Università
Ca'Foscari
Venezia

ANNO ACCADEMICO / ACADEMIC YEAR

2024 / 2025

Corso di Dottorato di Ricerca

in Science and Technology of Bio and Nanomaterials

ciclo 38

Analisi di sistemi di reti nella citta' e loro ottimizzazione per la sostenibilita'

SSD: FIS/02

Coordinatore del Dottorato
ch. prof. Flavio Rizzolio

Supervisore
ch. prof. Guido Caldarelli

Co-Supervisore
ch. prof. Stefano Mancuso

Dottorando / Dottoranda
Jacopo Moi
Matricola 847329

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

«Fa molto più rumore un albero che cade che una foresta che cresce.»

— Lao Tzu

Contents

List of Figures

Abstract

Cities concentrate population, infrastructure, and economic activity, generating structural vulnerabilities that include environmental stress, spatial inequalities, and pressure on public services. Understanding how urban form shapes social and ecological outcomes requires quantitative tools grounded in complexity science and network theory, framing cities as spatially embedded systems with emergent collective behavior. This doctoral thesis develops a quantitative framework for the analysis of urban accessibility to public green areas, grounded in statistical physics and network science. Green infrastructure contributes to environmental regulation, public health, and social cohesion, yet access to it is not uniformly distributed and depends on the interplay between spatial topology, demographic heterogeneity, and behavioral competition. The conceptual background draws on the multilayer network perspective on smart and resilient cities discussed in a review article published in Nature Cities, where urban systems are framed as interdependent networked structures. While the present thesis focuses on a single infrastructural layer — the pedestrian street network and its interaction with green areas — this broader systemic view motivates the adoption of physics-based tools capable of capturing global urban properties beyond traditional proximity indicators. The methodological contribution combines probabilistic path ensembles with agent-based dynamics. Accessibility is first quantified through the bag-of-paths formalism, which assigns Boltzmann-weighted probabilities to low-cost trajectories between urban blocks and green areas, yielding a structural measure of average farness. This metric is then extended through a biased random walk model in which agents compete for finite-capacity green spaces, producing a behavioral measure of realized accessibility. The comparison between the two quantities reveals how congestion and demographic pressure reshape the effective accessibility landscape: structural proximity alone can overestimate the accessibility of high-density areas, where competition for limited green capacity degrades realized access. The framework is applied to the city of Mestre as a case study. Census data at block level and empirical pedestrian street networks are integrated to compute a composite accessibility score combining structural exposure with nonlinear population weighting. Temporal comparisons across census years show that peripheral blocks with growing population density experience the largest accessibility losses, while network bottlenecks amplify spatial inequalities beyond what Euclidean distance would predict. Congestion effects are found to disproportionately penalize areas that appear well-served under purely topological metrics, revealing a systematic gap between nominal and effective green access. These results establish a reproducible, scalable methodology for accessibility assessment and provide an operational basis for evidence-based urban planning and sustainability policy.

Chapter 1

Introduction

1.1 Why a Science of Cities is Needed

The global shift toward urban living is one of the defining transformations of the contemporary era. Estimates consistently indicate that the majority of the world's population will reside in urban regions within the coming decades, with continuous expansion at the metropolitan and peri-urban scale (Ritchie et al., 2023; United Nations Department of Economic and Social Affairs, 2023). This demographic movement is not simply a matter of growth; it changes the material structure of cities and the dynamics of the populations living within them. As cities enlarge and densify, the pressures on transport systems, ecological resources, and social infrastructures become increasingly evident. These observations reinforce a structural paradox: even as available free space shrinks, urban populations require more spatial resources to cope with emerging pressures and maintain acceptable living conditions.

Urban expansion exposes structural vulnerabilities that traditional sectoral approaches fail to address. Climate dynamics further accentuate these weaknesses: intensifying heatwaves, more frequent flooding, and the steady sealing of soil show how the urban fabric amplifies environmental stress rather than absorbing it (Kovats and Kristie, 2006; Nicholls, 1995). At the same time, spatial inequalities do not attenuate with growth; they reorganize. The mechanisms of residential sorting and segregation described in early formal models (Schelling, 1971a) continue to operate, producing contemporary patterns in which access to services, green areas, and mobility remains uneven and systematically stratified (Clark and Fossett, 2008).

These patterns indicate that cities cannot be interpreted as linear systems in which local interventions produce proportional outcomes. Instead, they behave as interdependent structures in which failures in one component propagate across multiple layers. This phenomenon is especially clear in infrastructural networks: transport, energy supply, and communication systems are so tightly coupled that perturbations often trigger

cascading effects (Buldyrev et al., 2010; Bashan et al., 2013). A problem arising in one subsystem rarely remains isolated; it affects the entire urban metabolism.

Understanding and governing cities therefore requires a scientific perspective capable of describing such interconnected dynamics. Several empirical regularities reinforce this point. Urban indicators do not scale linearly with population size. Infrastructure networks tend to grow sublinearly, whereas socio-economic outputs such as innovation or social interactions scale superlinearly (Bettencourt et al., 2007a, 2010). These nontrivial relations indicate that the macroscopic behaviour of cities emerges from their internal interaction structure rather than from the properties of isolated components. They also show that growth, density, and spatial concentration create feedback loops that shape both opportunities and vulnerabilities.

The attempt to understand cities through systematic and analytical principles is not new. Throughout history, urban settlements have been interpreted in ways that reflect the scientific and cultural frameworks of their time. Classical authors already recognised that cities possess distinctive collective behaviours: Aristotle described the polis as a form of association that exceeds the individuals composing it,while Roman urbanism produced spatial layouts whose underlying regularities—exemplified by the orthogonal structure of the cardo and decumanus—remain recognisable in many Italian cities today.

A shift occurred between the nineteenth and early twentieth centuries, when rapid industrialisation transformed cities into unprecedented concentrations of population and infrastructure. Scholars began to treat the city as an empirical object whose properties could be measured and compared. The work of the Chicago School introduced early forms of spatial analysis, linking urban structure to social organisation and recognising that proximity, density, and mobility shape patterns of interaction. At the same time, geographers and regional scientists laid the foundations for a more formal interpretation of spatial organisation, from central place theory to the study of urban hierarchies.

Yet only in the second half of the twentieth century did cities begin to be treated explicitly as complex systems. The rise of quantitative geography, network science, and statistical physics introduced tools capable of analysing large, heterogeneous systems composed of many interacting units. Urban phenomena such as segregation, congestion, and spatial clustering were reinterpreted as collective patterns arising from simple local rules rather than from individual choices alone. This marked a conceptual shift: the city became understood as a system shaped by interactions, feedbacks, and self-organisation.

The contemporary science of cities builds on this historical trajectory. Advances in data availability, computational modelling, and network theory have made it possible to describe urban systems with a level of precision that was previously unattainable. The city can now be analysed as a multilevel structure whose spatial form, infrastructural networks, and population dynamics coevolve. This perspective aligns urban studies with the broader field of complexity science, which provides a unifying framework for understanding how macroscopic behaviours arise from microscopic interactions.

Complexity science offers a coherent framework for interpreting these observations. In recent years, the integration of statistical physics, network theory, and data-driven analyses has provided a foundation for describing cities as multilevel systems characterized by heterogeneity, feedbacks, and emergence (Barthelemy, 2019a; Caldarelli et al., 2023). This perspective does not treat the city as a collection of independent subsystems but as a coupled structure in which form, function, and population dynamics interact. The same approach is now central in the study of smart cities and urban digital infrastructures, where the interplay between technology, behaviour, and spatial constraints becomes increasingly relevant (Caldarelli et al., 2025a).

The need for a science of cities thus stems from practical considerations as much as from theoretical ones. Without quantitative tools that describe interactions, spatial structure, and collective dynamics, it is not possible to assess urban resilience, to address inequalities in access to resources, or to anticipate how population shifts modify the pressure on infrastructures.

1.2 Cities as Complex Systems

Understanding cities as complex systems offers a coherent way to interpret the patterns and tensions outlined earlier. An urban environment cannot be reduced to the sum of its components: it is a configuration in which people, infrastructures, and spatial form evolve together. Their interactions create dependencies that extend across multiple scales—local decisions shape metropolitan flows, while regional dynamics reshape neighbourhood conditions. The result is a system whose behaviour cannot be deduced by examining individual elements in isolation.

A defining feature of cities is their internal variety. Urban space exhibits contrasts that arise from historical trajectories, socio-economic differentiation, and the evolving morphology of the built environment. These contrasts are not static; they shift as populations relocate, economic activities reorganise, and infrastructures adapt. Through this continuous adjustment, local transformations produce cumulative effects that reorganise the urban structure at larger scales. Macroscopic patterns therefore emerge from interactions, not from predetermined design.

Statistical physics has proved particularly useful in describing such phenomena. Originally developed to study systems composed of many interacting particles, it provides concepts and tools that translate naturally to the urban context. Cities display the same key traits—large numbers of interacting units, nonlinearity, collective effects, and the emergence of global patterns from local rules—that make statistical physics an appropriate analytical framework (Barthelemy, 2019a). Models drawn from this tradition help explain how aggregation, competition, diffusion, and spatial constraints generate the observed distribution of functions and densities. In this perspective, the city becomes a macroscopic outcome of microscopic decisions, much like thermodynamic properties

arise from molecular interactions.

Network theory represents a central bridge between statistical physics and urban analysis. It recasts physical, social, and infrastructural relations into a unified representation, where nodes and edges encode the channels through which interactions occur. This approach allows one to analyse the structural organisation of the city using the same formal tools developed for complex systems: connectivity distributions, centrality measures, propagation dynamics, and robustness properties (Caldarelli et al., 2023). Because urban processes unfold along these relational structures—mobility along streets, energy through grids, information across communication layers—the network view provides direct access to the mechanisms that shape accessibility, resilience, and spatial inequality.

Crucially, urban networks are embedded in physical space. Geometric constraints, construction costs, and the morphology of neighbourhoods impose limits on how connections form and how movement unfolds. Spatial embedding reduces the set of feasible configurations and generates characteristic patterns of accessibility and isolation. It also creates dependencies between topology and geography: the organisation of the street network influences how people move, how services are reached, and how resources circulate through the urban fabric. Analysing these relations is essential to understand why some areas remain structurally advantaged while others persist at the margins.

This systemic perspective has become even more relevant as cities incorporate digital infrastructures and data-rich technologies. The interaction between human behaviour, sensing devices, and automated processes creates new couplings between physical and digital layers of urban life (Caldarelli et al., 2025a). Yet the availability of high-frequency data, in itself, does not clarify how cities function. Only when interpreted through the lens of complexity science—where statistical physics, network theory, and urban analytics converge—do these data reveal the underlying mechanisms governing collective dynamics.

Recognising cities as complex systems thus provides the conceptual basis for studying how local modifications propagate across the urban structure. It also motivates the quantitative tools adopted in this thesis, which focus on the interplay between spatial organisation, accessibility, and collective behaviour.

1.3 Scope of this Thesis

The conceptual framework developed so far provides the background for the research questions addressed in this thesis. If cities are understood as complex systems—shaped by interactions rather than by isolated components—then accessibility emerges as one of the domains where these interactions become visible and measurable. Accessibility to public resources, and in particular to green areas, offers a clear illustration of how spatial structure, demographic distribution, and behavioural dynamics intertwine. Green spaces

are not only environmental infrastructures; they reflect inequalities in urban form, the organisation of mobility, and the capacity of residents to adapt to spatial constraints. For these reasons, they provide a concrete setting in which to study how collective behaviours unfold on the underlying urban network.

The focus on green areas has an additional significance. Among all urban amenities, they represent a resource whose value depends on both physical proximity and actual usability. A park that is theoretically close may become practically unreachable due to the configuration of the street network or the concentration of population around it. Conversely, areas that appear peripheral may function as important anchors of accessibility once the relational structure of the city is taken into account. This dual nature—geometric and collective—makes green areas an appropriate testbed for investigating how complex spatial systems operate.

This thesis develops a quantitative perspective on these issues through three research papers that represent distinct steps of a unified argument. The first contribution situates the problem within the broader context of complexity science and urban studies. It discusses why traditional approaches to accessibility, often based on simple distance metrics or administrative thresholds, fail to capture the way in which urban structure shapes everyday spatial practices. This initial step clarifies the necessity of viewing accessibility as an emergent outcome of interactions occurring on networks, where topology, spatial embedding, and behavioural tendencies jointly determine how people use the city. By framing the problem in these terms, the first paper establishes the theoretical motivation for adopting tools inspired by statistical physics and network science.

The second paper moves from this conceptual ground to an explicit formulation. It introduces a diffusion-based perspective for analysing accessibility to public green areas using real street networks. The aim is to translate the idea of collective movement into a framework that captures how spatial organisation directs flows, how paths compete, and how local geometries influence large-scale patterns. This contribution demonstrates that accessibility cannot be inferred from shortest-path distances alone; it is affected by the multiplicity of possible routes, by the degree of connectivity of each area, and by the overall structure of the urban network. By applying this framework to a concrete case study, the paper shows how spatial networks impose constraints that are not immediately visible through traditional measures.

The third paper extends the analysis by introducing the demographic dimension. Cities evolve not only because their infrastructures change but also because the distribution of their inhabitants shifts over time. This population dynamics modifies the pressure exerted on public resources and alters the effective accessibility of different areas. The third contribution analyses how accessibility patterns respond to demographic transformations, and develops a synthetic score that integrates population distribution with structural properties of the network. By comparing different temporal configura-

tions, it becomes possible to identify stable structural inequalities, to detect areas where accessibility is particularly sensitive to demographic change, and to assess whether improvements in infrastructure or population redistribution translate into more balanced spatial conditions.

Taken together, the three papers advance the argument that accessibility must be interpreted as a systemic property. It is not the result of a single factor—neither geometric proximity nor infrastructural design nor demographic load—but emerges from the continuous interaction among these dimensions. The thesis therefore proposes a perspective in which the analysis of green-area accessibility parallels the broader study of cities as complex systems. Rather than offering a catalogue of indicators, it aims to show how the organisation of the city constrains the everyday possibilities of its inhabitants and how these possibilities change as the system evolves.

The structure of the manuscript reflects this progression. Following the introduction, each paper is presented as a chapter accompanied by a short preamble explaining how it fits into the overall trajectory of the thesis. This is followed by a concluding chapter that integrates the findings, discusses their implications for urban analysis, and outlines possible directions for future developments. In this way, the thesis offers both a set of specific results and a conceptual standpoint from which accessibility and spatial inequalities can be analysed within the wider scientific effort to understand cities through the lens of complexity.

Chapter 2

Tools

2.1 From networks to spatial networks

As introduced, complex systems are characterized by the unfathomability of correlational phenomena: the dynamics that emerge from the interaction of multiple elements cannot be explained by analyzing the single parts in isolation. This implies the need for a method that can take into account quantitatively the system coordination into its fundamental actors and codify the relationships among them through symbols and formal structures.

In this regard, complex networks play a central role: they provide a mathematical and conceptual language capable of representing the interactions between heterogeneous components, highlighting both structural patterns and emergent properties that would otherwise remain hidden. The network representation is particularly powerful because the structure that emerges is not imposed from outside, but arises from the aggregation of many pieces of information about the system, distilled into nodes and links.

In this chapter, we begin from the general framework of network theory—its definitions, measures, and structural models—and move toward the specific case of *spatial networks*, where topology and geometry are intertwined and physical constraints shape the structure of the system (Barthelemy, 2011, 2018; Newman, 2010; Barabási, 2016; Watts and Strogatz, 1998; Barabási and Albert, 1999).

2.1.1 Basic Concepts in Networks

A network, or graph, can be denoted as $G = (V, E)$, where V is the set of $N = |V|$ nodes (or vertices) and E the set of $M = |E|$ edges (or links). Edges may be undirected, when connections are symmetric, or directed, when orientation matters. They may also be unweighted, indicating only the existence of a relation, or weighted, when each link is associated with a cost or length $w_{ij} \in \mathbb{R}^+$.

A graph can be represented algebraically by its *adjacency matrix* $A = (a_{ij})$, where

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

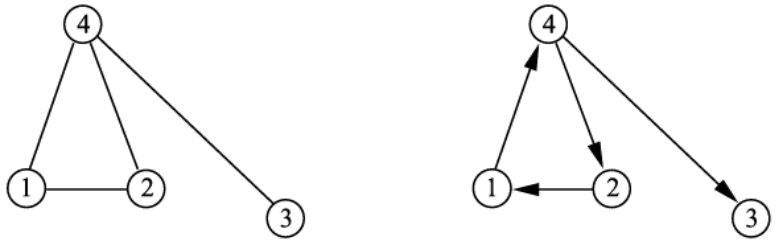


Figure 2.1: Two simple graphs and their adjacency matrices.

$a_{ij} = 1$ if $(i, j) \in E$ and $a_{ij} = 0$ otherwise (or $a_{ij} = w_{ij}$ in the weighted case). The degree of a node i is then $k_i = \sum_j a_{ij}$, with in-degree and out-degree distinguished in directed graphs.

