x\}$.



(b) Scatter plot for the evaluation of income elasticity of travel demand

Fig. 12. Using the dataset, for New York in 2019, in (a) a 3D scatter plot shows the relation among attractiveness x as land-use, the income level of visitors and travel demand as distance traveled for each visit. The projection planes show three 2D scatter plots for pairs of the previous variables. In particular, the Q-I plane shows the projected regression analysis an elasticity of $\varepsilon \simeq 0.76$ consistently with the theoretical prediction from Eq. (20) and from Eq. (20) where the income $I_x \sim k_x^{1+\delta_I/\alpha_0}$. This indicates that a 10 per cent increase in income leads to about 6.5 percent increase in distance traveled or, conversely, a 10 percent increase in the demand of travel distance requires an increase of the income by about 15.4 percent.

Using Eqs. (19) and (20), and visit elasticity of travel demand ε' can be defined through:

$$Q_x \sim k_x^{\varepsilon'} \qquad \text{where} \quad \varepsilon' = \left(1 + \frac{\theta_I}{\alpha_0}\right)\varepsilon = (1+\delta_I)\varepsilon \tag{21}$$

where the income scales as $I_x \sim k_x^{1+\theta_I/\alpha_0}$ respect to the number of trip arrivals. The visit elasticity of travel demand is a measure of how sensitive the number of trips to a certain destination is to changes in key travel attributes, such as fare level, service quality, journey time components, income and car ownership, and price of competing modes. The previous scaling relations of income elasticity, it is possible to range from the microeconomics perspective of transport economics to the macroeconomic outputs such as employment and the economic growth. For example, during a period of an economic growth (recession) where incomes are rising (falling) the distribution and the magnitudes of attractiveness can change and then modify the mobility pattern which, in its turn, has an impact on the global economic performance as well.

6. Conclusions and perspectives

In conclusion, the study presents a data-driven model for human mobility network based on an origin–destination structure, which serves as the foundation for understanding mobility visitation flows. The model utilizes latent variables associated with each location, representing attractiveness and productiveness, to capture the intrinsic characteristics of destinations and origins. The contributions of this research are twofold. Firstly, it provides a theoretical framework that describes and reproduces a visit generation stochastic process through the use of integral–differential equation for the evolution of the visit probability density and degree correlations in a trip mobility network. Consequently, analytical, numerical, and computational solutions are provided for important network characteristics, such as the strength distribution, assortativity, and clustering coefficient. A collateral impact of the visit generation model in travel dynamics is the introduction of a mathematical formalism of compound renewal processes commonly used in financial and actuarial science literature. Such stochastic models in a network perspective are useful for capturing the randomness and dependence between the arrival times and event sizes, providing a flexible framework for modeling and analyzing various mobility dynamics and financial and actuarial risks as well. The second contribution of this research has to do with the empirical analysis of real world phenomena. By analyzing origin–destination data, the study reveals the presence of scale-free behaviors in visit frequencies and identifies correlations between visits and trip costs.

The research also explores the statistical characteristics that latent variables should have to reproduce the observed patterns in the trip mobility network. Essentially, given the model's assumptions validated by empirical analysis, I anticipate observing inverse power law shapes in both the degree arrival distribution, characterized by coefficient μ_0, and the strength visit distribution, characterized by coefficient μ. Specifically, the degree distribution follows $P(k) \sim k^{-\mu_0}$, while the strength distribution follows $P(\kappa) \sim \kappa^{-\mu_0}$. Furthermore, the topological relationship between node degree and strength conforms to the scaling relation $\kappa = k^{1+\delta}$. An important topological scaling relation emerges, expressed as $\mu = \frac{\mu_0+\delta}{1+\delta}$. Leveraging our latent variable framework, we can link the topological factor δ with the visitation process mechanism, given by $\delta = \frac{\theta}{\alpha_0}$, where α_0 represents the coefficient of attraction rate and θ represents the coefficient of trip cost dependence on attractiveness. The network visitation model presented in the article generalizes some aspects of traditional mobility models, particularly their ability to maintain accuracy and relevance across different spatial and temporal scales. However, the principal advantage of the mobility network model lies in its parameter selection through temporal network analysis, utilizing measures such as centrality, correlations, and clustering. In some cases, the mobility network model may align more closely with popular mobility models, but in other cases, none of these models may be suitable for describing mobility flows, such as when dealing with degree correlations in the graph. In this situation, the mobility network model can still provide a description of commuting flows.

Moreover, the model permits to disentangle the effect of attractiveness (as land-use), population, trip costs and economic features of travelers on the visit dynamics in a mobility network. Clear scaling laws emerges between latent variables and travel demand. So, the model points out the effect of income on travel behavior depends strictly from the latent-variables that, therefore, can be considered as decision variable from a economic policy viewpoint. The possibility that human mobility belongs to the class of scale-free networks has impacted on the economic, engineering and mathematical communities in the multidisciplinary field of sustainable urban transportation, smart cities and world trade webs. As future outlooks, from a modeling side, the mathematical formalism discussed in the paper could also be extended to more general mobility graphs by considering more granular interactions on a time interval much shorter than the one used in the data explored here. For example, one can investigate more on the distribution of inter-arrival times of new trips to be non-Poissonian, and the trip size are events with intensity that does not have finite moments (Levy-jumps). Additionally, I have depicted the visitation process as cumulative counting of new visits, without accounting for link destruction in network dynamics due to the need for understanding visit duration mechanisms. A potential model extension will involve analyzing dwell times spent at each location during a visit.

Another upcoming study will focus on assessing emission mobility reductions for sustainable city planning. This approach aims to inform decisions regarding the optimal allocation of economically attractive urban areas based on transportation mode and land-use policies, while simultaneously minimizing pollutant emissions and the impact of mobility trips between origins and destinations. Furthermore, a broader panel analysis across other regions is necessary to enhance the robustness and validity of the model. Ultimately, this study illuminates potential applications in urban, transportation, and environmental economics. The focus of the research outlined in the paper can be seen as an activity-based modeling approach, which can more closely replicate actual traveler decisions and thus may provide better forecasts of future travel patterns. The research outcome of such an approach can have important implications for real-world issues. For example, the attractiveness of urban areas, such as land use, is shaped by factors like urban planning, policy regulations, and the economic dynamics of firm allocations and aggregations, see [9,75]. In particular, the findings in the present paper can help enhance the use of land-use zoning as a policy method to classify land and restrict its development. Future research in urban planning policies could focus on mixed-use development, blending residential, commercial, cultural, institutional, or entertainment uses into one space, in line with the scaling relations outlined in the present article. Spatial planning and territorial governance are crucial for sustainable land use by assessing the quality and characteristics of different locations relative to competing objectives and interests. As discussed in [107–109], for example, urban form can result from historical evolution or policy outcomes: compact urbanization often stems from containment policies directing development inward through regeneration, infill, or redevelopment. Polycentric urbanization is pursued by spatial planning policies like transit-oriented development, while diffuse urban form often results from policies encouraging private car use and homeownership. These insights provide valuable guidance to policymakers and urban planners, facilitating the understanding and prediction of mobility patterns for informed decision-making and sustainable development.

CRediT authorship contribution statement

Fabio Vanni: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

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

Data availability

Origin–Destination mobility data has been retreived by SafeGraph for academic research program [19], and they cannot made public available. Social, demographic and geographical data are public accessible at [41,83,91,110]. Codes for the main network analysis are povide in the Github repository [111].

Acknowledgments

This work has been supported by the French government, through the UCAJEDI Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01. The author wants to acknowledge to be a member of GNAMPA-INdAM.