Connectivity is described through sequences of nodes. A *walk* of length ℓ is a sequence $(v_0, v_1, \ldots, v_\ell)$ such that each pair $(v_{m-1}, v_m) \in E$. A *path* is a walk in which all vertices are distinct. The *shortest path distance*, or geodesic, between two nodes i and j is defined as

$$d(i, j) = \min_{P \in \mathcal{P}_{ij}} \ell(P),$$

where $\mathcal{P}_{ij}$ is the set of all paths from i to j. This metric is fundamental because many structural properties of networks are derived from it.

A graph is said to be connected if for all pairs of nodes $i, j \in V$ there exists at least one path with finite $d(i, j)$. Otherwise, it splits into disconnected components. Another important object is the graph Laplacian $L = D - A$, where D is the diagonal matrix of degrees ($D_{ii} = k_i$). The Laplacian plays a key role in processes such as diffusion and random walks on networks.

2.1.2 Topological Measures

Building on these definitions, one of the most important ideas in network theory is that of *centrality*. Centrality measures provide a way to quantify how relevant a node is in the structure of the graph, depending on its position relative to others. Different notions of centrality capture different aspects of influence or accessibility.

A first definition is the *betweenness centrality*, which measures how frequently a node lies on the shortest paths connecting other nodes. If σ_{st} denotes the total number of shortest paths between s and t, and $\sigma_{st}(i)$ the number of those paths that pass through

i, then the betweenness of node i is

$$B(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}.$$

Nodes with high $B(i)$ act as intermediaries, controlling the flow between different regions of the network. In a city, intersections or bridges with high betweenness are often critical points: their closure can fragment the urban fabric or create severe congestion.

Another way to evaluate centrality is through distances. The *farness* of node i is defined as

$$F(i) = \sum_{j \in V} d(i, j),$$

that is, the total distance from i to all other nodes. A node with low farness is, on average, close to the rest of the network, while a node with high farness is more peripheral. This notion captures the degree of isolation of an element: in a city, neighborhoods at the outskirts typically display high farness, reflecting their limited accessibility compared to central districts.

The inverse of farness defines the *closeness centrality*,

$$C(i) = \frac{1}{F(i)},$$

which highlights nodes that are on average at small distances from everywhere else. This corresponds to the classical definition introduced by Freeman. In disconnected networks, an alternative formulation known as harmonic centrality, $C_H(i) = \sum_j d(i, j)^{-1}$, is often preferred because it remains finite when some distances are infinite. In the present study, the street networks are connected , so the standard closeness definition is well defined. In urban networks, squares, central stations or hubs with high closeness are natural reference points, since they can be reached quickly from many different parts of the city.

While these measures are widely used, their application to cities also reveals some limitations. Betweenness identifies bottlenecks but ignores alternative longer routes that may still be attractive; farness and closeness quantify distance but neglect factors such as traffic, land use, or population density. In other words, classical centrality measures provide important structural information, but they do not necessarily reflect the effective accessibility experienced by inhabitants. It is precisely in this gap between topology and lived reality that more refined, stochastic and dynamic approaches become necessary.

2.1.3 Community Detection

Networks often exhibit a higher-level organization, where nodes group into communities (or modules) that are more densely connected among themselves than with the rest of

the system. Detecting such communities is a way to uncover the mesoscopic structure of a network.

A standard measure to evaluate community partitions is the *modularity* Q, defined as

$$Q = \frac{1}{2M} \sum_{i,j} \left(a_{ij} - \frac{k_i k_j}{2M} \right) \delta(c_i, c_j),$$

where c_i is the community of node i. High values of Q indicate that the partition captures a significant community structure compared to a random null model.

Among the many algorithms for community detection, the Louvain method has become a widely used standard. It operates by iteratively merging nodes into communities to maximize modularity, then constructing a new network of communities and repeating the procedure. Its efficiency and scalability make it suitable for large empirical networks.

2.1.4 Classes of Networks

Besides empirical studies, simple theoretical models have been developed to explain how common network patterns emerge. Before introducing these models, it is useful to clarify one fundamental property: *clustering*.

Clustering quantifies the tendency of neighbors of a node to be connected among themselves. Formally, the local clustering coefficient of a node i is defined as

$$C_i = \frac{2e_i}{k_i(k_i - 1)},$$

where k_i is the degree of node i and e_i is the number of links among its neighbors. A high clustering coefficient indicates the presence of tightly connected groups (triangles), a feature that is typical of social and spatial networks, including urban street systems.

The Erdős–Rényi random graph $G(N, p)$ is the simplest generative model: each edge is added independently with probability p. It generates homogeneous degree distributions and well-defined connectivity thresholds. However, it produces very low clustering for large N, failing to reproduce the strong local cohesion observed in real systems.

The Watts–Strogatz model was introduced precisely to reconcile high clustering with short characteristic path lengths. Starting from a regular lattice (which has strong local clustering) and rewiring a fraction of edges at random, the model generates networks that retain a high clustering coefficient while drastically reducing the average shortest-path length. This combination defines the *small-world* property, frequently observed in social, biological, and infrastructural systems.

Finally, the Barabási–Albert model introduced the mechanism of preferential attachment. By combining network growth with the rule that new nodes connect preferentially

to already well-connected ones, the model produces scale-free degree distributions of the form

$$P(k) \sim k^{-\gamma},$$

with the spontaneous emergence of hubs.

These models are not intended to reproduce all the details of real urban systems. Rather, they isolate fundamental mechanisms— randomness, local cohesion, and growth with reinforcement—that generate the structural patterns later observed in spatial and urban networks.

2.1.5 Spatial Networks

In many real systems, networks are not purely abstract but embedded in a physical space. A *spatial network* can be defined as a graph $G = (V, E)$ in which each node $i \in V$ is associated with a position in a metric space (for instance, coordinates in $\mathbb{R}^2$), and edges $(i, j) \in E$ carry a weight w_{ij} that often corresponds to the Euclidean distance between i and j (Caldarelli, 2007). This embedding introduces geometric constraints that strongly influence the topology of the network, differentiating it from purely abstract models.

A distinctive property of spatial networks is their approximate *planarity*: in many cases, such as road networks or power grids, edges can be drawn on a two-dimensional plane without crossings except at nodes. Planarity imposes structural restrictions—for example, the average degree of a planar graph satisfies $\langle k \rangle < 6$—and has direct implications for how connections can be arranged.

Connections are typically local: most links connect nearby nodes, while long-range edges are rare and usually costly. This results in a distribution of edge lengths concentrated on short distances and in strong correlations between topology and geometry.

Centrality measures also acquire a natural interpretation in this context. The shortest-path distance $d(i, j)$ reflects actual travel length or transport cost, and nodes with high betweenness or closeness often coincide with major junctions, crossroads, or hubs in the physical infrastructure.

Another important aspect is the trade-off between efficiency and cost. While adding long-range connections reduces typical path lengths and increases robustness, such links are often too expensive or unfeasible. As a consequence, real spatial networks tend to evolve towards sparse but well-connected topologies that balance construction costs with accessibility requirements (Caldarelli, 2007).

Spatial networks therefore constitute the natural framework for studying infrastructures, transport systems, and urban layouts, where geometry and topology are inseparably linked.

2.2 Stochastic Processes

Many physical systems evolve in a way that is not fully deterministic but influenced by randomness. The mathematical framework of stochastic processes provides the tools to describe such dynamics. This section introduces the main definitions and notations that will be employed later in the thesis (van Kampen, 2007; Newman, 2010).

2.2.1 General Definition

A *stochastic process* is a family of random variables $\{X_t\}_{t\in T}$ defined on a probability space, where t represents a discrete or continuous time index. Each X_t takes values in a state space S, and the evolution of the process is described in terms of probability distributions over S (van Kampen, 2007).

For example, consider the tossing of a fair coin. If X_t is defined as the outcome of the t-th toss, then $S = \{\mathrm{H}, \mathrm{T}\}$ and $\Pr(X_t = \mathrm{H}) = \Pr(X_t = \mathrm{T}) = 1/2$. Another classical example is Brownian motion, where X_t represents the position of a particle at time t, with distribution Gaussian of variance $\langle X_t^2 \rangle \sim t$, a prototype of diffusion in physics (van Kampen, 2007).

2.2.2 Markov Chains

A central class of stochastic processes is the *Markov chain*, characterized by the property that the future depends only on the present state, not on the full history (Norris, 1997). Formally,

$$\Pr(X_{t+1} = j \mid X_t = i, X_{t-1}, \ldots, X_0) = \Pr(X_{t+1} = j \mid X_t = i).$$

The process is encoded in a transition probability matrix $P = (p_{ij})$, with

$$p_{ij} = \Pr(X_{t+1} = j \mid X_t = i), \qquad \sum_j p_{ij} = 1.$$

2.2.3 Random Walks on Graphs

When the state space is the set of nodes of a network $G = (V, E)$, the Markov chain corresponds to a *random walk on a graph* (Newman, 2010; Masuda et al., 2017). At each step, the walker moves to one of the neighbors of the current node.

In an unweighted, undirected graph the natural choice is uniform:

$$p_{ij} = \begin{cases} \frac{1}{k_i}, & \text{if } (i,j) \in E, \\ 0, & \text{otherwise,} \end{cases}$$

where k_i is the degree of node i.

For instance, on a three-node line 1–2–3, from node 2 the walker has probability $1/2$ to move to 1 or 3, while from the extremes it deterministically returns to the center. This is the discrete analogue of diffusion on a line segment.

When the graph is weighted, each edge (i, j) is associated with a weight w_{ij} that may represent distance, capacity, or cost. The transition probabilities can then be defined as normalized weights, so that edges with larger w_{ij} are favored. This induces a *bias* in the random walk, which no longer explores the network uniformly:

$$p_{ij} = \frac{w_{ij}}{\sum_m w_{im}}.$$

2.2.4 Occupation Probability in Random Walks

Let $\pi_j(t) = \Pr(X_t = j)$ denote the probability that the walker is at node j at time t. The time evolution follows the master equation

$$\pi_j(t+1) = \sum_i \pi_i(t)\, p_{ij},$$

or, in matrix notation,

$$\pi(t+1) = \pi(t)P,$$

where P is the transition matrix.

This formalism describes how probability mass flows across the network as the random walk unfolds. At each step the distribution $\pi(t)$ is redistributed according to the local transition rules, encoding both the topology of the network and the stochastic nature of the process.

In many applications of statistical physics and network science one is interested in the long-term limit $t \to \infty$, which leads to the stationary distribution. In our case, however, the focus lies on the *transient regime*: how probability spreads through the system before being absorbed or terminated. This transient occupation distribution provides direct information on accessibility and flow patterns, and will later serve as the basis for the measures introduced in this thesis.

2.2.5 Absorbing Random Walks

A random walk becomes *absorbing* when a subset of nodes $A \subset V$ is designated as absorbing states. Once the walker reaches a node $a \in A$, the process terminates and the walker remains there forever. Formally, the transition probabilities are defined so that

$$p_{aa} = 1, \qquad p_{aj} = 0 \;\; \forall j \neq a, \quad a \in A,$$

while for non-absorbing nodes $i \notin A$ the probabilities follow the usual rule

$$p_{ij} = \frac{w_{ij}}{\sum_m w_{im}}.$$

Because probability mass is progressively absorbed, the process is non-conservative: if we restrict attention to the non-absorbing nodes $T = V \setminus A$, the total probability in T decreases with time.

The dynamics can be expressed by partitioning the transition matrix P into blocks corresponding to transient (non-absorbing) and absorbing states,

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix},$$

where Q encodes transitions among transient states, R transitions from transient to absorbing states, and I is the identity matrix for the absorbing part (Norris, 1997).

A central object is the *fundamental matrix*,

$$Z = (I - Q)^{-1} = I + Q + Q^2 + \cdots,$$

whose entry Z_{ij} gives the expected number of times the walk visits state j starting from i before being absorbed.

From Z one can derive the absorption probabilities. Let B denote the matrix

$$B = ZR,$$

then B_{ij} is the probability that a walk starting at transient node i will eventually be absorbed at absorbing node j (Masuda et al., 2017).

Absorbing random walks thus provide a natural framework to model processes with sinks, exits, or targets. They are widely used in network science to quantify the likelihood of reaching specific nodes or sets of nodes under stochastic dynamics.

2.3 Agent-Based Models

One of the most versatile approaches to the study of complex systems is provided by *agent-based models* (ABM). In this framework, the system is represented as a collection of autonomous agents, each endowed with a simple set of rules of behavior. Rather than writing down macroscopic equations of motion, the dynamics of the system are generated bottom-up through the repeated application of local rules at the level of individual agents (Bonabeau, 2002; Epstein, 2006).

A central characteristic of ABM is that the rules are often stochastic. Agents make decisions or move according to probability distributions, so that the system's evolution

is not deterministic but follows a stochastic process. Formally, the state of an agent at time t can be described by a random variable X_t, and its transition to a new state depends on conditional probabilities $\Pr(X_{t+1} = j \mid X_t = i)$. This places ABM in continuity with the theory of Markov chains and random walks discussed earlier, but extended to many interacting entities evolving in parallel.

What makes ABM powerful is the emergence of collective phenomena. Even when the rules governing individual agents are elementary—such as moving to a neighboring site with a given probability or imitating the behavior of neighbors—the simultaneous dynamics of many agents can produce unexpected macroscopic patterns. A paradigmatic example is the Schelling model of segregation (Schelling, 1971b), where agents placed on a grid follow simple stochastic relocation rules depending on the composition of their neighborhood. Despite the simplicity of the local behavior, the model produces large-scale spatial segregation, showing how strong emergent structures can arise from weak microscopic preferences.

Another important aspect of ABM is the embedding of agents in structured environments. Agents may inhabit a lattice, a continuous space, or a network, and their possible actions are constrained by this underlying geometry. When agents are placed on a graph, their motion reduces to a stochastic process on the network: at each step they choose one of the adjacent nodes according to a probability distribution, possibly biased by weights or costs. In this way, ABM directly connect with the theory of stochastic processes, while adding the richness of multiple, interacting walkers.

From a methodological perspective, ABM are flexible and transparent. They allow heterogeneity among agents, incorporate randomness in a natural way, and make it possible to simulate phenomena that would be intractable with analytical models. At the same time, their stochastic nature implies that simulations must often be repeated many times to obtain reliable averages, and results may be sensitive to the specific rules chosen. For this reason, calibration and validation against empirical data are crucial when applying ABM to real-world systems.

Despite these caveats, agent-based modeling has become a standard tool in statistical physics, computational social science, and urban studies. It provides a language to explore how local stochastic interactions can give rise to global structures, making it especially suitable for the study of accessibility, mobility, and resource distribution in cities.

Chapter 3

Cities and Complexity: A Perspective from Network Science

This chapter elaborates on the perspective article published in Nature Cities (Caldarelli et al., 2025b), to which the candidate contributed by performing the analysis of the state of the art and its synthesis. The goal is to show how concepts and methods from complexity science and network theory can inform the study and design of urban systems. While the published article offered a concise viewpoint, the discussion is here expanded with additional context, examples, and figures.

3.1 Cities as complex systems

Cities are paradigmatic examples of complex systems. They are made of millions of heterogeneous elements—individuals, infrastructures, institutions—whose local interactions give rise to emergent collective patterns. Traditional approaches to urban analysis often emphasize geometry (land use, distance, density) or administrative categories. While useful, these perspectives rarely capture the nonlinear dynamics that characterize urban life. Complexity science provides an alternative framework: cities can be studied as evolving networks of interacting elements, where topology and dynamics combine to generate global phenomena such as congestion, segregation, innovation, or resilience.

A first lesson from complexity science is the importance of scale. Phenomena that appear trivial at the individual level—daily commuting, energy consumption, social contacts—aggregate into nontrivial neighborhood dynamics, such as accessibility or segregation. These in turn shape systemic patterns at the scale of entire cities: traffic jams, economic specialization, or uneven resilience. Finally, systems of cities display large-scale properties such as innovation hubs, migration flows, or regional hierarchies. Figure 3.1 illustrates this nested organization. Crucially, emergent behavior at each scale requires dedicated models: local observations cannot be simply extrapolated to

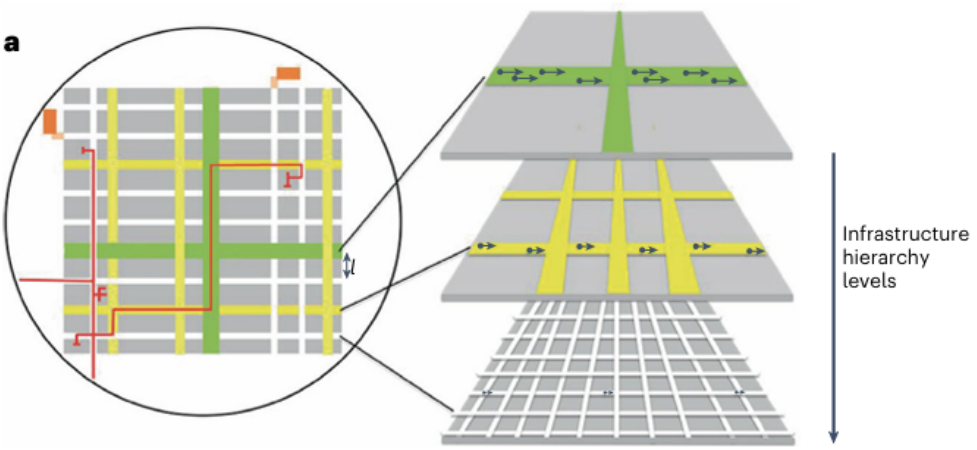


Figure 3.1: Urban scales: from individuals to neighborhoods, to cities, and ultimately to systems of cities. Each level displays emergent properties that cannot be reduced to the sum of local components.

predict global outcomes. Another central contribution of network science is the concept of multilayer networks. A city is not a single graph but the superposition of many interdependent layers: the street network, the public transport system, logistics chains, energy grids, social and digital connections. Each layer has its own structure and dynamics, but they are tightly coupled. A transport failure can disrupt the economy; a blackout can impact communications; the structure of streets can constrain access to healthcare. Multilayer network theory formalizes these dependencies, making it possible to identify vulnerabilities, predict cascading failures, and design more resilient infrastructures.

3.2 Scaling laws and universal patterns

Empirical studies have shown that many urban indicators scale with population size in systematic ways. Economic outputs (e.g. GDP, patent production) grow superlinearly with population, reflecting increasing returns to scale. Infrastructural requirements (e.g. road length, power grids) grow sublinearly, reflecting economies of scale. Household services tend to scale close to linearly. These scaling laws point to universal mechanisms of self-organization across cities, independent of geography or history. They provide quantitative evidence that cities share common principles rooted in network interactions.

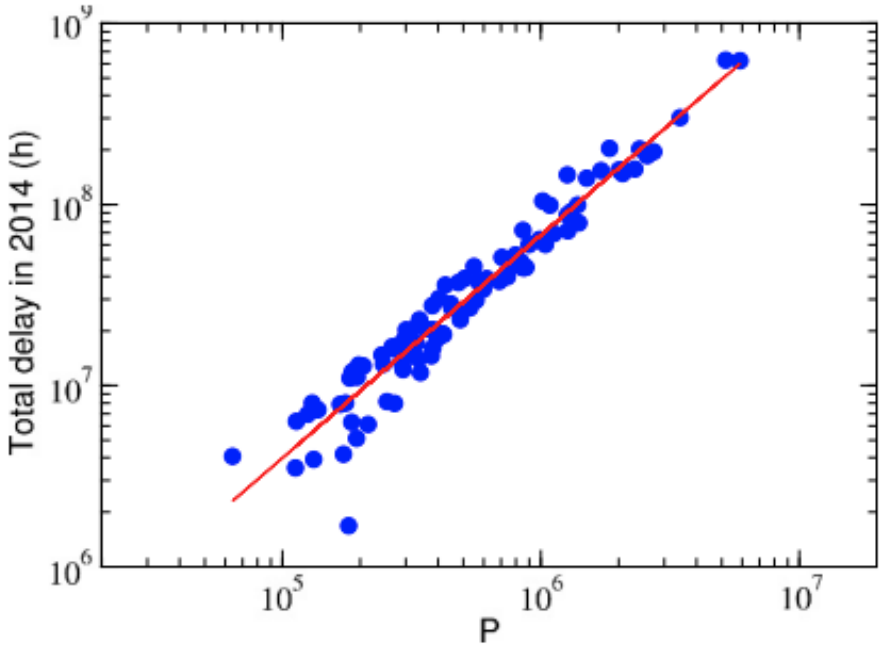


Figure 3.2: Illustrative scaling laws: Plot of the annual delay $\delta\tau$ vs. the number of drivers P for all cities in 2014 ((Depersin and Barthelemy, 2018)

.

3.3 From data to decisions

The next challenge is to operationalize these insights. Data must be transformed into networks, networks into indicators, and indicators into decisions. Today this process is enabled by unprecedented availability of fine-grained data: census surveys, mobile phone traces, transport smartcards, environmental sensors. Once transformed into network representations—mobility graphs, bipartite land use networks, temporal communication graphs—these data can be analyzed with complexity metrics such as centrality, modularity, or percolation thresholds. These indicators provide interpretable measures of access, resilience, and inequality. The critical step is to embed them into policy pipelines.

3.4 Discussion and Conclusion

This chapter develops the perspective introduced in the *Nature Cities* article by showing that cities can be understood more clearly when they are treated as complex systems shaped by relations rather than by isolated elements. In this view, the structure of connections within the urban fabric—between places, infrastructures and social activi-

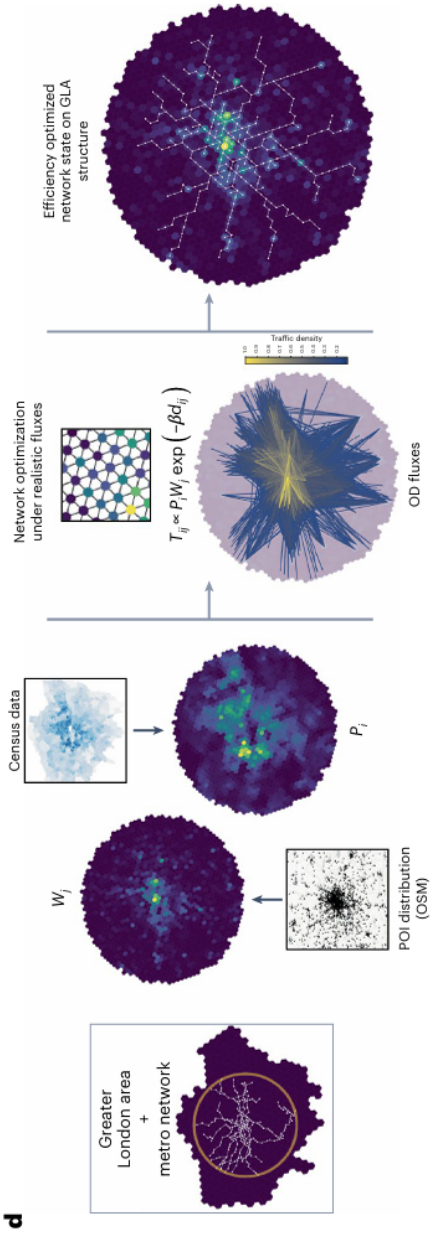


Figure 3.3: From data to decisions: a conceptual pipeline.

ties—plays a decisive role in shaping collective outcomes. Phenomena often interpreted as independent issues, such as congestion or unequal access, appear instead as manifestations of how the city is organised and how interactions accumulate across scales.

Interdependence is a defining feature of contemporary urban environments. The street layout, transport services, energy provision, digital communication and social exchanges influence each other in ways that cannot be separated without losing essential information. Describing the city as a multilayer network makes these dependencies explicit and allows their consequences to be analysed more systematically. It becomes possible to understand why certain areas are more exposed to disruption, why others maintain stability, and how local changes can produce system-wide effects.

Empirical regularities reinforce this conceptual approach. Scaling laws show that many urban indicators vary with population in predictable ways, revealing that cities evolve under constraints generated by their connective structure. These regularities indicate that global behaviour is not simply the sum of local conditions. It emerges instead from the organisation of interactions, and this organisational aspect becomes visible only through a network-based description.

The increasing availability of detailed urban data strengthens this analytical framework. Once spatial and behavioural information is expressed through networks, underlying patterns of accessibility, pressure and imbalance become observable. Crucially, this representation provides a way to explore how the system would respond to alternative configurations and how interventions might propagate through the existing structure. It offers a coherent method to compare scenarios and to evaluate potential trade-offs before changes are implemented.

The analysis presented here points to a methodological shift. Network science provides a language for linking the physical form of the city with the dynamics that unfold upon it. It offers a way to combine theoretical insights from complexity science with the practical demands of urban planning. By adopting this perspective, it becomes possible to reason about cities not as static objects but as evolving systems whose behaviour depends on their structure and on the interactions it enables.

Chapter 4

A Diffusive Model to Quantify Green Area Accessibility

This chapter presents the work developed in the article Urban topology and dynamics can assess the importance of green areas, published in Physical Review E (Moi, 2024), of which the candidate is the first author. The candidate designed the analytical model, implemented the numerical simulations, and carried out the data analysis. The aim is to provide a quantitative framework for measuring the accessibility of public green areas (PGAs) within cities, using tools from complex networks and statistical physics.

4.1 Introduction

The motivation for this research stems from a gap between the acknowledged importance of green spaces and the difficulty of evaluating their actual impact in practice. While it is well established that urban greenery improves biodiversity, mitigates climate change, and enhances the health and well-being of citizens, planners and decision-makers often lack robust metrics to compare accessibility across different neighborhoods or to predict the effects of new interventions. Traditional measures based only on geographic proximity or area size fail to capture the dynamics of how citizens effectively reach and use PGAs.

The approach introduced here combines two elements: the topology of the urban network and a dynamical model of biased diffusion. By representing the city as a spatial network of blocks and connections, and by simulating how agents move toward green spaces under realistic constraints, it becomes possible to define new centrality-based measures of accessibility that outperform purely geometric indicators. The model incorporates both population distribution and the limited capacity of PGAs, allowing us to evaluate not just the presence of green spaces but their actual usability.

This chapter therefore marks the transition from the general framework of network science for smart cities, presented in the previous chapter, to its application to a concrete

urban issue: the equitable distribution and accessibility of public green areas.

4.2 Why Public Green Areas Matter

In the context of complex urban systems, public green areas (PGAs) should be regarded as structural nodes of the socio-spatial network rather than as isolated amenities. Their presence and distribution influence systemic properties such as robustness, resilience, and equity. Unlike classical approaches in urban ecology that focus on vegetation cover or biodiversity, the perspective here emphasizes PGAs as resources embedded in the connectivity patterns of cities, where their importance can be quantified through accessibility.

Accessibility in this framework is not defined solely by geometric proximity, such as the Euclidean distance to the nearest park, but by the dynamics of movement in the underlying urban network. A resident seeking a PGA can be modeled as an agent exploring the network along paths whose likelihood depends on both distance and network topology. Thus, the effective usability of PGAs depends on the ensemble of possible routes and the probability that each route is actually taken.

This systemic viewpoint highlights why PGAs matter beyond their ecological role. They act as absorbing targets in diffusion processes, shaping flows of movement, social interaction, and opportunities for health and well-being. Their importance therefore lies in their position within the network: a centrally located green area accessible through multiple low-cost paths contributes disproportionately to urban equity, whereas a peripheral or isolated one may be underutilized despite its size.

The framework adopted in this chapter treats PGAs as critical resources whose accessibility can be measured with the mathematical tools of network science. By considering biased random walks, hitting probabilities, and path ensembles, it becomes possible to capture how PGAs influence the distribution of opportunities in cities. This approach shifts the focus from static provision of green space to the dynamic processes that govern how citizens actually reach and use them.

4.3 Accessibility: Concepts and Literature

Accessibility has long been a central concept in urban and transport studies. At its core, accessibility measures the ease with which individuals located at i can reach *opportunities* distributed in space.

In this framework, an *opportunity* O_j represents a resource or activity available at location j: jobs, schools, hospitals, commercial services, or public green areas. The precise definition depends on the policy question. In labour economics, O_j corresponds

to employment positions; in transport studies, to activity centres; in urban sustainability research, to the presence and characteristics of green spaces.

The second key component is *cost* c_{ij}. Cost is not merely Euclidean distance. It usually represents a generalized impedance, including travel time, waiting time, transfers, monetary costs, congestion, or physical barriers. Two destinations at equal geometric distance may have very different costs if one is served by rapid transit and the other requires indirect routes through a fragmented street network.

Hansen's seminal formulation (Hansen, 1959) defined accessibility as the weighted sum of opportunities discounted by cost:

$$A_i = \sum_j O_j f(c_{ij}),$$

where $f(c_{ij})$ is a decreasing function of cost. For example, if O_j represents jobs and c_{ij} travel time, zones with shorter commuting times exhibit higher employment accessibility.

Logsum accessibility, derived from discrete choice theory, extends this idea:

$$A_i^{\text{logsum}} = \frac{1}{\lambda} \ln \Big(\sum_j e^{\lambda U_{ij}} \Big),$$

where U_{ij} represents the systematic utility of reaching destination j from origin i. This formulation captures substitution effects: if several attractive destinations are available, overall accessibility increases even if each single alternative provides only moderate utility.

A complementary network-based perspective is provided by the analysis of London and Paris by Travençolo et al. (Travençolo et al., 2009), who quantified accessibility through a diversity entropy computed over self-avoiding random walks on the urban street network, with and without the underground system. In that framework, opportunities correspond to the set of reachable locations at a given topological scale, while cost is encoded in path length and network structure. The introduction of the Underground increases both the magnitude and the spatial homogeneity of accessibility, showing how infrastructure reshapes effective urban access.

However, all these measures share a fundamental limitation: they are essentially static. They evaluate the spatial configuration of opportunities and costs, but they do not explicitly describe how individuals dynamically move, interact, or compete for limited resources.

Consider the case of public parks. Suppose a neighbourhood has three parks within 10 minutes of walking. In a classical cumulative or Hansen-type formulation, accessibility would be high simply because multiple opportunities lie within a low cost threshold. Yet this measure implicitly assumes that each park can absorb an unlimited number of visitors. It does not account for crowding, capacity constraints, or behavioural adaptation. If one park is small and already saturated, or if residents preferentially

choose the same attractive destination, effective accessibility may be much lower than what the static indicator suggests.

In other words, classical accessibility metrics assume that opportunities are infinitely available and that agents do not influence one another's access. When resources are limited and population density is heterogeneous, this assumption becomes problematic. Accessibility is no longer only a function of distance and opportunity, but also of competition and dynamic use.

This limitation motivates the need for a stochastic and interaction-based framework, where accessibility emerges from the ensemble of possible movements of agents over the urban network, rather than being imposed solely by geometric proximity.

Stochastic Approaches in Complex Networks

Network science introduced a dynamic perspective. Random walks provide a natural way to model accessibility as an exploration process. Let P be the transition matrix of a random walk, and $H(i, j)$ the mean first-passage time (MFPT) from i to j. Then

$$H(i, j) = 1 + \sum_u P_{iu} H(u, j), \quad H(j, j) = 0,$$

and the associated closeness measure

$$C_i^{RW} = \frac{n}{\sum_j H(j, i)}$$

quantifies how efficiently node i can be reached on average.

Building on this, (Travençolo and Costa, 2008) introduced an entropy-based measure of accessibility. If $p_j^{(h)}(i)$ denotes the probability that a walker starting at i is at j after h steps, then

$$\kappa_i^{(h)} = \exp\Big(- \sum_j p_j^{(h)}(i) \log p_j^{(h)}(i) \Big),$$

representing the effective number of distinct destinations reachable in h steps. This connects accessibility directly to diversity: high values reflect balanced reachability, low values indicate concentration.

The connection between accessibility and diffusion becomes even clearer in studies of influential spreaders in networks. Arruda et al. (Arruda et al., 2014) demonstrated that the identification of key nodes in epidemic or information spreading depends crucially on centrality metrics derived from diffusion processes, rather than on purely structural measures. A node's accessibility can thus be interpreted as its capacity to channel flows in spreading dynamics. In the urban context, this means that accessibility measures are analogous to influence measures: blocks or neighborhoods with high accessibility act

as "spreaders" of opportunity, ensuring that residents can reach essential resources such as green areas.

Finally, the notion of effective distance proposed by Brockmann and Helbing (Brockmann and Helbing, 2013) reformulated accessibility in terms of probabilities. Defining

$$d_{ij} = -\log P_{ij},$$

where P_{ij} is the transition probability, accessibility emerges as the likelihood of flows, aligning measures with actual movement patterns rather than geometric distances.

Despite the wide variety of formulations, from Hansen's potential accessibility to logsum and mean first-passage times, these approaches remain essentially static: they quantify opportunities embedded in space but do not reflect the stochastic mechanisms by which individuals effectively reach public green areas. Moreover, all destinations are implicitly assumed to have unlimited capacity and every feasible path is considered equally available, regardless of cost. This abstraction is unrealistic in urban contexts, where accessibility depends on both the geometry of the network and the scarcity of resources. Step by step, accessibility has evolved from static spatial indicators to dynamic and probabilistic measures that connect efficiency, diversity, and diffusion. This theoretical progression—from Hansen's models to entropy-based accessibility, from the bag-of-paths framework to spreading and diffusive centralities—forms the conceptual basis of the approach developed here. In the following section, these principles are extended to model public green areas as absorbing targets within biased diffusion processes, allowing us to describe not only where green spaces are located, but how they are, according to a diffusive model, reached and used by urban populations.