Appendix A

A.1. Proof of Proposition 1

Proof. Let us notice that the attraction rate ν_x is transition rate of new visits per unit of time interval, as the chance of having a new arrival in a destination of type x originated from any location, so that the attraction rate is the mean intensity as in [10,11,49,84]:

$$\nu_x = \int_{\Omega_y} d\mathcal{V}_x = \int_{\Omega_y} \mathcal{K}(x,y) d\mu_y(y) = \int_{\Omega_y} \mathcal{K}(x,y)\phi(y)dy \tag{A.1}$$

The master equation for the evolution of the conditional probabilities $p_x(\kappa,n)$ for locations with attractiveness x, where the step size is the correspondent $\Delta\kappa = \iota$ which represents the weight of each link i.e. the decision heuristic variable to move in the selected destination. It can be written as an integro-differential Kolmogorov–Feller equation:

$$p(\kappa,t+\tau|x) = (1-\nu_x)p(\kappa,t|x) + \int_0^\kappa \nu_x p(\kappa-\iota,t|x)\varrho_x(\iota)d\iota$$

If we also assume that p_x is slow, so that it changes only slightly during this time step $\tau = \Delta t$ and redefine $\nu_x := \nu_x/\tau$, then we can write the continuum master equation like:

$$\frac{\partial}{\partial t}p_x(\kappa,t) = \dot{p}_x(\kappa,t) = \int_0^{\kappa} \nu_x \varrho_x(\tau)\,[p_x(\kappa-\tau,t) - p_x(\kappa,t)]d\tau \tag{A.2}$$

and $\varrho_x(\tau)$ is the distribution of the trip distances covered to reach destination blocks of attractiveness x. At this point let us apply the Laplace transformation in the variable κ so $\mathcal{L}\{p(k,t)\} \equiv \hat{p}(s,t)$ and the master equation transforms as:

$$\dot{\hat{p}}(s,t) = \nu_x\hat{\varrho}_x(s)\hat{p}_x(s,t) - \nu_x\hat{p}_x(s,t) = \nu_x\left(1-\hat{\varrho}_x(s)\right)\hat{p}_x(s,t)$$

where the convolution product has been used. The solution can be written as:

$$\hat{p}(s,t) = ce^{-\nu_x(1-\hat{\varrho}_x(s))t} \tag{A.3}$$

with the initial condition $\hat{p}(s,0) = \mathcal{L}\{\delta(k)\} = e^0 = 1$ so that $c = 1$. One can notice that the characteristic function is equivalent to the Laplace transform as expressed before. It is possible to write the Laplace transform of the jump distribution $\varrho_x(\tau)$ in terms of its moments as:

$$\hat{\varrho}_x(s) = \sum_{n=0}^{\infty}(-1)^n\frac{s^n}{n!}\mathbb{E}(\kappa^n) \tag{A.4}$$

If one assumes that $\varrho(\tau)$ is a peaked distribution, following the central limit theorem rationale, one can assume it is described by the first two (finite) moments so that the solution in Eq. (A.3) can be approximately:

$$\hat{p}(s,t) \approx e^{-\nu_x t\langle\tau\rangle_x s + \frac{1}{2}\nu_x t\langle\tau^2\rangle_x s^2} \tag{A.5}$$

its inverse Laplace transform can be considered [112] the case asymptotic case of $s \ll \frac{\langle\tau\rangle_x}{\langle\tau^2\rangle_x}$ of the truncated normal distribution as stated in the Eq. (2)

For the second point, the visit (in-strength) distribution of the mobility network is expressed as a compound probability distribution that results from assuming that a random variable κ is distributed according to some parametrized distribution with the latent parameter x distributed according to some attractiveness distribution [113,114]. So, the (unconditional) visiting in-strength distribution results from marginalizing the conditional distribution $p(k,t|x)$ of the non-negative real-valued random variable x. So, the probability density function of the visiting distribution is given by the following the mixture density:

$$P(\kappa,t) = \mathbb{E}\left[p(\kappa,t|x)\right] = \int_{\Omega_x} p(\kappa,t|x)d\mu_x(x) = \int_{\Omega_x} p(\kappa,t|x)\rho(x)dx \tag{A.6}$$

Here, $p(\kappa|x,t)$ is the distribution of κ when we know x at time t, in which the relation between κ and x can be seen as deterministic, i.e. $\kappa = F(x,t) = \mathbb{E}_t[\kappa|x] = \nu_x\langle\tau\rangle_x t$, defining the distribution by its expected degree value through the moment-generating function from the Laplace transform above. So, κ can only be single value, whose distribution is represented by dirac-delta functions $\delta(\kappa - F(x,t))$. Consequently, the empirical visiting density probability function can be written as:

$$P(\kappa,t) \sim \int_{\Omega_x}\delta(\kappa - F(x,t))\rho(x)dx = \int_{\Omega_x}\delta(\kappa - \nu_x\langle\tau\rangle_x t)\rho(x)dx \tag{A.7}$$

which consists in approximating the visiting probability density by means of a Dirac mixture [115,116], where $\rho(x)$ is the attractiveness probability density. Such procedure is equivalent to a change of variable respect to the deterministic one-to-one function in the static model as in [16]. At this point we use the property: $\delta(z(x)) = \sum_{i=1}^{m}\frac{\delta(x-x_0^{(i)})}{|\partial z(x)/\partial x|}$ where $x_0^{(i)}$ are the m-roots of $z(x) = 0$ where in the transport model $z(x) = \kappa - \nu_x\langle\tau\rangle_x t$ where z is a continuously differentiable function with z' nowhere zero. So:

$$P(\kappa,t) \sim \sum_{i=1}^{m}\left|\frac{\partial z(x)}{\partial x}\right|^{-1}_{x_0^{(i)}(\kappa)}\int\delta\left(x - x_0^{(i)}(\kappa)\right)\rho(x)dx \sim \sum_{i=1}^{m}\left|\frac{\partial z(x)}{\partial x}\right|^{-1}_{x_0^{(i)}(\kappa)}\rho\left(x_0^{(i)}(\kappa)\right)$$

which represent a general formula for the tail behavior of the degree distribution for a mobility network with a generic attraction rate ν_x. □

A.2. Proof of Proposition 2

Proof. The conditional origin–destination probability is the conditional probability that a destination block of attractiveness x is connected to an origin block of population y is:

$$\chi(y|x) = \frac{\phi(y)\mathcal{K}(x,y)}{\int\phi(y)\mathcal{K}(x,y)dy} = \frac{\partial\mathcal{V}_x(x,y)}{\partial y}\frac{1}{\nu_x(x,y)} = \frac{\partial}{\partial y}\log\mathcal{V}_x(x,y) \tag{A.8}$$

where $\mathcal{V}_x$ is the primitive function of the attraction rate ν_x. Io order to write the explicit expression of the origin–destination correlation it is necessary to know the pairing rule, for example through the connection kernel $\mathcal{K}(x,y)$ so that $\nu_x = \nu_0\int_{\Omega_y}\mathcal{K}(x,y)\phi(y)dy$, or imagining a generic primitive function for the attraction rate. So the conditional origin–destination probability is an important indicator for correlations between origins and destinations in the degree-distribution of the visiting mobility network. The conditional average-nearest-neighbor's in-degree for destinations of attractiveness x can be written for directed multigraph in a continuous limit as in [16,84]:

$$k_{nn}(x) = \int\mathbb{E}[\kappa|y]\,\chi(y|x)dy \tag{A.9}$$

where in the mobility model the conditional expected strength is $E[\kappa|y] \propto \nu_y t$. Since, for the markovian degree property, the degree two point correlation can be fully determined by the conditional probability $P(\kappa'|\kappa)$, the average degree of neighbors of an in-degree κ destination is known to be calculated as [48,59]:

$$\begin{aligned}k_{nn}(\kappa,t) &= 1 + \frac{1}{P(\kappa,t)}\int p(\kappa,t|x)\rho(x)k_{nn}(x)dx\\ &= 1 + \frac{t}{P(\kappa,t)}\iint\nu_y\chi(y|x)p(\kappa,t|x)\rho(x)dydx\end{aligned}$$

which is an in-out (origin–destination) assortativity measure, which is independent of κ for $\chi(y|x) = \chi(y)$ as in the case of multiplicative separable linking function $\mathcal{K}(x,y)$. Let us notice that since the network is directed, other than weighted, one could define, similarly, other three average nearest neighbor's degree functions: destination–origin, origin–origin and destination–destination.