4.3.1 Spatial networks as a representation of the city

In order to study accessibility within a complex urban environment, we require a spatial representation that is both geometrically grounded and analytically tractable. We construct a weighted spatial graph $G = (V, E)$ in which nodes correspond to the centroids of census blocks and public green areas (PGAs). The spatial partition is obtained through a Voronoi tessellation generated by these centroids.

The Voronoi tessellation should not be interpreted as a regular hexagonal grid. It is not an imposed discretization of space, but a partition induced by the empirical distribution of block centroids. Each cell consists of the set of points closer to its generating centroid than to any other. Because census blocks differ in size and are irregularly distributed across the city, the resulting Voronoi cells are heterogeneous in both shape and area.

By construction, the tessellation covers the entire urban domain without gaps. In a strict geometric sense, there are no "holes": every point in space is assigned to its nearest centroid. However, the spatial density of centroids is not uniform. In dense central

districts, blocks are smaller and centroids are closely spaced; in peripheral or functionally specialized areas (e.g., large residential blocks, industrial zones, infrastructure corridors), centroids may be farther apart. This produces Voronoi cells of significantly different sizes.

Edges in the spatial network connect nodes whose Voronoi cells share a common boundary. Each edge is weighted by the Euclidean distance between the two corresponding centroids. Because centroid spacing is irregular, adjacent edge weights are not uniform. Even in relatively dense areas, small variations in block geometry produce variations in centroid distances.

It is important to stress that the network representation is not equivalent to a simple discretization of Euclidean space. In continuous geometry, the distance between two locations is given directly by the metric $d(i, j)$. In the spatial graph, movement is constrained to occur through sequences of adjacent Voronoi cells. The shortest-path distance $d_{SP}(b, p)$ is therefore a path-dependent quantity defined on a discrete and heterogeneous structure.

Moreover, the graph is not topologically regular. Urban Voronoi tessellations do not produce constant node degrees: the number of adjacent cells varies across space, and adjacency patterns are locally idiosyncratic. Even in a hypothetical region where adjacent centroid distances were approximately equal, the variation in node degree and neighborhood configuration would still generate non-uniform multi-step transition dynamics in a Markov process.

The notion of "emptiness" must therefore be understood in a functional rather than geometric sense. While there are no literal holes in the tessellation, regions with low centroid density behave as sparse areas of the graph. Such regions are more common in the suburbs, but they can also occur within central districts wherever large urban artifacts or low-density land uses are present. These variations affect both local connectivity and global path structure.

A simple illustrative example clarifies this point. Consider two blocks located at similar Euclidean distance from a given park. If one block lies in a dense district with many small neighboring cells and multiple adjacency options, while the other lies in a low-density area with larger cells and fewer neighbors, the sets of admissible paths differ. The corresponding shortest-path distances and stochastic path ensembles may therefore diverge, even though the geometric distances are comparable.

The spatial network thus encodes two complementary layers of information: geometric distance (through edge weights) and urban morphology (through adjacency structure and centroid density). This representation provides a principled discretization of the city that preserves physical scale while enabling path-based and diffusion-based analyses of accessibility.

4.3.2 From deterministic to stochastic accessibility: the bag-of-paths

In reality, people do not follow a unique shortest path, nor do they always choose the absolute minimum distance. Movements are heterogeneous: some trajectories are longer, others deviate due to preferences or obstacles. To capture this variety, we introduce the *bag-of-paths* (BoP) formalism. The idea is to consider an ensemble of all possible paths P between nodes, and assign to each a probability

$$\mathbb{P}(P) \propto \exp(-\gamma\, c(P)), \tag{4.1}$$

where $c(P)$ is the cost of the path (proportional to distance) and $\gamma > 0$ is an inverse temperature controlling how strongly the ensemble is biased toward short routes. The form of this distribution is not arbitrary: it is the result of a variational principle in which one minimizes the expected path cost subject to an entropy constraint. Solving this constrained optimization with Lagrange multipliers yields exactly the Boltzmann weight. In other words, the BoP distribution is the maximum-entropy probability measure consistent with a prescribed expected cost. A full derivation is presented in Appendix A.

At the matrix level, the BoP formalism modifies the reference random walk P^{ref} by incorporating an exponential penalty on costs. The weighted transition matrix is

$$W \;=\; P^{\text{ref}} \circ e^{-\gamma C}, \tag{4.2}$$

where C is the cost matrix and $\circ$ denotes the Hadamard product. The *fundamental matrix* then collects all weighted walks:

$$Z \;=\; (I - W)^{-1}. \tag{4.3}$$

The matrix Z admits the equivalent series representation

$$Z = \sum_{t=0}^{\infty} W^t, \tag{4.4}$$

which makes explicit its probabilistic meaning. Each term W^t collects all walks of length t weighted by their Boltzmann cost, so that Z_{ij} sums the total contribution of all feasible walks from node i to node j. Hence Z plays the role of a partition function over paths. From Z, one obtains the hitting-path probabilities by first normalizing the columns,

$$Z_h \;=\; Z\,\mathrm{diag}(Z)^{-1}, \tag{4.5}$$

and then restricting the absorbing set to PGA nodes. This yields

$$Z_{b,p}^{\text{PGA}} \;=\; \frac{Z_{b,p}}{\sum_{q\in\{\text{PGA}\}} Z_{q,p}} \quad \text{and} \quad Z_{b,p}^{\text{PGA}} \leftarrow \frac{Z_{b,p}^{\text{PGA}}}{\sum_{q\in\{\text{PGA}\}} Z_{b,q}^{\text{PGA}}}, \tag{4.6}$$

where the last step corresponds to a row normalization restricted to the set of PGAs.These successive normalizations transform weighted path counts into proper hitting probabilities. The column normalization compensates for the total incoming path weight at each PGA node, while the row normalization ensures that for each origin block b the probabilities over all PGAs sum to one. In this way, $Z^{\mathrm{PGA}}_{b,p}$ defines a bona fide probability distribution over absorbing destinations. In other words, each entry $Z^{\mathrm{PGA}}_{b,p}$ gives the probability that a biased random walker starting from block b eventually reaches the PGA p, taking into account the multiplicity and relative likelihood of paths.

Hence the *average farness* refines the static definition (**??**) by weighting distances with these hitting probabilities:

$$\langle C_f(b) \rangle \;=\; \sum_{q \in \{\mathrm{PGA}\}} d_{SP}(b,q)\; Z^{\mathrm{PGA}}_{b,q}. \tag{4.7}$$

Here $d_{SP}(b,q)$ denotes the shortest-path distance between block b and park q on the weighted street network. Importantly, d_{SP} is not the expected path length under the BoP distribution. The BoP probabilities $Z^{\mathrm{PGA}}_{b,q}$ encode the likelihood of reaching each destination through the ensemble of feasible walks, while the metric component d_{SP} provides a structural distance scale. In this way, the formulation separates geometry (network distances) from stochastic routing (path probabilities). This formulation makes explicit the probabilistic nature of accessibility: the measure is no longer a simple sum of distances, but an expectation value over the ensemble of feasible walks.

4.3.3 Finite capacity and agent competition

Although more refined, the formulation in (4.7) still assumes that PGAs have infinite capacity. In practice, each green area has a limited walkable surface, which constrains the number of people who can use it at the same time. Accessibility is therefore also a function of competition among agents.

To capture this aspect, we introduce a simple agent-based model. Each PGA q is assigned a maximum occupancy proportional to its area:

$$\mathrm{max\_occ}_q = \frac{A_q}{f_{mq}}, \tag{4.8}$$

where A_q is the surface area and f_{mq} is the minimum area required per individual. Agents start from blocks according to population density and perform biased random walks guided by a reward function φ_i that increases towards PGAs:

$$\varphi_i = \mathbf{e}_i^{\top} Z\, \mathbf{e}_{\{\mathrm{PGA}\}}, \tag{4.9}$$

where $\mathbf{e}_i$ is the canonical vector of node i. The transition probabilities are then sampled with a soft bias:

$$P_{i \to j} = \frac{\exp[-\beta(\varphi_j - \varphi_i)]}{\sum_{k \in \mathcal{N}(i)} \exp[-\beta(\varphi_k - \varphi_i)]}, \tag{4.10}$$

where β interpolates between pure diffusion ($\beta = 0$) and greedy descent towards high-reward nodes ($\beta \to \infty$).

When agents reach a PGA, they attempt to occupy it. If the capacity is already saturated, the agent is forced to continue walking, possibly toward a more distant destination. Collecting these outcomes yields the occupancy distribution $P^b_{\mathrm{PGA}}(q)$, the probability that residents from block b end up using green area q. Accessibility then becomes

$$\langle C_f(b)\rangle_{\mathrm{PGA}} = \sum_{q\in\{\mathrm{PGA}\}} d_{SP}(b,q)\, P^b_{\mathrm{PGA}}(q). \tag{4.11}$$

The sequence from $C_f(b)$ (**??**) to $\langle C_f(b)\rangle$ (4.7) and finally to $\langle C_f(b)\rangle_{\mathrm{PGA}}$ (4.11) illustrates a conceptual progression. The first step is purely geometric, based only on network distances. The second adds stochasticity, acknowledging that urban mobility follows multiple paths with different likelihoods. The third incorporates social competition, recognizing that PGAs are finite resources for which residents must compete. In this way, accessibility evolves from a static index into a dynamic, probabilistic, and socially grounded measure.

4.4 Methods: Spatial networks and model setup

The framework introduced in the previous section can only become operational once it is coupled with a spatial representation of the city and a simulation protocol that connects theoretical quantities to observable indicators. The starting point is the construction of the urban network. Each city is represented as a weighted planar graph $G = (V, E)$ derived from a Voronoi tessellation of its built environment. The tessellation divides the territory into polygons, each associated with a centroid that is taken as a node of the network. Two nodes are connected if their cells share a boundary. This construction ensures that every portion of the urban surface belongs to a single polygon and that the resulting network covers the entire city without gaps or overlaps. The advantage of this representation is its neutrality: unlike administrative boundaries, which depend on historical conventions, Voronoi cells provide a geometric partition that is homogeneous and suitable for quantitative analysis.

Within this network, two kinds of nodes are identified. The first are the blocks b, associated with census data on the resident population. These nodes represent the demand side of the system, i.e. the origin points of agents. The second are the public green areas (PGAs) p, defined as the polygons corresponding to walkable parks, gardens, urban forests and other open green spaces extracted from OpenStreetMap data. These nodes represent the supply side of the system, i.e. the potential destinations of agents. Edges between nodes carry weights given by the Euclidean distance between centroids, which form the cost matrix C. This same matrix is used both for shortest-path calculations and for the definition of biased random walks.

The model includes several parameters that tune how agents move and how accessibility is evaluated. The first is the inverse temperature γ, which regulates the bag-of-paths ensemble. In the limit $\gamma \to 0$, all paths have nearly equal probability, while in the opposite limit $\gamma \to \infty$ only the shortest paths remain relevant. In practice, we select γ so that the characteristic scale of the ensemble matches the 15-minute walkability threshold. Assuming an average walking speed of $v \simeq 1.33\ \mathrm{m\,s^{-1}}$, this corresponds to approximately 1200 meters.

A second parameter is f_{mq}, the minimum walkable surface that each agent search to occupy a green area. This value determines the maximum number of individuals that a PGA can accommodate:

$$\mathrm{max\_occ}_q = \frac{A_q}{f_{mq}},$$

where A_q is the surface area of the PGA. By varying f_{mq} across different values, one can explore a spectrum of scenarios, ranging from abundant availability of green areas (small f_{mq}, large capacity) to scarcity (large f_{mq}, small capacity). In this way the model does not only describe the geometry of access but also the competition for limited resources.

The third parameter is β, which appears in the biased diffusion of agents. Each agent follows transition probabilities of the form (4.10), where $\beta = 0$ corresponds to a simple random walk, and increasing values of β bias movements toward nodes of higher reward φ_i. This interpolates between purely diffusive exploration and greedy search for the closest PGA, providing a flexible description of individual strategies.

The simulations proceed as follows. Each block node b receives an initial population of agents proportional to its census size. Agents then move along the network according to the biased random walk dynamics. A trajectory ends when one of three events occurs:

1. the agent reaches a PGA that still has free capacity and occupies it;
2. the agent reaches a PGA that is already saturated, is rejected, and continues walking in search of another destination;
3. the agent reaches the maximum allowed path length (energy budget) without accessing any PGA, in which case it is considered as unserved.

The last case is particularly relevant when f_{mq} is large and green areas are scarce relative to demand. By repeating this process for a large number of independent agents, one obtains the empirical occupancy distribution $P^b_{\mathrm{PGA}}(q)$, i.e. the probability that residents of block b end up in green area q. This distribution is then used in the definition of the occupancy-based accessibility (4.11). To ensure statistical robustness, simulations are repeated multiple times and averaged over different realizations.

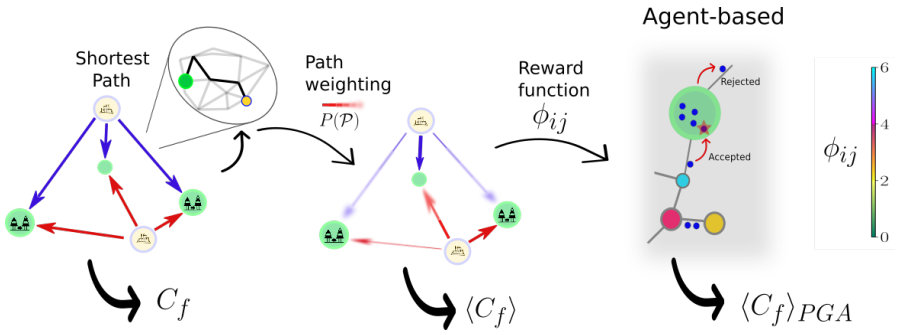


Figure 4.1: (color online) Main elements of the method devised in this work. From left: farness centrality calculated from shortest paths; average farness obtained by path weighting (bag-of-path approach); reward function and agent-based simulations.

4.4.1 Pair correlation analysis on spatial networks

Prior to presenting the results of the agent-based model, it is essential to understand the spatial disposition and clustering of Public Green Areas (PGAs), which characterize the scales at which occupation and spatial interactions emerge within the urban network.

To this aim, we introduce a discrete pair correlation function defined directly on the network topology, denoted as $g_{\text{net}}(m)$. The function extends the concept of the pair correlation $g(r)$, widely used in statistical physics to quantify spatial correlations in continuous domains (Gavagnin et al., 2018).

In the case of urban networks, the Euclidean distance is replaced by the network distance, defined as the length of the shortest path connecting two nodes along the graph. This formulation enables the quantification of spatial aggregation or dispersion among discrete entities—here, the PGAs—embedded in a networked urban structure. The function $g_{\text{net}}(m)$ is defined as

$$g_{\text{net}}(m) = \frac{c_d(m)}{\left(\frac{N}{Z}\right)\left(\frac{N-1}{Z-1}\right) s_d(m)}, \tag{4.12}$$

where $c_d(m)$ represents the number of pairs of PGA nodes whose network distance lies within the m-th distance bin, N is the total number of PGAs, $Z = |V|$ is the total number of vertices in the network, and $s_d(m)$ is the normalization factor obtained from a uniform reference distribution of nodes.

Values of $g_{\text{net}}(m) > 1$ indicate clustering of PGAs at the corresponding distance scale, while $g_{\text{net}}(m) < 1$ reflects a tendency toward spatial dispersion. This function thus provides a scale-dependent measure of spatial correlation directly defined over the graph, enabling the identification of characteristic aggregation scales and the comparison of the spatial organization of green areas across different urban morphologies. The analysis with the aforemntioed function is illustrated in Fig. 4.2. Excluding correla-

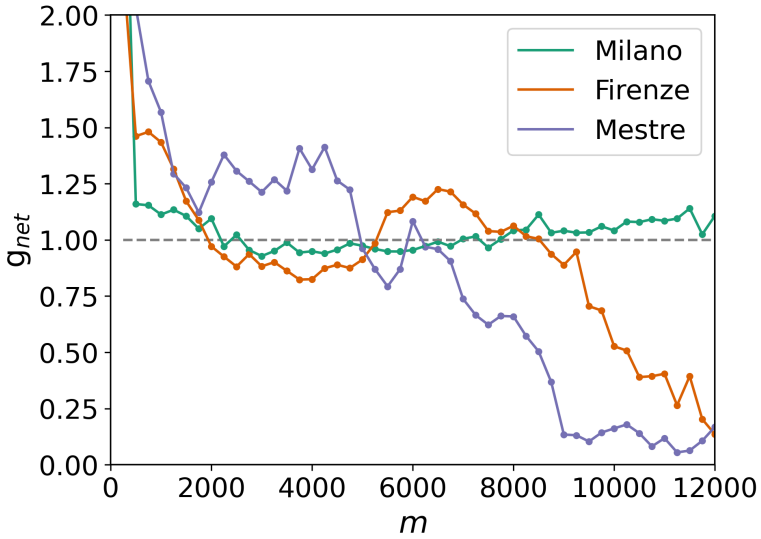


Figure 4.2: Pair correlation function $g_{\text{net}}(m)$ computed for the spatial networks of Mestre, Florence, and Milan. The function quantifies the degree of clustering among Public Green Areas (PGAs) as a function of network distance m. Short-range peaks indicate local aggregation of green areas, while flatter profiles correspond to more uniform spatial distributions. The distinct correlation patterns highlight different urban planning strategies and spatial organizations of PGAs across the three cities. For a visual mathcing see Fig. 4.3.

tions below 200 m, which may arise from uncertain PGA area definitions, all three cities—especially Mestre and Florence—exhibit significant short-range clustering. This indicates a planning approach focused on optimizing total green surface without excessive dispersion.

In the 2000–4000 m range, Mestre shows higher correlation peaks, revealing a small-cluster organization of PGAs consistent with the city's population density. Conversely, Florence displays significant clustering at 6000–8000 m, coupled with a statistically significant void in the corresponding Mestre range, which explains the presence of two large clusters in Florence, visible in Fig. 4.2. Milan, in contrast, demonstrates a nearly uniform correlation profile, suggesting a spatial strategy aimed at dispersing PGAs more evenly across the territory. This configuration theoretically maximizes the probability of reaching a green area from any location while reducing the likelihood of long exploratory paths in low-density regions.

Overall, the pair correlation analysis reveals that each city expresses a distinct pattern of green-space organization: Mestre favors localized clusters aligned with population distribution, Florence exhibits large-scale aggregation constrained by morphology, and Milan achieves near-uniform dispersion. These results provide a structural basis for interpreting the outcomes of the agent-based simulations, where the spatial arrangement of PGAs is expected to influence accessibility and movement efficiency.

4.5 Results: Centrality and accessibility measures

The framework can now be applied to real urban networks in order to compare different ways of quantifying accessibility. The progression is natural: we start from the simplest geometric definition of farness, then incorporate stochasticity through the bag-of-paths ensemble, and finally add the effects of finite capacity and competition among agents.

The first measure is the farness centrality (**??**), which sums the shortest-path distances from each block to all PGAs. This quantity provides a baseline description of accessibility: blocks close to many green areas have low farness, while peripheral or segregated blocks display high values. Visually, this calculated metric is akin to computing the sum of distances within a circle that stretches to encompass the farthest nodes, treating each path equally as contributing to the node's centrality. However, in an urban context, it is not necessarily taken for granted that the actual contribution of the shortest path length is proportional to its distance. In practice, maps of $C_f(b)$ reproduce the expected patterns: in Mestre, where the urban fabric is relatively uniform, the distribution of values is narrow; in Florence the historical structure creates sharp contrasts between accessible and inaccessible areas; and in Milan the large metropolitan scale produces a broad range of values reflecting its polycentric organization.

The second step is to refine this picture by introducing stochasticity. In the bag-of-paths formulation, accessibility is not determined by a single geodesic but by an ensemble of feasible routes. Distances are weighted by hitting probabilities $Z^{\mathrm{PGA}}_{b,q}$ derived from the fundamental matrix (4.3) and normalization (4.6). The resulting measure, the average farness (4.7), is an expectation value over paths. Short routes dominate, but alternative paths contribute with smaller weights. This smooths the extremes of the purely geometric case: blocks that were penalized by long unique geodesics may benefit from secondary routes, while densely connected areas no longer appear disproportionately favoured. The parameter γ regulates this balance, interpolating between uniform weighting and strict minimization of distance. The result is a picture of accessibility closer to real mobility, where individuals choose short but not necessarily unique paths.

The last step is to recognize that PGAs are not infinitely elastic destinations. Each green area has a finite capacity $\mathrm{max\_occ}_q = A_q / f_{mq}$, determined by its surface area and the reference demand f_{mq}. Agents reaching a saturated PGA are forced to continue their walk, which reshapes the distribution of access probabilities. Collecting these dynamics yields the occupancy distribution $P^b_{\mathrm{PGA}}(q)$ and the corresponding accessibility measure (4.11). In this case, even blocks surrounded by many PGAs may experience reduced accessibility if those PGAs are systematically overcrowded. Conversely, peripheral areas may gain relative importance when their local parks are less used.

Taken together, the three measures form a hierarchy. The geometric definition $C_f(b)$ highlights spatial inequalities but overestimates access in dense areas. The stochastic refinement $\langle C_f(b) \rangle$ incorporates mobility diversity and provides a smoother distribution of values. The occupancy-based version $\langle C_f(b) \rangle_{\mathrm{PGA}}$ adds the social dimension of

competition for scarce resources. Moving from the first to the last corresponds to an increasing degree of realism: from static geometry, to stochastic mobility, to capacity-constrained dynamics. Below the comparative analysis of different cities is reported.

4.6 Comparative analysis of case studies

The three definitions of accessibility described above can be visualized on the spatial networks of Mestre, Florence, and Milan. This comparison is not only useful to illustrate the progression from static to dynamic measures, but also to highlight how differences in urban form and scale translate into inequalities in access to green areas. The case studies were selected to provide three distinct examples: Mestre, a medium-sized modern development; Florence, a historical city with irregular and constrained morphology; and Milan, a large metropolitan area with a polycentric structure. The following analysis reproduces and extends the results reported in (Moi, 2024), with additional discussion tailored to the thesis context.

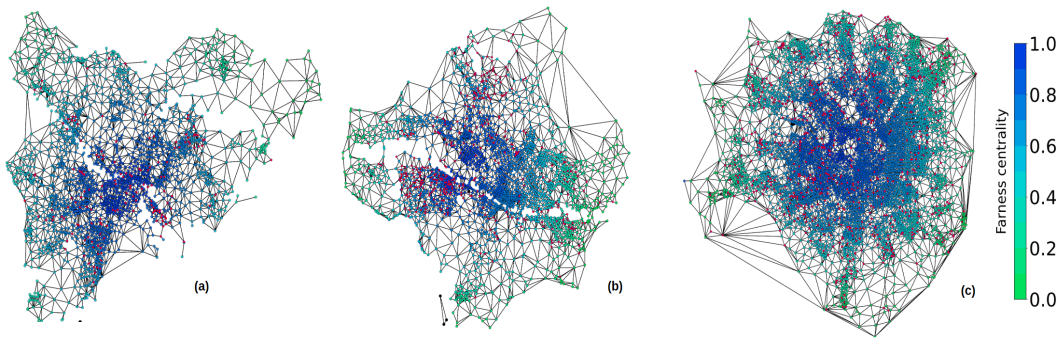
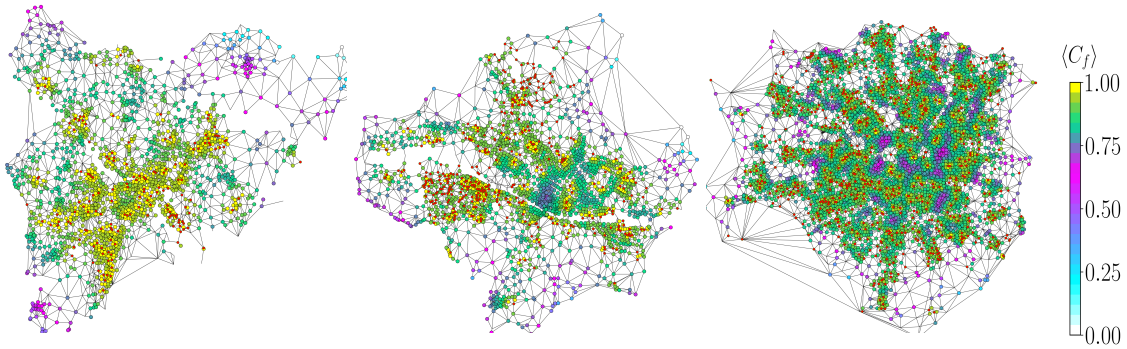


Figure 4.3: Farness centrality $C_f(b)$ for Mestre (a), Florence (b), and Milan (c). PGA nodes are shown (red).

The farness maps (Fig. 4.3) provide the most direct geometric picture of accessibility. In Mestre the values are relatively homogeneous: the distribution of public green areas across the compact urban fabric ensures that no district is extremely disadvantaged. This contrasts with Florence, where the historical city center and its dense urban morphology result in sharp gradients. Central blocks, located close to numerous small green areas, display very low farness, while peripheral districts, particularly those separated by the Arno river or by the surrounding hills, are much less accessible. Milan, with its much larger scale, exhibits a broader range of values: central and semicentral areas benefit from a dense network of parks, while peripheral neighborhoods, especially those along the outer ring, face distances that are significantly larger.

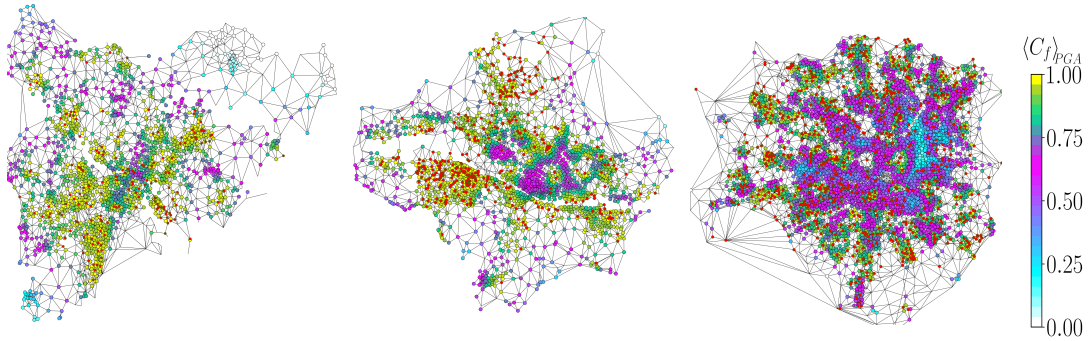
When stochastic paths are considered through the bag-of-paths formalism (Fig. 4.4), the picture changes substantially. Average farness $\langle C_f(b) \rangle$ smooths the extreme contrasts



(a) Mestre (b) Florence (c) Milan

Figure 4.4: (color online) Average Farness centrality $\langle C_f \rangle$ mapped onto the respective spatial networks. PGA nodes are shown (red).

observed in the deterministic case. Blocks penalized by unique long geodesics in Florence recover some accessibility through secondary paths, even if less probable. This effect is also visible in Milan, where peripheral districts retain some level of access through indirect connections. In Mestre, where the geometry was already homogeneous, the difference between deterministic and stochastic measures is minimal. This confirms that the bag-of-paths framework effectively captures mobility diversity, reducing the dependence of accessibility on single privileged routes.

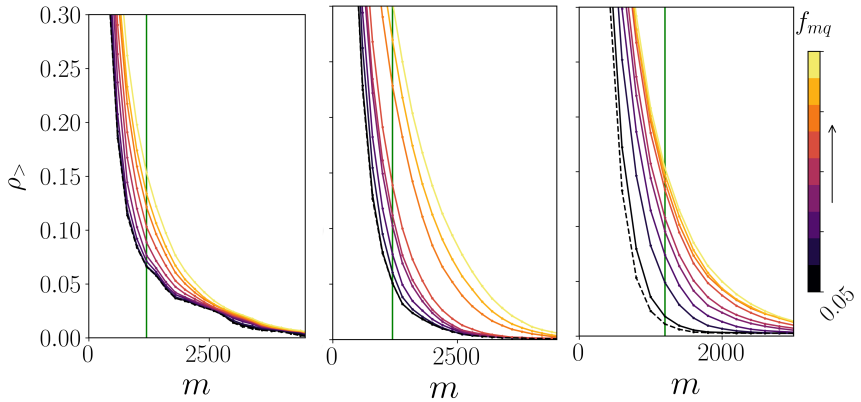


(a) Mestre (b) Florence (c) Milan

Figure 4.5: (color online) Average Farness centrality $\langle C_f \rangle_{PGA}$ from agent-based model mapped onto the respective spatial networks. PGA nodes are shown (red).

The most significant differences emerge when finite capacities are introduced (Fig. 4.5). The occupancy-based farness $\langle C_f(b) \rangle_{\text{PGA}}$ reveals the interplay between topology and competition. This metric resolves the lack of valuable information on both farness centrality and average farness centrality. In Fig.4.5, the Average Farness centrality calculated as in Eq.(4.11) is reported from the agent-based simulations described in

Methods. Compared to Fig.4.4, it is possible to note a partial overlap and expansion of low-quality clusters of blocks for the three cities. For Mestre, as for Florence, it is possible to note the appearance of enlarged low-quality groups of nodes (Magenta). While for Florence, it was also evident from the Average Farness $\langle C_f \rangle$ the presence of local fragility given by the low presence of PGAs, for Mestre, such fragility can be reported mainly for the modeled competition, where it appears clear that green areas are not sufficient to ensure equitable access for downtown residents.



(a) Mestre (b) Florence (c) Milan

Figure 4.6: (color online) Cumulative density for the three cities at different f_{mq}.

To characterize how far agents travel before being absorbed by a park, we define the cumulative density

$$\rho_>(m) = \frac{1}{N_a} \sum_{p \in \{\text{PGA}\}} \sum_{j \in \mathcal{V}_b^p} \mathbf{1}\{d_{SP}(p, j) \leq m\}\,, \tag{4.13}$$

where N_a is the total number of agents, $\mathcal{V}_b^p$ denotes the set of blocks from which park p has attracted agents, $d_{SP}(p, j)$ is the shortest-path distance between block j and park p, and $\mathbf{1}\{\cdot\}$ is the indicator function.

The quantity $\rho_>(m)$ measures the fraction of agents that reach a PGA within a distance not larger than m. It therefore provides a cumulative description of the effective spatial catchment of parks under competitive dynamics. In the absence of congestion, $\rho_>(m)$ would be primarily determined by the structural path distribution. Deviations from this baseline reflect saturation effects: as f_{mq} decreases (stronger competition), agents are progressively pushed toward more distant parks, leading to a slower growth of $\rho_>(m)$ with m. To understand how this property scales concerning the 15-minutes distance, we show in Fig. 4.7 the fit of the data reported by the cumulative density at increasing f_{mq}. These plots suggest a linear relationship between the two quantities,

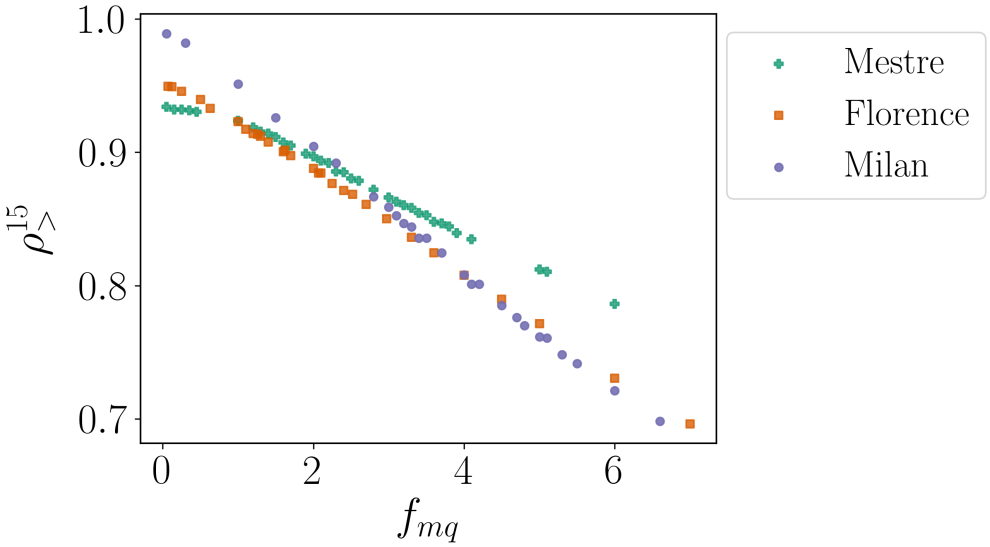
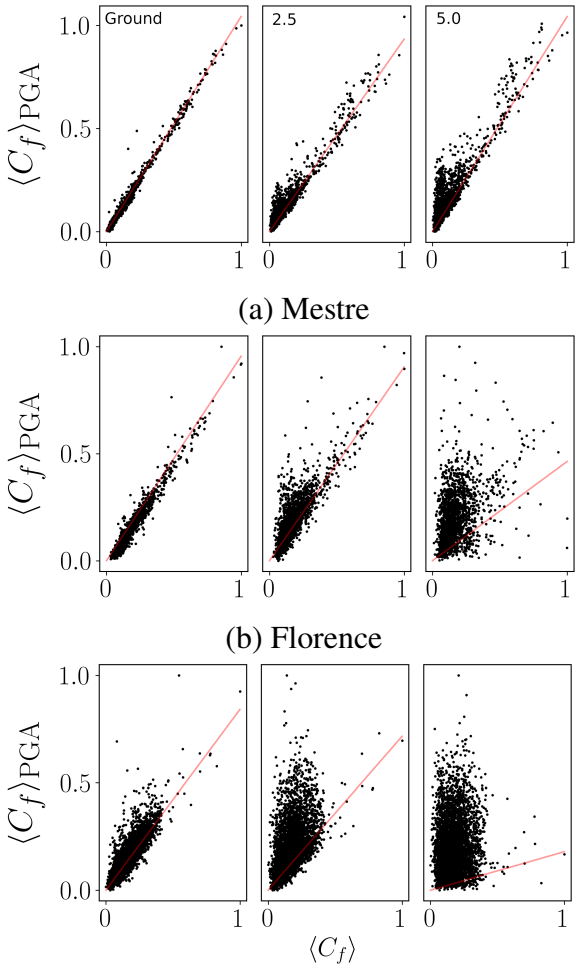


Figure 4.7: (color online) Fit of cumulative density at the 15-minutes distance vs different f_{mq}. Fitted points correspond to the ones that intersect the 15-minutes distance (green line).

which decrease with a specific coefficient. The linearity follows the diffusion/adsorption mechanism since the simulated copies of the system vary only by the parameter f_{mq}. The flux of agents on edges entering PGAs is proportional to the probability of observing a given path to them, and the fraction of agents accepted (*i.e.*, removed from the flux) depends on the actual occupancy of the designated PGA. A negative change in f_{mq} thus produces a linear decrease in the fraction of accepted agents. The negative slope then indicates globally the amount of crowd present in PGA's walkable areas. To explore to which extent agents compete for PGA access, the agent-based Average Farness centrality $\langle C_f \rangle_{\text{PGA}}$ is plotted versus the Average Farness $\langle C_f \rangle$ in Fig. 4.8. In particular, the figure shows a distinguishable pattern: at f_{mq} = ground, which corresponds to high PGA capacity, the Average Farness shows a linear correlation with agent-based one, which is broken at increasing f_{mq}. First, this emphasizes that the assumption underlying the simulations method is verified because under high-capacity conditions agents sample paths proportional to the probability of accessing the nearest ones, thus recovering the Average Farness as a proportion of path probability. The second instance reinforces the observation that, according to this model, the population gradient and the increasing scarcity of green areas modify the frequency with which PGAs accept agents. Indeed, histograms reported in Fig. 4.9 show how events considered 'rare,' (*i.e.* traveling long distances to access a PGA), become increasingly frequent as the area sought is constrained by the actual provision of green areas. This results in a distribution that shifts more toward higher distance values; this effect is prominent in the case of Milan. This points out that PGAs become increasingly correlated with blocks at long distances.