The clustering coefficient of a destination with attractiveness x can be interpreted as the probability that two randomly chosen edges from x are origin-neighbors. The clustering of a destination of degree one or zero is defined as zero. In the space of latent variables, consider a destination i of attractiveness x_i and population y_i, which is connected with probability $p(y_j,y_k|x_i)$ through trips originated from two other locations j and k which have attractiveness x_j and x_k and population y_j and y_k respectively. Since the network is markovian at the latent variable level, $p(y_j,y_k|x_i) = p(y_j|x_i)p(y_k|x_i)$. Thus, similarly to the definition in [48,70] together with modifications [60,86] for the directed and weighted case, the local origins-destination clustering for locations of attractiveness x_i can be written as:

$$\begin{aligned}c(x_i) &= \sum_{j,k}p\big((x_j,y_j),(x_k,y_k)\big)\,p(y_j,y_k|x_i)\\ &= \frac{\sum_{j,k}\frac{1}{2}\left(\mathcal{K}(x_j,y_k)+\mathcal{K}(x_k,y_j)\right)\mathcal{K}(x_i,y_j)\mathcal{K}(x_i,y_k)}{\sum_{j,k}\mathcal{K}(x_i,y_j)\mathcal{K}(x_i,y_k)}\end{aligned}$$

where $p\big((x_j,y_j),(x_k,y_k)\big)$ is the probability that the two origin nodes are connected one to the other in both directions. Now, in the asymptotic

continuous regime the clustering coefficient can be rewritten:

$$c(x) = M \iiiint \tfrac{1}{2}(\mathcal{K}(x',y'')+\mathcal{K}(x'',y'))\,\mathcal{K}(x,y')\,\mathcal{K}(x,y'')\,\rho(x')\rho(x'')\phi(y')\phi(y'')dx'dx''dy'dy''$$

where $M = (\int \mathcal{K}(x,y')\phi(y')dy' \int \mathcal{K}(x,y'')\phi(y'')dy'')^{-1}$, so it is possible to write:

$$c(x) = \iiint \chi(y'|x)\mathcal{K}(x',y'')\chi(y''|x)\rho(x')dx'dy'dy'' = \frac{1}{2v_0}\iint \left(v_{y'}+v_{y''}\right)\chi(y'|x)\,\chi(y''|x)dy'dy''$$

knowing that $v_y = v_0 \int_{\Omega_x} \mathcal{K}(x,y)\rho(x)dx$ and the definition of $\chi(y|x)$. Let us notice that for independent origins and destinations then $c(x) = c_0 = const$. Moreover in the case of clustering coefficient in a multigraph one can calculate the number of triangles repeated κ times, which in a markovian graph can be approximated on average as $\bar{c}_t(x) = tc(x)$, since at each time step a possible link is considered as a bernoullian trail, so that the observed number of links in t trials follows a binomial distribution and so the expected value is $t\mathcal{K}(x,y)$. Consequently, since the clustering coefficient has values in $[0,1]$, we normalize the adjacency matrix respect to t, so that the average local clustering coefficient of a node with strength κ, denoted by $c(\kappa)$, [48,58,59,117] is given by:

$$c(\kappa,t) = \frac{1}{tP(\kappa,t)}\int p(\kappa,t|x)\bar{c}_t(x)\rho(x)dx = \frac{1}{P(\kappa,t)}\int p(\kappa,t|x)c(x)\rho(x)dx$$

which represents the local in-clustering spectrum for destination locations and it is independent of κ for $\chi(y|x) = \chi(x)$ as in the case of multiplicative separable linking function $\mathcal{K}(x,y)$, where $P(\kappa,t)$ represents the in-strength distribution. Let us notice that in the case of the clustering coefficient the adjacency matrix and the latent variables are needed to be normalized in order to transform a multigraph in a weighted graph and from there a clustering coefficient no larger than 1 is guaranteed. □

A.3. Proof of Remark 3

Proof. The equations for the average in-strength of nearest neighbors and the in-clustering coefficient can be derived directly from Proposition 2 after some algebraic manipulations considering the production rate v_y and the conditional probability $\chi(y|x)$ are independent from x as, in the case when the attraction and production rates are recovered from a linking function that is multiplicative separable, i.e. $\mathcal{K}(x,y) = g(x)h(y)$. In fact, when origins and destinations are independent, that is $\chi(y|x) = \chi(y)$, then the average-nearest-neighbor's strength function $k_{nn}(\kappa,t) = 1 + t\int v_y\chi(y)dy$ which is a constant over the strength degrees κ and where $E[g(x)]$ and $E[h(y)]$ are the expectation values of the function $g(x)$ and $h(y)$ under the circumstances they have finite values, which occurs even for fat-tail distributions in a finite set for the latent variables [84]. In the case of infinite moments then the equation above represents a unreliable estimation but still shows the neutral assortativity of the graph. As regard with the clustering coefficient, the result in (12) is straightforward by using the multiplicative separable linking function as above, with the only difference that it has been normalized in order to provide a weighted matrix with link weights not larger that one. Another approaches for the estimates of the assortativity and clustering coefficient can be derived in terms of the strength degrees of the nodes as provided in [48,57,58] with proper modifications for directed weighted graph, see [84,118,119]. Their expected values for $k_{nn}(\kappa)$ and $c(\kappa)$ are analytically known in literature for neutral networks, i.e. no degree correlations, as derived in terms of degrees the expected average in-strength of nearest neighbors can be written:

$$k_{nn}(\kappa_{in}) = \sum_{\kappa_{out}} \kappa_{out}P(\kappa_{out}|\kappa_{in}) = \sum_{\kappa_{out}} \kappa_{out}\frac{\kappa_{out}P(\kappa_{out})}{\langle\kappa_{out}\rangle} = \frac{\langle\kappa_{out}\rangle^2}{\langle\kappa_{out}\rangle} = \langle k_{nn}\rangle = \mathbb{E}[k_{nn}^{(u)}]$$

where in the absence of correlations $P(\kappa_{out}|\kappa_{in}) = k_{out}P(\kappa_{out})/\langle\kappa_{in}\rangle$ has been used. In the case of clustering coefficient, after normalization of the multigraph, the in-clustering coefficient can be written as in [58,120]:

$$c(\kappa) = \sum_{\kappa'_{out},\kappa''_{out}} \frac{(\kappa'_{out}-1)(\kappa''_{out}-1)}{tN\kappa''_{out}P(\kappa''_{out})}P(\kappa''_{out}|\kappa'_{out})P(\kappa''_{out}|\kappa_{in})P(\kappa'_{out}|\kappa_{in}) = \frac{\langle\kappa_{in}\rangle^3}{tN\kappa_{in}^2P^2(\kappa_{in})}\sum_{\kappa'_{out},\kappa''_{out}} \frac{(\kappa'_{out}-1)(\kappa''_{out}-1)P(\kappa''_{out},\kappa'_{out})P(\kappa''_{out},\kappa_{in})P(\kappa'_{out},\kappa_{in})}{\kappa'_{out}\kappa''_{out}P(\kappa'_{out})P(\kappa''_{out})}$$

That in the case of uncorrelated networks:

$$c(\kappa) = \frac{\langle\kappa_{in}\rangle^3}{tN\kappa_{in}^2P^2(\kappa_{in})}\sum_{\kappa'_{out},\kappa''_{out}} \frac{(\kappa'_{out}-1)(\kappa''_{out}-1)}{\kappa'_{out}\kappa''_{out}P(\kappa''_{out})P(\kappa''_{out})}\cdot \frac{\kappa''_{out}P(\kappa''_{out})}{\langle\kappa''_{out}\rangle}\frac{\kappa'_{out}P(\kappa'_{out})}{\langle\kappa'_{out}\rangle} \cdot \frac{\kappa''_{out}P(\kappa''_{out})}{\langle\kappa''_{out}\rangle}\frac{k_{in}P(k_{in})}{\langle k_{in}\rangle}\cdot\frac{\kappa'_{out}P(\kappa'_{out})}{\langle\kappa'_{out}\rangle}\frac{k_{in}P(k_{in})}{\langle k_{in}\rangle}$$

$$= \frac{\langle\kappa_{in}\rangle}{tN\langle\kappa_{out}\rangle^4}\sum_{\kappa''_{out}}(\kappa''_{out}-1)\kappa''_{out}P(\kappa''_{out})\sum_{\kappa'_{out}}(\kappa'_{out}-1)\kappa'_{out}P(\kappa'_{out}) = \langle c\rangle = \mathbb{E}[c^{(u)}]$$ □

Appendix B. Inhomogeneous graph compound processes

The visit generation process of the temporal random graph is demonstrated to be equivalently formalized in terms of a mixture of compound Poisson counting processes, commonly used in financial and actuarial science literature [121,122]. In the following appendix, the visit generation process of the trip mobility network will be re-framed in terms of a mixture of many compound Poisson counting processes, as follows:

Proposition 3. *The probability distribution of the visit generation process can be restate as a mixture of compound counting Poisson processes.*

- *Let us define the evolution of the trip arrival process looking at the generation of new visits as a counting stochastic process* $\{\mathcal{C}_x(t), t \geq 0\}$, *defined on the probability space* $(\mathbb{R},\mathcal{B}(\mathbb{R}),\mu_x)$, *for all the locations of attractiveness* x *where* $\mathcal{C}_x(t)$ *represents the number of visits which have arrived between time* 0 *and* t *at the destination nodes of attractiveness* x, *and let the attraction rate* $v_x = \text{Prob}\{\mathcal{C}_x(t+\tau) = k|\mathcal{C}_x(t) = k-1\}$ *be the probability (per unit of infinitesimal time* τ*) that a destination of attractiveness* x *increase its visits from* $k-1$ *to* k. *Then, let* $\mathcal{Z}_k$ *be a sequence of independent, identically distributed square-integrable random variables, which represent jumps as trip-costs of the trip arrivals and they are distributed as a common random variable* z_x *with cumulative distribution function* $F_{\mathcal{Z}}(z_x)$.

Then, the conditional visit probability density function $p_x(k,t)$ *can be derived by a continuous-time stochastic process with jumps. In particular:*

1 *The compound Poisson process* $\{Y_x(t) : t > 0\}$ *is continuous-time stochastic process, adapted to a filtration* $\mathcal{F}_t$, *with random jumps of intensity* v_x, *defined by:*

$$Y_x(t) = \sum_{k=1}^{\mathcal{C}_x(t)} \mathcal{Z}_k \tag{B.1}$$

Consequently, it turns out that the probability density function can be explicitly written for a compound Poisson process as $p(\kappa,t|x) = f_{Y_x(t)}(\kappa,t|x)$ *and the cumulative distribution function is given by:*

$$F_{Y_x(t)}(\kappa,t|x) = e^{-v_xt}\sum_{n=0}^{\infty}\frac{(v_xt)^n}{n!}F_{Y_{n,x}}^{*n}(\kappa|x) \tag{B.2}$$

where $F_{Y_{n,x}}^{*n}$ *is the n-fold convolution of* $F_{\mathcal{Z}}(z)$. *Moreover, the characteristic function and the first two central moments of* $Y_x(t)$ *are:*

$$\mathbb{E}[e^{isY_x(t)}] = \exp\{v_xt(\mathbb{E}[e^{is\mathcal{Z}}-1])\}$$
$$\mathbb{E}[Y_x(t)] = \mathbb{E}[\mathcal{C}_x(t)]\,\mathbb{E}[\mathcal{Z}] = \mathbb{E}_t[\kappa|x] = v_x\langle z\rangle_x t$$
$$\mathbb{V}ar[Y_x(t)] = \mathbb{V}ar[\mathcal{C}_x(t)]\,\mathbb{E}[\mathcal{Z}^2] = v_x\langle z^2\rangle_x t$$

2 *The visit distribution is the mixture of stochastic Poisson distributions* $p_x(\kappa,t)$ *[123] i.e. the conditional visit probability density function where the intensity of the (simple) counting process,* ν_x*, is a random variable with a probability density function* $\rho(x)$*. The number of locations with visit strength* κ*, which we denote by* N_κ *satisfies, see [11,124]*

$$N_\kappa/n \xrightarrow{\mathbb{P}} f_\kappa \equiv \mathbb{E}_x\left[F'_{Y_x(t)}(\kappa,t|x)\right] \tag{B.3}$$

where $\xrightarrow{\mathbb{P}}$ *denotes convergence in probability. We recognize the limiting distribution* $\{f_k\}$ *as a so-called mixed Poisson distribution with mixing distribution* $\mu_x(x)$*.*

Proof. The proof is recovered from [52,53,125]. In particular, a compound Poisson counting process is equivalent to a Levy process and its realizations are piecewise constant cadlag functions. As a consequence of the above results, the compound Poisson process enjoys all the properties of Levy processes, including the Markov property. Compound Poisson processes are commonly used in insurance to model the number of claims and the size of each claim. Insurance companies can use this model to estimate their expected losses and set their premiums accordingly. □

Moreover, in finance, Compound Poisson processes can be used to model stock prices, where the arrival times of price changes follow a Poisson process and the size of each change is a random variable. Finally, in risk management, Compound Poisson processes can be used to model risk in various industries, such as energy trading, where the arrival times of price changes and the size of each change are both uncertain.

Appendix C. From kernel functions to visit distributions and correlations

For a given connection kernel $\mathcal{K}(x,y)$ it is possible to associate the attraction rate ν_x that can be normalized in order to define a transition probability as $\nu_x = \frac{\int \mathcal{K}(x,y)\phi(y)dy}{N\iint \mathcal{K}(x,y)\rho(x)\phi(y)dxdy}$ where the denominator is the total degree of the network of size N. as shown in Table C.4, by using the asymptotic approach for the visiting distribution. First, it is clear how for a completely multiplicatively separable kernel $\mathcal{K}(x,y)$ $(\cdot M\cdot)$ the rate only depend on the attractiveness and the origin–destination correlation is independent from the attractiveness variable x. Secondly, for a completely additively separable kernel $\mathcal{K}(x,y)$ $(\cdot A\cdot)$ the attraction rate does depend also on generation function for the population variable y. Consequently te origin–destination correlation does depend on both x and y. Furthermore, four particular cases of additively and multiplicatively separable kernels are presented with correspondent visiting distribution and origin–destination correlations. Ultimately, three cases are presented where the connection kernel is neither additively nor multiplicatively separable. In particular the last examples are different threshold-logistic kernels that can be approximated by simple attraction rates and so obtaining a straightforward expression for the visiting distribution and origin–destination correlation. Some derivations are presented in the Supplementary Material.

Appendix D. Interpretation of collective human mobility models

Gravity model. The gravity model [25,72] assumes that the interaction between two locations is directly proportional to their masses (e.g., population size) and inversely proportional to the distance between them. The gravity model can be interpreted in terms of the visitation model through the following a mean field approximation:

$$\mathcal{K}(x_i,y_j) = \frac{x_i^a y_j^b}{d_{x_iy_j}} \longrightarrow \mathcal{K}(x,y) = \frac{x^a y^b}{d(x,y)}$$

so that gravity model estimates the travel probability from origins of type y to destinations of attractiveness x. Consequently, the attraction rate in the visitation model can be written:

$$\nu_x = \nu_0 \int \mathcal{K}(x,y)\phi(y)dy = \nu_0 \int \frac{x^a y^b}{d(x,y)}\phi(y)dy$$

We assume an homogeneous resident population over the region and isotropic direction of origin–destination flows.[11] The distance can be approximately written as $d(x,y) \simeq \langle d\rangle_x + d_o(y)$, where $\langle d\rangle_x$ is the mean distance respect to center of population of origins of type y and $d_o(y)$ is the distance of such center of mass from the center of the overall residential populations. If the center of residential population is independent from y, one can reasonably say $d_o(y) \approx 0$. At this point, the attraction rate can be written as:

$$\nu_x = \nu_0 \int \frac{x^a y^b}{\langle d\rangle_x + d_o(y)}\phi(y)dy = \nu_0 \int x^a y^b\left(\frac{1}{\langle d\rangle_x} - \frac{d_o(y)}{\langle d\rangle_x^2}\right)\phi(y)dy$$
$$\sim \nu_0 \mathbb{E}[y^b]\,\frac{x^a}{\langle d\rangle_x} \tag{D.1}$$

Let us notice that the rate for gravity model in Eq. (D.1) is interpreted in terms of visitation model where the trip-cost is $\langle \tau\rangle_x = 1/\langle d\rangle_x$ which explains as the distance traveled is related to attractiveness of the destination location. In this circumstances, the gravity model produce neutral correlation networks. Let us notice that, however, the functional form of the distance $d(x,y)$ in general depends on the spatial geometry of the region. In fact, if asymmetry (inhomogeneous resident population and anisotropic movements) is present, the rate ν_x could generally depend on both latent variables, consequently the mobility networks can show degree correlations since it is straightforward to see that $\chi(y|x) = \chi(y)$.