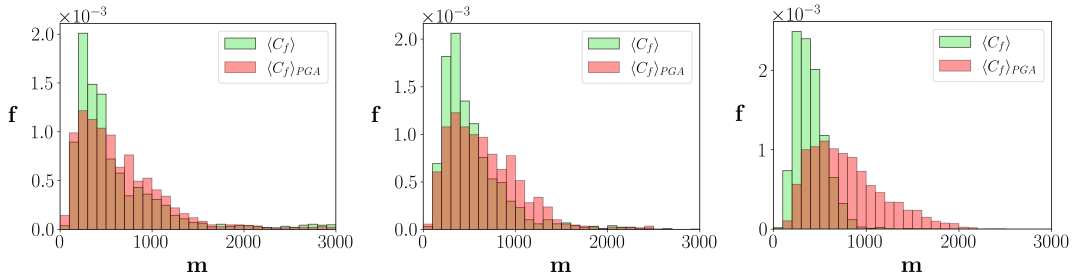


(a) Mestre

(b) Florence

(c) Milan

Figure 4.8: (color online) Scatter plot of the agent-based Average Farness centrality vs the Average farness per block at different f_{mq} (ground, 2.5, 5.0). The red line corresponds to the linear fit of the points.



(a) Mestre (b) Florence (c) Milan

Figure 4.9: (color online) Histogram of non normalized $\langle C_f \rangle_{\text{PGA}}$ and $\langle C_f \rangle$ per block expressed in meters at different distance bins m (*i.e.* probability density of blocks that fall in the i-th bins). In brown, there is an overlap between the two quantities.

Overall, the comparative analysis highlights a consistent progression: Farness centrality captures the role of spatial distribution, stochastic averaging accounts for path diversity, and occupancy-based farness reveals the constraints of finite resources. The cumulative density function $\rho_>(m)$ offers a synthetic indicator to compare cities, while sensitivity tests show that results are stable under reasonable parameter variations. Together, these findings demonstrate that the proposed framework captures both the geometric and social dimensions of accessibility, offering insights beyond classical measures. To dig deeper, we use the well-established correlation function from statistical mechanic framework to di dig into the findings.

4.7 Discussion

The analysis presented in this chapter shows how different definitions of accessibility capture complementary aspects of the relation between urban form, mobility and the provision of public green areas. The progression from the geometric farness $C_f(b)$ to the stochastic average farness $\langle C_f(b) \rangle$, and finally to the occupancy-based measure $\langle C_f(b) \rangle_{\text{PGA}}$, reveals a gradual enrichment of the concept of accessibility. Each step corrects an implicit assumption of the previous one: the purely geometric view ignores the multiplicity of feasible paths, the stochastic ensemble assumes unlimited capacity, and the agent-based extension introduces competition for scarce resources.

The comparison across Mestre, Florence and Milan illustrates how urban morphology mediates the effect of these definitions. In Mestre, the compact and relatively homogeneous fabric ensures that all three measures produce similar outcomes, with limited inequalities. In Florence, the historical structure amplifies the role of stochastic paths: peripheral areas regain partial accessibility when secondary routes are accounted for, but lose it again when capacity constraints saturate central parks. In Milan, the polycentric scale makes the transition between measures more gradual, but competition

effects reveal hidden inequalities even in areas rich in green spaces. These results confirm the conclusions of (Moi, 2024), showing that capacity-aware metrics are indispensable to capture effective accessibility.

The cumulative density function $\rho_{>}(m)$ offers a synthetic way to summarize these differences, while the sensitivity analysis confirms that the qualitative ranking of cities is robust under variations of the parameters γ, β and f_{mq}. This robustness suggests that the framework is not overly dependent on arbitrary modeling choices, but instead reflects structural properties of the cities under study.

Looking ahead, several extensions are possible. First, the present framework could be combined with temporal information, to account for diurnal and seasonal variations in the use of green areas. Second, multi-modal accessibility could be incorporated by integrating public transport and cycling networks, thus extending the analysis beyond purely pedestrian mobility. Third, socio-demographic heterogeneity could be explicitly modeled, weighting agents by age, income or vulnerability, in order to connect accessibility with equity more directly. Finally, accessibility to green areas could be studied in relation to other urban functions, such as climate adaptation, health benefits, or biodiversity corridors, embedding the analysis within the broader agenda of resilient and sustainable cities.

Chapter 5

Walking the City: Realistic Diffusion of Green Areas Accessibility

This chapter presents original work by the candidate, not included in the published articles. It develops a preliminary spatial and socio-morphological analysis of the Mestre urban area, providing the territorial context for the stochastic accessibility models discussed in the thesis. The study examines the distribution of census blocks, demographic gradients, green-area clustering, and street-network connectivity, identifying structural patterns that influence accessibility outcomes. The candidate performed the spatial data acquisition and preprocessing, conducted the network and demographic analysis, and designed the exploratory territorial framework. These results establish a territorial baseline for a planned full-scale extension in collaboration with Confindustria Veneto

5.1 Introduction

The previous chapter introduced a statistical physics framework for assessing the accessibility of public green areas (PGAs) in cities (Moi, 2024). There, urban space was abstracted through a Voronoi-like tessellation (geometric partitioning) where each block was represented as a polygonal cell and distances between blocks and parks were computed between centroids. This representation provided a useful first approximation, allowing us to define modified centrality measures and to simulate competitive dynamics of agents reaching PGAs. However, it also carried an important limitation: distances were calculated in Euclidean space, that is, as *straight lines across the tessellation*, ignoring the actual geometry of the pedestrian network. Similar limitations of "as the crow flies" distances in urban accessibility have long been discussed in transport and networked-urban studies (Hansen, 1959; Travençolo and Costa, 2008; Barthelemy, 2011). As a result, the model overlooked crucial spatial constraints such as barriers, detours, and the connectivity patterns of the street system (Barthelemy, 2011, 2018).

In short, accessibility was measured in an abstract geometric space, while real citizens reach parks by walking along streets.

In this chapter we overcome this limitation by extending the stochastic accessibility framework to the real pedestrian street network of Mestre, reconstructed from OpenStreetMap data. Within this framework, accessibility is no longer an abstract property of geometric cells, but the outcome of feasible routes embedded in the actual urban topology (Barthelemy, 2011). Every block and park is connected to its nearest intersections, and all movements are constrained by the street graph. This allows us to capture effects of peripherality, bottlenecks, and the hierarchical structure of the network (Barthelemy, 2018).

Beyond this core improvement, two further refinements are introduced. First, we define a revised accessibility score that combines two complementary components: (i) the structural potential of accessibility, obtained from the Bag-of-Paths ensemble of walks (Saerens et al., 2009), and (ii) the realized behaviour, captured through agent-based diffusion with competitive dynamics (van Kampen, 2007; Norris, 1997). Their combination produces an exposure index, which is then used to construct a composite accessibility score, highlighting discrepancies between theoretical potential and practical achievement of accessibility. Second, we integrate demographic information from the 2011 and 2021 censuses, introducing a nonlinear weighting of block populations. This ensures that accessibility deficits in highly populated areas weigh more heavily in the overall evaluation, embedding a notion of equity into the metric (Hansen, 1959).

The present study therefore represents the natural continuation of the previous work. Where the Voronoi-based model revealed the conceptual power of statistical physics for quantifying accessibility, the street-network-based model grounds these insights in urban reality. By comparing the two census years, we also move from a static picture to a longitudinal one, quantifying how demographic shifts and infrastructural changes translate into accessibility gains or losses. This methodological progression makes the framework more suitable for practical applications in urban planning, where equity, realism, and demographic sensitivity are essential (Bettencourt et al., 2007b; Barthelemy, 2011; Caldarelli et al., 2025b).

5.2 Methods

The methodological framework developed here extends the one introduced in the previous chapter (Moi, 2024), where accessibility was defined on Voronoi tessellations and distances were computed in Euclidean space. In that work we detailed the agent-based biased diffusion and the Bag-of-Paths formalism (Saerens et al., 2009; Francoisse et al., 2018). In this chapter we do not repeat those derivations, but focus instead on the extensions required to apply the model to the real street network of Mestre and to integrate demographic weighting.

5.2.1 Datasets

We employed two primary sources of data. First, official population counts were obtained from ISTAT for the 2011 and 2021 censuses, at the resolution of census blocks. Each block was associated with its resident population, allowing a temporal comparison. Second, the pedestrian network and the polygons of public green areas (PGAs) were extracted from OpenStreetMap (OpenStreetMap contributors, 2017). Park nodes were defined as the centroids of polygon geometries, while street nodes correspond to intersections of pedestrian-accessible segments.

Network Construction

We define an undirected spatial graph composed of three classes of nodes: city blocks, parks, and street intersections. The underlying street network consists of intersections as nodes and pedestrian-accessible segments as edges.

Each block node corresponds to the centroid of a census block and represents a population source. Each park node represents the centroid of a contiguous public green area. These block and park nodes are not part of the original street network but are connected to it via additional edges. For each block or park, we determine the nearest street intersection(s)—defined as the set of street nodes at minimum Euclidean distance from the centroid—and connect the block or park node to all of these nearest intersections. This choice defines a unique and comparable access point between areal units and the street network. Blocks and parks are treated as abstract sources and sinks rather than traversable regions; connecting them to all internal intersections would introduce artificial shortcuts and size-dependent biases. Moreover, the diffusion model assumes that walkers preferentially select paths of minimal cost (or energy), so the nearest intersection represents the most likely and physically consistent access to the street network.

The resulting graph is spatially embedded and supports the definition of walks and path-based accessibility metrics, as used in the subsequent agent-based and stochastic modeling procedures.

5.2.2 Metrics

Exposure index The previous work also pushed for the need of a baseline score to assess how much the competition is biased toward agents competition or toward the structural organization of streets. To this scope, the goal is to separate *structure* from *use*. BoP-based farness $\langle C_f \rangle_{\text{BoP}}$ summarizes the *structural potential* imposed by the street topology and park locations (what the network would afford in the absence of competition and dynamics), whereas simulation farness $\langle C_f \rangle_{\text{sim}}$ reflects *realized access* under biased exploration, congestion, and finite search budgets (what agents actually

experience). A direct difference, $\langle C_f \rangle_{\text{sim}} - \langle C_f \rangle_{\text{BoP}}$, is hard to compare across places because it depends on the absolute scale of distances and on city size. The ratio

$$E_i = \frac{\langle C_f(i) \rangle_{\text{sim}}}{\langle C_f(i) \rangle_{\text{sim}} + \langle C_f(i) \rangle_{\text{BoP}}} \in (0, 1) \tag{5.1}$$

is scale-free and bounded, enabling cross-block and cross-year comparisons without additional normalization. Interpretation is immediate: $E_i \approx 0.5$ indicates parity between realized and structural access; $E_i < 0.5$ signals that realized access outperforms the structural baseline (e.g., redundant routes or effective local park assignment); $E_i > 0.5$ indicates a penalty due to dynamics (e.g., detours, bottlenecks, competition). Using the symmetric denominator $\langle C_f \rangle_{\text{sim}} + \langle C_f \rangle_{\text{BoP}}$ (instead of, e.g., a pure quotient) bounds the index, reduces sensitivity to small denominators, and preserves monotonicity with respect to both components. In practice, E_i functions as a *diagnostic of mismatch*: it highlights where behavioural frictions (search bias, congestion, limited capacity) prevent the network's structural potential from being realized, or conversely where the dynamics exploit topology unusually well.

Population weighting Another missing piece in the previous work was an implicit weighting of the population factor, completely driven by the number of allocated agents. Intuitively, a block with a small number of inhabitants is less contributing to the global overall accessibility across the city. To ensure that blocks with larger populations contribute proportionally more to the overall accessibility assessment, we modulate the exposure index E_i with a population weighting factor. This adjustment is crucial to reflect that accessibility deficits in more densely populated areas affect a larger number of residents and should carry greater weight in an equity-oriented model. The correction takes the form of a nonlinear transformation:

$$P_i = \left(\frac{\text{pop}_i}{N_{\text{ref}}} \right)^{\beta_p}, \qquad 0 < \beta_p \leq 1, \tag{5.2}$$

where pop_i is the ISTAT-reported resident count for block i, N_{ref} is the maximum block population across the two census years, and $\beta_p = 0.8$ introduces a mild nonlinear scaling. This exponent is inspired by the Box–Cox family of transformations, often used to stabilize variance and reduce skewness in population-distributed datasets (Box and Cox, 1964; Sakia, 1992). The choice of a single reference value N_{ref} across both census years is deliberate and ensures temporal comparability of accessibility scores. Using a common normalization allows differences between 2011 and 2021 to reflect absolute changes in demographic pressure and accessibility conditions, rather than year-dependent rescaling. A year-specific normalization would remove information about population growth and would only support relative, within-year comparisons. Such concave scaling ensures that highly populated blocks are appropriately prioritized, while still allowing less

populated areas to contribute meaningfully to the analysis. Similar nonlinear transformations are common in urban-systems modeling to balance the difference weights and to avoid statistical dominance by the outliers (Batty, 2013; Barthelemy, 2019b). This approach is also aligned with accessibility justice and equity frameworks in urban planning, which emphasize population-sensitive metrics for fair access to infrastructure and services (Pereira et al., 2017; Foth et al., 2020).

The resulting composite accessibility score

$$s_i = \alpha E_i + \beta P_i, \tag{5.3}$$

with $\alpha + \beta = 1$ (here $\alpha = \beta = 0.5$), yields a normalized value $s_i \in [0, 1]$ that simultaneously captures behavioral efficiency relative to structural potential and the demographic importance of each block. The linear combination is adopted because the two terms play different roles. The exposure index E_i is not a direct measure of accessibility, but a diagnostic of how structural accessibility is realized under dynamics and competition. The population term P_i does not make a block more accessible by definition; it weights the importance of the exposure condition in terms of affected residents. The key advantage of combining simulation with analytical models lies in cross-validating structural potential against emergent behavioral outcomes. While the Bag-of-Paths metric assumes access in a low-competition, frictionless regime, the agent-based results reflect congestion and limited park attractiveness. Comparing the two provides a robust indicator of whether accessibility is being realized in practice, or remains theoretical.

5.3 Results

We applied our accessibility model to the pedestrian network of Mestre, using census population data from 2011 and 2021 to compute block-level accessibility scores. The results, shown in Fig. 5.2, reveal a clear spatial pattern of improvement over the decade.

In the 2011 map, accessibility scores show greater heterogeneity, with a substantial number of peripheral blocks exhibiting low values. In contrast, the 2021 map presents a smoother distribution, with more blocks attaining medium-to-high accessibility scores. These differences reflect not only changes in population distribution but also the dynamic behavior captured by the agent-based model. The difference map in Fig. 5.1 (upper panel) shows a widespread positive shift in scores across the urban area. Most blocks experience a net increase in accessibility ($\Delta s_i > 0$), especially in zones that previously faced limited connectivity or saturation effects. The right panel confirms this trend: the scatter of 2021 scores is skewed above the identity line, with relatively few blocks experiencing a decline.