Intervening opportunity model. Intervening opportunities models posit that mobility flow depends on the number of potential destinations between locations. In the intervening opportunities approach the probability of commuting between an origin with feature y and a destination of attractiveness x in the Schneider's version [24] is:

$$\mathcal{K}(x,y) = e^{-\gamma\langle s_{xy}\rangle} - e^{-\gamma(\langle s_{xy}\rangle - x)} = \frac{1-e^{-\gamma x}}{e^{\gamma\langle s_{xy}\rangle}}$$

given that there are s_{xy} is the number of opportunities (approximated by the population in this case) in a circle of radius d_{xy} centered in the origin location (excluding the source and destination). The parameter γ can be seen as a constant probability of accepting an opportunity destination. In the case that the number of opportunities depend on how far the origins are from destinations $\langle s_{xy}\rangle \sim d(x,y)$, [30,72], the attraction rate becomes:

$$\nu_x = \nu_0 \int \mathcal{K}(x,y)\phi(y)dy \approx \nu_0 \int \frac{1-e^{-\gamma x}}{e^{\gamma d(x,y)}}\phi(y)dy$$

moreover, in the case of homogeneous opportunities and isotropic movements $d(x,y) \approx \langle d\rangle_x$, and the attraction rate simplifies as:

$$\nu_x \sim \nu_0(1-e^{-\gamma x})e^{-\gamma\langle d\rangle_x}$$

which, in an homogeneous case, the attraction rate is of the exponential form, with neutral graph correlations.

[11] Defining $d(x,y)$ the distance between origins of feature y and destinations of feature x, isotropy guarantees that $d(x,y) \simeq \langle d\rangle_x + d_o(y)$, i.e. where $\langle d\rangle_x$ is the mean distance between the destination of attractiveness x and the center of mass of all the origins with y, and $d_o(y)$ is the distance between the center of mass of population-type y and the region centroid. Homogeneity of locations of different origin-types y, guarantees that $d_o(y)$ is independent from the destination's attractiveness x.

Table C.4
The relation between different connection kernels and attraction rates gives different visiting distributions in the asymptotic regime ($n \to +\infty$). The expressions are given as asymptotic approximations of tail distributions where ν_0 is a proper normalization factor. The attractiveness distribution is $\rho(x)$ and the population distribution $\phi(y)$ where $\bar{\Phi}(y)$ is its inverse cumulative distribution function.

	$K(x,y)$	ν_x	x_0	$P(k,n)$	$\chi(y\|x)$
·M·	$g(x)\cdot h(y)$	$\frac{1}{N\mathbb{E}[g(x)]}g(x)$	$g^{-1}\left(\frac{k}{\nu_0 n}\right)$	$\frac{n^{-1}}{\alpha\nu_0 x_0}\rho(x_0)$	$\frac{h(y)}{\mathbb{E}[h(y)]}\phi(y)$
·A·	$g(x)+h(y)$	$\frac{1+\frac{1}{\mathbb{E}[h(y)]}g(x)}{N\left(1+\frac{\mathbb{E}[g(x)]}{\mathbb{E}[h(y)]}\right)}$	$g^{-1}\left(\frac{k}{\nu_0 n}-\mathbb{E}[h(y)]\right)$	//	$\frac{g(x)+h(y)}{g(x)+\mathbb{E}[h(y)]}\phi(y)$
·1·	f_0	ν_0	1	$\frac{(n\nu_0)^k}{k!}e^{-n\nu_0}$	$\phi(y)$
·2·	$x^\alpha y^\beta$	$\nu_0 x^\alpha$	$\left(\frac{k}{\nu_0 n}\right)^{\frac{1}{\alpha}}$	$\frac{k^{\frac{1}{\alpha}-1}n^{-\frac{1}{\alpha}}}{\alpha\nu_0^{1-\frac{1}{\alpha}}}\rho(x_0)$	$\frac{y^\beta}{\mathbb{E}[y^\beta]}\phi(y)$
·3·	$e^{\alpha x+\beta y}$	$\nu_0 e^{\alpha x}$	$\frac{1}{\alpha}\log\frac{k}{\nu_0 n}$	$\frac{k^{-1}}{\alpha\nu_0}\rho(x_0)$	$\frac{e^{\beta y}}{\mathbb{E}[e^{\beta y}]}\phi(y)$
·4·	$x^\alpha+y^\beta$	$\nu_0 x^\alpha+\nu_0\mathbb{E}[y^\beta]$	$\left(\frac{k}{\nu_0 n}-\mathbb{E}[y^\beta]\right)^{\frac{1}{\alpha}}$	$\frac{n^{-1}}{\alpha\nu_0 x_0^{\alpha-1}}\rho(x_0)$	$\frac{x^\alpha+y^\beta}{x^\alpha+\mathbb{E}[y^\beta]}\phi(y)$
·5·	$\Theta(x^\alpha+y^\beta-C)$	$\nu_0\bar{\Phi}\left[(C-x^\alpha)^{1/\beta}\right]$	$\left(C-\bar{\Phi}^{-1}\left[\frac{k}{\nu_0 n}\right]^\beta\right)^{\frac{1}{\beta}}$	$\frac{n^{-1}}{\nu_0\bar{\Phi}'\left[(C-x_o^\alpha)^{1/\beta}\right]}\rho(x_0)$	$\frac{\Theta(x^\alpha+y^\beta-C)}{\bar{\Phi}[(C-x^\alpha)^{1/\beta}]}\phi(y)$
·6·	$\frac{axy}{1+axy}$ [a]	$\nu_0 x(1-ax\mathbb{E}[y])$	$\{1\ ,\ \frac{2a\mathbb{E}[y]}{\nu_0}\frac{k}{n}\}$	$\frac{n^{-1}\sum_i\rho(x_0^{(i)})}{\nu_0}$	$\frac{1-axy}{1-ax\mathbb{E}[y]}y\phi(y)$
·7·	$\frac{1}{1+e^{-(x+y+b)}}$ [b]	$\nu_0\left(1-\frac{\mathbb{E}[e^{-y}]}{e^b}e^{-x}\right)$	$-\log\left(\frac{e^b}{\mathbb{E}[e^{-y}]}(1-\frac{k}{\nu_0 n})\right)$	$\frac{n^{-1}}{\nu_0\left(1-\frac{k}{\nu_0 n}\right)}\rho(x_0)$	$\frac{e^{(x+b)}-e^{-y}}{e^{(x+b)}-\mathbb{E}[e^{-y}]}\phi(y)$
·R·	$\frac{xy}{c^2}(1-\frac{x}{c})$	$\nu_0 x(1-cx)$	$\{1\ ,\ \frac{2ck}{\nu_0 n}\}$	$\frac{n^{-1}\sum_i\rho(x_0^{(i)})}{\nu_0}$	

[a] $a \ll \frac{1}{\max\{xy\}}$.
[b] $b \gg \min\{x+y\}$.

Radiation model. The radiation model [26,72] can be embedded in the trip-mobility network model by considering the travel probability kernel as:

$$\mathcal{K}(x,y)=\frac{xy}{(y+\langle s_{xy}\rangle)(y+\langle s_{xy}\rangle+x)}=\frac{xy}{(y+\langle s_{xy}\rangle)^2(1+\frac{x}{y+\langle s_{xy}\rangle})}$$

By following [72], for a region dense of opportunities among two locations and under homogeneous population assumption $x, y \ll \langle s_{xy}\rangle \sim \langle d\rangle_x$, the attraction rate can be written as:

$$\nu_x \sim \nu_0\int\mathcal{K}(x,y)\rho(y)dy \sim \nu_0\frac{x}{\langle d\rangle_x^2}\left(1-\frac{x}{\langle d\rangle_x}\right)$$

Another extreme condition is the case of an opportunity-sparse region so that $\langle s_{xy}\rangle \approx 0$, and the attraction rate becomes $\mathcal{K}(x,y)\approx\frac{x}{x+y}$, so that the attraction rate is not trivial and the network correlations are not neutral.

Impedance model. According to the impedance model [73], the number of trips flow from origin location of feature y to destination with attractiveness x can be formulated, in the continuous mean-field version, as:

$$\mathcal{K}(x,y)=\frac{x+y}{d(x,y)}$$

and it can be seen that the attraction rate can be written as $\nu_x = \nu_0 x/\langle d\rangle_x + \nu_0\mathbb{E}[y]/\langle d\rangle_x$ and the graph correlations show up as showed in Appendix C.

Other models also engage with the discourse on the universality of scaling relations in visitation patterns, a topic recently emphasized in numerous studies such as [3]. These models suggest that the number of visitors to any given location decreases inversely with the square of the product of their visit frequency and travel distance. In contrast, the model presented here does not focus on individual mobility behavior, hence it does not account for recurring movements with varying visit frequencies as explored in studies like [1,20,99]. Essentially, this model operates as a macroscopic network model of mobility, where the units represent locations and the links represent travelers.

Appendix E. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.chaos.2024.115175.

References

[1] Alessandretti Laura, Aslak Ulf, Lehmann Sune. The scales of human mobility. Nature 2020;587(7834):402–7.
[2] Batty Michael. The new science of cities. MIT Press; 2013.
[3] Schläpfer Markus, Dong Lei, O'Keeffe Kevin, Santi Paolo, Szell Michael, Salat Hadrien, Anklesaria Samuel, Vazifeh Mohammad, Ratti Carlo, West Geoffrey B. The universal visitation law of human mobility. Nature 2021;593(7860):522–7.
[4] Pappalardo Luca, Manley Ed, Sekara Vedran, Alessandretti Laura. Future directions in human mobility science. Nat Comput Sci 2023;1–13.
[5] Barbosa Hugo, Barthelemy Marc, Ghoshal Gourab, James Charlotte R, Lenormand Maxime, Louail Thomas, Menezes Ronaldo, Ramasco José J, Simini Filippo, Tomasini Marcello. Human mobility: Models and applications. Phys Rep 2018;734:1–74.
[6] Solmaz Gürkan, Turgut Damla. A survey of human mobility models. IEEE Access 2019;7:125711–31.
[7] Arcaute Elsa, Ramasco José J. Recent advances in urban system science: Models and data. PLoS One 2022;17(8):e0272863.
[8] Bettencourt Luís MA. The origins of scaling in cities. Science 2013;340(6139):1438–41.
[9] Bettencourt Luís MA. Introduction to urban science. 2021.
[10] Van Der Hofstad Remco. Random graphs and complex networks, vol. 43, Cambridge University Press; 2016.
[11] Bollobás Béla, Janson Svante, Riordan Oliver. The phase transition in inhomogeneous random graphs. Random Structures Algorithms 2007;31(1):3–122.
[12] Kim Bomin, Lee Kevin H, Xue Lingzhou, Niu Xiaoyue. A review of dynamic network models with latent variables. Statist. Surv. 2018;12:105.
[13] Rastelli Riccardo, Friel Nial, Raftery Adrian E. Properties of latent variable network models. Netw Sci 2016;4(4):407–32.
[14] Hartle Harrison, Papadopoulos Fragkiskos, Krioukov Dmitri. Dynamic hidden-variable network models. Phys Rev E 2021;103(5):052307.
[15] Balogh Sámuel G, Pollner Péter, Palla Gergely. Generalised thresholding of hidden variable network models with scale-free property. Sci Rep 2019;9(1):1–10.
[16] Caldarelli Guido, Capocci Andrea, De Los Rios Paolo, Munoz Miguel A. Scale-free networks from varying vertex intrinsic fitness. Phys Rev Lett 2002;89(25):258702.
[17] Feller William. An introduction to probability theory and its applications, Volume 2, vol. 81, John Wiley & Sons; 1991.
[18] Denisov SI, Bystrik Yu S. Exact stationary solutions of the Kolmogorov–Feller equation in a bounded domain. Commun Nonlinear Sci Numer Simul 2019;74:248–59.
[19] SafeGraph. 2019, Free sample of neighborhood patterns data, retrieved on 2019. https://www.safegraph.com/.
[20] Song Chaoming, Koren Tal, Wang Pu, Barabási Albert-László. Modelling the scaling properties of human mobility. Nat Phys 2010;6(10):818–23.

[21] Fabio DR, Fabio D, Carlo P. Profiling core-periphery network structure by random walkers. Sci Rep 2013;3:1467.

[22] Schläpfer Markus, Dong Lei, O'Keeffe Kevin, Santi Paolo, Szell Michael, Salat Hadrien, Anklesaria Samuel, Vazifeh Mohammad, Ratti Carlo, West Geoffrey B. The universal visitation law of human mobility. Nature 2021;593(7860):522–7.

[23] Nie Wei-Peng, Zhao Zhi-Dan, Cai Shi-Min, Zhou Tao. Understanding the urban mobility community by taxi travel trajectory. Commun Nonlinear Sci Numer Simul 2021;101:105863.

[24] Schneider Morton. Gravity models and trip distribution theory. Pap Reg Sci 1959;5(1):51–6.

[25] Masucci A Paolo, Serras Joan, Johansson Anders, Batty Michael. Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. Phys Rev E 2013;88(2):022812.

[26] Simini Filippo, González Marta C, Maritan Amos, Barabási Albert-László. A universal model for mobility and migration patterns. Nature 2012;484(7392):96–100.

[27] Kang Chaogui, Liu Yu, Guo Diansheng, Qin Kun. A generalized radiation model for human mobility: spatial scale, searching direction and trip constraint. PLoS One 2015;10(11):e0143500.

[28] Ren Yihui, Ercsey-Ravasz Mária, Wang Pu, González Marta C, Toroczkai Zoltán. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. Nat Commun 2014;5(1):5347.

[29] Stouffer Samuel A. Intervening opportunities: a theory relating mobility and distance. Am Sociol Rev 1940;5(6):845–67.

[30] Liu Er-Jian, Yan Xiao-Yong. A universal opportunity model for human mobility. Sci Rep 2020;10(1):4657.

[31] Le Néchet Florent. Urban spatial structure, daily mobility and energy consumption: a study of 34 European cities. Cybergeo: Eur J Geogr 2012.

[32] Wang Anqi, Zhang Anshu, Chan Edwin HW, Shi Wenzhong, Zhou Xiaolin, Liu Zhewei. A review of human mobility research based on big data and its implication for smart city development. ISPRS Int J Geo-Inf 2020;10(1):13.

[33] Bassolas Aleix, Barbosa-Filho Hugo, Dickinson Brian, Dotiwalla Xerxes, Eastham Paul, Gallotti Riccardo, Ghoshal Gourab, Gipson Bryant, Hazarie Surendra A, Kautz Henry, et al. Hierarchical organization of urban mobility and its connection with city livability. Nat Commun 2019;10(1):4817.