Taken together, these findings suggest that accessibility to public green areas in Mestre has improved between 2011 and 2021, both in terms of the number of well-served blocks and the overall spatial equity of access. To assess temporal changes in

spatial accessibility between 2011 and 2021, we performed a paired t-test on accessibility ratios measured at the census unit level ($n = 1001$). The test revealed a statistically significant increase in accessibility over the decade ($t = 35.44$, $p < 0.0001$), with an average change of $\Delta = 0.0418$ (SD = 0.0373). These results confirm a widespread and systematic improvement in accessibility conditions across the study area. Several clusters of accessibility improvement align with known areas of population redistribution, suggesting that increased demand in newer residential zones has been met with usable green areas. However, some areas remain structurally disadvantaged despite modest improvements, thereby highlighting the importance of balancing infrastructure deployment with behavioral modeling in urban analytics.

In some blocks, improvements were driven more by changes in reachability rather than simple proximity. According to the model, this highlights how the addition or connection of just one well-positioned green area can redistribute access opportunities for hundreds of residents. From the whole-urban perspective, this trend appears to be driven by a combination of structural and behavioral factors, including a noticeable shift in residential patterns and changes in population composition that can be observed through the comparison of spatial clustering effects: adjacent blocks often experienced similar changes in accessibility. These spatial spillovers suggest that accessibility improvements propagate through the network, benefiting neighboring areas even without direct infrastructural changes. The evolving urban form has increasingly encouraged populations to settle in areas outside the traditional urban core, potentially resulting in a more theoretical or underutilized accessibility to peripheral PGAs .While newly developed peripheral areas benefit from additional or improved parks, some inner-city areas risk experiencing relative decline due to competition and limited walkable access (Seto et al., 2012).

From a theoretical viewpoint, the spatial accessibility landscape behaves analogously to a scalar potential defined over the urban grid. Local gradients in accessibility can thus be interpreted as zones of "mobility pressure", where marginal changes in topology or population would yield disproportionately large accessibility shifts. This reveals candidate areas for infrastructural interventions that maximize equity gains under minimal investment.

5.4 Discussion and Conclusion

This study demonstrates how the integration of statistical physics principles with agent-based modeling can provide nuanced insights into equitable access to urban green spaces. By applying our stochastic accessibility framework to the pedestrian network of Mestre, we quantified how spatial structure, demographic shifts, and competitive dynamics collectively shape green space accessibility. The analysis reveals a clear improvement between 2011 and 2021, with previously fragmented accessibility patterns

giving way to a more equitable distribution. The agent-based simulations highlight how congestion effects and population redistribution influence accessibility, while the composite, population-weighted accessibility score captures both connectivity advantages and demographic relevance.

Our findings also show that equitable access is not solely determined by proximity but also by the interplay of urban form, demographic pressures, and behavioral competition. Compared to earlier formulations, this study makes three key methodological extensions.

First, it applies the accessibility framework to the actual pedestrian street network of Mestre, rather than to a generalized or idealized urban graph. This allows the model to incorporate real spatial constraints such as intersection layout, peripheral fragmentation, and the hierarchical structure of street connectivity. These features significantly affect how pedestrians navigate and determine which blocks are functionally isolated or well-integrated— aspects that cannot be captured by topologies lacking geographic grounding.

Second, the analysis introduces a revised accessibility score that captures both behavioral and structural components. The score is constructed as a convex combination of two terms: an exposure index derived from agent-based simulations, and a baseline from a path ensemble model. The exposure index reflects the relative ease or difficulty agents face in reaching green areas under crowding and competition, while the Bag-of-Paths farness centrality captures the purely topological potential of the same environment. Their combination enables a comparison between realized and latent accessibility, offering a more nuanced assessment of spatial performance.

Third, the score incorporates a nonlinear transformation of population to ensure that blocks with higher demographic weight exert greater influence on the final evaluation. Rather than using raw population counts—which would risk overemphasizing a small number of highly populated units—the model applies a concave exponent that compresses extreme values while preserving relative differences. This approach balances representational fairness with demographic relevance, allowing the accessibility metric to reflect both spatial structure and population distribution without being skewed by outliers.

The physics-inspired Bag-of-Paths approach, combined with real-world census data, offers a scalable method for diagnosing spatial inequalities and evaluating planning interventions. Future research could extend this framework to incorporate temporal dynamics in park usage or assess accessibility under different urban development scenarios. Ultimately, this work bridges statistical mechanics and urban analytics, providing planners with actionable tools to foster more inclusive and resilient green infrastructure.

The methodology developed in this study opens several promising avenues for both research and practical urban planning. Over the next decade, the increasing availability of high-resolution mobility data—such as anonymized GPS traces or smart city sensor networks—could refine the agent-based model, allowing real-time simulations of park usage under different urban growth scenarios. If current trends continue, the ac-

cessibility scores in Mestre may further converge toward equity, provided that urban development prioritizes pedestrian infrastructure and distributes green spaces in sync with demographic shifts.

In this respect, we believe that the use of this methodology can allow policymakers to optimize the resources of green areas for the whole population. On the other hand, one of the challenges that cannot be optimized within this framework is the set of emerging pressures introduced by climate change and rising urban densities. Heatwaves and extreme weather events may alter park visitation patterns, while gentrification could displace vulnerable populations to less accessible areas, counteracting recent gains. Embedding predictive climate and socioeconomic models into the framework could help planners preemptively identify at-risk neighborhoods.

5.4.1 Future perspectives

On a broader scale, this physics-inspired approach could be adapted to other urban systems where spatial competition arises, such as public transit access or healthcare distribution. If integrated into municipal planning tools, such models could dynamically evaluate policy impacts, turning reactive urban management into proactive, equity-driven design. The future of equitable cities may hinge on such interdisciplinary fusion—where statistical physics, data science, and participatory planning collaboratively shape more livable environments.

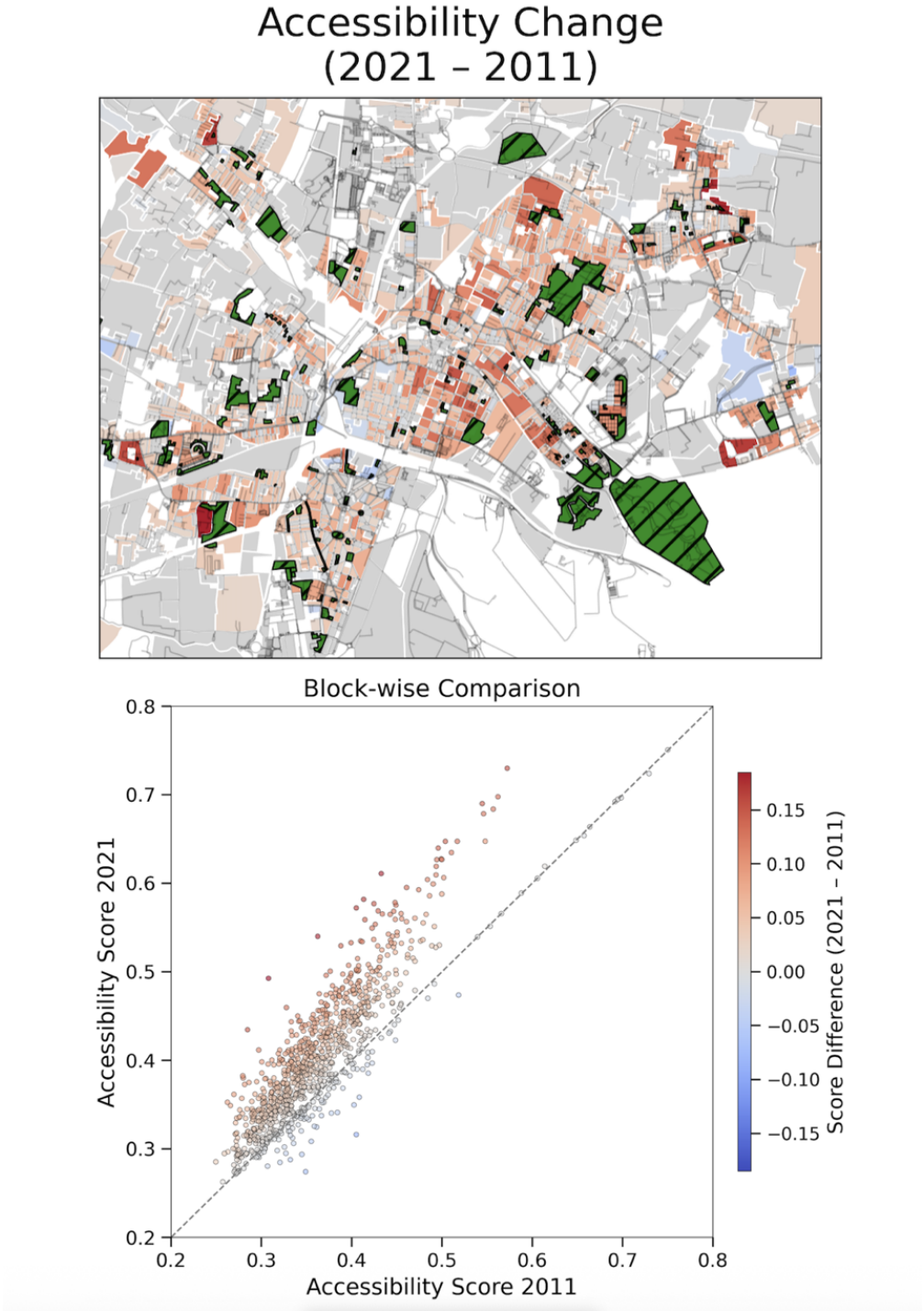


Figure 5.1: Top: spatial distribution of the change in block-level accessibility scores between 2011 and 2021. Bottom: scatter plot comparing the 2011 and 2021 scores for every block; the dashed line marks the identity $y = x$. Colours in both panels encode the score difference $\Delta S_i = S_i^{2021} - S_i^{2011}$.

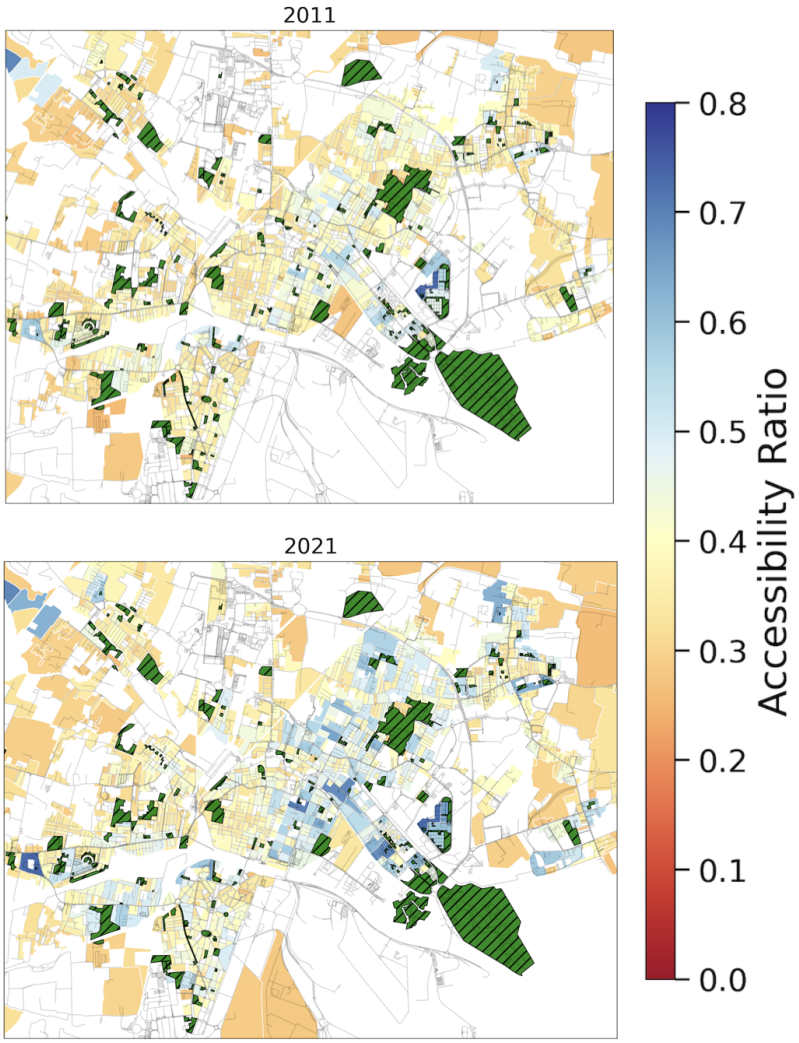


Figure 5.2: Block-level accessibility scores for (a) 2011 and (b) 2021, displayed using a common colour scale. Blue tones indicate higher accessibility, red tones lower accessibility.

Chapter 6

Take home messages

6.1 Conclusion

This thesis has analysed cities through the conceptual and quantitative tools of complexity science, showing how network theory, stochastic processes and statistical physics can be used to study spatial accessibility and, in particular, the distribution and use of public green areas. The work has developed along a clear trajectory: from the need for a scientific framework able to describe urban complexity, to the construction of spatial models, and finally to the application of these models in three research articles.

The introductory chapters have argued that cities behave as complex systems: their properties emerge from interactions, spatial constraints and demographic heterogeneity. Classical descriptions are insufficient because they cannot account for nonlinearity, feedbacks or multi-scale organisation. Network theory provides a structure in which these features can be quantified, and spatial networks extend the analysis to real geometries, distances and constraints relevant for mobility and access.

The first article has placed these ideas within the broader context of complexity science applied to cities. It has shown how concepts such as centrality, robustness, path structure and multilayer interactions offer a precise way to interpret urban phenomena, shifting the perspective from static descriptions to dynamic and interaction-driven systems.

The second article has introduced a diffusive accessibility model based on biased random walks modeled by the bag-of-paths formalism. This framework replaces simple shortest-path metrics with a probabilistic description of access, where low-cost paths dominate the effective relation between blocks and green areas. The approach has been applied to three Italian cities and has produced a refined centrality measure of farness that captures both topology and path likelihood.

The third article has extended this framework in three directions. First, it has incorporated the real pedestrian street network of Mestre, allowing the model to operate on

the actual urban geometry instead of an abstract tessellation. Second, it has combined analytical path statistics with an agent-based model under competition, producing a new accessibility score capable of distinguishing structural potential from realised accessibility. Third, it has introduced a nonlinear population weighting, ensuring that accessibility deficits in highly populated blocks contribute more to the evaluation. Applying this score to population data from 2011 and 2021 has revealed measurable, statistically significant improvements in accessibility and has shown how demographic redistribution, street connectivity and competitive dynamics shape the spatial distribution of advantages and disadvantages.

Taken together, the results show that accessibility is not determined solely by proximity, but by the interplay between infrastructure, demographic pressure and stochastic movement. The models developed in this thesis operate with limited data requirements, generalise across heterogeneous urban forms, and provide a quantitative basis for evaluating spatial equity.

6.2 Future perspectives

The thesis demonstrates how network theory and statistical-physics models can produce operational indicators for analysing cities as complex systems. Several directions emerge for extending this work, involving integration with global urban agendas, methodological generalisation, and translation into tools usable by institutions.

6.2.1 Cities, institutions, and data infrastructures

Urban policy frameworks increasingly recognise the need for quantitative, network-based approaches. UN–Habitat coordinates global efforts on sustainable urbanisation, including SDG monitoring and the New Urban Agenda (UN–Habitat, 2020; United Nations, 2016). City networks such as C40 (C40 Cities, 2020) and global programmes like the World Bank's GPSC (World Bank, 2018) promote integrated planning for climate mitigation and adaptation. In Europe, the Mission "100 Climate-Neutral and Smart Cities by 2030" and the NetZeroCities platform embed experimentation and data-driven tools in urban transitions (European Commission, 2022; NetZeroCities Consortium, 2022).

The indicators developed in this thesis—particularly the diffusive accessibility framework—can be aligned with these institutional data infrastructures. Embedding such metrics in SDG 11 monitoring systems or in city-network dashboards would make them operational beyond the academic domain. This requires co-design with municipalities and alignment with programmes such as UN–Habitat's SDG Cities initiative (UN–Habitat, 2021).

6.2.2 Model generalisation and integration

The proposed models remain intentionally parsimonious. Several extensions are now technically feasible and necessary for broader applicability. A first direction is the transition from single-layer spatial networks to multilayer representations that couple pedestrian accessibility with transport, energy, and communication systems. This allows one to study how perturbations propagate across interdependent urban infrastructures.

A second direction is enriching behavioural heterogeneity. Current agents differ only by origin, energy, and stochastic choice. Future versions may include heterogeneous preferences, socio-demographic mobility budgets, and explicit dependence on climatic or temporal conditions. Integrating climate projections would make the model suitable for evaluating accessibility under heat waves or flood risk, issues central to contemporary resilience planning.