[34] Greenwald Michael J. The relationship between land use and intrazonal trip making behaviors: Evidence and implications. Transp Res D 2006;11(6):432–46.

[35] Lee Minjin, Holme Petter. Relating land use and human intra-city mobility. PLoS One 2015;10(10):e0140152.

[36] Akiba Yuri, Yamasaki Masaya, Shima Hiroyuki, Sato Motohiro. Correlation analysis between land use and urban street patterns. 2022, Available at SSRN 4111251.

[37] Fouquet Roger. Trends in income and price elasticities of transport demand (1850–2010). Energy Policy 2012;50:62–71.

[38] Le Vine Scott, Chen Bingqing Emily, Polak John. Does the income elasticity of road traffic depend on the source of income? Transp Res A 2014;67:15–29.

[39] Litman Todd Alexander. Understanding transport demands and elasticities-how prices and other factors affect travel behavior. 2021.

[40] Ghoddusi Hamed, Rodivilov Alexander, Roy Mandira. Income elasticity of demand versus consumption: Implications for energy policy analysis. Energy Econ 2021;95:105009.

[41] Keplergl. TIGER/Line shapefiles. 2021, data retrieved by Esri, https://www.arcgis.com/apps/mapviewer/index.html?url= https://services5.arcgis.com/GfwWNkhOj9bNBqoJ/ArcGIS/rest/services/NYC_Census_Blocks_for_2010_US_census_Water_Included/FeatureServer/0.

[42] Zhong Chen, Schläpfer Markus, Müller Arisona Stefan, Batty Michael, Ratti Carlo, Schmitt Gerhard. Revealing centrality in the spatial structure of cities from human activity patterns. Urban Stud 2017;54(2):437–55.

[43] Janson Svante, Rucinski Andrzej, Luczak Tomasz. Random graphs. John Wiley & Sons; 2011.

[44] Bollobás Béla. Random graphs. In: Modern graph theory. Springer; 1998, p. 215–52.

[45] Ben-Eliezer Omri, Hefetz Dan, Kronenberg Gal, Parczyk Olaf, Shikhelman Clara, Stojaković Miloš. Semi-random graph process. Random Structures Algorithms 2020;56(3):648–75.

[46] Hanneke Steve, Fu Wenjie, Xing Eric P. Discrete temporal models of social networks. Electron J Statist 2010;4:585–605.

[47] Zhang Xiao, Moore Cristopher, Newman Mark EJ. Random graph models for dynamic networks. Eur Phys J B 2017;90(10):1–14.

[48] Boguná Marián, Pastor-Satorras Romualdo. Class of correlated random networks with hidden variables. Phys Rev E 2003;68(3):036112.

[49] Hoppe K, Rodgers GJ. Mutual selection in time-varying networks. Phys Rev E 2013;88(4):042804.

[50] Barrat Alain, Barthelemy Marc, Vespignani Alessandro. Dynamical processes on complex networks. Cambridge University Press; 2008.

[51] Geography Division US Census Bureau. Centers of population computation for the United States 1950 - 2020. Washington, DC: US Department of Commerce; 2021.

[52] Privault Nicolas. CRC Press; 2022.

[53] Tankov Peter. Financial modelling with jump processes. Chapman and HallCRC; 2003.

[54] Armenter Roc, Koren Miklós. A balls-and-bins model of trade. Amer Econ Rev 2014;104(7):2127–51.

[55] Riccaboni Massimo, Rossi Alessandro, Schiavo Stefano. Global networks of trade and bits. J Econ Interact Coord 2013;8:33–56.

[56] Casiraghi Giona, Nanumyan Vahan. Configuration models as an urn problem. Sci Rep 2021;11(1):13416.

[57] Latora Vito, Nicosia Vincenzo, Russo Giovanni. Complex networks: principles, methods and applications. Cambridge University Press; 2017.

[58] Serrano M Angeles, Boguná Marian, Pastor-Satorras Romualdo, Vespignani Alessandro. Correlations in complex networks. In: Large scale structure and dynamics of complex networks: From information technology to finance and natural sciences. 2007, p. 35–66.

[59] Murase Yohsuke, Jo Hang-Hyun, Török János, Kertész János, Kaski Kimmo. Sampling networks by nodal attributes. Phys Rev E 2019;99(5):052304.

[60] Fardet Tanguy, Levina Anna. Weighted directed clustering: Interpretations and requirements for heterogeneous, inferred, and measured networks. Phys Rev Res 2021;3(4):043124.

[61] Fagiolo Giorgio. Clustering in complex directed networks. Phys Rev E 2007;76(2):026107.

[62] Boguná Marián, Pastor-Satorras Romualdo. Epidemic spreading in correlated complex networks. Phys Rev E 2002;66(4):047104.

[63] Jiang Bin, Duan Yingying, Lu Feng, Yang Tinghong, Zhao Jing. Topological structure of urban street networks from the perspective of degree correlations. Environ Plan B: Plann Des 2014;41(5):813–28.

[64] Chung Fan, Lu Linyuan. Connected components in random graphs with given expected degree sequences. Ann Combin 2002;6(2):125–45.

[65] Chung Fan, Lu Linyuan. The average distances in random graphs with given expected degrees. Proc Natl Acad Sci 2002;99(25):15879–82.

[66] Servedio Vito DP, Caldarelli Guido, Buttà Paolo. Vertex intrinsic fitness: How to produce arbitrary scale-free networks. Phys Rev E 2004;70(5):056126.

[67] Masuda Naoki, Miwa Hiroyoshi, Konno Norio. Analysis of scale-free networks based on a threshold graph with intrinsic vertex weights. Phys Rev E 2004;70(3):036124.

[68] Fujihara A, Uchida M, Miwa H. Universal power laws in the threshold network model: A theoretical analysis based on extreme value theory. Phys A 2010;389(5):1124–30.

[69] Di Gangi Domenico, Bormetti Giacomo, Lillo Fabrizio. Score driven generalized fitness model for sparse and weighted temporal networks. 2022, arXiv preprint arXiv:2202.09854.

[70] Van Der Hofstad Remco, Janssen AJEM, Van Leeuwaarden Johan SH, Stegehuis Clara. Local clustering in scale-free networks with hidden variables. Phys Rev E 2017;95(2):022307.

[71] Kaiser Marcus. Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks. New J Phys 2008;10(8):083042.

[72] Hong Inho, Jung Woo-Sung, Jo Hang-Hyun. Gravity model explained by the radiation model on a population landscape. PLoS One 2019;14(6):e0218028.

[73] Sallah Kankoé, Giorgi Roch, Bengtsson Linus, Lu Xin, Wetter Erik, Adrien Paul, Rebaudet Stanislas, Piarroux Renaud, Gaudart Jean. Mathematical models for predicting human mobility in the context of infectious disease spread: introducing the impedance model. Int J Health Geogr 2017;16:1–11.

[74] Martin David, Gale Christopher, Cockings Samantha, Harfoot Andrew. Origin-destination geodemographics for analysis of travel to work flows. Comput Environ Urban Syst 2018;67:68–79.

[75] Rodrigue Jean-Paul. The geography of transport systems. Routledge; 2020.

[76] Kang Yuhao, Gao Song, Liang Yunlei, Li Mingxiao, Rao Jinmeng, Kruse Jake. Multiscale dynamic human mobility flow dataset in the US during the COVID-19 epidemic. Sci Data 2020;7(1):1–13.

[77] Scheaffer Richard L, Mendenhall III William, Ott R Lyman, Gerow Kenneth G. Elementary survey sampling. Cengage Learning; 2011.

[78] Census bureau fips code. Understanding Geographic Identifiers (GEOIDs). https://www.census.gov/programs-surveys/geography/guidance/geo-identifiers.html.

[79] SafeGraph. What about bias in the SafeGraph dataset? https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset.

[80] Li Zhenlong, Ning Huan, Jing Fengrui, Lessani M Naser. Understanding the bias of mobile location data across spatial scales and over time: a comprehensive analysis of SafeGraph data in the united states. PLoS One 2024;19(1):e0294430.

[81] U.S. Census Bureau Survey. American Community Survey Data population for census block group, ACS 1-year estimates. https://www.census.gov/data/developers/data-sets/acs-5year.html.

[82] Keplergl. New York City drop-off car rides. 2016, data retrieved from Kepler.gl data, https://github.com/uber-web/kepler.gl-data.

[83] U.S. Census Bureau TIGER. TIGER/Line Files and Shapefiles geographical data. https://www.census.gov/geo/maps-data/data/tiger-line.html.

[84] Vanni Fabio. Incremental formation of scale-free fitness networks. 2021, arXiv preprint arXiv:2106.02168.

[85] Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. Proc Natl Acad Sci 2004;101(11):3747–52. http://dx.doi.org/10.1073/pnas.0400087101, URL https://www.pnas.org/doi/abs/10.1073/pnas.0400087101.
[86] Clemente Gian Paolo, Grassi Rosanna. Directed clustering in weighted networks: A new perspective. Chaos Solitons Fractals 2018;107:26–38.
[87] Clemente Gian Paolo, Grassi Rosanna. DirectedClustering: A package for computing clustering coefficient in weighted and directed networks. 2018.
[88] Shi Shuyang, Wang Lin, Xu Shuangdie, Wang Xiaofan. Prediction of intra-urban human mobility by integrating regional functions and trip intentions. IEEE Trans Knowl Data Eng 2020;34(10):4972–81.
[89] Wagner Peter, Wegener Michael. Urban land use, transport and environment models: Experiences with an integrated microscopic approach. DisP Plan Rev 2007;43(170):45–56.
[90] Acheampong Ransford A, Silva Elisabete A. Land use–transport interaction modeling: A review of the literature and future research directions. J Transp Land use 2015;8(3):11–38.
[91] NewYork open data. Primary Land Use Tax Lot Output (PLUTO). https://data.cityofnewyork.us/City-Government/Primary-Land-Use-Tax-Lot-Output-PLUTO-/64uk-42ks.
[92] Gardiner Crispin. Stochastic methods, vol. 4, Springer Berlin; 2009.
[93] McCauley Joseph L. Stochastic calculus and differential equations for physics and finance. Cambridge University Press; 2013.
[94] Mathieu Edouard, Ritchie Hannah, Rodés-Guirao Lucas, Appel Cameron, Giattino Charlie, Hasell Joe, Macdonald Bobbie, Dattani Saloni, Beltekian Diana, Ortiz-Ospina Esteban, Roser Max. Coronavirus pandemic (COVID-19). Our World Data 2020. https://ourworldindata.org/coronavirus.
[95] Vanni Fabio, Lambert David, Palatella Luigi, Grigolini Paolo. On the use of aggregated human mobility data to estimate the reproduction number. Sci Rep 2021;11(1):23286.
[96] Bonaccorsi Giovanni, Pierri Francesco, Cinelli Matteo, Flori Andrea, Galeazzi Alessandro, Porcelli Francesco, Schmidt Ana Lucia, Valensise Carlo Michele, Scala Antonio, Quattrociocchi Walter, et al. Economic and social consequences of human mobility restrictions under COVID-19. Proc Natl Acad Sci 2020;117(27):15530–5.
[97] Galeazzi Alessandro, Cinelli Matteo, Bonaccorsi Giovanni, Pierri Francesco, Schmidt Ana Lucia, Scala Antonio, Pammolli Fabio, Quattrociocchi Walter. Human mobility in response to COVID-19 in France, Italy and UK. Sci Rep 2021;11(1):13141.
[98] Liang Xiao, Zhao Jichang, Dong Li, Xu Ke. Unraveling the origin of exponential law in intra-urban human mobility. Sci Rep 2013;3(1):2983.
[99] Gallotti Riccardo, Bazzani Armando, Rambaldi Sandro, Barthelemy Marc. A stochastic model of randomly accelerated walkers for human mobility. Nat Commun 2016;7(1):12600.
[100] Van Acker Veronique, Witlox Frank, Van Wee Bert. The effects of the land use system on travel behavior: a structural equation modeling approach. Transp Plan Technol 2007;30(4):331–53.
[101] Geurs Karst T, Van Wee Bert. Accessibility evaluation of land-use and transport strategies: review and research directions. J Transp Geogr 2004;12(2):127–40.
[102] Felbermayr Gabriel J, Tarasov Alexander. Trade and the spatial distribution of transport infrastructure. J Urban Econom 2022;130:103473.
[103] Carra Giulia, Mulalic Ismir, Fosgerau Mogens, Barthelemy Marc. Modelling the relation between income and commuting distance. J R Soc Interface 2016;13(119):20160306.
[104] Mayeres Inge. The efficiency effects of transport policies in the presence of externalities and distortionary taxes. J Transp Econom Policy 2000;233–59.
[105] Goodwin Phil, Dargay Joyce, Hanly Mark. Elasticities of road traffic and fuel consumption with respect to price and income: a review. Transp Rev 2004;24(3):275–92.
[106] Börjesson Maria, Fosgerau Mogens, Algers Staffan. On the income elasticity of the value of travel time. Transp Res A 2012;46(2):368–77.
[107] Wang Yinhai, Zeng Ziqiang. Data-driven solutions to transportation problems. Elsevier; 2018.
[108] Oliveira Vítor. Urban morphology: an introduction to the study of the physical form of cities. Springer; 2016.
[109] O'Sullivan A. Urban Economics. McGraw-Hill series in economics, McGraw-Hill Education; 2018, URL https://books.google.it/books?id=_bjcnAEACAAJ.
[110] U.S. Census Bureau Survey. American Community Survey Data at tlevel of ceneus block groups, B08303 TRAVEL TIME TO WORK, ACS 1-year estimates. https://data.census.gov/table?q=travel+time&tid=ACSDT1Y2019.B08303.
[111] Fabio Vanni. Project title. 2023, https://github.com/fxfabio/tripmobnet.
[112] Berberan-Santos Mário N. Computation of one-sided probability density functions from their cumulants. J Math Chem 2007;41(1):71–7.
[113] Everitt Brian. Finite mixture distributions. Springer Science & Business Media; 2013.
[114] Sundt Bjørn, Vernic Raluca. Recursions for convolutions and compound distributions with insurance applications. Springer Science & Business Media; 2009.
[115] Tulsyan Aditya, Gopaluni R Bhushan, Khare Swanand R. Particle filtering without tears: A primer for beginners. Comput Chem Eng 2016;95:130–45.
[116] Handschin Johannes Edmund, Mayne David Q. Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. Int J Control 1969;9(5):547–59.
[117] Stegehuis Clara, van der Hofstad Remco, Janssen AJEM, van Leeuwaarden Johan SH. Clustering spectrum of scale-free networks. Phys Rev E 2017;96(4):042309.
[118] Boguná Marián, Serrano M Ángeles. Generalized percolation in random directed networks. Phys Rev E 2005;72(1):016106.
[119] Serrano M Ángeles, Boguñá Marián, Pastor-Satorras Romualdo. Correlations in weighted networks. Phys Rev E 2006;74(5):055101.
[120] Dorogovtsev SN. Clustering of correlated networks. Phys Rev E 2004;69(2):027104.
[121] Jang Jiwook, Oh Rosy. A review on Poisson, Cox, Hawkes, shot-noise Poisson and dynamic contagion process and their compound processes. Ann Actuar Sci 2021;15(3):623–44.
[122] Privault Nicolas. Stochastic finance: An introduction with market examples. CRC Press; 2013.
[123] Willmot Gord. Mixed compound Poisson distributions. ASTIN Bull J IAA 1986;16(S1):S59–79.
[124] van der Hofstad Remco. Critical behavior in inhomogeneous random graphs. Random Structures Algorithms 2013;42(4):480–508.
[125] Scalas Enrico. A class of CTRWs: compound fractional Poisson processes. In: Fractional dynamics: recent advances. World Scientific; 2012, p. 353–74.