A third direction concerns multimodality and temporal dynamics. Extending the framework to cycling, public transport, or time-varying networks would produce accessibility indicators that better reflect real urban mobility conditions.

The diffusive and centrality-based approach used for public green areas is service-agnostic and can be extended to health care, childcare, emergency infrastructures, cooling centres, or food retail. In each case, congestion, capacity limits, and spatial constraints shape accessibility inequalities that the model can quantify.

City networks such as C40 explicitly integrate nature-based solutions and green–blue infrastructures into climate adaptation strategies (C40 Cities, 2021). Extending the framework from recreational access to protective functions (cooling, shading, flood mitigation) would support decisions on where to prioritise investments in climate-resilient infrastructure.

6.2.3 From models to decision-support tools

The analytical and simulation tools developed in the thesis can evolve into operational instruments through three steps: (i) open-source pipelines that take standard datasets (street networks, census blocks, land use) and automatically produce accessibility fields and inequality metrics; (ii) integration with institutional programmes such as SDG Cities; and (iii) collaborative case studies with city networks (GPSC, C40, EU Mission Cities) to test indicators against real planning constraints.

A medium-term objective is interactive interfaces allowing planners to explore hypothetical interventions—new green areas, demographic shifts, network changes— and receive immediate feedback on accessibility and inequality.

Appendix A

Bag-of-paths biased random walk

This appendix summarises the derivation of the biased random walk induced by the bag-of-paths (BoP) model, following (Francoisse et al., 2018, 2016). The derivation is presented in a compact form and includes the optimisation problem defining the Gibbs-biased path distribution.

Let $G = (\mathcal{V}, \mathcal{E})$ be a strongly connected directed graph with affinity matrix $\mathbf{A} = (a_{kl})$ and cost matrix $\mathbf{C} = (c_{kl})$, with $c_{kl} = +\infty$ whenever $a_{kl} = 0$. The reference random walk is

$$p^{\text{ref}}_{kl} = \frac{a_{kl}}{\sum_{l'} a_{kl'}}, \qquad \mathbf{P}^{\text{ref}} = (p^{\text{ref}}_{kl}).$$

A path $\wp = (k_0, \ldots, k_t)$ starting in $k_0 = i$ has reference likelihood

$$\tilde{\pi}_{\text{ref}}(\wp) = \prod_{\tau=1}^{t} p^{\text{ref}}_{k_{\tau-1}k_\tau},$$

and additive cost

$$\tilde{c}(\wp) = \sum_{\tau=1}^{t} c_{k_{\tau-1}k_\tau}.$$

The BoP model assigns a probability $P(\wp)$ to each path by solving the entropy-regularised optimisation problem

$$\begin{aligned} \min_{P} \quad & \sum_{\wp} P(\wp)\, \tilde{c}(\wp) \\ \text{s.t.} \quad & \sum_{\wp} P(\wp)\, \log \frac{P(\wp)}{\tilde{\pi}_{\text{ref}}(\wp)} = J_0, \\ & \sum_{\wp} P(\wp) = 1. \end{aligned}$$

This problem seeks the distribution that minimises the expected cost while keeping a fixed Kullback–Leibler divergence with respect to the reference random walk. The parameter β appears as the Lagrange multiplier associated with the Kullback–Leibler divergence constraint J_0. The two quantities are therefore not independent: fixing the admissible divergence from the reference random walk uniquely determines the value of β, and vice versa. In practice, β is used as the control parameter of the model, as it directly regulates the balance between adherence to the reference dynamics and penalisation of high-cost paths. Introducing Lagrange multipliers and optimising yields the Gibbs–Boltzmann distribution

$$P(\wp) = \frac{1}{Z_i(t)}\,\tilde{\pi}_{\mathrm{ref}}(\wp)\,e^{-\beta\tilde{c}(\wp)}, \qquad \beta > 0,$$

where β is the inverse temperature and $Z_i(t)$ the normalising constant.

Define the biased transition matrix

$$\mathbf{W} = \mathbf{P}^{\mathrm{ref}} \circ e^{-\beta\mathbf{C}}, \qquad w_{kl} = p^{\mathrm{ref}}_{kl}\,e^{-\beta c_{kl}},$$

so that the unnormalised weight of a path is the product of the w_{kl} along the path.

Summing over all paths of fixed length t gives

$$Z_i(t) = \mathbf{e}_i^{\top}\mathbf{W}^t\mathbf{e}.$$

For arbitrary-length paths, the geometric series converges for all $\beta > 0$ because $\mathbf{W}$ is substochastic:

$$\sum_{t=0}^{\infty}\mathbf{W}^t = (\mathbf{I}-\mathbf{W})^{-1}.$$

This defines the fundamental matrix

$$\mathbf{Z} = (z_{ij}) = (\mathbf{I}-\mathbf{W})^{-1}.$$

The non-hitting bag-of-paths probability that a sampled path begins at i and ends at j is therefore

$$P(s=i, e=j) = \frac{z_{ij}}{\sum_{u,v} z_{uv}}.$$

The fundamental matrix $\mathbf{Z}$ collects the total statistical weight of all paths connecting pairs of nodes under the Gibbs-biased ensemble. In particular, the entry z_{ij} represents the cumulative contribution of all paths starting at node i and visiting node j, summed over all possible lengths and weighted by both the reference random walk and the exponential cost penalty. Equivalently, z_{ij} can be interpreted as the expected number of visits to j by a walker starting from i before its trajectory is suppressed by the cost-induced

dissipation. To restrict the ensemble to first-passage paths ending in node j, the j-th row of $\mathbf{W}$ is set to zero, obtaining $\mathbf{W}^{(-j)}$. The associated fundamental matrix is

$$\mathbf{Z}^{(-j)} = (\mathbf{I} - \mathbf{W}^{(-j)})^{-1}.$$

A central identity (Appendix B in (Francoisse et al., 2016)) shows that

$$z_{ij}^{(-j)} = \frac{z_{ij}}{z_{jj}}.$$

This motivates the hitting-path definition

$$z_{ij}^{h} = \frac{z_{ij}}{z_{jj}}.$$

Collecting all entries yields the normalised matrix

$$\mathbf{Z}_h = \mathbf{Z}\mathbf{D}_h^{-1}, \qquad \mathbf{D}_h = \mathrm{Diag}(\mathbf{Z}), \qquad z_{jj}^{h} = 1.$$

The probability of drawing a hitting path from i to j is then

$$P_h(s = i, e = j) = \frac{z_{ij}^{h}}{\sum_{u,v} z_{uv}^{h}} = \frac{z_{ij}/z_{jj}}{\sum_{u,v} z_{uv}/z_{vv}}.$$

An important remark concerns the stochastic nature of the biased walk induced by $\mathbf{W}$. Because each transition is weighted by the factor $\exp(-\beta c_{kl})$, we have

$$\sum_{l} w_{kl} = \sum_{l} p_{kl}^{\mathrm{ref}}\, e^{-\beta c_{kl}} < 1$$

whenever at least one outgoing cost from k is strictly positive. Therefore $\mathbf{W}$ is a *substochastic* matrix: the total transition probability from a node is strictly less than one, meaning that the walker can "disappear" with probability

$$1 - \sum_{l} w_{kl}.$$

Consequently, the biased BoP random walk does *not* admit a stationary distribution on the nodes, since probability mass is not conserved at the level of node-to-node dynamics. This dissipative behaviour is not an artefact of the construction, but a direct consequence of the statistical level at which the BoP probability measure is defined.

In the BoP formulation, probabilities are assigned to complete paths through a Gibbs weight depending on their total cost, and normalisation is imposed globally on the space of paths. No normalisation constraint is enforced at the level of single-step transitions. As a result, the exponential factor $e^{-\beta c_{kl}}$ does not define a conditional

transition probability, but a multiplicative weight contributing to the total statistical weight of all path extensions beyond node k. When node-level quantities are computed by summing over all admissible continuations, one obtains

$$\sum_l w_{kl} = \sum_l p_{kl}^{\mathrm{ref}} e^{-\beta c_{kl}} < 1,$$

because path extensions with large cumulative cost contribute negligibly to the path ensemble. The missing mass therefore encodes the suppression of statistically unlikely continuations.

If the exponential weights were normalised locally, one would instead define conditional transition probabilities

$$\tilde{p}_{kl} = \frac{p_{kl}^{\mathrm{ref}} e^{-\beta c_{kl}}}{\sum_m p_{km}^{\mathrm{ref}} e^{-\beta c_{km}}},$$

thereby enforcing probability conservation at each node and inducing a conservative Markov chain on $\mathcal{V}$. This operation amounts to conditioning the path ensemble on continuation at every step and removes information on the absolute statistical weight of future path extensions, retaining only their relative preference.

The statistical difference between the two constructions can be illustrated by a fluid analogy. In the locally normalised case, the process resembles an incompressible flow, in which all mass entering a node is redistributed among its outgoing edges. In the BoP case, where the flow represents the statistical weight (or "energy") of the walker, the flow is compressible: part of the mass is progressively absorbed as paths extend, because the ensemble weight decreases with cumulative cost. This absorption is not imposed dynamically, but emerges from the statistical weighting of long trajectories.

The temperature parameter β controls the rate of this mass loss. Large values of β strongly penalise costly transitions, concentrate the path measure on short trajectories, and enhance dissipation. Conversely, small values of β yield weaker suppression and broader exploration. Stationarity would require local mass conservation, which is incompatible with a model designed to encode information on absolute path costs rather than solely on relative transition probabilities.

Bibliography

Arruda, G. F. d., Barbieri, A. L., Rodrigues, F. A., da Fontoura Costa, L., and Moreno, Y. (2014). The role of centrality for the accessibility of complex networks. *Physical Review E*, 90(3):032812.

Barabási, A.-L. (2016). *Network Science*. Cambridge University Press, Cambridge.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Barthelemy, M. (2011). Spatial networks. *Physics Reports*, 499(1-3):1–101.

Barthelemy, M. (2018). *Morphogenesis of Spatial Networks*. Springer, Cham.

Barthelemy, M. (2019a). The statistical physics of cities. *Nature Reviews Physics*, 1:406–415.

Barthelemy, M. (2019b). The statistical physics of cities. *Nature Reviews Physics*, 1(6):406–415.

Bashan, A., Berezin, Y., Buldyrev, S. V., and Havlin, S. (2013). The extreme vulnerability of interdependent spatially embedded networks. *Nature Physics*, 9:667–672.

Batty, M. (2013). The new science of cities. *MIT Press*.

Bettencourt, L. M., Lobo, J., Helbing, D., Kuhnert, C., and West, G. B. (2007a). Growth, innovation, scaling, and the pace of life in cities. *PNAS*, 104:7301–7306.

Bettencourt, L. M., Lobo, J., Strumsky, D., and West, G. B. (2010). Urban scaling and its deviations: revealing the structure of wealth, innovation and crime across cities. *PLoS One*, 5(11):e13541.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., and West, G. B. (2007b). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104(17):7301–7306.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7280–7287.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, 26(2):211–252.

Brockmann, D. and Helbing, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342(6164):1337–1342.

Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., and Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature*, 464:1025–1028.

C40 Cities (2020). About c40 cities. https://www.c40.org/about. Global network of mayors taking climate action.

C40 Cities (2021). Water, heat and nature programme. https://www.c40.org/what-we-do/scaling-up-climate-action/water-heat-nature.

Caldarelli, G. (2007). *Scale-Free Networks: Complex Webs in Nature and Technology*. Oxford University Press.

Caldarelli, G., Arcaute, E., Barthelemy, M., Batty, M., Gershenson, C., Helbing, D., Mancuso, S., Moreno, Y., Ramasco, J. J., Rozenblat, C., et al. (2023). Cities as complex systems. *Nature Computational Science*, 3:374–383.

Caldarelli, G., Chiesi, L., Chirici, G., Galmarini, B., Mancuso, S., Moi, J., and De Domenico, M. (2025a). Lessons from complex networks to smart cities. *Nature Cities*.

Caldarelli, G., Moi, J., et al. (2025b). Lessons from complex networks to smart cities. *Nature Cities*. in press.

Clark, W. A. and Fossett, M. (2008). Understanding the social context of the schelling segregation model. *PNAS*, 105:4109–4114.

Depersin, J. and Barthelemy, M. (2018). From global scaling to the dynamics of individual cities. *Proceedings of the National Academy of Sciences*, 115(1):1–10.

Epstein, J. M. (2006). *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press.

European Commission (2022). Eu mission: 100 climate-neutral and smart cities by 2030. https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/mission-areas/climate-neutral-and-smart-cities_en.

Foth, N., Manaugh, K., and El-Geneidy, A. (2020). Towards equitable transit: examining transit accessibility and social need in montreal. *Transport Policy*, 96:31–38.

Francoisse, K., Fouss, F., and Saerens, M. (2018). Bag-of-paths node criticality measures. *Neurocomputing*, 273:18–29.

Francoisse, K., Kivimäki, I., Mantrach, A., Rossi, F., and Saerens, M. (2016). A bag-of-paths framework for network data analysis. *arXiv preprint arXiv:1302.6766*.

Gavagnin, E., Owen, J. P., and Yates, C. A. (2018). Pair correlation functions for identifying spatial correlation in discrete domains. *Physical Review E*, 97(6):062104.

Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2):73–76.

Kovats, R. S. and Kristie, L. E. (2006). Heatwaves and public health in europe. *European Journal of Public Health*, 16(6):592–599.

Masuda, N., Porter, M., and Lambiotte, R. (2017). Random walks and diffusion on networks. *Physics Reports*, 716-717:1–58.

Moi, A. F. N. i. A. (2024). A diffusive model to quantify green area accessibility at urban level. —.

NetZeroCities Consortium (2022). Netzerocities initiative. https://netzerocities.eu.

Newman, M. (2010). *Networks: An Introduction*. Oxford University Press, Oxford.

Nicholls, R. J. (1995). Coastal megacities and climate change. *GeoJournal*, 37:369–379.

Norris, J. (1997). *Markov Chains*. Cambridge University Press.

OpenStreetMap contributors (2017). Openstreetmap. https://www.openstreetmap.org. Accessed: 2025-01-01.

Pereira, R. H. M., Schwanen, T., and Banister, D. (2017). Distributive justice and equity in transportation. *Transport Reviews*, 37(2):170–191.

Ritchie, H., Roser, M., and Samborska, V. (2023). Urbanization. https://ourworldindata.org/urbanization. Accessed: 2024-01-10.

Saerens, M., Achbany, Y., Fouss, F., and Yen, L. H. (2009). Randomized shortest-path problems: Two related models. *Neural Computation*, 21(8):2363–2404.

Sakia, R. M. (1992). *The Box–Cox transformation technique: a review*, volume 41.

Schelling, T. C. (1971a). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1:143–186.

Schelling, T. C. (1971b). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(2):143–186.

Seto, K. C., Güneralp, B., and Hutyra, L. R. (2012). A meta-analysis of global urban land expansion. *Proceedings of the National Academy of Sciences*, 109(40):16083–16088.

Travençolo, B. A. N., Costa, L. d. F., Viana, M. P., and Strano, E. (2009). On the efficiency of underground systems in large cities. *arXiv preprint arXiv:0911.2028*.

Travençolo, B. A. N. and Costa, L. F. (2008). Accessibility in complex networks. *Physics Letters A*, 373(1):89–95.

UN–Habitat (2020). Un–habitat and the sustainable development goals. https://unhabitat.org/sdg-goals. Accessed: 2025-01-15.

UN–Habitat (2021). Sdg cities: Delivering sustainable urbanisation. https://unhabitat.org/programme/sdg-cities.

United Nations (2016). The new urban agenda. https://habitat3.org/the-new-urban-agenda. Adopted at Habitat III, Quito.

United Nations Department of Economic and Social Affairs (2023). World urbanization prospects: 2023 revision.

van Kampen, N. (2007). *Stochastic Processes in Physics and Chemistry*. Elsevier.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.

World Bank (2018). Global platform for sustainable cities (gpsc). https://www.worldbank.org/en/topic/urbandevelopment/brief/gpsc. Supported by the Global Environment Facility.

Università
Ca'Foscari
Venezia

Borsa di dottorato cofinanziata con risorse dell'Unione europea-*NextGeneration EU*
Piano Nazionale di Ripresa e Resilienza Missione 4 – Componente 1 – Riforma 4.1 Riforma dei Dottorati

M4C1 - Investimento 3.4 - "Didattica e competenze universitarie avanzate" - Sub Investimento - "Dottorati dedicati alle transizioni digitali e ambientali"
M4C1 - Investimento 4.1 - "Estensione del numero di dottorati di ricerca e dottorati innovativi per la Pubblica Amministrazione e il patrimonio culturale"
M4C2 – Investimento 3.3 - "Introduzione di dottorati innovativi che rispondono ai fabbisogni di innovazione delle imprese e promuovono l'assunzione dei ricercatori dalle imprese"

CUP H73C22000450004

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA