Socioeconomic patterns in human mobility and social networks

PhD Thesis

by

Ludovico Napoli

Submitted to
Central European University
Department of Network and Data Science

In partial fulfilment of the requirements for the degree of Doctor of Philosophy in Network Science

Supervisor: Prof. Márton Karsai

Vienna, Austria
2024

Abstract

Socioeconomic status (SES) has a profound influence on human life, shaping health outcomes, social mobility, and access to resources. Beyond these well-recognized effects, SES also impacts how people navigate physical spaces and build social connections, contributing to patterns of segregation and social cohesion. Recent advancements in large-scale digital data offer an opportunity to explore these interactions in unprecedented detail. This thesis combines digital, traditional, and innovative data sources to analyze the connections between SES, mobility, and social networks.

First, we introduce a methodology for observing SES patterns in large-scale mobility and social network data by aligning two sources of individual digital traces with socioeconomic maps. Building on the first source, we explore how the COVID-19 pandemic influenced social and spatial interactions, finding that while mobility segregation increased as expected under movement restrictions, social segregation decreased, with individuals maintaining broader social ties to offset reduced physical contact. Finally, using the second source of combined digital and socioeconomic data, we investigate individual deviations from the Exploration and Preferential Return (EPR) model, revealing that the model's accuracy varies across different populations, with significant biases emerging among specific sociodemographic groups.

In sum, this thesis contributes to understanding SES's role in mobility and social networks, providing insights for policies that promote inclusion, resilience, and equitable access to resources.

Acknowledgements

I would like to thank my supervisor, Márton Karsai, for his constant support and encouragement, especially through the most challenging times of my PhD journey. His confidence in me remained steady, even when I struggled to see my own path clearly. I am also grateful for his unique ability to bring order to my disorder and rigor to my confusion, guiding my work with knowledge, patience, and clarity. Márton's kindness and attentiveness extended beyond academic guidance; his sincere dedication to the well-being of each of his students was always evident, and I am profoundly grateful to have experienced this support. His incredible passion for his work and tireless efforts have often left me wondering how he manages it all, yet they remain a source of great inspiration to me. I look forward to continuing our relationship in the years to come, both professionally and personally.

I would like to thank my collaborators, Vedran Sekara, Manuel García-Herranz, and Esteban Moro, for their invaluable insights, support, and dedication throughout this research. I would also like to thank UNICEF, the Directorate of Science, Technology, and Innovation of Sierra Leone, and Cuebiq for providing access to the data. Thanks to the Accelnet-Multinet project for its support during my research visit. Moreover, I would like to thank the MIT Media Lab and especially Esteban Moro for hosting and welcoming me during my research visit and making it an incredible experience.

I would like to thank Prof. Ronaldo Menezes and Prof. Marta González for their constructive and insightful feedback on my thesis, which greatly enhanced its quality and clarity.

I would like to thank all the Faculty of the Department of Network and Data Science and Olga Peredi for their support.

Finally, a huge thank you, in no particular order, to Viola e Axelle, Adri, Sofi e Didine, i cugini e gli zii, Luca e Mino, gli Scarra, Fra e Peppo, Fralla, i Tosti, Mamma e Papà, and all the past and present students, postdocs, and visitors of DNDS.

Declaration of Authorship

I, the undersigned, Ludovico Napoli, candidate for the PhD degree in Network Science declare herewith that the present thesis is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright. I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Vienna, January 29, 2025

Ludovico Napoli

Signature

Copyright Notice

[1]

[1]Icon by Font Awesome

List of Publications

Napoli, L., Sekara, V., García-Herranz, M., & Karsai, M. (2023). Socioeconomic reorganization of communication and mobility networks in response to external shocks. Proceedings of the National Academy of Sciences, 120(50), e2305285120.

Napoli, L., Karsai, M., & Moro, E. (2024). One rule does not fit all: deviations from universality in human mobility modeling. arXiv-2410.23502

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context

Socioeconomic inequalities remain one of the most pressing global challenges, influencing a wide range of outcomes, including health, education, and quality of life. Addressing these inequalities is essential for promoting inclusive societies and achieving sustainable development, as recognized in the United Nations Sustainable Development Goals (SDGs), particularly Goals 1 (No Poverty), 3 (Good Health and Well-being), 4 (Quality Education), and 10 (Reduced Inequality) [6]. Understanding socioeconomic disparities and their impacts is crucial, as they can create feedback loops that reinforce existing social divisions and hinder social mobility. Studies across disciplines emphasize that socioeconomic inequalities not only affect individual prospects but also shape collective outcomes by impacting social cohesion, access to resources, and overall economic stability [7–9].

Socioeconomic status (SES) is a construct that represents an individual's or group's economic and social position relative to others. The conceptualization and measurement of SES are among the most contested topics in social science due to its broad and often elusive definition [10–12]. Indeed, there is no consensus on a single, definitive definition of SES. In some instances, SES is measured based on a theoretical framework that associates it with perceived social position and the personal evaluation of occupational prestige [13, 14]. In other cases, a more objective definition of SES is used, relying solely on direct measurements of well-defined characteristics such as income and education level [15, 16]. This thesis does not aim to resolve these complex debates or take a particular stance. Instead, we acknowledge that the appropriate definition of SES often depends on the specific research question or the data available [17]. Moreover, determining the most explanatory definition of SES for a given outcome can be the central focus of a study. Here, when referring to high and low SES, we primarily consider the most commonly used variables in the literature, such as income, wealth, education, and employment status. However, we also adopt a broader definition of SES that includes demographic traits like gender, age, and ethnicity. These traits are not only interrelated with socioeconomic variables but also play complex roles in shaping individual behavior and collective phenomena, such as social segregation. In some contexts, these demographic attributes amplify socioeconomic disparities, intensifying their effects on mobility and social networks. In others, they act more as correlates of SES, reflecting associated patterns without necessarily driving them. This distinction underscores the need to carefully consider their roles depending on the specific research question and context.

The variables constituting SES are generally neither equally nor equitably distributed, often interconnected through complex and multifaceted mechanisms. Wealth [18] and income [19, 20] are unequally distributed across all population scales, with clear indications of rising inequalities [21]. The share of economic resources owned by the wealthiest individuals is increasing globally [22]. Economic inequalities have direct consequences for education and opportunities, where the more advantaged individuals have better access to higher quality education [23, 24]. This, in turn, leads to unequal access to employment and the job market, creating a vicious cycle that hinders social mobility and meritocracy [25, 26]. Moreover, persistent social norms and discrimination exacerbate existing inequalities, particularly in terms of gender and race. Women and ethnic minorities face greater barriers to education, fewer work opportunities, and pay gaps [27–31]. The complex interplay between these variables prevents a straightforward stratification into social classes, resulting in nontrivial definitions of vulnerable groups.

The study of inequalities is crucial not only for understanding their relationship with sociodemographic characteristics and the mechanisms that reinforce these disparities but also, and perhaps more importantly, for examining their impact on various outcomes of human life [9]. Among these outcomes, physical and mental health stand out as particularly significant [32]. On average, individuals with higher socioeconomic status enjoy better health across multiple dimensions and for various reasons. They tend to have a longer life expectancy [33–35], lower mortality rates [36], and a lower prevalence of chronic diseases [37]. Additionally, they are less likely to suffer from mental disorders and mental illnesses [38–40]. Health outcomes are influenced by SES through both structural factors—such as the ability to afford quality housing [41] and greater capacity for health expenditures [42]—and socialization factors, including diet and health habits [43] and social support [44].

Beyond health outcomes, SES is intricately linked to various aspects of human behavior. The goal of this thesis is to investigate, through extensive analysis of digital data, the association between SES and two critical facets of human behavior: human mobility, i.e., how people move and explore physical space, and social networks, i.e., the creation, structure, and dynamics of social relationships.

This chapter provides the theoretical, empirical and methodological foundation necessary to introduce the results presented in the following chapters. We begin by contextualizing the study introducing the concepts of human mobility and and social networks in Section 1.2. Next, we examine the direct connections between SES and individual-level behaviors in mobility and social networks in Section 1.3, highlighting how personal SES influences physical movements and social interactions. Moving beyond individual patterns, the chapter explores emerging collective phenomena in Section 1.4, such as segregation, which arise from the aggregation of individual behaviors. In Sections 1.5 and 1.6 we address the observation, modeling and measurement of these processes, discussing key data sources, models and metrics used to study SES-related patterns in mobility and social networks. Finally, in Section 1.7, we briefly introduce the content the following chapters. This structure aims to provide a comprehensive framework for understanding the interplay between SES, mobility, and social networks.

1.2 Human mobility and social networks

Human movement has long been essential to societal development, from migration for resources to daily commutes for work and education. From the earliest days of human history, mobility has been crucial for survival, as people migrated to escape inhospitable environments, find resources, and adapt to changing conditions [45]. As societies evolved, the purpose of mobility expanded beyond mere survival to include the pursuit of opportunities, escape from conflicts, territorial expansion, and the exploration of new frontiers. The movement of people played a central role in the rise of civilizations, enabling the spread of cultures, trade, and ideas across vast distances. On a more personal level, daily movements, such as commuting for work, accessing education, or seeking leisure activities, form a vital part of modern routines, though access to these activities can differ significantly depending on individual possibilities and constraints. These movements are not only about navigating physical space but also about engaging with economic, social, and cultural opportunities [46]. The patterns of human mobility, shaped by both necessity and choice, have long influenced the way societies are structured and how individuals interact within them.

Equally fundamental to human existence is the need for social connection. From the earliest human groups to the intricate societies of today, forming and maintaining social relationships has been essential not just for survival but for the flourishing of communities [47]. Early human societies were built on close-knit ties of kinship and cooperation, which gradually expanded into broader networks that included alliances, trade partners, and cultural exchanges. These social networks have always been more than just a web of personal relationships; they are the channels through which knowledge is shared, resources are distributed, and social norms are established. In modern times, the role of social networks has grown even more complex, encompassing everything from family bonds to global professional connections. These networks not only provide emotional support and a sense of belonging but also serve as a means for economic opportunities and social mobility. The nature and strength of these connections, influenced by both individual agency and structural factors, play an important role in shaping the social fabric of communities and the opportunities available to individuals [48, 49].

Human mobility and social networks are often deeply intertwined, influencing and shaping each other [50–52]. People frequently move to places where they have pre-existing social connections, reinforcing their network ties [53–55]. Conversely, mobility itself can foster the creation of new social relationships. When individuals travel or commute for work, education, or leisure, they encounter others, and these physical interactions can lead to the formation of new social ties, which can then be predicted from their movements [56, 57].

SES plays a pivotal role in both human mobility and social networks, influencing individual behaviors and broader societal patterns. On a personal level, SES directly impacts how people move and form social connections. Additionally, the influence of SES on the places people frequent and the relationships they establish leads to notable collective outcomes, particularly in the form of segregation observed in both mobility and social networks. The following sections will explore the direct individual connections between SES and human mobility, as well as SES and social networks, while also examining the emergence of segregation. In the following sections, we will first explore the relationship between SES and aspects more closely tied to individual behavior, such as relocation patterns, city exploration habits, and the formation of personal networks and social capital. Then, we will shift to a broader perspective, examining the emergence of social segregation in mobility and social networks as a property of the system as a whole, shaped by the interplay and aggregation of individual behaviors.

1.3 Individual patterns

1.3.1 Socioeconomic patterns in human mobility

Individual movements are closely linked to both personal SES and the SES of the places visited. Whether considering long-distance relocations driven by large migration flows or short-distance movements typical of daily routines, distinct socioeconomic patterns shape how individuals navigate physical space.

Long-distance mobility. Economic opportunities and social relations are crucial driving forces behind global migrations and relocations [58]. Major centers of commerce and industry have historically attracted significant population flows [59]. Long-distance movements are driven by the availability of opportunities at a destination and can be influenced by the number of intervening opportunities encountered along the journey [60]. Wage differences between regions are a primary reason for labor migrations, often leading to flows from low-income to high-income areas [61, 62]. Higher economic inequalities tend to increase awareness of existing opportunities, motivating people to relocate [63]. Similarly, trade with more advanced economies can incentivize the start of migration processes [64]. Additionally, a variety of social linkages between locations, such as cultural affinities, former colonial connections, and family or personal networks, stimulate, direct, and sustain the movement of people [58, 65]. Beyond the socioeconomic contexts of both origin and destination, the decision to migrate is also influenced by an individual's SES. Migration intentions are closely tied to individual wealth [66], gender [67], and ethnic similarities between origin and destination regions, as well as within migrant networks [58, 68].

Urban mobility. At a different spatial scale, SES also shapes human mobility within cities, which are becoming increasingly central to social development and daily life as the global urban population continues to surge. Socioeconomic factors such as income, education, employment status, and ethnicity significantly influence urban mobility patterns.

Depending on the spatial arrangement of housing and job opportunities, wealthier individuals travel shorter distances in some cities but longer distances in others [69]. Ethnic differences also impact mobility, with variations in the number of locations visited and the extent of activity spaces [70, 71]. Additionally, individuals with lower SES, particularly in less developed areas, tend to have smaller activity spaces and travel shorter distances [72, 73]. Higher SES individuals, on the other hand, display greater diversity in their location visits, though not necessarily traveling longer distances, and tend to concentrate their activities at different times of the day [74–76].

In some cases, the poor travel longer distances but visit fewer places, indicating a complex relationship between travel distance and the diversity of locations visited [77, 78]. Transportation modes are also influenced by SES, with factors such as employment, income, car ownership, and education determining how people move within cities [79].

Gender further complicates mobility patterns, with women generally traveling shorter distances, making fewer trips, and showing less spatial diversity than men, influenced by higher levels of daytime fixity constraints and by the quality of infrastructures [80–84]. The unemployed tend to travel less overall, reflecting their limited economic engagement [85]. Additionally, areas with higher education levels and better access to urban centers see more frequent mobility [86].

Urban mobility patterns also reflect socioeconomic inequalities, as regions with more diverse mobility fluxes and earlier diurnal rhythms tend to have lower unemployment rates [87]. The interplay between SES and urban infrastructure further influences mobility, with cities varying in how strongly SES correlates with movement patterns; in some cities, this correlation is less pronounced due to equitable access to public transportation and amenities [88]. Moreover, SES affects where people spend their time, with certain points of interest, such as dining establishments, being frequented differently depending on one's income and employment status [89].

The relationship between SES and human mobility is complex, influencing movement patterns across various scales, from international migrations to movements in urban environments. Socioeconomic factors such as income, education, and employment status shape where, how, and why individuals move, reflecting broader social and economic inequalities. Moreover, as previously mentioned, mobility patterns are shaped not only by an individual's SES but also by their interactions with social contacts, which can introduce additional dynamics [50–55]. For instance, engaging with individuals from a different SES can influence mobility decisions, altering typical movement patterns in response to social, temporal, and spatial factors. Recognizing the interplay between all these dimensions is essential for addressing mobility-related disparities and creating policies that promote equitable access to opportunities and resources.

1.3.2 Socioeconomic patterns in social networks

SES is intricately linked to the structure and dynamics of social networks, profoundly influencing the resources and opportunities available to individuals. Social networks are not merely a reflection of individual social behavior; they play a pivotal role in determining access to information, economic mobility, and career advancement, thereby perpetuating or mitigating social stratification.

Personal networks. A key aspect of personal social networks related to SES is the structure of the ego network itself, though the impact of SES on network structure can vary significantly across cultural and societal contexts. Most observations on how SES influences the flow and quality of information and subsequently affects economic decisions and career outcomes stem from studies conducted in Western, urbanized societies. In these contexts, diversity is among the main structural features of social networks shaped by social status and, vice versa, facilitates access to opportunities [90]. Diversity relates both to the type and intensity of connections, ranging from intense personal relations to loose professional links, and to the ability to connect with a range of different social circles that are loosely connected with each other. Being able to reach multiple communities through various connections increases the chances of accessing diverse information and resources. Indeed, individuals with higher SES in these settings often maintain more diverse social networks and extend their connections outside their immediate community [91]. This diversity is positively correlated with economic development, as it provides access to a broader range of information and resources that can enhance individual and community socioeconomic outcomes [92]. The geographical spread of social connections, or spatial diversity, further enhances economic opportunities by allowing individuals to span across different regions. Conversely, lower SES in these societies is often associated with more homogeneous networks, which may limit access to diverse resources and reinforce existing inequalities. Age also plays a role in the diversity of personal networks, with younger people in urbanized settings being more prone to establish new and diverse connections [93].

The concepts of network diversity can be analyzed through the lenses of structural holes and weak ties. Structural holes refer to gaps between non-redundant contacts within a network [94, 95], while weak ties refer to loose connections that are important for accessing disparate information [48]. Weak ties and structural holes are strongly related, as the former serve as bridges to connect unconnected disparate groups. On the other hand, strong ties usually develop within communities, characterized by dense local networks. Individuals who bridge the gaps between various groups, known as brokers, gain access to diverse information and control over resources, which can be leveraged for socioeconomic success. Networks rich in structural holes enable access to unique information and resources, facilitating higher performance, faster career advancement, and greater rewards. Conversely, networks characterized by closure, where everyone is interconnected, may foster trust and cooperation but often limit exposure to new information, potentially restricting upward mobility. Thus, the balance between weak and strong ties in a network is crucial for leveraging the benefits of both diversity and trust. Entropy provides a well-established approach to quantifying this balance and, more generally, the diversity and heterogeneity of connections [92].

In the job market, weaker ties are more likely to connect individuals to different parts of the social structure, offering access to more diverse job opportunities [96]. In organizational settings, individuals with entropic personal networks rich in weak ties and structural holes are better positioned to connect with higher organizational levels and bridge different functions within an organization, leading to greater career success [97, 98]. However, in some contexts, strong ties are equally important; having a friend within the organization, especially a high-ability individual, can significantly increase negotiated salary outcomes and enhance career success [99, 100]. Additionally, in job allocation systems that rely heavily on personal networks, where trust and obligations are critical, strong personal connections may be more valuable than the information exchanged through weak ties [101]. The advantages of weak ties may also vary depending on SES, with different outcomes for individuals based on their initial socioeconomic position [102].

Another important structural property is centrality, i.e. the position an individual holds within a social network, which also determines access to resources and influence. An individual's centrality in a social network can be a strong indicator of their financial status [91]. Those who are central within networks, especially in organizational settings, tend to have greater influence and are often better positioned for promotions [97, 103]. However, the effects of network centrality and the associated benefits are often dependent on SES aspects, particularly gender, leading to disparities in career advancement [104]. Therefore, while centrality can offer significant advantages, these benefits are not evenly distributed and may compromise fairness and equal opportunities.

Social capital. To fully understand the interplay between social networks and SES, it is essential to consider not only the structure of connections but also the personal characteristics of those connections. Social resources—defined as the advantages one derives from one's social network—are crucial for status attainment and socioeconomic mobility. The accumulation of these resources shape one's social capital, which, along with economic and cultural capital, influences an individual's position and potential within society, where individuals leverage connections to gain advantages aligned with their social standing [49]. Transitions between these forms of capital are possible, where, for instance, social capital gained through cross-class connections can foster access to educational and employment opportunities, thereby translating into economic or cultural capital over time. For example, the degree to which individuals from low SES backgrounds are connected to those with higher SES is a strong predictor of upward economic mobility [105, 106]. Areas with greater economic connectedness exhibit higher rates of income mobility, emphasizing the importance of cross-class social ties in breaking cycles of poverty. Consequently, the social capital developed through relations and connections has significant implications for status outcomes and well-being [107–109].

The SES of an individual's contacts significantly influences their own socioeconomic outcomes, starting from childhood. Social capital within families and communities is critical for developing human capital, particularly in education. Indeed, the relationships among parents and their interactions with schools and other institutions provide a supportive environment for children's education [108]. Additionally, the SES of neighbors during childhood significantly influences educational outcomes, underscoring the long-term impact of social networks on social mobility [110]. Moreover, children who live near peers involved in early childhood intervention programs experience significant spillover effects on educational and developmental outcomes [111]. Even in later educational stages, the presence of high-achieving peers positively influences educational performances, although there is considerable heterogeneity in the effects [112].

After education, the SES of contacts continues to be important for employment status and career development. The occupational status of contacts and weak ties to higher-status individuals significantly influence occupational prestige [113, 114]. Employment status itself has a significant positive impact on the likelihood of being employed, with the effect being particularly prominent for veterans and refugees [115, 116]. Additionally, peer effects contribute to economic behaviors, such as housing market decisions, where individuals are more likely to transition from renting to homeownership if their social contacts experience positive housing price changes [117].

In conclusion, as with human mobility, the relationship between SES and social networks is complex and multifaceted. SES influences not only the structure and diversity of social networks but also the resources and opportunities that individuals can access through these networks. In turn, the interplay between network structure, the strength of ties, and the distribution of social resources shapes socioeconomic outcomes, as social capital can, under certain conditions, be leveraged or transformed into economic or cultural capital, influencing one's overall social mobility and well-being.

1.4 Emerging collective phenomena

Up to this point, we have focused on how SES influences individual behavior or population patterns, affecting both physical mobility and the structure of personal social networks. However, we have yet to explore the role of SES in the emergence of collective phenomena, i.e., patterns that arise from interactions among individuals. These patterns cannot be fully understood by examining single units alone but require a perspective that considers the system formed by interacting individuals as a whole. This approach aligns with the principles of complex systems, where collective behaviors emerge from the interplay of many interconnected elements, giving rise to new properties at the population level [118].

Specifically, the microscopic mechanisms related to status homophily and environmental constraints that limit movements and social opportunities contribute to segregation in social systems. This collective phenomenon is characterized by the social or physical isolation of individuals belonging to different social groups, who predominantly interact with others similar to themselves. In this section, we will examine how segregation forms and manifests in neighborhoods, social networks, and mobility.

1.4.1 Mechanisms of segregation formation

Homophily. Homophily is a fundamental principle in the formation and evolution of social networks, wherein individuals are more likely to form connections with others who are similar to them [119, 120]. This similarity-based connection pattern extends to human mobility, as physical contacts are a specific type of social connection, and the type of people encountered in different places strongly influences movement patterns [121]. In fact, the principle of homophily applied to human mobility results in people being more likely to visit places where they expect to meet others similar to themselves. For the purposes of this discussion, we will refer to homophily primarily as a mechanism in social network formation, while acknowledging that it similarly applies to physical spaces and human mobility, leading to comparable segregation dynamics, as we will explore further.

All dimensions of SES play a significant role in homophily-induced link formation. These include ascribed characteristics such as race [122, 123], age [124, 125], and gender [104, 122, 126, 127], as well as achieved characteristics like education [126, 128], income [129], and occupation [130]. Homophily based on these dimensions often leads to the creation of relatively homogeneous communities, as individuals with similar SES are more likely to interact and form enduring ties. Moreover, when multiple status dimensions are correlated, individuals are more likely to form relationships with those who are similar across several characteristics, leading to even greater homophily [131]. This homogeneity is further reinforced by the tendency for ties between dissimilar individuals to dissolve more quickly, thereby deepening the homophilous nature of social networks [119, 132, 133].

Homophily affects a wide range of relationships, including marriage, professional connections, friendships, and even casual interactions in physical spaces. In marriage and close friendships, homophily is particularly pronounced, often reinforcing social ties within similar SES groups [123, 124, 134]. Both work and school or college networks are shaped by homophily, where co-workers or students are more likely to engage, study, or collaborate with those who share similar backgrounds or characteristics [122, 127, 135]. This pattern extends to weaker ties and even casual encounters, where brief interactions are influenced by common SES traits [136]. This latter effect closely connects with human mobility, as such encounters are not entirely random; people are more likely to spend time in places they know are frequented by others with similar backgrounds [121]. Moreover, the impact of homophily intensifies when multiple types of relationships overlap between individuals, leading to stronger homophilous ties in multilayer networks compared to single-layer ones, a cumulative effect that further entrenches the social boundaries shaped by SES [137].

Opportunities and constraints. Despite the undeniable role of homophily in link formation, attributing the observation of homophilic links and, more broadly, social segregation solely to personal preferences would be misleading. While other link formation mechanisms like preferential attachment [138, 139] or triadic closure [140] can also lead to homophilic links, the focus here is on the environmental constraints that shape the available opportunities for contact [141, 142]. First of all, there is a cognitive limit to the number of people with whom one can maintain stable social relationships [143], and individuals tend to exhibit unique and stable ”social signatures” in how they allocate time and resources across these relationships, even as specific contacts change over time [144]. Moreover, the opportunities for forming connections are neither infinite nor evenly distributed; they are significantly shaped by the context in which we live and grow, offering limited and often biased opportunities [131]. This distinction underpins the difference between choice homophily, which reflects individual preferences for associating with similar others, and baseline homophily, which arises from the distribution of characteristics within the available pool of potential ties [119].

Geography is one of the strongest constraints on social connections; physical distance acts as a major barrier because it requires more effort to connect with those who are farther away compared to those who are nearby [145]. As a result, people are more likely to form connections with those who live close by, simply because of the higher exposure to nearby individuals [146, 147]. Residential segregation further amplifies this effect, as neighborhoods tend to be highly homogeneous in traits like race or income, leading to a much higher likelihood of connecting with similar others—not due to personal choice, but due to environmental factors [148–150]. Another strong constraint is family, where individuals have no control over the SES of their parents, siblings, or extended relatives, who influence their early social networks from birth. Family ties are typically homogeneous across many traits, though they do offer diversity in gender due to the prevalence of heterosexual couples and the equal likelihood of male and female children [119]. Additionally, settings such as schools, workplaces, and voluntary organizations naturally foster interpersonal ties through shared activities, with their demographic compositions influencing the level of similarity among those who connect [122, 131, 141]. Group size effects also drive the level of homophily in a network; larger groups offer more opportunities for heterogeneous links, while an imbalance between minority and majority groups can lead to strong baseline homophily and different perception biases [131, 151].

While the distinction between individual preferences and environmental constraints is an important consideration in understanding the formation of social networks, our focus is not on determining whether homophilic networks arise mostly from personal choices or from the opportunities available within a given context [152]. Instead, our interest lies in examining the outcome of these processes: social segregation. In fact, it is also crucial to recognize that homophily and opportunities are deeply interlinked, as the contexts shaping opportunities often reinforce homophilous tendencies, creating feedback loops that intensify segregation. By focusing on the resulting patterns of segregation, we can better understand how these dynamics manifest in various social and physical spaces, influencing the broader structure of society. In the next section, we will delve into the phenomenon of social segregation, exploring how it emerges and the implications it has for neighborhoods, social networks, and mobility.

1.4.2 Segregation

Segregation is a deeply ingrained phenomenon that manifests across various aspects of social life, influencing where people live, whom they interact with, and the experiences they encounter daily. It tends to create environments where individuals with similar traits cluster together, often leading to significant social isolation from other groups.

Residential segregation. Residential segregation, one of the most prominent forms, is characterized by the physical separation of groups within urban environments, observable at various spatial scales [153]. Quantifying this phenomenon accurately remains a challenge due to its complex, multidimensional nature [149, 154, 155]. This separation is not merely a reflection of economic disparities, which restrict residential choices, but is also deeply influenced by historical and ongoing patterns of discrimination [148]. Even as economic opportunities improve, allowing for some degree of residential integration, the persistence of racial and ethnic divides underscores the complex interplay between socioeconomic mobility and deeply rooted biases [148, 156, 157].

Seen at a broader scale, residential segregation is closely tied to economic inequalities [158, 159] and is linked to broader trends in globalization and economic restructuring [158, 160–162]. Despite some recent reductions in racial segregation [163–165], the divide along economic lines has become more pronounced, not only in the United States but also in Europe [154, 160–162].

The implications of residential segregation are far-reaching, as it perpetuates inequality and limits social mobility, particularly across generations. Children growing up in high-poverty neighborhoods often face significant barriers to economic advancement, reinforcing cycles of poverty and limiting their potential for upward mobility [166, 167]. Consequently, it exacerbates disparities in access to quality education, healthcare, and employment opportunities, further amplifying social divides [168–173].

Experienced segregation. Beyond residential segregation, which primarily reflects the physical separation of groups within urban environments, experienced segregation captures the social and economic divides that individuals encounter throughout their daily lives across various contexts, such as workplaces, shopping areas, and public spaces. Unlike residential segregation, experienced segregation considers the dynamic interactions that occur as people move through different spaces and interact with diverse groups [121, 174–176]. Importantly, the segregation people experience in their daily encounters is typically lower than residential segregation, as daily mobility and interactions in more socially mixed areas often reduce segregation [177–183].

However, experienced segregation can also be high and is not uniform across socioeconomic classes and ethnic groups. Higher SES individuals, as well as majority ethnic groups, often have more exclusive and dispersed activity spaces, reducing their chances of interacting with lower SES individuals or ethnic minorities [70, 184]. This segregation is evident in both offline activities, such as where people shop and socialize, and online interactions, where wealthier and majority groups remain more segregated [185–187]. Conversely, ethnic minorities and lower SES individuals tend to have more concentrated activity spaces, limiting their exposure to diverse groups and reinforcing divisions [188–190].

While residential segregation remains relatively static, experienced segregation is highly dynamic, fluctuating not only throughout the day but also across weekdays, weekends, and even seasons. Segregation tends to decrease during the day when individuals engage in various parts of the city, leading to temporary reductions in social and economic divides [191, 192]. These reductions are generally more pronounced on weekdays when work and other activities draw people out of their residential areas into more diverse urban spaces. In contrast, during nights, weekends, and in suburban areas with concentrated wealth or poverty, segregation often persists, reflecting the uneven impact of daily and weekly mobility patterns [177, 193].

Social network segregation. Building on the concepts of residential and experienced segregation, segregation in social networks reflects the separation between social groups in social ties, with people forming connections primarily within their socioeconomic and ethnic groups [187, 194, 195]. Core friendship networks, for instance, are more segregated by ethnicity and gender than broader, weaker ties, especially among ethnic majority members who maintain more homogeneous networks [196]. However, this pattern also extends to acquaintanceship networks, which show significant segregation by race, political ideology, and religiosity, similar to what is observed in close ties [197]. Even in broader online networks, ethnic segregation remains pronounced, with smaller ethnic groups often displaying greater diversity in their social connections due to their smaller size [198, 199].

The interaction between social network segregation and residential segregation can either exacerbate or mitigate social isolation. Social networks often reinforce spatial segregation, as individuals tend to form connections within their socioeconomic class, reflecting the composition of their neighborhoods [200, 201]. Physical barriers, such as those found in urban environments, can further fragment social networks and contribute to greater income inequality over time [202]. However, mobility patterns, such as longer commutes, have the potential to diversify social networks by connecting individuals across different socioeconomic strata, potentially reducing segregation in social interactions [53, 203]. Despite these potential mitigating factors, social networks often remain highly stratified, particularly in contexts where structural divisions like caste, gender, or income are deeply entrenched [57, 185, 204]. This segregation is particularly pronounced in wealthier individuals' communication networks, where interactions tend to remain within specific socioeconomic strata [187].

Although the composition of social ties often mirrors physical segregation, behavioral observations illustrate how social network segregation can persist independently of physical proximity. This concept is evident in communication patterns, where people frequently interact with others who share similar characteristics, and where behavioral patterns are often ascribable to common traits. For instance, in Istanbul, significant differences were found in the communication patterns between Syrian refugees and the local Turkish population, highlighting how behavioral segregation can exist even when spatial segregation is moderate [205–207]. These patterns are also observed in gender-segregated networks within organizations, where men and women form separate networks, influencing career advancement and social influence [104, 208].

1.5 Types of data and data integration

Studying the effects of SES on human mobility and social networks requires comprehensive data collection across three primary elements: (1) socioeconomic status (SES), (2) human mobility, and (3) social networks. Accurately observing these elements demands data sources that can reliably capture the nuances within each category. The accurate collection and integration of different sources is itself a broad and complex area of research, as discussed in Chapter 2. This section first defines each observational category, then examines the major data sources, highlighting their strengths, limitations, and suitability for analyzing SES, mobility, and social networks. Finally, we explore methods for combining these data sources to create integrated datasets, enabling a more complete empirical analysis of mobility and social network patterns related to SES.

1.5.1 Key Observational Categories

Socioeconomic Status (SES): SES encompasses a range of indicators that reflect the social and economic positioning of individuals or groups, including income, wealth, education, employment, and demographic factors (e.g., gender, ethnicity, age). Collecting SES data is essential, as it provides a foundational context to understand how socioeconomic factors may shape patterns in both mobility and social network structures.

Human Mobility: Studying mobility patterns is key to understanding how individuals navigate physical space, access resources, and engage in economic and social activities. Mobility data sheds light on spatial behaviors and movement patterns, which often vary based on SES, revealing insights into access to opportunities and spatial segregation.

Social Networks: Social networks capture the structure, diversity, and strength of individuals' connections. Observing these networks allows for analysis of how SES influences social connectivity, the types of relationships individuals maintain, and access to social and economic opportunities within their networks.

Data on SES, human mobility, and social networks can be collected at varying levels of granularity, from large administrative regions to individual-level data. A primary challenge in this process is balancing the need for high-resolution individual data with the requirement for a large sample size. Achieving both precision and scale simultaneously is often difficult, as they tend to involve trade-offs. As a result, a variety of data collection methods have been developed to capture information on SES, human mobility, and social networks, along with strategies for integrating these components effectively, as discussed in detail in the following subsection.

The different observational categories are inherently interconnected, with strong correlations observed between SES and the ways individuals move or connect socially, as we mentioned in the previous sections. This interconnectedness becomes particularly complex when examining the socioeconomic and behavioral aspects (mobility or social networks or both) together. If the same data source is used to observe both aspects, it can lead to trivial correlations unless handled with care. For example, using a mobile phone dataset to infer the spatial distribution of wealth and subsequently analyzing the correlation between mobile phone behavior (e.g., social connections or mobility activity) and the inferred wealth risks introducing circular reasoning or producing correlations that are not meaningful.

To address this challenge, it is preferable to use data sources that are independent when measuring these different aspects, thereby avoiding trivial correlations. An even more robust approach involves adopting multimodal data collection, where a variety of distinct data sources are used to measure each of the different observational categories. This approach not only mitigates the risk of circular reasoning but also leads to better proxies for each category, enhancing the validity and richness of the analysis.

1.5.2 Data sources

Surveys and Interviews

Surveys and interviews are among the most direct and customizable methods for collecting data on SES. They allow researchers to ask targeted questions that capture the specific SES indicators relevant to the study, such as income, education, employment status, and demographic information. A key advantage of surveys is the control they provide over data collection design, enabling the collection of detailed, nuanced, and individual-level data. Additionally, study groups can be carefully selected to ensure that the sample aligns closely with research objectives, making surveys a valuable tool for precise socioeconomic data collection [209]. However, surveys are resource-intensive, requiring significant time and financial investment, especially for large, representative samples [210]. They are also prone to participation and response biases, where respondents may give socially desirable rather than truthful answers [211].

When it comes to social network data, surveys and interviews allow researchers to gather detailed information on the types and quality of relationships, including family ties, friendships, professional connections, and the frequency or intensity of interactions. Surveys can capture important nuances, such as emotional closeness, relationship types (e.g., support, advice, companionship), and other qualitative details that are often missing from digital data sources [123, 212]. This high level of control makes surveys particularly useful for exploring specific social phenomena [122, 131]. However, social network surveys face unique challenges, including scalability and issues like recall bias, which can affect the accuracy of reported relationships, and the time-intensive nature of collecting connected social network data across large samples [96]. Recently, platforms like Amazon Mechanical Turk have enabled the collection of survey data from large, diverse, and geographically distributed samples, addressing some of the scalability limitations of traditional survey methods [213].

Research surveys are less commonly used for directly collecting human mobility data, as such data typically require either institutional efforts through census travel surveys or, more commonly, digital sources. In fact, the digital data revolution has fundamentally changed the study of human mobility, greatly expanding research possibilities. The advantages of using digital data sources often outweigh their limitations. Indeed, research on human mobility surged following the availability of GPS and mobile phone traces, as detailed later in this section [214]. Just as with mobility data, the collection of social network data has evolved significantly, transitioning from traditional survey methods to more sophisticated digital data collection techniques. However, while digital data has been transformative, its impact on social network studies has not been as groundbreaking as in the field of human mobility, even though it has played a key role in the rise of computational social science [215]. The complexity of social networks, which includes aspects such as tie strength, relationship types, and actor attributes, remains fundamental to measure and is often better captured through controlled studies and surveys.

Despite all the challenges, surveys and interviews remain essential tools for capturing both SES and social network information, though their cost, time demands, and certain biases can limit their applicability for large-scale or real-time monitoring.

Census and Administrative Data

Censuses and administrative data provide a comprehensive, population-level view of socioeconomic conditions, contrasting with the customizable but limited scope of research surveys. National and international agencies conduct censuses and large-scale surveys, gathering extensive data on key indicators such as income, poverty indices, education, employment, and demographics across entire populations. These sources are invaluable for identifying large-scale trends; for instance, the American Community Survey (ACS) provides annual updates on socioeconomic conditions in the U.S., while the Demographic and Health Surveys (DHS) offer detailed demographic and health data across many developing countries, supporting longitudinal SES studies [216, 217]. Administrative data sources, including tax records, social security databases, and healthcare utilization data, are collected regularly and provide high-accuracy insights into specific socioeconomic areas [167, 218, 219]. Additionally, real estate prices serve as an indirect but powerful indicator of SES, reflecting neighborhood wealth, access to resources, and general economic conditions [220].

In addition to socioeconomic data, census data and government-led travel surveys have historically been used to capture human mobility patterns over extended periods and across large populations. These datasets provide a macro-level view of mobility, capturing information on long-term residential changes and commuting behaviors, often detailing work locations or changes in residence [221–223]. This macro-level insight is especially useful for understanding broader mobility trends, although these sources are less effective for capturing dynamic or real-time changes in movement patterns.

While both censuses and administrative data offer extensive coverage and reliability, they are not without limitations [224, 225]. Censuses are typically costly and conducted infrequently, which can result in outdated data, particularly in developing regions. Administrative data, though more timely, is often restricted to predefined categories and geographic areas, limiting its flexibility compared to surveys. Additionally, census and travel survey data generally have low temporal resolution and may focus primarily on work-related movements, making them less suitable for detailed, dynamic mobility analysis. Despite these constraints, they remain essential for large-scale, longitudinal analysis, providing critical context for understanding socioeconomic patterns and long-term mobility trends.

Mobile Phone Data: Call Detailed Records (CDRs)

Call Detail Records (CDRs) from mobile phone data offer an innovative, scalable approach to studying both socioeconomic status (SES) and behavioral patterns, providing extensive, individual-level information on a large scale. Collected primarily for billing purposes, CDRs capture information on call frequency, duration, location, and airtime purchases, offering insights into socioeconomic indicators at individual and regional levels [2, 226, 227]. By analyzing these usage patterns, researchers can estimate wealth and SES, making CDRs particularly valuable in contexts where traditional data sources are unavailable or outdated. Additionally, the high temporal resolution of CDRs enables near real-time monitoring, offering a dynamic view of socioeconomic changes and shifts in behavior over time.

In the study of human mobility, CDRs provide spatially detailed insights by logging user locations each time a call or text is made, enabling researchers to model movement patterns across large populations [228–230]. Furthermore, CDRs are a powerful tool for mapping social networks, as they capture who communicates with whom, along with the frequency and duration of interactions, thus providing a large-scale, real-time view of social ties and network dynamics [231, 232]. This makes CDRs particularly useful for understanding the strength and structure of social connections [233, 234].

Mobile phone data presents several challenges. CDRs typically lack information on the context, content, and quality of interactions, limiting the depth of analysis, particularly in social network studies. Privacy concerns are also significant; although anonymization can help, the individual-level granularity of CDRs raises ethical issues around data confidentiality [235]. Access to CDRs is often restricted, as they are owned by telecommunication companies, which may only provide data through costly and restrictive agreements. Additionally, mobile phone ownership and usage vary across population segments, especially in low-income or rural areas, introducing potential biases and limiting representativeness [236–239]. Moreover, associating a mobile device with a unique individual can be misleading, as multiple users may share a device, further complicating data interpretation [240]. Finally, while CDRs offer extensive mobility insights, they remain incomplete, as user locations are only logged during a call or message activity, potentially overlooking periods of inactivity or travel.

Despite these limitations, CDRs remain an invaluable tool for studying SES, mobility, and social networks, offering detailed, large-scale insights that support dynamic and spatially refined analyses.

Remote Sensing

To overcome the limitations of traditional socioeconomic data collection methods, particularly in terms of outdated information and resolution, satellite imagery has emerged as a valuable tool for inferring socioeconomic conditions. Nighttime light (NTL) data, for example, has been widely used as a proxy for economic activity, with brighter areas typically indicating higher levels of development [241, 242]. However, NTL data can be less effective in capturing variations in poorer areas, which are often uniformly characterized by low levels of lighting. To address this limitation, combining NTL with daytime satellite imagery has proven effective, as daylight images provide rich information about landscape features [243]. For instance, objects such as buildings, roads, and other infrastructure can be detected from daytime satellite imagery and used as proxies for wealth and development [244–246]. These remote sensing approaches offer the advantage of potentially continuous, real-time data collection across extensive areas at a relatively low cost, due to the availability of temporally and spatially extensive remote sensing data.

Such methods have been particularly effective in low-income regions, where traditional data collection methods are often expensive and logistically challenging [2]. However, the indirect nature of remote sensing requires combining these data with ground-truth information from surveys or administrative records to enhance accuracy and capture the nuances of local socioeconomic conditions. As mentioned earlier, such ground-truth data is often outdated, coarse-grained, or entirely unavailable in many regions. Furthermore, despite the high accuracy achieved through deep learning models trained on remote sensing data, their lack of interpretability presents challenges in policy-making contexts, emphasizing the need for more interpretable models that can inform decision-makers effectively [3].

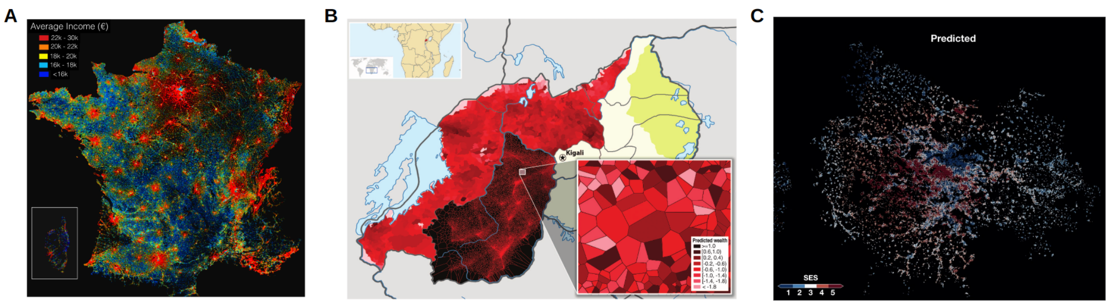

Figure 1.1: Socioeconomic data from different sources. A) Average income in France coming from the census (figure taken from [1]. B) Wealth predicted from CDRs data in Rwanda (figure taken from [2]). C) Average income predicted from satellite images in Paris (figure taken from [3]).

Social Media Data

Social media platforms provide an innovative means of deriving socioeconomic data and analyzing human behavior on both individual and aggregate levels. By examining aspects such as language use, sentiment in posts, network structures, and online behavior, researchers can infer various demographic and socioeconomic characteristics, including economic well-being and unemployment rates [1, 87, 247–250]. Compared to mobile phone data, social media data is generally more accessible and cost-effective, making it an attractive option for research. For instance, Twitter's API, widely used for human behavior studies, has historically offered broad access, though recent restrictions have limited availability. Additionally, the Facebook Marketing API offers unique insights for socioeconomic research by providing aggregated demographic data on users, which can be useful for examining socioeconomic distributions and trends across different regions [251].

Social media platforms also enable researchers to observe mobility patterns through geotagged posts, check-ins, and location-sharing services, which provide insights into urban mobility and place-based activities. Platforms such as Twitter and Foursquare allow researchers to infer movement patterns, though these insights can be biased due to selective user engagement and platform-specific behavior [252, 253].

In terms of social networks, social media data allows for large-scale mapping of social interactions, facilitating the tracking of information exchange, analysis of interaction patterns, and even testing of traditional social network theories, such as the "six degrees of separation" [202, 254, 255]. Social media data is often rich in detail, capturing not only who is connected to whom but also the context and content of interactions. For example, 'likes,' comments, shares, and tags can reveal relationship types and strengths. However, this type of data presents challenges: it is typically platform-specific, subject to privacy constraints and access restrictions, and may not accurately reflect offline social ties [256, 257]. Furthermore, the nature of interactions is influenced by each platform's specific norms and functionalities, which may not fully capture real-world social behaviors.

Social media data also faces representativeness issues that are more pronounced than those seen in mobile phone data. Internet access and social media adoption vary significantly, particularly in low-income areas, introducing potential biases in socioeconomic inference. As a result, social media data may underrepresent populations in poorer regions or those with limited internet access, thus limiting its comprehensiveness for assessing SES across diverse communities.

Emails

Email exchanges provide a unique opportunity to analyze large-scale networks of social interactions, offering insights not only into the frequency and structure of communication but, in some cases, the content itself. This can allow researchers to infer the nature of relationships through textual analysis techniques, identifying patterns of information exchange and social ties [258]. However, email data is highly sensitive and confidential, which presents significant challenges for collection and analysis due to stringent privacy concerns. Despite these barriers, some notable exceptions exist, such as the Enron email dataset, which has served as a valuable resource for studying organizational communication networks under special circumstances [259].

GPS Data

GPS data offers highly accurate and continuous tracking of individual movements, surpassing the resolution and detail of CDRs, and enabling in-depth study of travel behaviors and patterns. This data can be collected from various sources, including smartphones, vehicle tracking systems, dedicated GPS devices, and increasingly through app aggregators, which compile GPS data from multiple mobile applications to create extensive datasets [260, 261]. GPS data is particularly valuable for analyzing short-term mobility patterns in urban areas, providing high spatial resolution that allows for detailed insights into urban mobility [262–264].

However, GPS data presents challenges similar to those of CDRs, especially concerning privacy and potential biases. The use of app aggregation raises additional privacy concerns, as data from multiple applications is collected and combined, potentially without full user awareness. Additionally, GPS tracking can be limited by battery constraints on mobile devices, and user behavior may vary based on factors such as device usage patterns and app permissions, which can introduce biases in data representativeness.

Banknote tracking

An interesting, though less common, method for studying human mobility is tracking the movement of banknotes. Projects such as the "Where's George?" initiative provide unique insights into human mobility by monitoring the circulation of paper currency [265]. This method has been effective in understanding long-distance travel patterns but lacks the spatial granularity and real-time accuracy offered by other data sources, making it less suitable for studying everyday mobility [266].

Experimental Studies and Sensor-Based Approaches

Recent advances in wearable technology and sensor-based data collection have enabled more direct and accurate measurement of social interactions and movements. Several notable studies have employed such methods to understand the dynamics of social networks:

- **Reality Mining** [267]: One of the earliest large-scale social network studies using digital data, the Reality Mining project tracked the interactions of 100 MIT students and staff over nine months using mobile phone data. By combining Bluetooth proximity data, CDRs, and survey information, the study provided a comprehensive view of both the physical and social networks of participants, demonstrating how digital traces can be leveraged to map social connections and study behavioral patterns.

- **SocioPatterns** [268]: This interdisciplinary project focuses on collecting high-resolution face-to-face interaction data using wearable sensors. These sensors, often embedded in badges, detect close-range interactions between individuals, providing detailed information about who interacts with whom, for how long, and in what context. SocioPatterns has been applied in various settings, such as schools, conferences, hospitals, and workplaces, enabling the analysis of how social networks form and change in different environments.

- **The Copenhagen Networks Study** [269]: This large-scale experiment collected data from over 700 university students using smartphones, capturing both social interactions and mobility patterns over an extended period. The study combined Bluetooth signals, location data, CDRs, and questionnaire data to map out participants' social networks with high temporal resolution, providing rich insights into how social ties evolve over time.

- **DyLNet** [270]: DyLNet is a large-scale longitudinal social experiment designed to study the relationship between socialization and language development in preschool children. Over three years, DyLNet tracked the proximity interactions of around

200 children and adults every five seconds using RFID sensors, along with socio-demographic and language survey data. This comprehensive dataset offers a unique view into the co-evolution of social and linguistic networks, capturing real-world interactions in classroom and play settings.

These experimental studies represent a significant leap forward in our ability to collect social network and mobility data, capturing real-world, face-to-face interactions with unprecedented granularity and accuracy. However, they also present challenges, such as the need for participant compliance, concerns about data privacy, and the potential for behavior to be altered by the awareness of being monitored.

Multimodal Data

As mentioned in Section 1.5.1, the different observational categories are deeply interconnected, so using one data source to observe both the socioeconomic and the behavioral aspects can be risky and may lead to circular correlations. One effective approach to mitigate this risk is to use multiple data sources as proxies not only for the different categories but also within each category. For behavioral aspects, combining multiple sources at the individual level is often very challenging, primarily due to privacy concerns that make it nearly impossible to deanonymize and link datasets. However, at an aggregate level, such combinations are feasible—for example, when analyzing aggregate mobility flows or social connections between spatial regions.

On the other hand, a combination of techniques and data sources has been extensively used to infer SES, aiming to balance the limitations of some methods with the strengths of others. By integrating data from multiple sources, such as high-resolution satellite imagery, social media, and mobile phone data, researchers can overcome the weaknesses inherent in individual methods, achieving more accurate and representative SES estimates [271–273]. Ground-truth data from accurate surveys, combined with complex multi-step machine learning models applied to diverse data sources, has further enhanced the accuracy and representativeness of SES inference.

1.5.3 Data combination

Now that we've discussed how to collect the two essential elements—SES data and behavioral data—it is crucial to explore how to combine these elements to effectively link observed behavior to SES. Ideally, the most direct and informative approach would involve access to complete data at the individual level, where each person's SES variables are paired with exhaustive details about their social contacts and movements. This would include comprehensive information on all SES indicators, a fully detailed social network with exact contact relationships and tie strengths, as well as an accurate log of all places visited along with the duration of stay.

However, such ideal data is rarely available due to practical limitations, including privacy concerns, data accessibility, and the challenges of collecting such extensive information. Consequently, various data merging and combination techniques are used to approximate this ideal scenario.

One example is fully survey-based studies, where respondents are asked to provide information on their SES and social network (or, more rarely, human mobility) patterns simultaneously [123, 131], or information on both their social connections and travel history [274]. These approaches offer detailed individual-level insights into both elements but, due to the inherent limitations of surveys, are typically restricted to smaller sample sizes.

Another approach involves combining digital behavioral data, such as mobile phone usage or social media activity, with survey responses on SES [226, 237, 275]. This method enables to pairing of precise behavioral information with individual-level SES indicators, although the reliance on survey data again limits the sample size and generalizability.

For larger-scale analyses, digital behavioral data may be linked with banking transaction records, from which SES can be inferred through expenditure patterns [187, 276]. This technique offers the advantage of capturing detailed financial behaviors alongside mobility or social interactions. However, such coupled datasets are exceedingly rare due to stringent privacy constraints and the limited availability of banking information.

A more widely used method, particularly for large-scale studies, involves combining digital behavioral data with socioeconomic maps [72, 74, 121, 185, 200]. In this approach, a user's home location is inferred from their digital traces, such as the geolocation of activities captured through mobile phones or social media posts [277, 278]. This inferred location is then matched with socioeconomic indicators associated with that geographical area. By utilizing median values from fine-grained socioeconomic maps, users' SES can be estimated and analyzed alongside extensive data on mobility and social networks.

This last method is often considered optimal for studying large populations, as it provides a balance between accuracy and privacy. It allows to work with broad, representative samples while maintaining user anonymity, as the exact SES values, metadata, and precise home locations remain undisclosed. The granularity of the socioeconomic map plays a pivotal role in ensuring the accuracy of this approach: the finer the map, the more precise the SES inference, allowing for robust analyses of the relationship between SES and human behavior.

1.6 Metrics and models

Beyond the collection and integration of relevant data, analyzing socioeconomic effects in human mobility and social networks requires both appropriate metrics and robust modeling frameworks. While data collection offers a foundation for observing behaviors and patterns, metrics provide the quantitative tools necessary for measuring key aspects of mobility and social interactions. These metrics help translate raw data into interpretable indicators that highlight differences in movement and social network structure.

Complementing these metrics, models formalize our understanding of the underlying principles and mechanisms that shape these patterns. Models enable us to link theoretical frameworks with observed data, allowing for the exploration of how people move, with whom they interact, and how these behaviors contribute to broader phenomena like segregation and inequality. Together, metrics and models deepen our insights, uncover latent patterns, and enable predictions about future behaviors based on observed dynamics.

In this section, we provide a brief overview of the metrics and models of human mobility and social networks that are most relevant to this thesis. For completeness, additional metrics and models relevant to the field can be found in Appendix A.

1.6.1 Human mobility metrics and models

Human mobility metrics and models together provide a comprehensive approach for analyzing how people navigate physical space, capturing both individual-level behaviors and broader, population-level movement patterns. Mobility metrics quantify concepts such as distance traveled, movement frequency, and spatial patterns, providing insights into both micro and macro-level mobility behaviors. On the other hand, mobility models help formalize our understanding of movement, with individual-based models emphasizing the stochastic and behavioral drivers of personal mobility, and population-level models focusing on aggregate flows between geographic locations. This section provides an overview of the most relevant metrics and models used in human mobility, particularly those pertinent to this thesis [214].

Metrics

Below, we outline the most significant metrics for studying human mobility:

- **Jump lengths:** The jump length refers to the straight-line distance between two consecutive locations visited by an individual. This metric is crucial for understanding the scale of human movement, as it reflects a combination of frequent short-distance trips and occasional long-distance journeys. The jump length between two positions i and $i+1$ can be described as:

  $$d_{i,i+1} = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$$

  where (x_i, y_i) and (x_{i+1}, y_{i+1}) are the coordinates of the respective locations. In many studies, the distribution of jump lengths follows a heavy-tailed form, indicating that long journeys, though rare, still occur with non-negligible frequency [266].

- **Radius of gyration:** The radius of gyration measures the spatial extent of an individual's movement by quantifying the average distance between visited locations and their center of mass. It captures how dispersed or localized a person's movements are. The radius of gyration R_g is defined as:

  $$R_g = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\vec{r}_i - \vec{r}_{\text{cm}})^2}$$

  where N is the number of locations visited, $\vec{r}_i$ is the position of the i-th location, and $\vec{r}_{\text{cm}}$ is the center of mass of the individual's movements, given by:

  $$\vec{r}_{\text{cm}} = \frac{1}{N} \sum_{i=1}^{N} \vec{r}_i$$

  This metric often follows a heavy-tailed distribution, indicating that some individuals exhibit highly localized movement patterns, while others travel widely across larger areas [228].

- **Most frequent locations:** Identifying the locations that an individual visits most frequently, such as home, work, or favorite social venues, helps to decode their daily routines. Typically, the frequency of visited locations, ranked from most visited to least visited, follows Zipf's law. More specifically, the visitation frequency f_k of the k-th most visited location is proportional to $k^{-\gamma}$, where γ is a parameter determined from empirical data [279].

- **Inter-event time:** Inter-event time refers to the time interval between two consecutive movements or transitions between locations. This metric captures the temporal dynamics of mobility by measuring how much time elapses between successive events, which can include both the time spent at a location and the travel time to the next one. Inter-event times often follow heavy-tailed distributions, indicating that short intervals of activity are common, but longer periods of inactivity also occur occasionally, reflecting the irregularity of human movement patterns [266].

Individual-level models

Individual-based models attempt to simulate and reproduce the paths and decision-making processes of individuals as they move through space. These models are especially helpful for analyzing short-term, spatially fine-grained, and local movements, and the patterns of day-to-day travel behaviors.

- **Brownian motion:** Brownian motion models random movement where each step is taken in a random direction [280]. While useful as a simple baseline, it often fails to capture the structured patterns of real human mobility. Lévy flights, on the other hand, offer a better fit by combining frequent short trips with occasional long-distance moves, reflecting observed human behavior [281].

- **Continuous-time random walk (CTRW):** Unlike Brownian motion, CTRW incorporates waiting times between movements, reflecting the fact that individuals typically spend time stationary between trips [266]. This model is well-suited for understanding urban commuting or daily mobility patterns, where people regularly alternate between periods of movement and stationary activity.

- **Exploration and preferential return (EPR):** While continuous-time random walks (CTRW) account for waiting times, they fall short of capturing individuals' strong tendency to revisit familiar locations frequently. The EPR model addresses this limitation by incorporating two core mechanisms: exploration and preferential return, enabling it to replicate scaling laws observed in human mobility patterns [279].

  In the EPR model, exploration reflects an individual's likelihood of visiting new locations. This probability decreases as the individual accumulates more unique locations in their travel history, indicating a saturation effect where the number of new places visited grows sublinearly with time. This diminishing exploration probability aligns with empirical evidence showing that people gradually settle into a routine of familiar places, with exploration playing a smaller role as this routine stabilizes.

  The preferential return mechanism captures the strong propensity for individuals to return to previously visited locations, such as home, work, or other frequently visited sites. Each location's attractiveness, or likelihood of revisitation, increases with the frequency of past visits. This mechanism allows the model to replicate cyclical travel routines and to account for the spatial and temporal regularity observed in empirical mobility data.

  By combining exploration and preferential return, the EPR model successfully reproduces notable scaling laws in human mobility at the aggregate level. Specifically, it explains the sublinear growth in the number of unique locations visited over time. Additionally, the model captures the Zipf distribution of visitation frequencies, where a small number of locations are visited very frequently while most locations are visited only occasionally. These scaling properties reflect a balance between novelty-seeking behavior and habitual routines, making the EPR model a robust tool for modeling real-world human mobility patterns. Due to its robustness, the EPR model has inspired extensions incorporating factors such as physical constraints [282], recency effects [283], ranking mechanisms [284], place relevance [285], routine behaviors [286], and social preferences [121]

Population-level models

Population-level models aggregate individual behaviors to understand collective flows across larger geographic regions. These models are better suitable for analyzing temporally aggregate movements between neighborhoods, cities, regions, or countries.

- **Gravity models**: Gravity models are widely used in human mobility studies to predict the flow of people W_{ij} between two locations i and j. Inspired by Newton's law of gravitation, these models assume that movement between two locations is positively influenced by the "mass" (or population size) of each location and inversely influenced by the distance between them [145]. The basic form of the gravity model is:

$$W_{ij} = C\frac{N_i^{\alpha}N_j^{\beta}}{d_{ij}^{\gamma}} \tag{1.1}$$

  where:

  - N_i and N_j represent the population sizes of locations i and j,
  - d_{ij} is the distance between the two locations,
  - C is a proportionality constant, and
  - α, β, and γ are exponents typically fitted from empirical data.

  In this model, population size acts as an attraction factor: the larger the populations N_i and N_j, the more likely people are to travel between these locations. Distance d_{ij} functions as a deterrent, with longer distances generally reducing the likelihood of travel between two places. The parameters α, β, and γ are crucial for capturing the specific dynamics of human mobility. Typically:

  - α and β capture the sensitivity of movement flow to the populations of the origin and destination locations, respectively. These values reflect how strongly population density or size influences the propensity for travel, often with larger exponents indicating a higher dependence on population size.
  - γ reflects the friction of distance, with larger values implying a sharper decline in travel likelihood as distance increases.

  By fitting these parameters to empirical data, gravity models can be adjusted to reflect the unique travel behaviors within different regions or countries.

  Gravity models have been successfully applied to analyze various types of human flows, including commuting patterns, migration, and even the spread of infectious diseases. Despite their simplicity, they provide an intuitive framework to understand how both urban centers and rural areas interact, often showing that highly populated locations (such as cities) exert a stronger "pull" on surrounding regions.

1.6.2 Social network metrics and models

Metrics and models of social networks provide a structured framework for analyzing how individuals form social ties and how these relationships scale up to create complex social structures. Social network metrics quantify key features such as connectivity, clustering, and centrality, while social network models capture the microscopic rules that govern tie formation and evolution, such as homophily and preferential attachment, which help explain the fundamental principles shaping social networks. Together, these tools allow us to explore both the formation of social networks and the emergence of macro-level patterns. This section reviews essential metrics and models used to study social networks [287].

Metrics

Below, we outline the most significant metrics for studying social networks.

- **Degree and degree distribution:** The degree of a node represents the number of direct connections (edges) it has to other nodes. In directed networks, this can be broken down into in-degree (incoming connections) and out-degree (outgoing connections). The degree distribution gives a probability distribution of degrees across the network and is crucial for understanding the network's structure. For example, random networks have a Poisson degree distribution, while scale-free networks follow a power law. The degree of a node k_i can be written as:

$$k_i = \sum_j A_{ij}$$

  where A_{ij} is the adjacency matrix. The degree distribution helps identify whether the network has highly connected hubs or is more evenly distributed in terms of node connectivity.

- **Link weight:** In many social networks, edges or links between nodes are not binary but weighted, representing the strength or intensity of a connection. For example, in communication networks, link weight might represent the number of calls, messages, or emails exchanged between two individuals. This adds an extra layer of information by not only counting the existence of a connection but also measuring its strength. The total weight of connections for a node i (weighted degree) is given by:

$$w_i = \sum_j w_{ij}$$

  where w_{ij} is the weight of the link between nodes i and j. Link weights provide insights into the intensity of interactions in a social network and can highlight strong ties versus weak ties.

Network models

Below, we review the most notable network models that are suitable for social network studies.

- **Erdős–Rényi (ER) random graph**: The ER model is one of the simplest models for generating random networks, where each pair of nodes is connected with a fixed probability [288]. This model assumes that all connections are equally likely and independent of each other, resulting in networks with a Poisson degree distribution. While it provides a useful baseline, it fails to capture the clustering and degree heterogeneity observed in real social networks, making it most valuable for benchmarking more sophisticated models.

- **Barabási–Albert (BA) scale-free network**: The Barabási–Albert model explains the emergence of hub nodes in networks through the principle of preferential attachment, where nodes with more connections are more likely to attract new links [138]. This process generates networks with a power-law degree distribution, meaning a few highly connected nodes dominate, while most nodes have few connections. This model accurately reflects real-world social networks, where a small number of individuals (e.g., influencers or leaders) accumulate a disproportionate number of connections.

- **Configuration model**: The configuration model generates random networks that preserve a given degree distribution, ensuring that each node retains its pre-assigned number of connections [287]. By maintaining the degree heterogeneity observed in real-world social networks, the configuration model provides an essential tool for studying the effects of degree distribution on network structure. It serves as a valuable benchmark for comparing how other factors, like homophily or social influence, contribute to network formation beyond simple structural properties.

- **Homophily models**: Homophily models explain how individuals tend to form connections with others who share similar attributes. This principle, as we have seen, is a key driver of clustering and network segregation. By modeling tie formation based on shared characteristics, homophily models help explain the emergence of tightly knit, homogeneous groups within social networks and the resulting social stratification [139, 289].

- **Gravity model**: Despite being mostly used in the context of mobility flows, the gravity model described in the previous section can be applied to social networks as well [290]. Indeed, as we have seen, social ties are strongly related to physical distance, as people living nearby are more likely to establish a link than people living away [234, 291, 292]. Therefore, the strength of social links between locations can be also modeled with Newton's gravity law.

1.6.3 Segregation metrics and models

Segregation metrics and models offer a quantitative and theoretical foundation for understanding how individuals or groups become separated within social or spatial environments. Segregation metrics measure the degree of separation across various dimensions, such as socioeconomic status, race, or geographic location, highlighting disparities within and across networks or regions. Complementing these metrics, segregation models provide insights into the underlying mechanisms that drive separation, such as homophily and resource inequality. Together, these tools enable a deeper exploration of how structural patterns of segregation emerge and persist, as well as their broader social and economic implications.

Metrics

Below we outline the main metrics used to quantify segregation, both in mobility and in social networks.

- **Assortativity:** Assortativity is a widely used network metric that measures the tendency of nodes to connect with others that share similar attributes, such as SES, degree, or other characteristics [293]. This metric can be applied to both social networks, where nodes represent individuals, and mobility networks, where nodes represent locations and links represent the number of trips between them. A positive assortativity coefficient indicates that nodes are more likely to connect with others who have similar attributes, a clear signal of segregation. Conversely, a negative assortativity coefficient suggests disassortative mixing, where nodes connect with others who have dissimilar attributes. The assortativity coefficient ρ is given by:

$$\rho = \frac{\sum_{i,j} A_{ij}(s_i - \bar{s})(s_j - \bar{s})}{\sum_{i,j} A_{ij}(s_i - \bar{s})^2} \tag{1.2}$$

  where A_{ij} is the adjacency matrix of the network, s represents a generic socioeconomic attribute of the nodes, and $\bar{s}$ is the mean value of s across the network. This formulation essentially calculates the Pearson correlation coefficient between the attributes s_i and s_j at either end of an edge (i, j).

- **Individual assortativity:** The assortativity defined above measures a unique value for a whole network. A natural extension for an individual (node-level) metric is the generalized assortativity [294], defined for every node u as:

$$r_u = \sum_{ij} w_{multi}(i; u) \frac{A_{ij}}{k_i} \tilde{x}_i \tilde{x}_j \tag{1.3}$$

  where $w_{multi}(i; u)$ is the multiscale distribution defined in [294], A_{ij} is the adjacency matrix, k_i is the degree of node i, and $\tilde{x}_i = (x_i - \tilde{x})/\sigma$ is the standardized socioeconomic attribute considered. A positive value indicates the tendency for a node to connect to others with similar SES, while a negative value indicates the tendency to connect to dissimilar nodes.

- **Diversity:** Diversity measures the heterogeneity of a group or community by quantifying the variety of attributes present [121, 295]. It can refer to the diversity of places visited by a person or the diversity of contacts in a social network. Entropy is often used to measure diversity, with higher entropy indicating a more diverse, less segregated group [296]. Given a group of places or contacts of a person u, the diversity is defined as:

$$D_u = -\sum_{c=1}^{n} P_u(c) \log P_u(c)$$

where c is a given discrete socioeconomic characteristic (e.g., the income class), and $P_u(c)$ is the normalized frequency of visits to places or connections to people that belong to class c.

Models

Below we provide an overview of the main segregation models.

- **Residential and experienced segregation:** The most foundational model for understanding residential segregation is the Schelling model of segregation, developed by Thomas Schelling in the 1970s [297, 298]. This agent-based model illustrates how individual preferences for neighbors of similar characteristics can lead to large-scale patterns of segregation. In the basic version of the model, agents are placed on a grid, representing a city, and each agent belongs to one of two groups. Agents assess their satisfaction based on the proportion of neighbors from their own group. If this proportion falls below a predefined threshold, agents move to a new location. Even when agents have relatively mild preferences, the system tends to evolve toward highly segregated outcomes, showing how micro-level behaviors aggregate into macro-level segregation.

  The most notable result of the Schelling model is the demonstration of how even small individual preferences for similarity can lead to widespread segregation. It highlights the non-linear and emergent nature of segregation, where individual choices, without strong discriminatory intent, can still result in significant spatial separation of groups. Since the original Schelling model was proposed, numerous variants have been developed to capture more realistic aspects of segregation dynamics. Relevant extensions have considered factors such as relocation distance and place relevance [299], neighborhood awareness of agents [300], agents' income inhomogeneity [301], rewards for interactions [302], actual geographical regions [303, 304], neighborhood sizes and shapes [305], aging effects [306], different integration or relocation policies [307, 308], and urban venues [309].

  To model experienced segregation in daily mobility, the most notable approach still builds on the Schelling model. Specifically, it incorporates social preferences into the Exploration and Preferential Return (EPR) model by introducing an additional parameter that quantifies an individual's propensity to visit places where they are part of a minority group [121]. Another model examines the role of multilayered transport systems, analyzed through the frameworks of random walks and Lévy flights, in generating experienced segregation in urban environments [310]. Additionally, depending on the distribution of SES in the sample population under consideration, any of the mobility models previously mentioned can lead to experienced segregation and can serve as null models for empirical observations. For instance, in cases where nearby places have similar SES configurations and neighborhoods with significantly different groups are physically separated by large distances (as is often the case in real-world scenarios), gravity models can also result in segregation. This is because individuals are more likely to visit nearby places and, therefore, predominantly interact with those of similar socioeconomic status, as diverse individuals tend to live and visit far locations.

- **Social network segregation models**: Social network segregation models explore how social ties form between individuals based on attribute similarities, such as race, ethnicity, or socioeconomic background. Schelling-like approaches to social networks demonstrate that even a mild bias against forming ties with dissimilar individuals can result in significant segregation over time [311]. Alternatively, segregation can also arise from the assumption that individuals form or sever social ties through a utility maximization process, where agents seek to maximize their benefits from social connections [312]. Moreover, extensions of Schelling's model to more complex network structures, like random and scale-free networks, reveal that segregation can emerge across various topologies, with lower connectivity amplifying segregation effects [313]. Furthermore, analogously to mobility models, many of the previously discussed social network models can also produce segregation under certain conditions. For instance, a configuration model with a specific degree-SES distribution can naturally lead to segregation. Such models, including the configuration model, can serve as null models to test the significance of observed segregation, helping to assess whether the level of segregation in real networks exceeds what would be expected by chance or inherent structural properties.

1.7 Structure of the thesis

This thesis delves into the complex relationship between SES and human behavior, focusing on mobility patterns and social networks at both the individual and collective levels. Specifically, the following research questions are explored in the subsequent chapters:

1. **Chapter 2**: *Observation of socioeconomic patterns in mobility and social networks*
   How can we obtain accurate and representative observations of large-scale populations that account for both SES and their mobility and social network behaviors?
   This chapter develops methodologies for inferring SES and observing human mobility and social networks using digital traces, and traditional and non-traditional data sources, ensuring accurate measurement across large, diverse populations.

2. **Chapter 3**: *Socioeconomic reorganization of mobility and communication networks in response to external shocks*
   How do external shocks, particularly emergency policies, affect social and mobility network segregation?
   This chapter investigates the impact of lockdown policies during the COVID-19 pandemic, analyzing how these measures reshape mobility and social networks and whether they exacerbate or mitigate pre-existing socioeconomic segregation.

3. **Chapter 4**: *Deviations from universality in human mobility modeling*
   What are the consequences of applying universal mobility models to populations with significant heterogeneity in movement behaviors?
   This chapter critically assesses individual deviations from the universal scaling laws of the Exploration and Preferential Return model, and how these are not uniformly distributed across the population but affect specific socioeconomic groups.

Chapter 2

Observation of socioeconomic patterns in mobility and social networks

2.1 Introduction

In Chapter 1, we emphasized that a critical aspect of studying the socioeconomic effects in human mobility and social networks is data collection. Specifically, two essential types of data need to be gathered and, more importantly, accurately integrated: socioeconomic data and behavioral data related to physical movements and social connections. As discussed, there are various methods and sources to obtain and combine these empirical observations, with the choice of approach strongly dependent on the research objectives and the desired balance between scale and accuracy.

This thesis aims to study these phenomena on a large scale while maintaining an individual-level perspective. Ideally, we would have access to extensive datasets, encompassing millions of individual observations, with the following characteristics:

- A comprehensive history of places visited over an extended period, with temporal and spatial resolution, including duration of stay, type of location, and reason for visit (e.g., work, transit transfer, leisure).
- Complete information on social connections for the same set of individuals during the same period, with temporal information, spatial geolocation for both ends of the connection, the type of interaction (e.g., face-to-face, phone, online), the nature of the relationship (e.g., family, friends, colleagues), and a self-assessed measure of the relationship's strength.
- Detailed personal socioeconomic variables, ranging from income and wealth to profession, education level, demographic features (such as ethnicity, gender, and age), and precise home and work locations.

In this chapter, we present the methodological and technical solution developed to approximate the ideal configuration. We describe the characteristics of the collected data and the methodology used to integrate multiple data sources. In particular, we outline how the spatial resolution of the socioeconomic data was aligned with that of the digital behavioral data, how the two sources were combined to infer individuals' SES, how the methodology can be adapted to different types of data and contexts, and how population representativity was maintained throughout the process.

2.2 Data Description

The behavioral data used in this thesis are derived from two major sources of large-scale digital data. The first source consists of GPS location data from mobile phone devices in the United States. The second source is a dataset of Call Detail Records (CDRs) obtained from a telecommunications company in Sierra Leone. Each of these digital data sources is paired with corresponding socioeconomic data, as detailed below.

2.2.1 Mobile Phone GPS Traces

Our first source of behavioral data consists of GPS location data from mobile phone devices in the United States, collected between October 2016 and March 2017. The data was provided by Cuebiq, a location intelligence company that gathers anonymized location data from mobile applications used by opted-in users. These users gave their consent to share data in compliance with the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). The data was shared through Cuebiq's Data for Good initiative, which ensures that access is limited to qualified academic or humanitarian researchers under strict contractual agreements that prohibit data sharing or any attempt to re-identify individuals.

The GPS dataset consists of location pings, with each ping containing the following details: *user ID, latitude, longitude, date, and time*. The raw dataset comprises approximately 70 billion pings from 14 million devices. While previous work has performed much of the data filtering and processing [121], we summarize the key steps here for completeness. Initially, pings were restricted to 11 core-based statistical areas (CBSAs) [314], specifically: New York-Jersey City (New York), Los Angeles-Long Beach-Anaheim (Los Angeles), Chicago-Naperville-Elgin (Chicago), Dallas-Fort Worth-Arlington (Dallas), Philadelphia-Camden-Wilmington (Philadelphia), Washington-Arlington-Alexandria (Washington), Miami-Fort Lauderdale-West Palm Beach (Miami), Boston-Cambridge-Newton (Boston), San Francisco-Oakland-Hayward (San Francisco), Detroit-Warren-Dearborn (Detroit), and Seattle-Tacoma-Bellevue (Seattle). Additionally, only devices with at least 2,000 pings were retained, resulting in a dataset of 67 billion pings from 4.5 million devices.

Pings are simple data points that can be recorded at any time, even during transit periods. However, for our study and mobility studies in general, what is most important is to detect *stays*, i.e., single stop-over locations where a user spent some time. If multiple consecutive pings are close in space, with at least some time spent between the first and last ping, we can infer that this collection of pings represents a stay. This is the logic behind the algorithm proposed by Hariharan and Toyama [315], which represents the state-of-the-art methodology. The algorithm is governed by two parameters that represent the maximum distance between each ping and the centroid of the stay and the minimum duration between the first and last ping. Since we are interested in leveraging the very high spatial resolution of our GPS data, stays were identified when the maximum distance from the cluster's centroid is under 50 meters and the time occurring between the first and last ping was at least 5 minutes. Moreover, to avoid extremely long stays that are meaningless for our study, only stays lasting less than 24 hours were retained. Pings not being assigned to a stay were ignored. To validate the stays identified by this algorithm, previous studies, including the original paper, have cross-verified its outputs with ground truth data from known locations or comparisons with other methods [315–319].

These stays form the primary units of analysis in this study. Each stay is defined by attributes including *user ID, centroid latitude, centroid longitude, start time, and duration.* After the stay detection step and after discarding users with less than 10 stays during nighttime, necessary for the home location inference step (see further sections for details), 976 million stays from 3.6 million anonymous individuals were obtained. However, only stays that can be associated with well-defined and categorized places are worth keeping for our analysis. To do this, Foursquare venue data were used, i.e., an extensive and detailed list of 1.1 million points of interest (POIs), each assigned to a well-defined category, such as shops, cafes, and offices. This matching was performed in [121] using a nearest-neighbor search with a maximum distance of 200 meters, i.e., assigning each stay to the closest Foursquare venue; if the closest venue was more than 200 meters away from the stay's centroid, the stay was discarded. Hence, all the remaining stays have an additional *Foursquare venue* feature. Despite the simplicity of the approach, this strategy has been shown to work well for mobility data [320]. Moreover, the robustness of the methodology has been tested in [121], using different values instead of 200 meters as the maximum distance for venue coupling.

Every Foursquare venue is already assigned to a place category by Foursquare classification, which is regularly updated and consists of 592 categories such as Coffee Shop, Art Studio, Office, Building, Department Store, and Park. Consequently, each remaining stay, which was assigned to a Foursquare venue, is automatically assigned to a Foursquare category. Moreover, these categories have also been manually grouped in [121] into a taxonomy of 13 groups: Art / Museum, City / Outdoors, Coffee / Tea, College, Entertainment, Food, Grocery, Health, Education, Service, Shopping, Sports, Transportation, and Work. The detailed mapping of each place category to the taxonomy of 13 groups can be found in the Supplementary Information of [121]. Due to privacy reasons, stays associated with sensitive venues have been discarded [321]. Only venues visited by at least 20 users were included in the analysis, and only users with valid Foursquare visits were considered. Finally, for the purpose of this study, users in the bottom 20% in terms of the number of distinct visited places (less than 11 distinct visited places) were discarded, as the individual-level analysis in Chapter 4 requires a minimum amount of distinct visited places.

At the end of this pipeline, we are left with 389 million stays associated with Foursquare venues and categories from 1.5 million users. This significant reduction—nearly two-thirds from the original 976 million stays—results from the cleaning and filtering steps outlined above. These steps ensure precise home location and SES inference, sufficient statistics for user-level analysis in Chapter 4, and the inclusion of only stays with accurately assigned POIs.

2.2.2 Call Detail Records

The second source of digital data consists of anonymized Call Detail Records (CDRs) provided by a major mobile phone operator in Sierra Leone, covering the mobile phone communication activity of 1,270,214 anonymized users (16% of the country's population) between March 17 and April 17, 2020. The raw data was made available specifically for this research, following best data protection practices, through a collaboration with UNICEF's Frontier Data Technology Unit.

Each data point in the CDRs contains the following information: *caller ID, callee ID, date and time, event type, and tower ID.* The caller is the user who initiated the call or sent an SMS, while the callee is the recipient of the communication. The event type indicates whether the communication was a call or an SMS, and the tower ID is a numerical identifier of the mobile network tower that handled the communication.

CDRs are collected by mobile operators for billing purposes, so the raw data includes other details that are not relevant to our study. The anonymization of customer IDs was performed via random hashing by the provider to ensure privacy protection.

The cell tower ID is key to analyzing user mobility. By associating it with its geographical coordinates (longitude and latitude), the tower ID can serve as a proxy for the user's location at the time of communication. Unlike GPS pings, which can be generated anywhere, CDRs are restricted to fixed tower locations, which are unevenly distributed across space. While this enhances user privacy by preventing precise location tracking, it also presents challenges for SES inference, as described later in the section on spatial matching.

Despite its lower geospatial accuracy compared to GPS data, CDR data offers a significant advantage. It can be viewed as both:

- A record of users' movements in physical space.
- A network of links between pairs of users in the communication space.

This dual property allows us to simultaneously track two behavioral aspects of the same set of individuals: their mobility and their communication patterns.

Although CDR data has lower geospatial accuracy compared to GPS data, it possesses a unique quality that makes it invaluable for our study. Specifically, CDR data provides simultaneous insights into two aspects of individual behavior for the same sample of users:

- Movements in physical space, allowing for the study of mobility patterns.
- Connections in the communication space, capturing social network interactions.

This dual property enables an integrated analysis of mobility and social networks, which is not possible with GPS data alone. While GPS data offers higher spatial resolution and greater precision for tracking movements, it lacks information about social interactions. The ability to analyze both mobility and communication behaviors within the same dataset makes CDR data uniquely suited for exploring the interplay between these dimensions of human activity.

2.2.3 Socioeconomic Data

As mentioned in Chapter 1, the most accurate technique for merging our digital data with individual SES information involves direct surveys of mobile phone device owners, asking specific questions about income, education, ethnicity, gender, profession, and more. However, in our case, this approach is not feasible for several reasons:

- The data is strictly anonymized, and any attempt to de-anonymize users would not only be unethical and dangerous for privacy protection but is also strictly prohibited.
- The large scale of the data makes it practically unfeasible to conduct traditional surveys on such a vast sample of users, even without anonymization.

Therefore, we must employ an alternative methodology to infer SES. The general approach follows these steps:

1. Obtain spatially aggregated data for the desired socioeconomic indicator (e.g., median income at the regional level).
2. Infer the home location of users from their digital traces.
3. Assign users the socioeconomic indicator corresponding to the spatial unit (e.g., the region) in which their inferred home location is situated.

Below, we describe the two sources of spatially aggregated socioeconomic data that are coupled with our behavioral data.

American Community Survey

The socioeconomic data used alongside the GPS data were obtained from the American Community Survey (ACS) for the year 2017, ensuring temporal alignment with the GPS data [216]. The ACS is an annual survey conducted by the U.S. Census Bureau, providing comprehensive demographic, social, economic, and housing data across the United States. For this study, we utilized data at the block group level, which is the most fine-grained spatial resolution available for socioeconomic indicators.

The variables extracted from the ACS include:

- **Median Income**: The median income for each block group, representing the central point of income distribution.
- **Education Level**: The number of individuals with secondary education, from which we derive the fraction relative to the total population in the block group.
- **Ethnicity**: Data on the number of individuals identifying as white, black, Asian, Native American, Hawaiian, or other ethnic groups. We calculate the proportion of each group relative to the total population within the block group.
- **Means of Transportation**: Information on how individuals commute to work, specifically the number of people using cars and those using public transportation. From this, we calculate the fraction of each mode of transport relative to the total commuting population.

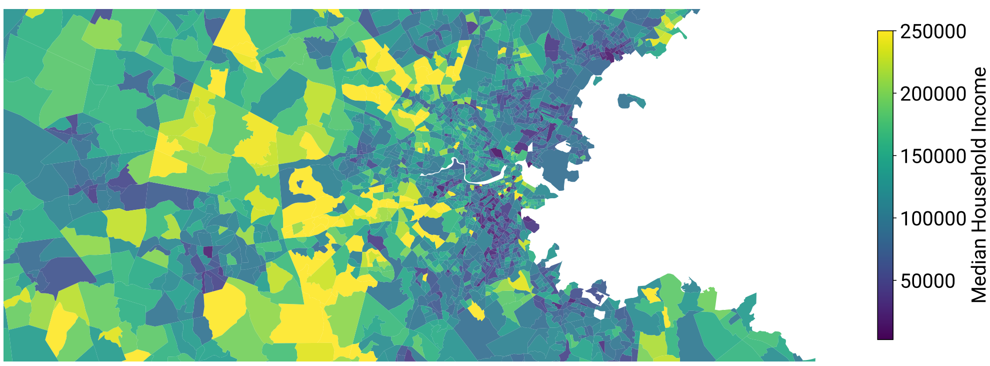

Figure 2.1: ACS data. Median household income in the Boston area at the Census Block Group level.

Although the ACS data is not at the individual level, it allows us to analyze socioeconomic patterns with a high degree of spatial precision. In urban environments, where most of our users are located, block groups often consist of only a few buildings, as seen in Fig. 2.1, ensuring minimal income variation within each block group. Thus, assigning the median income of a block group to individuals residing there provides a reliable estimate of their actual income. Similarly, for ethnicity and transportation means, the assigned values can be interpreted as probabilities—representing the likelihood of belonging to a particular ethnic group or using a specific mode of transport.

Relative Wealth Index for Sierra Leone

To complement the behavioral data derived from Call Detail Records (CDRs) in Sierra Leone, we utilized a dataset offering microestimates of wealth for low- and middle-income countries, including Sierra Leone [271]. This dataset provides a Relative Wealth Index (RWI), inferred from nontraditional data sources such as satellite imagery, mobile phone usage patterns, and topographic maps. Machine learning algorithms trained on nationally representative household surveys were used to generate wealth estimates at a spatial resolution of 2.4 km.

The RWI provides an asset-based measure of relative wealth, capturing economic disparities across small geographical areas. This approach yields a more up-to-date and granular understanding of wealth distribution compared to traditional census data. In Sierra Leone, existing census data on wealth are outdated and available at much coarser spatial resolutions, which would lead to imprecise socioeconomic indicators at the individual level. In contrast, the RWI enables a more accurate assignment of socioeconomic characteristics, reflecting local economic conditions, as illustrated in Fig. 2.2 B, where the raw data is displayed.

As discussed in Chapter 1, using nontraditional data sources such as the RWI helps overcome the limitations of traditional institutional surveys, especially in contexts where data collection is less frequent or granular. While this type of inference may not be necessary in countries like the United States, where the ACS provides regularly updated and fine-grained data, it is invaluable in regions like Sierra Leone, where traditional data collection mechanisms are limited [272].

2.3 Spatial Matching

In this section, we explain how we spatially combined the socioeconomic and behavioral data to ensure alignment. Achieving this alignment is critical, as the spatial resolution of the socioeconomic data must match that of the digital behavioral data to preserve the accuracy of our analysis. For example, if the digital data has a coarse spatial resolution, such as tens of kilometers, using a highly detailed socioeconomic map would be ineffective. It would be unclear which small socioeconomic area to assign to a user. On the other hand, if the digital data has very high spatial resolution, using low-resolution socioeconomic data would result in significant information loss, as distinct user locations would be aggregated into broad socioeconomic regions.

In the case of GPS pings, as previously mentioned, space is not restricted to predefined areas, meaning that users can, in theory, be located anywhere. While measurement uncertainties exist, they are minor and hard to quantify, especially when compared to the granularity of any available socioeconomic data. Therefore, we can confidently utilize the most fine-grained socioeconomic data available from the ACS, specifically at the block group level.

Conversely, for CDRs, space is not uniformly distributed, and users' exact locations are inherently uncertain due to the spatial distribution of cell towers. As such, we need to adopt a different methodology to match the socioeconomic data to the spatial resolution of the CDR data, which involves applying specialized geospatial analysis techniques.

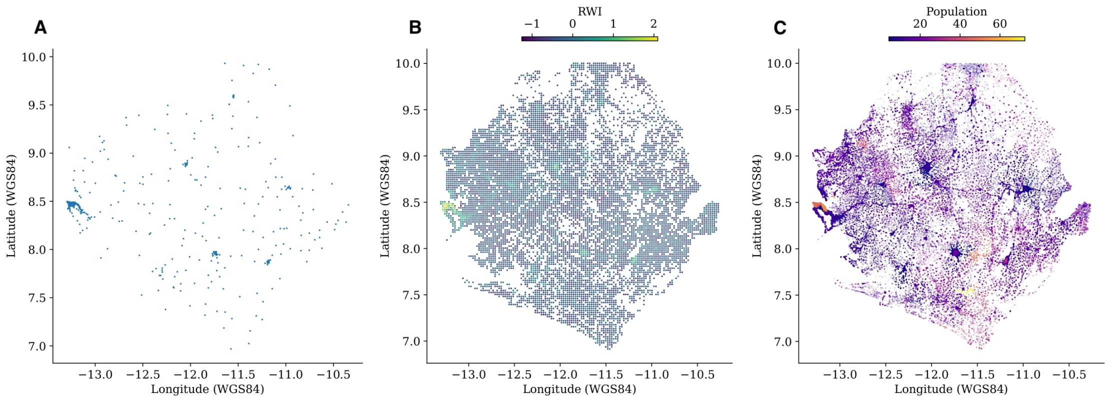

Figure 2.2: Raw geospatial data. A) Spatial distribution of cell towers. B) Spatial distribution of RWI values. C) Spatial distribution of population. All the data shown here (raw data) are points georeferenced in WGS84 (longitude and latitude).

Cell tower locations are georeferenced using geographic coordinates (longitude and latitude) in the WGS84 system. By clustering cell tower IDs with nearly identical coordinates (less than 1 meter apart), we identified 405 unique tower locations, as shown in Fig. 2.2 A. Our objective is to assign a unique socioeconomic value to each of these cell tower locations. The RWI dataset covers Sierra Leone with 8,435 grid points, each representing a 2.4 km by 2.4 km patch. These points are also georeferenced using WGS84 but are arranged on a regular square grid in the WGS84/Pseudo-Mercator projection. This means that each RWI value corresponds to a square area of 2.4 km on a planar surface. Fig. 2.2 B displays the raw RWI values, color-coded to show wealth distribution across these patches.

Our goal is to aggregate the RWI data so that each cell tower location is assigned a single corresponding RWI value. To achieve this, we need to account for population density distribution across the region. For this purpose, we utilize high-resolution population density data from Meta's Data for Good initiative [4]. This dataset provides population estimates at a 1 arcsecond grid resolution, where each point represents the number of people living within a 1 arcsecond-sided cell, which is approximately 30 meters across on a spherical surface. Fig. 2.2 C illustrates this population distribution.

To summarize, we work with three key georeferenced datasets:

1. Cell tower locations: points in WGS84.
2. RWI cells: points in WGS84, each representing a 2.4 km square in the WGS84/Pseudo-Mercator projection.
3. Population density: points in WGS84, corresponding to 1 arcsecond grid cells.

We approximate the area covered by each cell tower using a Voronoi tessellation [322]. A Voronoi tessellation divides space such that the boundary between any two adjacent regions is equidistant from the corresponding cell towers. Consequently, each tower's Voronoi cell contains all the locations that are closer to that tower than to any other. To generate this tessellation, we first project the cell tower coordinates into the UTM

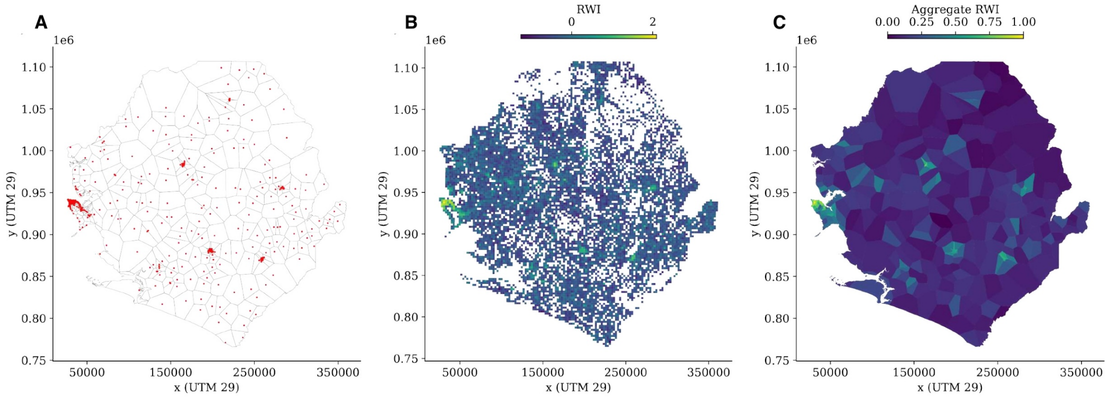

Figure 2.3: Processed geospatial data. A) The intersection between the Voronoi tessellation and the country's border (grey polygons) based on cell tower locations (red points). B) RWI square cells. C) Aggregated RWI map. All data are polygons georeferenced in UTM 29.

29 coordinate system (the system covering much of Sierra Leone). Using these projected coordinates, we apply Euclidean distance calculations to construct the Voronoi diagram. For each Voronoi cell, we then clip the boundaries to match Sierra Leone's geographic borders, as shown in Fig. 2.3 A.

The next step is to aggregate the RWI values across these Voronoi cells, taking population density into account. The RWI data points are projected into the WGS84/Pseudo-Mercator coordinate system (the RWI's native system). Each RWI point is treated as the center of a square polygon with a 2.4 km side, which we project into UTM 29 to align it with the cell tower locations, as illustrated in Fig. 2.3 B. Given the disproportionately high resolution of the population data (one arcsecond at the equator is around 30 meters, and the maximum value of a population cell is 73 people), we did not find it necessary to consider the spatial extension of the population cell grids. Instead, we treated them as points. There are two main reasons behind this choice:

1. 30 meters is several orders of magnitude higher resolution than both the RWI map (2.4 km) and the Voronoi partitions. Further, as mentioned earlier, population cells, due to their high spatial resolution, are populated by at most 73 people. Since we use the population data to weight the RWI values inside a single Voronoi cell to obtain an aggregate RWI value for the Voronoi cell (as explained in detail below), considering the area of the population cells would add or remove at most a few dozen people in the weight of a single RWI cell, which would marginally affect the computation of the aggregate RWI.

2. Treating them as points instead of polygons drastically decreases the computational cost, while representing population cells as approximately 30 m^2 areas instead of points would add only an insign

For these reasons, we treat the population data as discrete points and project them into UTM 29, ignoring the area each grid cell covers.

Now that all three datasets are aligned within the same coordinate system (UTM 29), we can aggregate the RWI values across each Voronoi cell v using the following steps. An example of this process for one Voronoi cell is illustrated in Fig. 2.4:

1. Identify the spatial intersections between the RWI cells and each Voronoi cell v (Fig. 2.4 A). These intersections are denoted as A_v, with their respective RWI values $R(A_v)$. If the square area of an RWI cell falls completely within a Voronoi

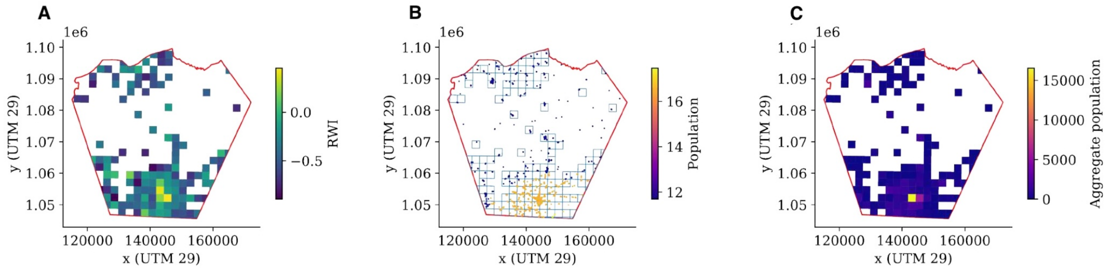

Figure 2.4: Spatial mapping. A) Spatial intersections of RWI cells (color-scaled) with a given Voronoi cell (red), located at the country's northern border. B) Population points within the intersections. C) Aggregate population within each intersection. All data are georeferenced in UTM 29.

cell v, like the cells in the middle of the Voronoi cell in Fig. 2.4 A, then A_v corresponds to the entire original square area of the RWI cell. If the square area of the RWI cell falls only partially within a Voronoi cell v, like the cells at the border of the Voronoi cell in Fig. 2.4 A, then A_v corresponds only to the portion of the RWI cell that falls within the Voronoi cell.

2. Identify the population points p within each RWI intersection A_v. In Fig. 2.4 B, population points p are colored according to their population count, and intersections A_v are also shown to visualize which points fall within each intersection.

3. Sum the population values w_p within each RWI intersection A_v to calculate the weight for RWI aggregation:

$$W(A_v) = \sum_{p \text{ within } A_v} w_p$$

   Fig. 2.4 C shows the population weights for the example cell, resulting from the sum of the population points shown in Fig. 2.4 B.

4. Compute the aggregate RWI value R_v for each Voronoi cell v by taking a population-weighted average of the RWI values:

$$R_v = \frac{\sum_{A_v} W(A_v) R(A_v)}{\sum_{A_v} W(A_v)}$$

To further clarify our methodology, we will illustrate it with a toy example. Let's consider a Voronoi cell v with only three RWI cells partially overlapping with it, having RWI values of 0.5, 0.7, and 0.9, respectively. The spatial intersections A_v of the three RWI cells are populated by weights $W(A_v)$ of 50, 100, and 120 people, respectively. The resulting aggregate RWI value R_v is given by:

$$R_v = \frac{\sum_{A_v} W(A_v) R(A_v)}{\sum_{A_v} W(A_v)} = \frac{50 \times 0.5 + 100 \times 0.7 + 120 \times 0.9}{50 + 100 + 120} = 0.75$$

The final aggregated RWI map is displayed in Fig. 2.3 C. This map now accurately reflects the spatial resolution of the CDR data, with RWI values appropriately aggregated based on population density distribution and cell tower coverage.

At this stage, we have generated two distinct socioeconomic maps that correspond to the respective digital traces. These maps offer socioeconomic indicators at the following spatial resolutions:

- **U.S. map:** Census block group
- **Sierra Leone map:** Mobile network cell tower

The next step is to couple these socioeconomic maps with the digital traces by inferring the home locations of mobile phone users and subsequently assigning SES based on the spatial units corresponding to the inferred home locations. The methodology for this coupling process differs slightly depending on the data source—GPS traces or CDRs—due to the differences in spatial resolution and the nature of the data.

2.3.1 GPS traces

For users in the GPS dataset, home locations were inferred based on the most frequently visited Census block group between 10:00 p.m. and 6:00 a.m., which is assumed to be when individuals are most likely at home. To improve the reliability of this inference, we applied a filter to include only users who had at least 10 recorded nights at their inferred home location. Additionally, to focus on users with sufficiently rich mobility patterns, we excluded those in the lowest quintile in terms of the number of distinct places visited (resulting in a minimum of 11 distinct visited places per user). This relatively strict filter, which by definition removed 20% of users, was necessary for our analysis in Chapter 4, where the focus is on understanding asymptotic mobility laws, which require robust data on the diversity of places visited.

Once the home locations were established, we assigned the corresponding socioeconomic indicators of each block group. The indicators included median income, education level, racial composition, and transportation data, which collectively serve as proxies for the users' SES.

2.3.2 CDRs

For users in the CDR dataset from Sierra Leone, the home location inference followed a similar approach. Since each communication event is linked to a cell tower rather than an exact GPS coordinate, we used the cell tower with the most activity during nighttime hours (9:00 p.m. to 6:00 a.m.) as a proxy for the user's home location. Due to the COVID-19 lockdowns and curfews in place during the data collection period, we gave double weight to activity recorded during those periods, as users were more likely to stay at home.

Given the smaller sample size and shorter observation period compared to the GPS dataset, we applied a less stringent activity filter. Users with fewer than two distinct geolocated communication events during nighttime were excluded. Additionally, we removed the top 0.5% of users who had highly anomalous behavior, such as receiving a large volume of incoming communication but recording no outgoing communication, or vice versa. Such patterns often indicate automated systems like call centers, which are unsuitable for reliable analysis.

We also introduced a filter based on spatial uncertainty, which accounts for potential errors in detecting home locations from CDR data [323]. Spatial uncertainty is measured by considering the distances between candidate home locations and the distribution of observations at each location. If the candidate locations are close together and one location has significantly more observations than the others, the spatial uncertainty is low, suggesting higher confidence in the home location inference. Conversely, if the candidate locations are far apart with similar observation counts, spatial uncertainty is high. In our analysis, users with spatial uncertainty greater than 25 km for their inferred home location were excluded.

Finally, as with users in the GPS dataset, we matched the home towers of CDR users to the RWI map created in the previous section. Each user was then assigned an RWI value based on their home tower, serving as a proxy for their SES.

2.4 Final outcome

After completing the entire data processing pipeline, we are left with the following clean, individual-level datasets that couple behavioral data with socioeconomic indicators:

1. 1,511,393 users in 11 major U.S. metropolitan areas (CBSAs), with six months of observation on their visited locations. Each user is associated with socioeconomic data, including income, education, ethnicity, and usual means of transportation.
2. 505,676 users in Sierra Leone, with one month of observation on both their physical movements and social communication history, alongside socioeconomic data representing their wealth.

We conclude this chapter by analyzing the representativity of our filtered sample of users in Sierra Leone. A similar analysis of the representativity for the GPS users in the U.S. has been previously done and can be found in the Supplementary Information of [121].

2.5 Representativity

To evaluate whether the spatial distribution of mobile phone users' home locations is representative of the actual population distribution, we compared it to aggregations of Facebook high-resolution population density maps at various spatial resolutions. The Facebook dataset is the same that has already been described and used to compute the population weights in Section 2.3, and can be seen in Fig. 2.2 C.

We first conducted this comparison at the most fine-grained resolution, the cell tower level. For the mobile phone data, we simply counted the number of users with inferred home locations at each tower. For the Facebook population data, we discarded the spatial areas of the data points, treating them as individual points (as we did in the spatial matching section), and summed the population values of all points falling within the Voronoi cell of each tower. To standardize the comparison, we divided both population counts by the area of the Voronoi cells to obtain population density values. The resulting comparison is shown in Fig. 2.5 A and B. As illustrated in Fig. 2.5 C, the two population densities are highly correlated ($\rho = 0.80$), demonstrating that the inferred home location population density for mobile phone users is a strong representation of the actual population density, even at the highest resolution.

Further validation of our home location inference is achieved by aggregating the data at coarser spatial resolutions. Specifically, we compared the data at three administrative census levels: chiefdoms, districts, and provinces, each offering increasingly lower spatial resolution. For both mobile phone users and the Facebook population data, we counted the number of individuals within each census area. As shown in Fig. 2.5 D, E, and F, our home location inference remains highly representative at the chiefdom level (with a correlation of $\rho = 0.94$). Chiefdoms have approximately half the resolution of cell towers

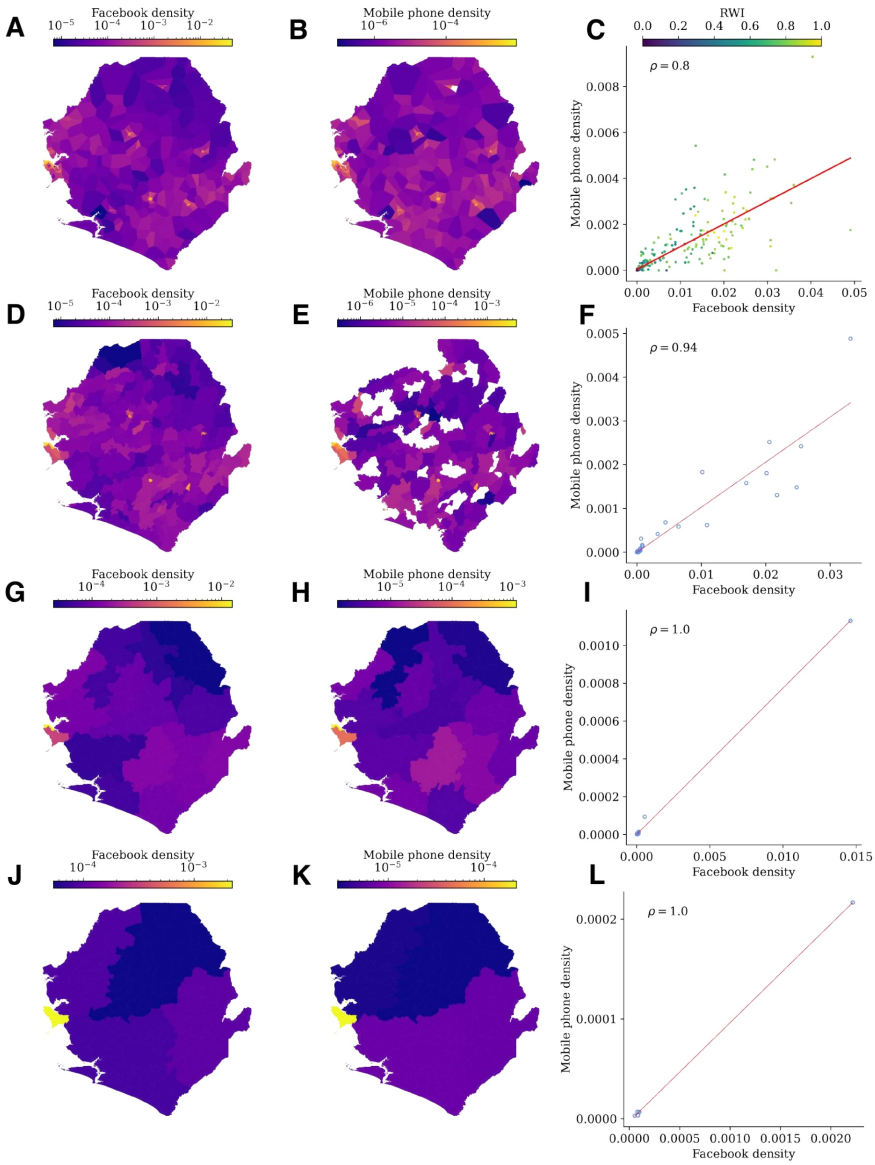

Figure 2.5: Population density validation. (Left panels) Population density in each Voronoi cell (A), chiefdom (D), district (G), and province (J) from Facebook high-resolution population density data [4]. (Middle panels) Population density in each Voronoi cell (B), chiefdom (E), district (H), and province (K) from mobile phone user's home locations. (Right panels) Relation between the population density from Facebook data (left panels) and mobile phone users (middle panels) at Voronoi cell (C), chiefdom (F), district (I), and province (L) level. The red lines are the results of OLS linear regressions. ρ is the Pearson correlation coefficient.

(with 405 towers and 207 chiefdoms). At even coarser resolutions, the correlations reach near-perfect levels at both the district (Fig. 2.5 G, H, and I) and province levels (Fig. 2.5 J, K, and L.

In conclusion, our home location inference is highly representative of the actual population density distribution, both at the tower level and across all administrative census levels.

2.6 Discussion

This chapter demonstrates the value of integrating behavioral and socioeconomic data to analyze mobility and social network patterns at both large and individual scales. By employing a consistent and coherent methodology, we have successfully combined GPS and CDR data with socioeconomic indicators, despite significant differences in the nature and resolution of these datasets. This underscores the versatility of our approach, which adapts to varying geographic and socioeconomic contexts. In particular, we have shown how to accurately aggregate socioeconomic data to match the spatial resolution of behavioral data, ensuring that socioeconomic indicators reflect the actual areas covered by each digital trace. Moreover, careful filtering of outliers and accurate home location detection were critical for achieving robust SES inferences. Ultimately, the representativeness of the inferred home locations, validated for both GPS and CDR data, highlights the robustness of our methodology.

However, our approach also has limitations. Digital data, whether collected for billing purposes (CDRs) or other reasons (GPS), do not directly measure physical movements or social connections, but rather serve as proxies for these aspects of human behavior. Additionally, the accuracy of our SES inferences is constrained by the spatial resolution of the data, which, while fine-grained in both the U.S. and Sierra Leone, still presents some uncertainty. Despite these challenges, the ability to analyze large-scale, fine-grained data on both mobility and socioeconomic factors represents a substantial opportunity, allowing us to overcome the inherent limitations in data precision for the purposes of this thesis.

In summary, this chapter highlights the significant potential of using digital traces and combining traditional with non-traditional data sources to infer SES and analyze individual mobility patterns and social behaviors at scale. The successful integration of behavioral and socioeconomic data sets the stage for exploring critical phenomena such as social segregation and inequalities, which we will investigate in the next chapters.

Chapter 3

Socioeconomic reorganization of mobility and communication networks in response to external shocks

3.1 Introduction

Segregation patterns among people stem from opportunity constraints and mechanisms that drive homophilic ties across various socioeconomic dimensions, as discussed in Chapter 1 [119, 198, 311]. While factors like gender, education, age, and ethnicity contribute to these patterns, income and wealth stand out as key contributors in social network segregation, characterized by the separation of different socioeconomic groups [158, 162, 187]. Similarly, economic background affects mobility patterns, shaping the places individuals visit, events they attend, and their transport choices. As a result, people tend to interact within their own socioeconomic spheres, reinforcing segregation in mobility networks [121, 185, 186, 200, 203, 297].

While segregation in social and mobility networks is generally regarded as stable and slow to change [148, 324], external shocks—like a global pandemic—can force abrupt behavioral shifts, leading to a sudden reorganization of socioeconomic networks. During the early phase of the COVID-19 pandemic, most countries implemented non-pharmaceutical interventions [325] to reduce mobility and social contact, aiming to slow viral spread. These measures, including lockdowns and curfews, effectively mitigated outbreaks [326–330], but also triggered severe disruptions to the economy [331], mental health [332, 333], and even food consumption [333].

These interventions inevitably impacted social and mobility networks from a socioeconomic perspective, as different groups adapted to varying degrees [326, 328, 334–336]. Wealthier individuals were more capable of adapting by avoiding public transportation and shifting to remote work [328, 337, 338], while lower SES groups faced greater challenges due to job insecurity and the need for physical presence in essential roles [339]. This disparity, alongside limited access to resources, resulted in disproportionate health impacts on poorer populations, including higher mortality rates during the pandemic [340, 341].

Given these observations, an essential question arises: *How did abrupt behavioral changes and varying adjustment capacities reorganize social and mobility networks in* *the short term*? This chapter addresses this question by analyzing the CDRs data preprocessed as outlined in Chapter 2, which was collected in Sierra Leone during the early phase of the COVID-19 pandemic. The dataset spans from before the interventions to the first national lockdown and curfew periods in April 2020. This enables us to track the mobility and communication behaviors of half a million anonymized individuals. Through this analysis, we investigate how social and mobility networks were reorganized in the short term and explore shifts in socioeconomic segregation patterns, revealing potential new phenomena driven by external shocks.

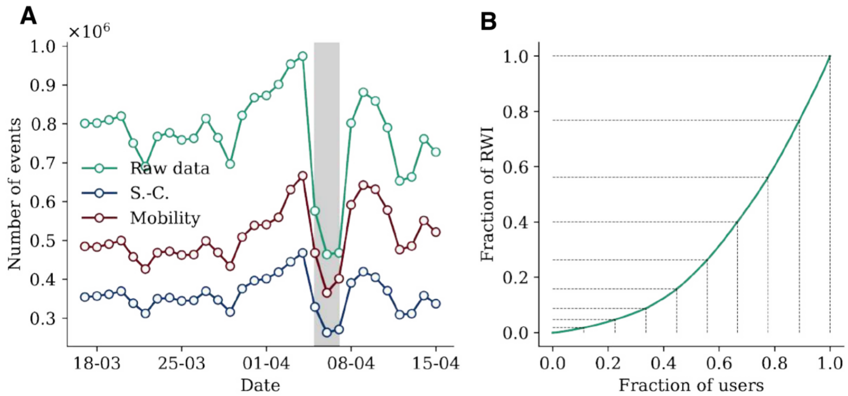

Figure 3.1: Processed data. A) Dynamics of the recorded data volume measured as the daily number of data points in the raw data (green curve), in the social-communication network (blue curve), and in the mobility network (red curve). Shaded area refers to the lockdown intervention period. B) Socioeconomic class segmentation based on the empirical Lorenz curve defined as the cumulative fraction of RWI of the sorted fraction of individuals by their inferred RWI.

3.2 Results

3.2.1 Socioeconomic networks

In response to the rapid global spread of COVID-19 in 2020, the government of Sierra Leone implemented a three-day full lockdown from April 5 to 7, followed by a 14-day nationwide nighttime curfew starting on April 9, alongside other travel and shopping restrictions (further details are provided in Appendix B). To understand the impact of these measures on mobility and communication behaviors, we use CDR data, cleaned and combined with RWI estimates, covering the activity of 505,676 individual mobile phone users with associated home locations and RWI. As mentioned in Chapter 2, this user sample is highly representative of the population distribution in Sierra Leone. Our observation spans from March 17, 2020, covering one month and including two weeks of a reference period (denoted as R1 and R2), followed by the lockdown (LD) and curfew (CF) periods. For accurate comparison, these periods correspond to the same weekdays (Sunday-Tuesday) across the observed weeks, aligning with the days of the LD period.

Using this combined dataset, we construct two socioeconomic networks: a social-communication network ($G_S(t)$), where nodes represent individuals and links reflect time-varying mobile communication between them, and a mobility network ($G_M(t)$), where nodes correspond to home and visited locations, with links capturing mobility patterns.

In the social-communication network $G_S(t)$, each node represents a mobile phone user, and link weights correspond to the number of communication events observed between pairs of nodes. We include only events (*caller, callee*) where both the caller and callee have inferred home locations and corresponding RWI values. This filtering also results in fewer communication events than those present in the raw data, as seen in Fig. 3.1 A.

The mobility network $G_M(t)$ is built by converting communication events (*caller, callee, tower*) into mobility events (*home tower, tower*), where the *home tower* refers to the caller's inferred home location, and the *tower* indicates the caller's location during the event. Thus, the nodes in $G_M(t)$ are tower locations, and the links represent home-to-location movement events. Due to the limited number of events for which tower IDs are available and the restricted number of towers with inferred aggregate RWI, the total number of movement events is also limited, as shown in Fig. 3.1 A.

To prevent spurious correlations due to spatial effects, we exclude links that originate and terminate at the same location (see Appendix B for results without this exclusion). Each node, whether representing an individual or a place, is assigned an RWI index following the methodology outlined in Chapter 2. For simplicity, we standardize RWI values from 0 (poorest) to 1 (richest). The Lorenz curve from our sample, shown in Fig. 3.1 B, yields a Gini coefficient of 0.38, closely matching the World Bank's value of 0.36 [342]. To create socioeconomic groups, we sort users by RWI and divide them into nine classes, ensuring the groups are as balanced as possible in size while keeping individuals with identical RWI within the same class. The resulting class divisions are illustrated in Fig. 3.1 B.

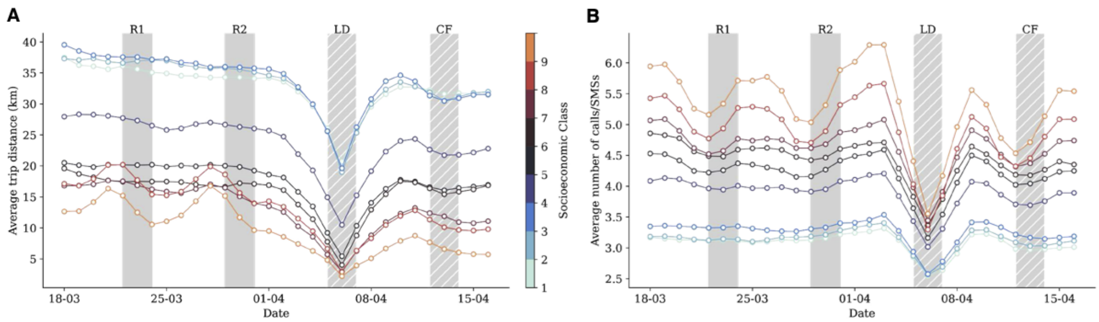

Figure 3.2: Impacts of interventions on social and travel patterns. Evolution of A) average trip length and B) number of mobile communication events for people belonging to different socioeconomic classes, from class 1 (poorest) to class 9 (richest).

$G_M(t)$ and $G_S(t)$ are two socioeconomic networks where nodes represent locations and users, respectively. Links represent home-to-location trips and communication events between users, respectively. Each node has an associated socioeconomic value (RWI) and, in turn, an associated socioeconomic class (calculated as described above).

3.2.2 Effects of lockdown on dynamics of social and mobility activities

We begin by analyzing the impact of the lockdown on mobility and social behavior across different socioeconomic classes, tracking the daily evolution of these activities. The lockdown significantly affected people's activities, but the extent of the impact varied across socioeconomic groups. For each class, we measured the average travel distance (across all trips recorded at time t) for mobility and the average number of communication events per person (among all active users at time t) for social interactions. These metrics (shown in Fig. 3.2 A, B) reveal distinct patterns of behavioral change in response to the lockdown (LD). While both travel distances and communication activities were significantly reduced for all groups during this period, the ability to adapt to the restrictions differed noticeably between socioeconomic classes.

For mobility, prior to the lockdown (R1 and R2 weeks in Fig. 3.2 A, B), higher socioeconomic classes exhibited regular weekly travel patterns, with an average daily travel distance of about 14 km. These individuals, likely white-collar workers with office jobs, were able to adjust more easily to the lockdown, reducing their daily travel distance to the bare minimum. In contrast, lower socioeconomic classes, who traveled greater distances (around 37 km on average) during the reference period, were less able to reduce their mobility during the lockdown, managing to cut travel distances to approximately 20 km. Although this represents a larger relative reduction, it was still far from the level achieved by higher SES individuals. During the curfew period, mobility gradually returned to near-normal levels, though none of the classes fully resumed their pre-lockdown travel behaviors. The large disparity between the travel distances of poorer and wealthier groups can be also attributed partially to the differences in the size of Voronoi cells in rich and poor areas.

A similar socioeconomic disparity is evident in communication dynamics, as shown in Fig. 3.2 B. Individuals from lower socioeconomic classes made fewer calls (approximately 4 calls on average) on reference days, with no distinct weekday-weekend pattern. During the lockdown, they experienced the smallest reduction in communication activities (about 25%), continuing to make the fewest calls in the population. In contrast, higher socioeconomic classes, who typically made more calls and displayed clear weekly cycles (with fewer calls during weekends), experienced a larger relative reduction in communication activities (around 40%) during the lockdown. However, they still remained the most active group in terms of communication. This suggests that wealthier individuals were able to adjust their mobile communication volume more readily, which may have influenced the structure of their social networks.

In summary, the lockdown disrupted both mobility and social interactions, but the extent of the disruption varied by socioeconomic class. While the reduction in mobility was expected due to stay-at-home orders [330, 343], the corresponding changes in social communication are more surprising, as these interactions were not directly restricted by physical movement constraints.

3.2.3 Dynamics of network segregation patterns

After assessing the unequal impact of lockdown policies on mobility and communication across socioeconomic classes, we now turn our focus to the effects on segregation. As discussed in Chapter 1, network segregation patterns can be examined through the concept of network assortativity [293], which measures connection preferences between similar nodes (whether people or places). We use the inferred RWI as a node characteristic in assortativity calculations to quantify segregation, ensuring that the results are not influenced by the adopted division into classes. Consequently, due to the scalar nature of the RWI, assortativity can be interpreted as the Pearson correlation between the RWIs of connected nodes. Segregation patterns are visualized through assortativity matrices [187], which depict the probabilities of connections (either through communication or visits) between people from different SES groups. The assortativity coefficient, ρ[293], defined in Chapter 1, summarizes the overall socioeconomic segregation observed in the assortativity matrix. This coefficient, which ranges from -1 (disassortativity or anti-segregation) to 1 (maximum segregation), indicates how much the assortativity matrix is concentrated around its diagonal. A value of $\rho = 0$ reflects no segregation in

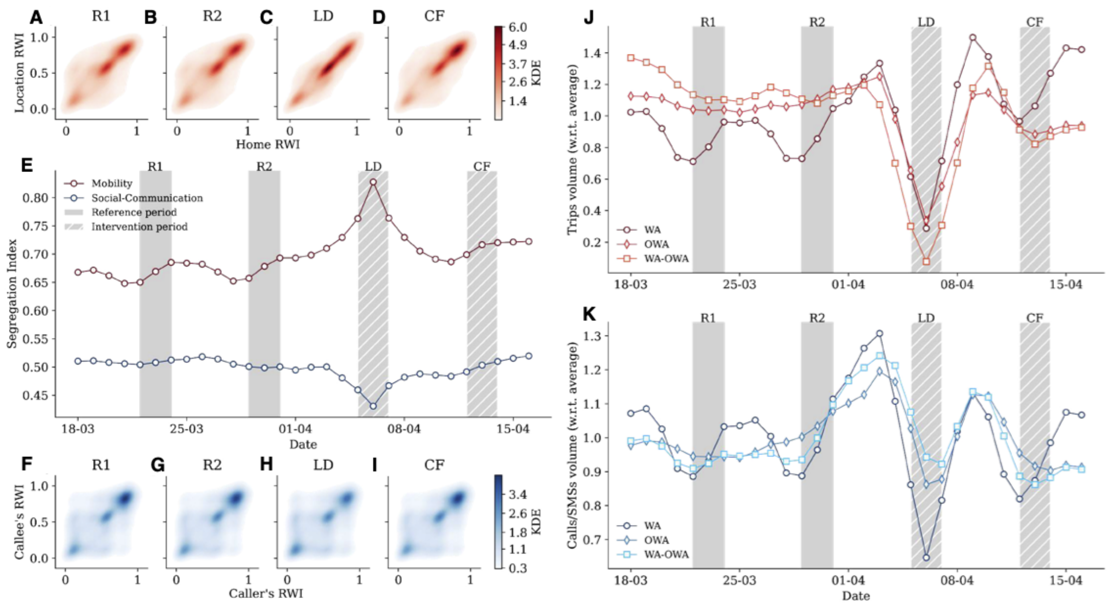

Figure 3.3: Socioeconomic segregation dynamics in mobility and social communication networks. A-D) Socioeconomic assortativity matrices (shown as the kernel density of the joint probability of RWIs) of the mobility network during the two reference (R1 and R2), lockdown (LD), and curfew (CF) periods. E) The dynamics of the ρ socioeconomic assortativity index computed for the mobility (red) and social communication (blue) networks. F-I) Same as A-D) but for the social communication network. J) Relative number of travels within WA, OWA, and between the areas WA-OWA (also accounting for OWA-WA trips). K) Number of communication events between people living in WA and OWA, or between the two geographic areas. All curves are normalized by their average computed over the full data period. For calculations on panels E, J and K we used 3-day symmetric rolling time windows with 1-day shift to obtain aggregated networks around the middle day at time t of the actual window. For the corresponding non-aggregated results see Appendix B

the network. While entropy-like metrics have also been used to quantify segregation in populations[92, 121], assortativity provides a clearer understanding of network effects, such as homophilic mechanisms and relative differences in socioeconomic diversity between individuals and their peers. It also addresses data sparsity issues more effectively (for a detailed analysis using entropy measures, see Appendix B).

The assortativity matrix for the two pre-lockdown reference periods (R1 in Fig. 3.3 A, F and R2 in Fig. 3.3 B, G) displays a strong diagonal component, indicating positive assortative mixing in both networks. This is further confirmed by the relatively high assortativity indices: $\rho_M \sim 0.65$ for the mobility network (red) and $\rho_S \sim 0.5$ for the social communication network (blue) during the R1 and R2 periods (Fig. 3.3 E). In other words, under normal conditions, both networks exhibit a high degree of segregation, where individuals tend to visit places and interact with peers from similar socioeconomic groups rather than engaging with other groups. These findings align with similar studies, such as Dong et al. [185], where $\rho \sim 0.4 - 0.8$ was observed in both mobility and online networks across multiple countries. As discussed in Chapter 1, such patterns can emerge simply due to confounding factors like working hours, spatial distance, RWI distribution, or because of network characteristics rather than meaningful socioeconomic interactions. To assess the significance of our results, we calculated the segregation levels generated by simple null models, including the gravity law and configuration model, in Appendix

B. None of these models replicated the segregation levels observed in the empirical data.

These segregation patterns changed dramatically during the national lockdown, which began on April 5, 2020 (LD period in Fig. 3.3 E). Once the lockdown was announced (April 1, 2020), the mobility segregation index began to rise, reaching its peak during the lockdown. This increase in mobility segregation was expected and has been observed in other studies [327, 344]. During the lockdown, non-essential workplaces were closed, and a stay-at-home order was enforced, which limited people's mobility and concentrated their movements around residential areas. This is reflected in the stronger diagonal component of the mobility assortativity matrix in Fig. 3.3 C, compared to the reference periods (Fig. 3.3 A, B). Interestingly, this heightened segregation did not persist after the lockdown and returned to near pre-lockdown levels during the curfew period (CF in Fig. 3.3 D, E). This is somewhat different from other studies, where increased mobility segregation remained residual in US cities even after lockdown periods [295].

In contrast, communication dynamics, which are not constrained by physical proximity, followed a different segregation pattern. Remarkably, we observed that socioeconomic assortativity in the social communication network decreased during the lockdown, reaching its lowest point. The network reorganized into a less segregated configuration, with increased communication between different socioeconomic classes compared to the pre-lockdown periods (R1 and R2 in Fig. 3.3 E). These opposing segregation trends in mobility and social communication suggest that, with mobility restricted, individuals compensated by increasing communication with peers from other socioeconomic classes. However, as with mobility, the altered segregation patterns in the social communication network were short-lived and soon returned to pre-lockdown levels. It is important to note that all results presented in Fig. 3.3 reflect averages aggregated using a 3-day sliding window with a 1-day shift. For the non-aggregated results, refer to Appendix B

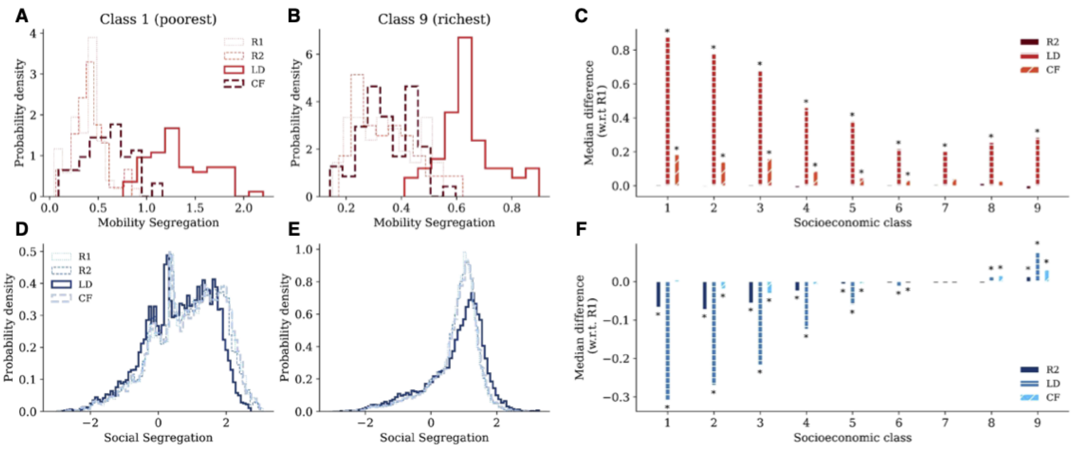

Figure 3.4: Dynamics of individual-level segregation patterns. The $P(r_u(t))$ individual assortativity index distributions computed from the mobility (panels A-B), in red) and social communication (panels D-E), in blue) networks for the poorest (class 1 in panels A) and D)) and the richest (class 9 in panels B) and E)) socioeconomic classes for the two reference periods (R1 and R2, thin dashed lines), the lockdown (LD, solid line), and curfew (CF, dashed thick line). Panels C) and F) depict the pairwise differences of median assortativity values of $P(r_u(t))$ for each nine socioeconomic group in the mobility and social networks (respectively). Differences are calculated between R1 and the R2, LD and CF periods. The asterisks symbols over the bars (when bars are positive, otherwise under them) in panels C) and F) indicate statistical significant differences computed with the one-tailed Mann-Withney U-test (with p-value < 0.01). The full list of p-values is shown in Appendix B.

3.2.4 Socioeconomic network reorganization

The baseline segregation levels and the contrasting network reorganization patterns observed during the lockdown are deeply rooted in Sierra Leone's socioeconomic structure and the sharp urban-rural divide. The capital, Freetown, located in the predominantly urban Western Area (WA) province, holds a disproportionately high concentration of wealthy individuals compared to the rest of the country, referred to as outside the Western Area (OWA), which is more rural and has, on average, twice the multidimensional poverty rates [345] (see Appendix B for the geographical division of WA and OWA). While local spatial factors do not fully account for the observed segregation patterns (as detailed in Appendix B), interactions between WA and OWA—whether through mobility or communication—play a key role in shaping overall network segregation. To investigate this, we classify each edge of the mobility network G_M (or social network G_S) into three categories: edges within WA, within OWA, or between WA and OWA. It's important to note that mobility links represent trips from individuals' home locations to other places, while social links represent communication between individuals with different home locations.

To capture network reorganization, we track the relative changes in mobility and communication volumes over time by measuring the number of interactions within or between WA and OWA areas, relative to the overall average throughout the observation period. Our results show that mobility patterns (Fig. 3.3 J) changed drastically during the lockdown. The number of trips began to drop even before the lockdown, likely in response to the early announcement of restrictions. During the lockdown itself, the largest relative decrease ($\sim$95%) was observed in trips between WA and OWA, while trips within WA and within OWA saw smaller relative declines compared to the reference periods. The sharp reduction in long-distance travel between urban and rural areas contributed significantly to the rise in mobility network segregation, as it amplified the relative share of short-distance trips within areas of similar socioeconomic status (i.e., within WA or within OWA).

Conversely, the dynamics of the social communication network (Fig. 3.3 K) exhibited a different pattern of reorganization. While communication volumes initially increased before the lockdown (possibly reflecting a form of coordination), they dropped significantly just before the restrictions took effect, across all categories. Notably, communication within WA decreased the most ($\sim$30%), while the decline within OWA was smaller ($\sim$10%). However, the relative volume of communication between WA and OWA remained relatively stable during the lockdown. This greater reduction in communication within each area, combined with the maintained communication between WA and OWA, accounts for the overall decrease in social network segregation observed in Fig. 3.3 E.

As a result, during the lockdown, the wealthiest (WA) and poorest (OWA) regions of Sierra Leone became less physically connected, with mobility largely restricted to local movements. However, while this led to an increase in mobility network segregation, the relatively higher importance of long-distance communication between urban and rural areas resulted in a decrease in social network segregation. This highlights the contrasting effects of lockdown on physical movements and social communications, with mobility becoming more localized, while communication remained an essential bridge between different socioeconomic regions.

3.2.5 Individual-level segregation

So far, we have examined segregation dynamics at the global network level, but this approach provides limited insight into individual behavioral responses. By analyzing how the personal networks of individuals reorganize, we can assess whether different socioeconomic classes responded to external shocks in distinct ways. To track the segregation changes of an individual u within the network (a location in the mobility network or a person in the social network), we compute the individual assortativity index, $r_u(t)$, as introduced by Peel et al. [294] and defined in equation 1.3 in Chapter 1. This index quantifies the homogeneity of an individual's local network based on the SES of their neighbors compared to their own SES. By calculating $r_u(t)$ for each node within a socioeconomic class, we can monitor the segregation dynamics of the class through the distribution of individual assortativity indices, $P(r_u(t))$. Note that $r_u(t)$ is unbounded and can take both positive and negative values, indicating assortative and disassortative mixing, respectively.

For mobility, both the lowest (class 1, Fig. 3.4 A) and highest socioeconomic classes (class 9, Fig. 3.4 B) displayed positive assortativity values during the reference periods, suggesting that mobility segregation is present under normal circumstances, albeit with considerable variability. The lowest class showed stronger mobility segregation, with a median $r_u(t)$ around 0.4, compared to the highest class, with a median of 0.32 (see Table S1 in the Appendix B for precise values across all classes). The assortativity values remained relatively stable between the reference periods (R1 and R2 in Fig. 3.4 A, B). However, the lockdown triggered a notable increase in the assortativity distributions for all classes, with the magnitude of change varying by class (the median differences from the R1 distribution are shown in Fig. 3.4 C, LD bar). Specifically, the lockdown caused more than a threefold increase in the median assortativity values for the poorest class (from $\sim$0.41 to $\sim$1.28) and almost doubled the value for the richest class (from $\sim$0.34 to $\sim$0.62). Consequently, individuals from lower socioeconomic classes experienced greater mobility segregation, while those from middle and higher classes also became more segregated, but to a lesser extent (all differences are statistically significant based on one-tailed Mann-Whitney U Tests compared to R1). During the curfew (CF) period, mobility segregation relaxed closer to pre-lockdown levels for the wealthier groups, while the lower socioeconomic classes exhibited some residual segregation (see CF bars in Fig. 3.4 C).

In the social communication network, individual assortativity indices were also mostly positive during the reference periods. However, the social network exhibited slightly smaller median values for the poorest class ($\sim$1.18) and stronger segregation for the richest class (with a median of $\sim$1.28, see Fig. 3.4 D, E, and Table A.10 in Appendix B). Although this indicates baseline assortativity in the social network, the $P(r_u(t))$ distribution also included negative values, signaling disassortative mixing for some individuals.

The lockdown led to unexpected changes in the social communication network. While global assortativity suggested an overall decrease in segregation, individual-level analysis revealed that this effect was not homogeneous across socioeconomic classes. As shown in Fig. 3.4 D, the reduction in global assortativity was primarily driven by individuals from lower socioeconomic classes. For the poorest class, the median assortativity decreased from 1.24 in the reference period to 1.02 during the lockdown (see Table S1 in Appendix B.10 for all values). This shift is reflected as negative median differences in Fig. 3.4 F (LD bar) compared to R1. During this period, poorer individuals interacted with a more diverse set of peers from higher socioeconomic classes, shifting their $P(r_u(t))$ distribution leftward (in Fig. 3.4 D). Conversely, the wealthiest classes became more segregated during the lockdown, as indicated by the rightward shift of their LD distribution in Fig. 3.4 E and the increase in their median assortativity from 1.28 to 1.38, resulting in positive median differences in Fig. 3.4 F. Thus, while the majority of individuals from lower classes experienced a significant reduction in network segregation, leading to a moderate decrease in global assortativity (as seen in Fig. 3.3 C), wealthier individuals remained more isolated. Interestingly, unlike mobility, the top socioeconomic classes displayed some residual positive assortativity during the curfew period (Fig. 3.4 E), as they remained relatively isolated even after the lockdown. Spatial effects alone cannot explain the observed assortativity levels in either the mobility or social communication networks (see Appendix B for more details).

These findings demonstrate that interventions can have vastly different effects on individuals and communities based on socioeconomic background. Focusing exclusively on mobility or social communication provides only a partial picture of how people adapt to external shocks. Our analysis highlights the importance of simultaneously tracking multiple aspects of human behavior to fully understand the socioeconomic determinants of responses to crises.

3.3 Discussion

This chapter explored how external shocks, such as the COVID-19 lockdown, affected segregation dynamics in both mobility and social communication networks in Sierra Leone. Our findings reveal significant differences in how socioeconomic classes responded to these measures, particularly in terms of mobility restrictions and communication patterns. Segregation patterns in mobility networks intensified during the lockdown, as individuals from lower socioeconomic backgrounds experienced a larger reduction in their mobility compared to wealthier groups. The lockdown constrained people's movements to local areas, increasing the relative segregation within both urban and rural regions.

In contrast to mobility, the lockdown led to a surprising decrease in social network segregation. Individuals from lower socioeconomic classes showed a reduction in their assortativity, indicating increased communication across class boundaries. This suggests that in response to physical isolation, individuals compensated by reaching out to contacts from different socioeconomic groups, especially in rural and urban regions. Wealthier individuals, however, became more socially segregated during the lockdown, further emphasizing the unequal ways different classes adapted to these external constraints. Interestingly, while mobility segregation patterns normalized after the lockdown, some residual social segregation remained among wealthier groups during the curfew period.

These contrasting dynamics highlight the complex interplay between mobility and communication behaviors in response to external interventions. While our findings provide novel insights into mobility and social reorganization during the COVID-19 pandemic, they also align with broader research showing how mobility shifts were influenced by socioeconomic disparities. Studies in other contexts have demonstrated similar patterns of mobility reduction among lower-income populations and an uneven capacity to adapt to pandemic restrictions [326, 328, 334–339, 346, 347]. Our study further confirms that, due to restrictions, the overall social mixing was reduced during the lockdown, which increased experienced segregation in mobility [295, 327, 348]. However, this effect was only observed during the lockdown phase, after which mobility segregation returned to previous levels, contrary to findings from other studies [295, 349]. Additionally, the observed increase in cross-class communication echoes findings from studies that emphasize the role of digital communication in mitigating the impacts of physical isolation [350]. The main contribution of our study within the existing literature is its ability to connect these two aspects—mobility and communication—through an individual-level and bidimensional analysis, providing a more comprehensive understanding of behavioral adaptations during the pandemic. By examining both the mobility and communication networks of the same sample of users, this chapter underscores the importance of a multi-dimensional approach to understanding segregation and adaptation in times of crisis.

Chapter 4

Deviations from universality in human mobility modeling

4.1 Introduction

Effective modeling of urban mobility is essential not only for urban planning [351] and infrastructure development [352], but also for optimizing public services [353], reducing traffic congestion [354, 355], enhancing social integration [121], and managing the spread of diseases [356, 357].

Human mobility models aim to identify and replicate universal laws governing movement patterns, such as the distribution of travel distances [266, 358], the scaling of mobility flows with distance and population size [145, 359], and the spatiotemporal dynamics of aggregate movements [360]. While these models have advanced our understanding of general mobility behaviors, their applicability across diverse populations is still uncertain. It remains unclear whether such models adequately capture the behaviors of all individuals or, conversely, may favor certain demographic characteristics, potentially overlooking others.

Among individual-level models, the Exploration and Preferential Return (EPR) model [279] is particularly notable for its simplicity and its ability to replicate key scaling laws in human mobility. By introducing lower stochasticity than random walk models [51, 266], it enhances predictability [228, 229] by focusing on two main principles: individuals explore new locations less frequently over time, and they revisit familiar locations more often. However, the model assumes uniform mobility mechanisms across all individuals, overlooking variations due to socioeconomic backgrounds and lifestyles. In Chapter 1, we discussed how factors like socioeconomic status influence travel behaviors, affecting daily travel distances, travel frequency, and the socioeconomic profiles of visited locations. Furthermore, different demographic groups exhibit distinct responses to mobility constraints during emergencies. Additionally, people's visitation patterns often align with identifiable lifestyle and activity profiles that cut across socioeconomic groups [361]. These factors likely affect the balance between exploration and recurrent visits, which are not considered by the EPR model, potentially leading to varying levels of accuracy across demographic and lifestyle groups.

In this chapter, we evaluate the EPR model's performance at the individual level, focusing on mobility scaling laws related to exploration dynamics and visitation frequency. Our analysis aims to determine whether the model accurately represents mobility across diverse individual profiles. We introduce metrics to quantify deviations from the expected scaling laws and investigate how these deviations are associated with violations of the model's assumptions, particularly in non-bursty exploration and preferential return. Furthermore, we examine the types of places visited when these assumptions are not met and analyze the socioeconomic and lifestyle characteristics of individuals who most frequently deviate from the model's predictions.

4.2 Results

To address our research question, we analyze the micro-scale movements of 1.5 million anonymized users across 11 core-based statistical areas (CBSAs) in the United States [314], with data spanning from October 2016 to March 2017. The dataset has been thoroughly cleaned and preprocessed, as detailed in Chapter 2. For each user, we capture a list of temporally and spatially fine-grained visits, each associated with a Foursquare venue and its corresponding category [362]. In the following analysis, we focus specifically on a subsample of 51,648 users within the Boston-Cambridge-Newton CBSA (referred to as Boston for simplicity). Results for other CBSAs are presented in Appendix C.

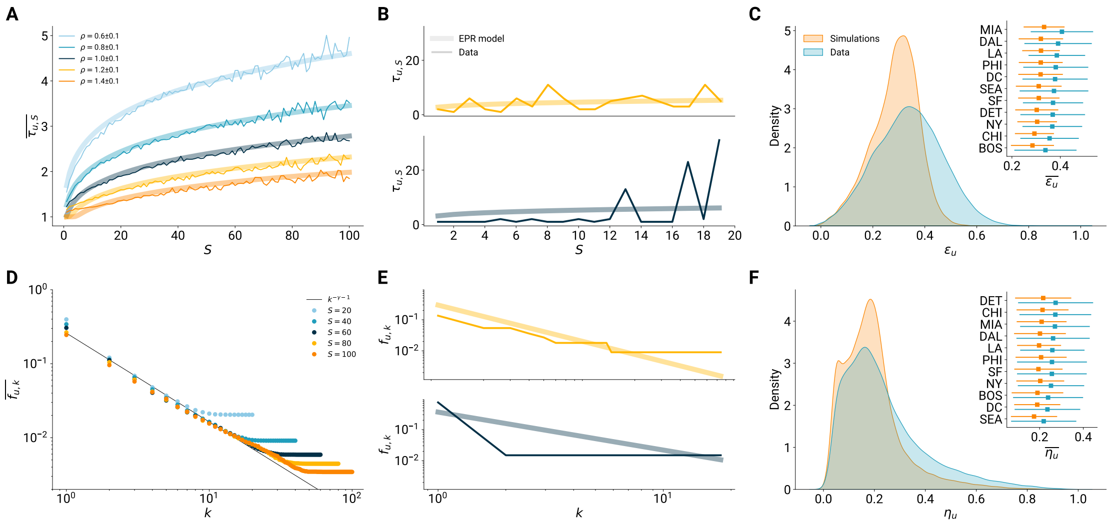

Figure 4.1: Individual deviations beyond population-wide accuracy. A) The average exploration inter-event time as a function of the number of distinct places S for users with different values of ρ_u (thin lines) and the corresponding predictions by the EPR model (thick lines). B) Individual-level observations of exploration inter-event time as a function of S (thin lines) and the predictions of the EPR model (thick lines) for two example users who visited 20 distinct places and that show, respectively, well-fitting (upper panel) and deviating (lower panel) behavior compared to the model. C) Distribution of individual deviations ϵ_u from the data (light blue) and from stochastic simulations of the EPR model (orange). The inset shows the mean and standard deviations of the same distributions across all CBSAs. D) Average visitation frequency of ranked locations according to the EPR model (thick black line) and for users with different values of S. E) Same as B) but for visitation frequency. F) Same as E) but for η_u.

4.2.1 Individual-level deviations

In the Exploration and Preferential Return (EPR) model, users move between locations with each step representing either a return to a previously visited place (return step) or an exploration of a new place (exploration step). The likelihood of a user u exploring, given they have already visited S distinct locations, is captured by the exploration probability $P_u(S) = \rho_u S^{-\gamma}$. This probability is controlled by the individual parameter ρ_u, calculated based on the total number of distinct places S_u visited by u relative to the total number of visits N_u, as:

$$\rho_u = \frac{S_u^{\gamma+1}}{(\gamma+1)N_u},$$

where the parameter $\gamma = 0.22$ has been computed on same dataset and for the same sample of users in [121] and it is fixed for all users.

The EPR model accurately reproduces two main aspects of human mobility: exploration dynamics and visitation frequency, at least on average. For exploration dynamics, we examine the inter-event time $\tau_{u,S}$, defined as the number of steps after which a user u visits a new location given that they have already visited S distinct locations. According to the model, given the exploration probability $P_u(S) = \rho_u S^{-\gamma}$, the probability that the next exploration step will occur after T steps is given by $P(\tau_{u,S} = T) = (1 - P_u(S))^{T-1} P_u(S)$, which is a geometric distribution. From this distribution, we get that the expected inter-event time for the EPR model $\langle \tau_{u,S} \rangle$ is given by $\langle \tau_{u,S} \rangle = P_u^{-1}(S) = S^\gamma / \rho_u$. For visitation frequency, the EPR model predicts that the frequency with which a user u visits their kth most frequented location follows Zipf's law, with $\langle f_{u,k} \rangle \sim k^{-\gamma}$.

While averaging $\tau_{u,S}$ or $f_{u,k}$ across users in our dataset (as shown in Fig.4.1 A and D) shows that the EPR model accurately captures both exploration inter-event time and visitation frequency at the population level, this does not guarantee precision at the individual level. Since these properties can be measured for each user, we can assess how well the model represents individual behaviors.

To illustrate, Fig.4.1 B and E (upper panels) show an example of a user whose $\tau_{u,S}$ and $f_{u,k}$ align well with the model's scaling laws, while Fig.4.1 B and E (lower panels) present a user whose behavior deviates from the model's predictions. To quantify these deviations, we define two individual metrics that measure the discrepancies between observed and expected scaling for both inter-event time and visitation frequency. For exploration inter-event times, we introduce ϵ_u, defined as the symmetric mean absolute percentage error (SMAPE) between $\tau_{u,S}$ and $\langle \tau_{u,S} \rangle$ for each user u:

$$\epsilon_u = \frac{1}{S_u} \sum_{S=1}^{S_u} \frac{|\tau_{u,S} - \langle \tau_{u,S} \rangle|}{|\tau_{u,S}| + |\langle \tau_{u,S} \rangle|},$$

where S_u is the total number of distinct places visited by user u. For visitation frequency, we define η_u as the Kullback-Leibler divergence between $f_{u,k}$ and $\langle f_{u,k} \rangle$:

$$\eta_u = \sum_{k=1}^{K_u} f_{u,k} \log \frac{f_{u,k}}{\langle f_{u,k} \rangle},$$

where K_u is the rank of the least visited location, excluding locations visited only once to reduce the tail effect visible in Fig.4.1 B. These metrics can be calculated for both observed data and EPR model simulations, allowing us to compare individual deviations between real and simulated trajectories.

The distributions of ϵ_u and η_u for users in Boston and for simulated trajectories based on the EPR model are shown in Fig.4.1 C and F. To enable direct comparison, we normalize these distributions to a 0–1 range. The results indicate that while the EPR model captures the average dynamics of exploration and visitation frequency well, it does not equally represent all individuals. The observed deviations are not solely attributable to the stochastic nature of the model. The orange distributions in Fig.4.1 C and F represent the outcomes from stochastic EPR simulations for the same user set (see Appendix C for details). Although stochasticity accounts for some of the variance (particularly in η_u), substantial differences remain, especially in the largest deviations, which cannot be explained by model stochasticity alone. This pattern is consistent across all CBSAs, as indicated in the subpanels of Fig.4.1 C and F, where mean and standard deviations of real and simulated distributions of ϵ_u and η_u are shown for all CBSAs.

Interestingly, despite capturing distinct mobility properties, ϵ_u and η_u are strongly correlated (Pearson correlation of 0.75). This suggests that users who deviate significantly from the model in terms of exploration dynamics are also likely to show deviations in visitation frequency (see Appendix C for additional details).

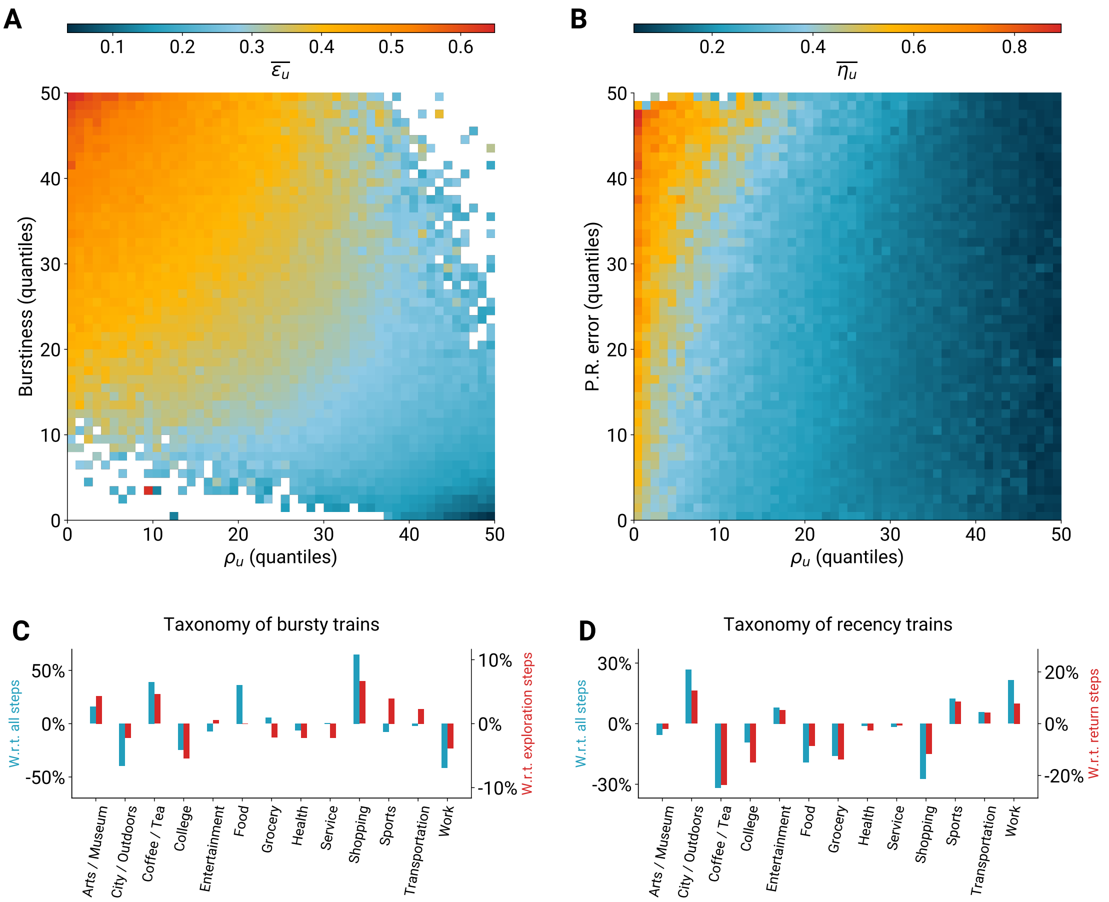

Figure 4.2: Deviations are related to violations of the EPR model's microscopic mechanisms. A) Average values of the deviation ϵ_u (color-coded as in the color bar) for users grouped in quantiles of ρ_u and exploration burstiness. Controlling for stochasticity with ρ_u, ϵ_u increases with burstiness. B) Average values of the deviation η_u (color-coded as in the color bar) for users grouped in quantiles of ρ_u and preferential return error (P.R. error). Controlling for stochasticity with ρ_u, η_u increases with P.R. error. C) Characterization of bursty trains, i.e., sequences of consecutive exploration steps, in terms of relative visits to defined categories, compared to all visits (blue bars, left y-axis) and to exploration steps only (red bars, right y-axis). D) Characterization of recency trains, i.e., sequences of consecutive visits to the same place, in terms of relative visits to defined categories, compared to all visits (blue bars, left y-axis) and to return steps only (red bars, right y-axis).

4.2.2 Microscopic mechanisms

Given the range of deviation metrics that cannot be fully explained by stochasticity alone, we aim to investigate the underlying microscopic characteristics associated with these deviations. Specifically, we seek to identify mobility behaviors that the EPR model does not account for due to its assumptions, which may contribute to poorer representativity for certain users.

We begin by examining individual exploration tendency, represented by the parameter ρ_u, which impacts both ϵ_u and η_u by influencing the stochastic component. As shown in Fig.4.1 C and F, ρ_u shapes the distribution of these deviations in model simulations (see Appendix C for details). Although ρ_u clearly relates to deviations in ϵ_u and η_u, it is already part of the EPR model, providing a baseline for variability. Here, we aim to isolate characteristics that fall outside the model's assumptions. By controlling for exploration tendency, we focus on deviations attributable to features beyond the model's stochastic framework, refining our question to: *for users with similar exploration tendencies (and thus similar stochastic uncertainty), what characteristics make some users more accurately represented by the model than others, with smaller observed deviations?*

Burstiness

For exploration inter-event times, the EPR model assumes a steady decrease in exploration probability as the number of distinct places grows, represented by $P_u(S) = \rho_u S^{-\gamma}$. This approach neglects *burstiness*—a common feature in human activity where periods of low activity alternate with short bursts of high activity [231], also present in human mobility [363]. We hypothesize that users who exhibit bursty exploration patterns, alternating between intense exploration phases and long inactive periods, tend to have higher ϵ_u values. Such users display mobility behaviors that the EPR model does not capture.

To test this hypothesis, we measure each user's exploration burstiness and analyze its relationship with ϵ_u, controlling for stochastic uncertainty via ρ_u. For a user u with N_u total visits and exploration inter-event times $\{\tau_{u,S}\}_{S=1}^{S_u}$, we compute the burstiness coefficient B_u for finite sequences as defined in [364]:

$$B_u = \frac{\sqrt{N_u + 1}r - \sqrt{N_u - 1}}{\left(\sqrt{N_u + 1} - 2\right) r + \sqrt{N_u - 1}},$$

where $r = \sigma(\tau_{u,S})/\mu(\tau_{u,S})$ is the coefficient of variation (with σ and μ as the empirical standard deviation and mean of the $\{\tau_{u,S}\}_{S=1}^{S_u}$ sequence).

The relationship between B_u and ϵ_u, controlling for ρ_u, is shown in Fig.4.2 A. Grouping users into quantiles of ρ_u and B_u and calculating the average ϵ_u within each group reveals that, beyond the dependency on ρ_u, ϵ_u is positively associated with burstiness. Specifically, for users with similar exploration tendencies, those with higher burstiness are less accurately represented by the EPR model in terms of exploration dynamics.

Preferential return

The EPR model's prediction of visitation frequency distribution, $\langle f_k \rangle \sim k^{-\gamma}$, is driven by the preferential return mechanism, a widely observed phenomenon in human systems [138]. However, this assumption may not equally apply to all individuals. According to the preferential return criterion, the probability Π_i of visiting a location i is proportional to the fraction of prior visits ϕ_i to that location. We hypothesize that users who follow this criterion less closely will display visitation frequency distributions that deviate more from the model's expected scaling.

To quantify adherence to the preferential return principle, we introduce a preferential return error (P.R. error). This error measures the deviation from $\Pi_i = \phi_i$ by analyzing the empirical likelihood of returning to previously visited locations based on their visitation frequencies. We bin ϕ (a continuous variable in $(0, 1]$) into 20 equally spaced intervals, ϕ_b, and for each return step, assign a 1 to the Π_{ϕ_b} list of the location returned to and 0 to others. For each bin, we compute the observed probability $\tilde{\Pi}_{\phi_b}$ as the empirical average of the binary values in Π_{ϕ_b}. We then calculate the preferential return error $P.R.E.$ as the symmetric mean absolute percentage error (SMAPE) of $\tilde{\Pi}_{\phi_b}$ from ϕ_b:

$$P.R.E. = \frac{1}{20} \sum_{b=1}^{20} \frac{|\tilde{\Pi}_{\phi_b} - \phi_b|}{|\tilde{\Pi}_{\phi_b}| + |\phi_b|}.$$

Like burstiness, we group users into quantiles of ρ_u and P.R. error, then compute the average η_u deviation within each group. Fig.4.2 B shows that η_u is positively associated with P.R. error, beyond its dependency on ρ_u, particularly for users with lower ρ_u. For users with similar low exploration tendencies, the more they deviate from the preferential return criterion, the less their visitation frequency aligns with the EPR model's expected distribution. However, for users with high exploration tendencies, adherence to preferential return does not strongly influence the model's accuracy.

To test the robustness of these findings, we replicated the analysis across all CBSAs, confirming that the roles of burstiness and preferential return in shaping deviations ϵ_u and η_u are consistent across locations (see Appendix C for details).

4.2.3 Characterization of assumptions' violations

After assessing the direct and individual-level relationship between deviations from scaling laws and violations of the EPR model's assumptions, we aim to characterize these violations from a behavioral perspective. In other words, we investigate the types of places people visit when they deviate from the smooth exploration dynamics of the EPR model, instead engaging in bursty exploration periods. Similarly, when people violate the preferential return criterion, where do they tend to return?

We can address these questions because each step in our dataset is associated with a Foursquare category, as described in Section 2.2.1. Foursquare categorization consists of 592 categories which are regularly updated. Additionally, categories have been manually grouped into 13 macro groups [121]: Art/Museum, City/Outdoors, Coffee/Tea, College, Entertainment, Food, Grocery, Health, Service, Shopping, Sports, Transportation, Work. Details on the precise assignment of each category to a macro group are provided in the Supplementary Information of [121].

To characterize exploration burstiness, we analyze the places visited during bursty exploration trains, i.e., sequences of consecutive exploration steps. In other words, bursty trains are sequences of visits to new places, all of which have been never explored before. Specifically, we compare the frequency of visits to a given category with its overall frequency, as well as its frequency during exploration steps only. For example, let's say a user visits Food places in 20% of their recorded stays. However, during bursty exploration trains, they visit Food places in 30% of their "bursty" stays. Finally, when they explore new places—not necessarily in bursty trains—they visit Food places in 23% of their exploration stays. This means that the user visits Food places 50% more often during bursty trains compared to any other time, and 15% more than when they normally explore new places. Both comparisons are important: the first reveals how much more or less a category is visited during bursty trains compared to any other step, while the second restricts the comparison to exploration steps only, acknowledging that some categories may generally be explored more but not necessarily during bursty periods. We measure both this relative frequencies for all the 13 taxonomy groups to characterize what places people visit more likely when they explore in a bursty manner.

As shown in Fig.4.2 C, bursty trains are indeed characterized by specific categories. In particular, places of amusement such as museums and art galleries, coffee shops, and shopping locations are visited significantly more often during bursty exploration trains. Conversely, routine locations such as workplaces are rarely explored during bursty trains. Additionally, the macro category *City/Outdoors*, which includes parks, neighborhoods, playgrounds, and residential places, is typically not explored in a bursty manner. The other categories show either small or varied differences depending on the comparison.

Regarding violations of preferential return, recency has been observed as a characteristic of human mobility not captured by the EPR model [283]. Recency describes a memory effect, indicating that humans tend to return not only to frequently visited locations but also more often to those visited recently. In contrast, the preferential return criterion assumes that all locations can be revisited based solely on their past visitation frequency, without considering the temporal order of visits. To capture this effect while drawing a parallel with bursty trains, we introduce the concept of recency trains, i.e., sequences of consecutive return visits to the same place. We characterize these recency trains using the same methodology applied to bursty exploration trains, comparing the frequency of visits to a given category with both its overall frequency and, in this case, its frequency during return steps only. Let's consider the same example of a user who visits Food places in 20% of their recorded stays. However, during recency trains, they return to Food places in 10% of their "recency" stays. Finally, when they return to previously visited places—not necessarily in recency trains—they visit Food places in 8% of their return stays. This means that the user returns to previously visited Food places 10% less often during recency trains compared to any other time, and 2% more often than when they normally return to previously visited places.

The results are shown in Fig.4.2 D, where we observe a pattern somewhat opposite to bursty trains. When people repeatedly return to the same place, they tend to do so at routine and habitual locations, such as residential areas, workplaces, as well as transportation hubs and sports venues. In contrast, amusement places like coffee shops, restaurants, and shopping malls are not typically revisited continuously. Interestingly, entertainment places also appear in recency trains, though not significantly. It is also notable that routine categories such as *College* and *Groceries* are absent from recency trains.

The results of this section are consistent across all CBSAs, as detailed in Appendix C.

4.2.4 Vulnerable groups

Finally, having identified the microscopic mechanisms whose violations are associated with deviations from the model, and how these violations are characterized from a behavioral perspective, we aim to characterize who is more prone to these patterns and, hence, at risk of not being modeled correctly. To answer this question, we test whether the deviations are related to any individual traits concerning the SES and life habits of people. Our goal is to explore whether deviations are uniformly distributed across the

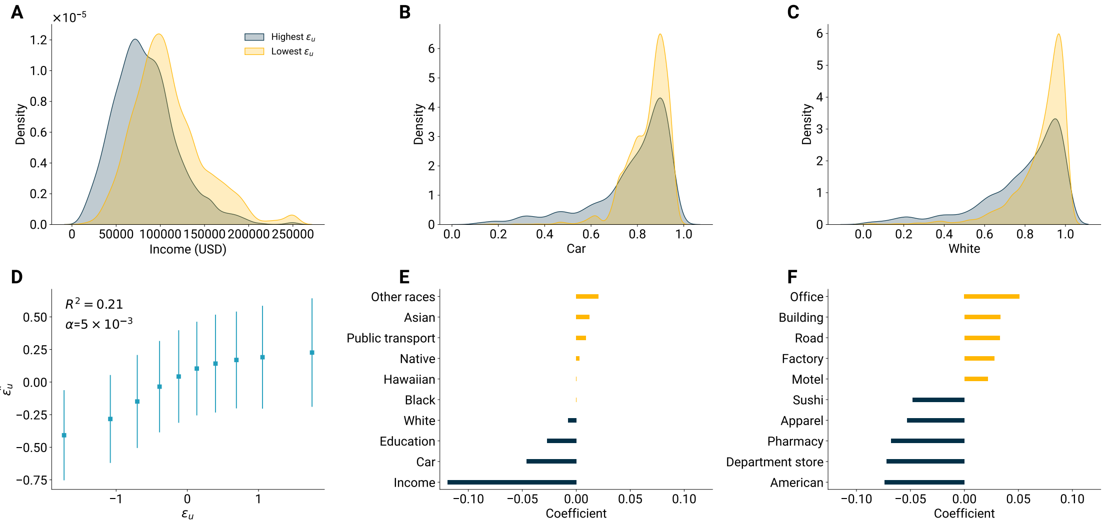

Figure 4.3: Deviations are biased towards SES and lifestyles. A) The income distribution of users in the highest (blue) and lowest (yellow) 10% quantiles of the deviation ϵ_u. B) Same as A) for car usage. C) Same as A) for the probability of being white. D) Results of the LASSO regression: true ϵ_u vs predicted $\hat{\epsilon_u}$ based on census and lifestyle features. R^2 is the coefficient of determination, and α is the regularization parameter. E) Coefficients of the census features from the LASSO regression. F) Coefficients of the life habit features (showing only the highest and lowest 5) from the LASSO regression.

population or if there are specific groups that show significantly larger deviations, meaning they are less well-represented by the EPR model. As detailed in Chapter 2, users are assigned SES indicators related to wealth, education, race, and means of transportation. Moreover, we assign each user a category score for each of the 592 Foursquare categories, based on the fraction of visits spent in places classified in each category. We use these features as lifestyle indicators.

Given an indicator obtained from census, like income, means of transportation, and race, we compute its distribution for people in the highest and lowest 10% quantiles of the ϵ_u and η_u distributions. These are users who are respectively described as the worst and best by the EPR model. In Fig.4.3, we show these distributions for the variables of income (panel A), car usage (B), and the probability of being white (C) for the two extreme quantiles of ϵ_u. The income distribution of the worst-described users in the highest 10% of ϵ_u is shifted to lower values compared to the best-described users in the lowest 10%. This implies a possible correlation between income and the ϵ deviation, suggesting that the EPR model predicts the mobility of people with lower incomes less accurately. Meanwhile, the same distributions computed for white people and car owners are less concentrated but skew towards lower values for those in the highest 10% of ϵ errors. Additionally, similar observations appear when investigating the η_u deviations (for details, see Appendix C). These findings suggest that users poorly represented by the EPR model, in terms of both exploration dynamics and visitation frequency, are more likely to be poorer, less likely to own a car, and less likely to be of white racial origin than those who are well-represented by the model.

To establish a more robust observation of these biases, not only for extreme quantiles but for all users, we perform a regression analysis. We use the deviation metrics (ϵ_u or η_u) as dependent variables and all census and life habit features as predictors. Moreover, we also include S_u and K_u, respectively for ϵ_u and η_u, as control variables, as they indicate the number of elements in the sum of either ϵ_u or η_u, respectively. In this way, we avoid measuring an effect due to varying sample sizes used for the computation of the dependent variables. Given the high number of features (one control, 10 census, 592 for life habits), we consider a Least Absolute Shrinkage and Selection Operator (LASSO) regression, whose objective function is:

$$\min_{\beta} \left(\sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \alpha \sum_{j=1}^{p} |\beta_j| \right),$$

where y_i is the dependent variable, β_i's are the coefficients, x_{ij} are the independent variables and α is the regularization parameter. There are numerous reasons to use this type of model in our case. Indeed, through the regularization parameter α, it performs automatic feature selection by shrinking the least important coefficients to zero, thus handling multicollinearity and preventing overfitting at the same time. We estimate the parameter α through 10-fold cross-validation with 3 repeated randomizations. All variables, including the dependent ones, have been standardized by subtracting the mean and scaling to unit variance.

Interestingly, we achieve $R^2 = 0.21$ correlation between the observed ϵ_u and predicted $\hat{\epsilon}_u$ ($R^2 = 0.22$ for η_u, see Appendix C), indicating that the deviations are not randomly distributed. On the contrary, they can be partially predicted by the socioeconomic and life habit features of users, as seen in Fig.4.3 D and in Appendix C. Similar scores are achieved for all 11 CBSAs (see Appendix C for details).

The different features ranked by their regression coefficients (in Fig.4.3 E) identify income as the most relevant socioeconomic characteristic in determining the ϵ error (and also η, see Appendix C). Indeed, this significantly negative coefficient of income in both regressions verifies our earlier observation that users who are poorly represented by the model are more likely to be poorer. Moreover, this observation is robust across all CBSAs analyzed (for details, see Appendix C), with income consistently showing the largest negative coefficient. Exceptions appear for η_u in New York and Los Angeles, where income has the second most negative coefficient (details in Appendix C). On the other hand, the coefficient for car use behaves differently for the two deviation metrics, being negatively associated with ϵ_u and positively with η_u, not only in Boston but in most CBSAs (details in Appendix C). Finally, contrary to our earlier conjecture, ethnicity does not strongly influence model performance in Boston, but it is significant in some CBSAs (details in Appendix C). In particular, the probability of being white is negatively associated with ϵ_u in many CBSAs, indicating that white individuals are often more likely to be well-represented by the model in terms of exploration dynamics. On the other hand, black people are often more likely to be well-represented by the EPR in terms of visitation frequency.

Similar to the census variables, model deviations are also strongly biased concerning certain life habits, as shown in Fig.4.3 F, where we present five features with the highest and lowest coefficients. A positive coefficient for a given category indicates that users who spend a large fraction of their time visiting that category are likely to exhibit higher deviations and, consequently, be poorly represented by the model. Conversely, a negative coefficient indicates that users who spend a large fraction of their time in that category are more likely to adhere to the model. Places like offices, factories, and buildings are primarily positively associated with larger deviations. In contrast, negatively correlated categories include restaurants like Sushi and American . In Appendix C, we show the results for all categories, grouping them by the macro category they are assigned to. Most place belonging to the "Food", "Shopping", and "Arts / Museum" macro categories are negatively related with deviations, while most places belonging to the "Work" and "City / Outdoor" macro categories are positively related with deviations. Similar patterns are observed across all CBSAs (see Appendix C for details). Overall, these results indicate that mobility constraints and routines have a strong impact on the adherence with the EPR scaling laws: users with more work-driven daily routines, who spend more time in offices and factories, are more likely to be poorly represented by the model, while users who spend more time in restaurants, museums and shops are less likely to experience deviations.

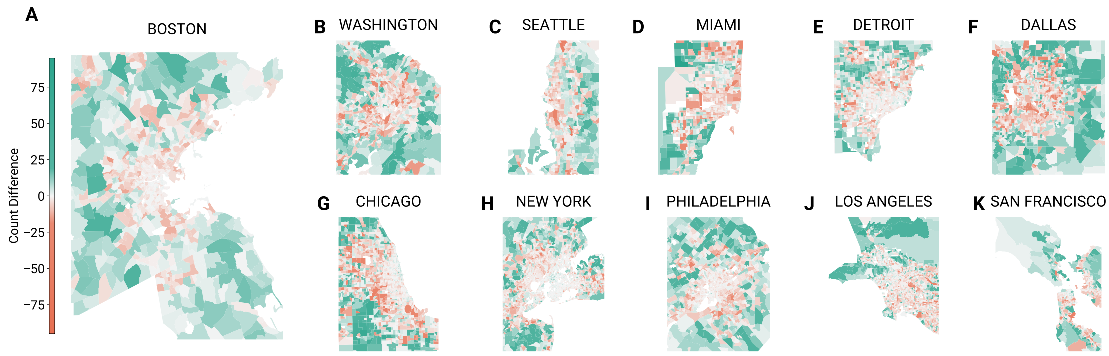

Figure 4.4: Urban-rural divide. A-K) U.S. census tracts division of the 11 CBSAs under consideration, where each tract is colored according to the difference count between the number of users in the highest and lowest 10% of $\epsilon_u + \eta_u$ with estimated home location in the tract.

4.2.5 Spatial distribution

In the previous section, we demonstrated how deviations from the EPR model are not random but biased towards certain groups of individuals, both in terms of socioeconomic indicators and life habits. A notable implication of these findings is the unequal spatial accuracy of the EPR model. Due to residential segregation and urbanization patterns in metropolitan areas, deviations follow certain spatial distributions.

Similar to the previous section, we examine this consequence by considering the distributions of the two deviation metrics, ϵ_u and η_u. However, as noted above, these variables are highly correlated, so we use their sum as a unique variable to provide a comprehensive picture of the model's combined accuracy at the individual level. We consider U.S. census tracts as spatial units and count how many users in the highest and lowest 10% of the combined deviation metric live in each tract. The difference between these values indicates the model's combined accuracy within a given tract. If this difference is close to 0, the number of well- and poorly-represented users are comparable, reflecting balanced model predictions. Conversely, if the difference is positive (or negative), well-represented users exceed poorly-represented ones (or vice versa).

The results for the Boston CBSA, focusing on Boston City and its surroundings, are shown in Fig.4.4 A. This map reveals a clear pattern: users who are best-represented by the model are predominantly located further from the city center, in larger and more rural census tracts. In contrast, the worst-represented users are mostly concentrated closer to the city center, in smaller and more urban tracts, though in the very central areas of the city, the two categories are relatively balanced. This finding has notable policy implications, as it indicates that the EPR model, widely used (with its many variations) to model urban mobility, performs better for individuals living outside urban areas, while it underperforms in describing the mobility of those living within urban areas. Examining spatial distributions in other CBSAs (see Fig.4.4 B to K), we consistently observe this urban-rural division pattern, more evident in some areas (e.g., Washington, Dallas) than in others (e.g., San Francisco, New York).

This spatial pattern reflects residential segregation by income, with wealthier individuals typically residing outside city centers. In conclusion, we paradoxically observe that due to socioeconomic biases the EPR model, designed to simulate urban mobility, performs better in rural settings.

4.3 Discussion

The results in this chapter highlight the limitations of evaluating human mobility models based solely on population-wide behavior and the risks of assuming universality across diverse populations. By analyzing the GPS traces described in Chapter 2, we compared individual exploration dynamics and visitation frequency with the predictions of the EPR model, revealing significant variations in adherence that cannot be fully explained by stochasticity.

Our results demonstrate that individuals who deviate from the EPR model share distinct behavioral patterns. They tend to visit habitual locations in consecutive sequences and rarely engage in exploration. When they do explore, it happens in short bursts, predominantly at non-routine and amusement places. These behaviors are not randomly distributed across the population but are strongly linked to socioeconomic factors, with lower-income individuals being particularly poorly represented by the model. The observed correlation between behavioral traits and income aligns with expectations: individuals with higher income face fewer spatial and temporal constraints in their daily mobility [72, 75, 365]. This advantage stems from their access to greater resources, enabling them to explore a wider variety of destinations, including those often inaccessible to less affluent individuals [74, 78, 87]. Their mobility patterns are less likely to be dictated by work routines and necessities; instead, they are frequently shaped by leisure activities [69, 81]. Moreover, wealthier individuals benefit from enhanced access to transportation, such as private vehicles, which further increases their flexibility and range of movement [79]. Our findings highlight that this association between socioeconomic status (SES) and behavioral constraints significantly contributes to the poorer representation of less affluent individuals by the EPR model. Additionally, spatial analysis reveals that individuals in suburban and rural areas are generally better represented by the model than those in urban centers, underscoring a spatial bias likely influenced by residential and income segregation.

While these findings highlight certain limitations of the EPR model, they should be understood in the context of its original design. As a minimalistic framework, the EPR model seeks to capture broad statistical patterns of mobility using universal parameters, rather than to reflect individual or group-level variability. Rather than viewing deviations as shortcomings of the model, these results reveal an opportunity to expand its scope. By incorporating factors such as socioeconomic or spatial heterogeneity, future adaptations of the EPR model could better capture the nuanced behaviors of specific population groups, particularly those in urban settings. While unified models can offer valuable insights at the population level, it is essential to recognize that their accuracy will not be uniform across all individuals, and may be particularly poor for some.

Chapter 5

Conclusion

5.1 Main contributions

This thesis has investigated the intricate relationships between socioeconomic status (SES), human mobility, and social networks, aiming to understand how SES shapes movement patterns and social connections. Through deep and extensive analysis of large-scale data, this work provides insights into both individual behaviors and collective phenomena, such as segregation and socioeconomic disparities within universal modeling frameworks. By leveraging a combination of digital, traditional, and non-traditional data sources, this research underscores the role of SES in influencing physical movements and social ties, revealing how these interactions contribute to broader social patterns.

Chapter 1 contextualizes this thesis with an extensive review of theoretical frameworks, methods, metrics, and models. It constructs a comprehensive background, drawing on a range of studies to highlight the complexity of the interplay between SES, physical movement, and social interactions. By integrating diverse approaches from social science, network theory, data science, and mobility studies, Chapter 1 provides a strong foundation for the data-driven analysis that follows.

Chapter 2 lays the methodological groundwork for observing socioeconomic patterns in mobility and social networks by establishing a cohesive approach for data collection, integration, and analysis. Addressing the core challenge of accurately and representatively capturing large-scale, fine-grained socioeconomic and behavioral data, this chapter seeks an optimal balance between scale and accuracy. It introduces general methodologies for inferring SES from behavioral and socioeconomic data, incorporating refined techniques for spatially matching diverse data sources. The result is a statistically representative sample of individuals with detailed SES indicators, alongside precise, granular proxies for their movements and social connections. This chapter thus provides a robust framework for the in-depth analyses in the following chapters on SES-related behaviors across diverse populations.

Chapter 3 delves into the socioeconomic reorganization of mobility and social networks in response to the COVID-19 pandemic, investigating how sudden and widespread mobility restrictions impacted social and spatial interaction patterns. Using Call Detail Records (CDRs) from Sierra Leone during the early lockdown and curfew periods, this chapter explores shifts in socioeconomic segregation as individuals from different socioeconomic backgrounds adapted to the restrictions at varying rates and capacities. The analysis reveals two key findings: First, while mobility segregation increased as expected due to travel restrictions, social segregation notably decreased. It appears that people compensated for reduced physical interactions by maintaining a more diverse set of social connections, creating a more integrated pattern in social communication networks. Second, the analyses show that, while some aspects of mobility and social interactions were universally affected, the degree of impact differed significantly based on SES, highlighting the complex dynamics and rich phenomenology that can arise from an external shock to human behavior.

Chapter 4 provides an in-depth examination of deviations from established models in human mobility, specifically focusing on the Exploration and Preferential Return (EPR) model. This chapter investigates individual mobility behaviors that diverge from the universal scaling laws of the EPR model, highlighting how these deviations are related to violations of the model's assumptions, particularly regarding bursty behavior and adherence to the preferential return mechanisms. The analysis further shows that these violations occur more frequently during visits to specific categories of places. Moreover, the chapter demonstrates how such deviating behavioral traits are not uniformly distributed across socioeconomic groups and underscores that socioeconomic factors play a critical role in shaping mobility behaviors, suggesting that models based solely on aggregate or population-wide metrics may fail to capture the nuanced variations present in heterogeneous populations. Consequently, Chapter 4 illustrates the importance of tailoring mobility models to better account for the diversity of socioeconomic factors influencing individual behaviors.

5.2 Implications

The findings in this thesis underscore the importance of socioeconomically disaggregated analyses in understanding mobility and social interactions. Examining human behavior through multiple, stratified behavioral traits provides critical insights with implications for public health, policy-making, and equity. These results highlight how mobility and social networks vary across socioeconomic groups, illuminating differences that, if overlooked, can deepen existing disparities.

One major implication is for epidemic modeling and emergency policy design. Socioeconomically stratified data can be instrumental in developing more equitable policies, particularly in public health. Recognizing the differing mobility and social patterns of lower-income groups—who often have limited flexibility to stay at home—can help prevent an uneven burden of disease exposure and secondary economic impacts on economically disadvantaged populations. Tailoring interventions based on stratified data helps ensure that policies do not disproportionately impact vulnerable groups, as observed during the COVID-19 pandemic, when lower-income individuals faced higher exposure risks due to constrained mobility options [326, 328, 335, 340].

The study also underscores the importance of diverse behavioral representation in mobility models. Conventional models that assume population-wide universality fail to capture the heterogeneity across socioeconomic groups, especially in urban environments where external constraints heavily influence mobility. Such assumptions risk obscuring the unique mobility needs and constraints of lower-income individuals. This adds an additional burden to vulnerable groups, who are also often underrepresented in digital data, resulting in models that may inadequately reflect their reality and overlook their specific needs in urban planning, transportation, and resource allocation [330, 366–368].

An additional implication involves residential segregation and its impact on model accuracy. Our findings reveal that the EPR model, commonly used to simulate urban mobility, demonstrates unequal spatial accuracy, performing better in suburban and rural areas than in densely populated urban centers. This discrepancy likely stems from socioeconomic residential segregation, as wealthier individuals, who generally live in less dense areas, align more closely with model predictions. This mirrors the WEIRD problem in social science, where findings disproportionately represent certain demographic groups (Western, Educated, Industrialized, Rich, and Democratic), leading to skewed conclusions and biases [369]. In this context, the EPR model similarly performs best for a subset of the population—those with greater flexibility in time and financial resources—while underrepresenting lower-income groups in urban settings. Rather than framing this as a fundamental limitation of the EPR model, these deviations highlight the need for adaptations or extensions of the model to account for the population heterogeneities revealed by our analysis. Factors such as SES, demographics, and cultural differences significantly influence mobility patterns and should be integrated into more flexible or context-specific versions of the model. This perspective reinforces the EPR model's validity as a baseline while emphasizing the importance of refining mobility models to capture the full spectrum of human behavior. For equitable policy-making, mobility models must be refined to represent the diversity of human behavior, particularly in cities where these patterns are most complex. Investments in public infrastructure and service allocation should reflect the actual mobility patterns of all groups, especially lower-income populations, to provide adequate support where it is most needed.

In summary, this thesis illustrates that a nuanced understanding of SES-related mobility and social interaction patterns is essential for informed policy-making. By moving beyond population-wide assumptions and being mindful of behavioral differences, future research can support a more equitable approach to urban planning and public health interventions.

5.3 Limitations

While this thesis makes notable contributions, several limitations must be acknowledged, especially concerning data quality. These limitations arise from various aspects of the study, particularly in the use of large-scale behavioral datasets. Recognizing these constraints is essential for accurately interpreting the findings and for guiding future research that may refine the approaches used here.

Data Quality Limitations

Data quality presents key limitations to this study, many of which were discussed in Chapter 1. Large-scale behavioral datasets, such as Call Detail Records (CDRs) and GPS data, offer undeniable advantages for studying populations at scale, yet they are also prone to issues related to privacy, accuracy, and the challenges of inferring SES at the individual level. Despite efforts to address these issues, certain constraints remain due to the sensitive nature of location and communication data, as well as the inherent uncertainties in using socioeconomic proxies.

Firstly, privacy concerns are a major consideration when analyzing CDRs and GPS data, given the sensitivity of individuals' location and communication patterns. To address this, user data are fully anonymized in both datasets to ensure that no identifiable information could be linked back to individuals. Additionally, we refrained from displaying any sensitive information about individual whereabouts or inferred home locations, presenting only aggregated results to maintain user privacy.

Secondly, mobile phone data serves as a proxy for social networks and mobility behaviors, but it cannot directly verify that users visited specific locations or maintained genuine social ties with their contacts. This is particularly relevant for social contacts in the CDR dataset from Sierra Leone. One way to limit the inclusion of non-social contacts could have been to set a threshold on the minimum number of calls between two users to consider their interaction a genuine social tie. However, such a threshold was not applied due to the short time window of the dataset, which covers only one month. Applying this threshold would have significantly reduced the dataset, which is already sparse due to the filtering steps and the SES inference process. Moreover, it would likely remove many interactions that may actually represent real social contacts, as only a small fraction of social ties are expected to share more than one call within a one-month period. Nevertheless, we mitigated the presence of non-social contacts by excluding users with anomalous behaviors or insufficient activity, thereby reducing data points that might distort the analysis.

A further limitation concerns the inference of home locations and the assignment of SES indicators. Home locations are estimated based on nighttime movement patterns, while SES is assigned using average or median values within specified geographic areas. These proxies naturally carry some uncertainty. To improve confidence in inferred home locations, we included only users with a high degree of spatial certainty, minimizing the likelihood of misclassification. For SES assignments, we relied on high-resolution socioeconomic maps, and in the case of CDR data, we matched their spatial resolution as closely as possible to that of communication towers. This approach provided representative SES values for each tower location. However, despite these efforts, the final SES map still shows variable spatial resolution, often resulting in larger SES tracts in rural areas due to the distribution of mobile cell towers, which may obscure finer socioeconomic distinctions in these regions.

Finally, the categorization of venues used in Chapter 4 relies entirely on Foursquare classification. While we have no control over this categorization and cannot test its validity, Foursquare is a leading company in POI data provision and regularly updates its categorization. Therefore, we hypothesize that this limitation does not significantly affect our results. Additionally, the further manual classification into 13 macro groups, as done in [121], is subject to some level of uncertainty. For some venues, the assignment to a specific macro group might be open to interpretation. However, we believe that such cases are limited to a small number of venues and are unlikely to significantly impact the results.

In sum, while these mitigation strategies helped address data quality concerns, the limitations reflect the inherent challenges of using large-scale behavioral proxies, underscoring the need for cautious interpretation of results based on indirect data sources.

Confounding Factors

Although this thesis identifies significant relationships between SES and social and mobility behaviors, establishing a causal relationship remains challenging. Specifically, it is difficult to assert that a user or a group behaves in a certain way solely because of their socioeconomic background. Similarly, determining whether the observed segregation is truly significant from a socioeconomic perspective—or if it emerges naturally from complex interactions between users—is hard. The observed patterns may result from confounding factors that correlate with socioeconomic indicators, which could lead to similar empirical patterns.

In our research, we undertook various measures to account for confounding factors. In Appendix A, we assessed segregation levels using different null models, including the gravity model, the configuration model, and a random SES-rewiring model. This approach helped isolate the contributions of physical distance, degree distribution, and SES distribution to observed segregation. Our findings indicate that segregation levels derived from these null models never reach the empirical levels observed, suggesting a significant degree of socioeconomic preferences in behavior. For instance, assortativity in the gravity model structures was relatively high but did not fully reproduce empirical patterns, indicating that distance effects contribute to but do not fully explain observed segregation. Additionally, in Appendix A, we confirm that the observed patterns are not influenced by work constraints or other confounding factors, such as public holidays or conflict events.

In Chapter 4, to strengthen the association between SES and deviations from the model, we implemented a LASSO regression. This method helps mitigate overfitting by automatically selecting statistically significant features, effectively ruling out variables that lack explanatory power regarding deviations from the model. Furthermore, we controlled for the inherent stochasticity of the EPR model to ensure that the observed patterns were not simply due to varying exploration tendencies. Additionally, we accounted for data size in the regression to prevent deviations that could arise merely from differences in the volume of data available for each individual.

Despite these efforts, establishing a statistically robust and causal link between behavior and SES remains challenging, partly due to the nature of the data. Despite some recent significant efforts [152], addressing this limitation will require further studies and potentially different data collection methodologies to determine the distinct and unambiguous role of SES in shaping behavior.

5.4 Future Research

This thesis sets the stage for future research that could extend and refine the findings presented here. Two main directions for future work are especially promising: enhancing the generalizability and robustness of these findings and developing models that more accurately capture the complexities observed.

Firstly, validating generalizability and robustness is crucial to assess whether the observed patterns between SES, mobility, and social interactions hold in different contexts. This could involve studying populations across varied geographic, economic, and cultural landscapes, as well as examining different time periods, to determine if the results are universally applicable or context-specific. Such studies would strengthen the findings and provide insights into how SES-driven mobility and social behaviors vary across settings, potentially revealing location- or time-specific influences that shape these dynamics.

Secondly, advancing mobility and social interaction models represents an important frontier. While this thesis linked various modeling frameworks to observations and critically examined their limitations (particularly in Chapter 4), there remains a need to develop new models that can more accurately replicate observed patterns while accounting for behavioral differences across socioeconomic groups. In terms of segregation, exploring the interplay between external constraints, activity levels, SES homophily, and spatial distribution could yield new insights, not only by explaining observed phenomena but by enabling scenario simulations through parameter tuning. Such models could help simulate potential outcomes under varying conditions, providing valuable foresight for policy applications.

Regarding the unequal representation of different groups in the EPR model, future work could involve developing a modified version of the EPR model that incorporates SES distinctions, allowing for memory effects like burstiness and recency. Such an extension would capture the diversity of movement patterns and increase the model's applicability in urban, socioeconomically diverse settings.

Importantly, future models should strive to balance simplicity with realism: moving beyond overly simplistic frameworks that fail to capture real-world complexity, while avoiding excessive parameterization that could lead to overfitting. This balanced approach would produce models that are both theoretically meaningful and practically useful, capable of revealing underlying mechanisms without merely replicating observations.

In summary, we hope that this thesis serves as a catalyst for future, exciting research exploring these complex phenomena. By continuing to investigate the nuanced relationship between SES, mobility, and social connections, future research can contribute to a deeper understanding of human behavior and inform policies that promote social equity.

Appendix A

Additional metrics and models

In this chapter, we provide a brief overview of the additional metrics and models of human mobility and social networks, that are not explicitly mentioned in Chapter 1 but are still relevant to the fields of human mobility and social networks.

A.0.1 Human mobility metrics and models

Metrics

- **Mean square displacement (MSD):** The mean square displacement (MSD) measures the average squared distance that individuals travel from their starting location over time. This metric is essential for capturing the spatial dispersion of movement. MSD at a given time t is calculated as:

$$MSD(t) = \frac{1}{N}\sum_{i=1}^{N}(\vec{r}_i(t) - \vec{r}_i(0))^2$$

  where N is the total number of individuals, $\vec{r}_i(t)$ is the position of individual i at time t, and $\vec{r}_i(0)$ is their initial position. MSD has been shown to follow a slower than logarithmic growth, indicating an ultra-slow diffusion process [279].

- **Origin-destination matrix:** An origin-destination (OD) matrix represents the flow of individuals between different geographical locations over a specific time period. The matrix M is defined as:

$$M_{ij} = \text{number of trips from location } i \text{ to location } j$$

  OD matrices are widely used in transportation planning, regional mobility studies, and traffic analysis, as they provide a comprehensive view of large-scale population movement patterns.

Individual-level models

- **Social-based models**: As previously mentioned, social connections and physical movements are often correlated. These models consider that human mobility is influenced not just by random factors but also by social connections, as people tend to visit places where their social ties are concentrated [51, 53, 370, 371].

Population-level models

- **Intervening opportunities model**: This model hypothesizes that the movement of people is determined not only by the distance between origin and destination but by the availability of opportunities (e.g., jobs, services) between these points [60]. People starting to travel tend to stop when they reach a destination that offers sufficient opportunities.

- **Radiation model**: The radiation model improves upon the gravity model by incorporating the distribution of opportunities between the origin and destination. Rather than just distance and population size, it considers how the availability of nearby opportunities affects movement decisions [359].

A.0.2 Social network metrics and models

Metrics

Below, we outline the most significant metrics for studying social networks.

- **Clustering coefficient:** The clustering coefficient quantifies the tendency of nodes to form tightly interconnected groups. Locally, the clustering coefficient of a node measures how many of its neighbors are connected to each other. The local clustering coefficient for a node i is:

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

  where e_i is the number of edges between the neighbors of node i and k_i is the degree of node i. The global clustering coefficient, which averages over all nodes, provides insights into whether the network exhibits high levels of clustering, as often found in social networks.

- **Path length and average shortest path length:** The path length between two nodes is the number of edges in the shortest route connecting them. By calculating the average shortest path length across all pairs of nodes, we can determine how efficiently information or influence can spread through the network. The shortest path between nodes i and j, denoted as L_{ij}, provides insights into network connectivity. The average shortest path length across all node pairs is:

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} L_{ij}$$

  where N is the total number of nodes in the network. Networks with a low average path length, like small-world networks, allow information to travel quickly between nodes.

- **Centrality measures:** Centrality measures are used to quantify the importance of nodes in the network. Degree centrality identifies the most connected nodes by counting the number of direct edges they have. Betweenness centrality measures how often a node lies on the shortest paths between other nodes, highlighting its role in controlling information flow. Eigenvector centrality extends this by considering not only a node's direct connections but also the influence of its neighbors.

Together, these centrality measures help identify influential individuals or key hubs in social networks, such as opinion leaders or connectors between different communities.

Network models

- **Watts–Strogatz small-world (SW) model**: The SW model was developed to explain the small-world phenomenon observed in real-world networks, where most nodes can be reached from any other in just a few steps despite high local clustering [372]. Starting with a regular lattice, the model randomly rewires some edges, introducing shortcuts between distant nodes. This approach preserves high clustering while ensuring short average path lengths, making it ideal for modeling the balance between local and global connectivity seen in social networks.

- **Exponential random graph models (ERGM)**: ERGM provides a flexible framework for modeling social networks based on various network statistics, such as reciprocity, clustering, and homophily [373]. Rather than treating tie formation as independent, this model incorporates dependencies between edges, allowing for the study of complex social structures. ERGM is particularly useful for empirical research where network data is available, helping to reveal the social forces driving tie formation, such as similarity or social influence.

- **Stochastic block models (SBM)**: As a generative model, the Stochastic Block Model (SBM) simulates networks by dividing nodes into different blocks or groups and specifying the probability of connections both within and between these blocks [374]. By controlling the probabilities of intra- and inter-group connections, SBM can replicate a variety of social configurations, from strong community structures to highly interconnected social ones. This flexibility allows it to model a wide range of social phenomena and the incorporation of other social tie formation mechanisms, like homophily and triadic closure [375].

- **Temporal network models**: Temporal network models account for the dynamic nature of social networks, capturing how ties form and dissolve over time [376]. These models simulate evolving relationships, enabling the study of how external factors, such as changes in social context or individual preferences, influence the growth and decay of networks. They are particularly valuable for understanding dynamic processes on social networks, such as how information or epidemics spread or how social cohesion changes over time.

- **Activity-driven model**: The activity-driven network model is a dynamic model used to represent networks where interactions are driven by the activity level of each node [377]. At each time step, nodes become active with a probability proportional to their activity potential, and when active, they generate a fixed number of random connections with other nodes. These links are temporary, existing only for a given time step. This model is particularly useful for studying time-varying social networks, such as communication networks, where the structure of connections changes over time, and like other temporal network models is widely applied in research on dynamic processes.

- **Multilayer network models:**: Multilayer network models are an extension of traditional single-layer networks, where nodes can be connected through different types of relationships or interactions across multiple layers [378]. Each layer represents a different type of connection or a different dimension of the system under study (e.g., work relationships, friendships, or family ties in social networks, or different transport modes in mobility networks). These models allow for a more realistic and comprehensive representation of complex social systems where interactions occur on multiple fronts simultaneously.

A.0.3 Segregation metrics

- **Coleman Index:** The Coleman index measures the degree of homophily in social networks, capturing how much individuals in a group tend to interact with others from the same group. It compares the observed number of ties within a group to what would be expected under random mixing. For a group i, it is calculated as:

$$h_i = \frac{\frac{e_{ii}}{E_i} - f_i}{1 - f_i}$$

where $\frac{e_{ii}}{E_i}$ represents the proportion of ties within group i, and f_i is the fraction of nodes belonging to group i. The Coleman index varies between -1 (complete heterophily) and 1 (complete homophily) [379].

- **External-Internal (E-I) Index:** The E-I index measures the balance of connections within and between subgroups in a network [380]. It is calculated as:

$$\text{E-I index} = \frac{E - I}{E + I}$$

where E is the number of external ties (between different subgroups), and I is the number of internal ties (within the same subgroup). The index ranges from -1 to 1, where values close to -1 indicate high internal connectivity, values close to 1 indicate high external connectivity, and a value of 0 represents an equal balance of internal and external ties.

- **Spectral Segregation Index (SSI):** The Spectral Segregation Index (SSI) measures segregation by analyzing the eigenvalues of the network's adjacency matrix. The largest eigenvalue λ_1 represents overall connectivity, while the second-largest eigenvalue λ_2 reflects the extent of community division. The SSI is given by:

$$SSI = \frac{\lambda_2}{\lambda_1}$$

A higher SSI indicates stronger segregation, where distinct groups are well-separated, while a lower SSI suggests more integration across the network [381].

- **Random-walk approaches:** Random-walk-based segregation measures offer a novel perspective on understanding segregation in both social and spatial networks. These approaches calculate segregation by analyzing the probability that individuals encounter members of their own or different groups during random walks on a network [155, 310, 382].

- **Residential Segregation:** The Dissimilarity Index (D) is a widely used metric for measuring residential segregation [149]. It quantifies the evenness with which two groups (e.g., a minority group and a majority group) are distributed across geographic units, such as neighborhoods. The index is computed as:

$$D = \frac{1}{2} \sum_{i=1}^{n} \left| \frac{x_i}{X} - \frac{y_i}{Y} \right|$$

where x_i and y_i are the population counts of the minority and majority groups, respectively, in geographic unit i, and X and Y represent the total populations of the minority and majority groups across all units. The index ranges from 0 (perfect integration) to 1 (complete segregation), indicating the proportion of one group that would need to move in order to achieve an even distribution across all areas.

Appendix B

Socioeconomic reorganization of mobility and communication networks in response to external shocks

B.1 COVID-19 timeline and restrictions in Sierra Leone

The first case of COVID-19 in Sierra Leone was confirmed on March 31, 2020. On April 1, after the second case was confirmed, the government announced a 3-d lockdown to be put in place on April 5-7. On April 9, the government announced a restriction on all inter-district travel for 14 days and a curfew from 9 PM to 6 AM. Also, face masks were strongly encouraged, only shops selling essential items were left open and people were asked to stay at home unless they had extremely urgent reason not to. There were also other types of restrictions, like school and workplace closures. The precise timeline and strength of the nine types of restrictions collected by [5] are shown in Fig. B.1, and the category codes are fully listed in Supp. Tables B.1-B.9

B.2 Reference periods and confounding events

In our study, we analyze the effect of restriction policies by comparing social behavior during lockdown and curfew to two reference periods:

- R1: March 22-24 2020
- R2: March 29-31 2020

We use these specific time windows because they are of the same length as the lockdown (the event of our greatest interest) and during the same days of the week (from Sunday to Tuesday).

Despite being relatively close in time to the lockdown, they are both (especially R1) representative reference periods, in that they are still periods in which people were not significantly affected yet by the outbreak of the pandemic and by response policies. Indeed, we can see from Fig. B.1 that the most socially impactful restriction policies (stay-at-home requirements, restriction to internal movements, restriction on gatherings) were put in place only after R2. However, relevant restrictions like school closures and

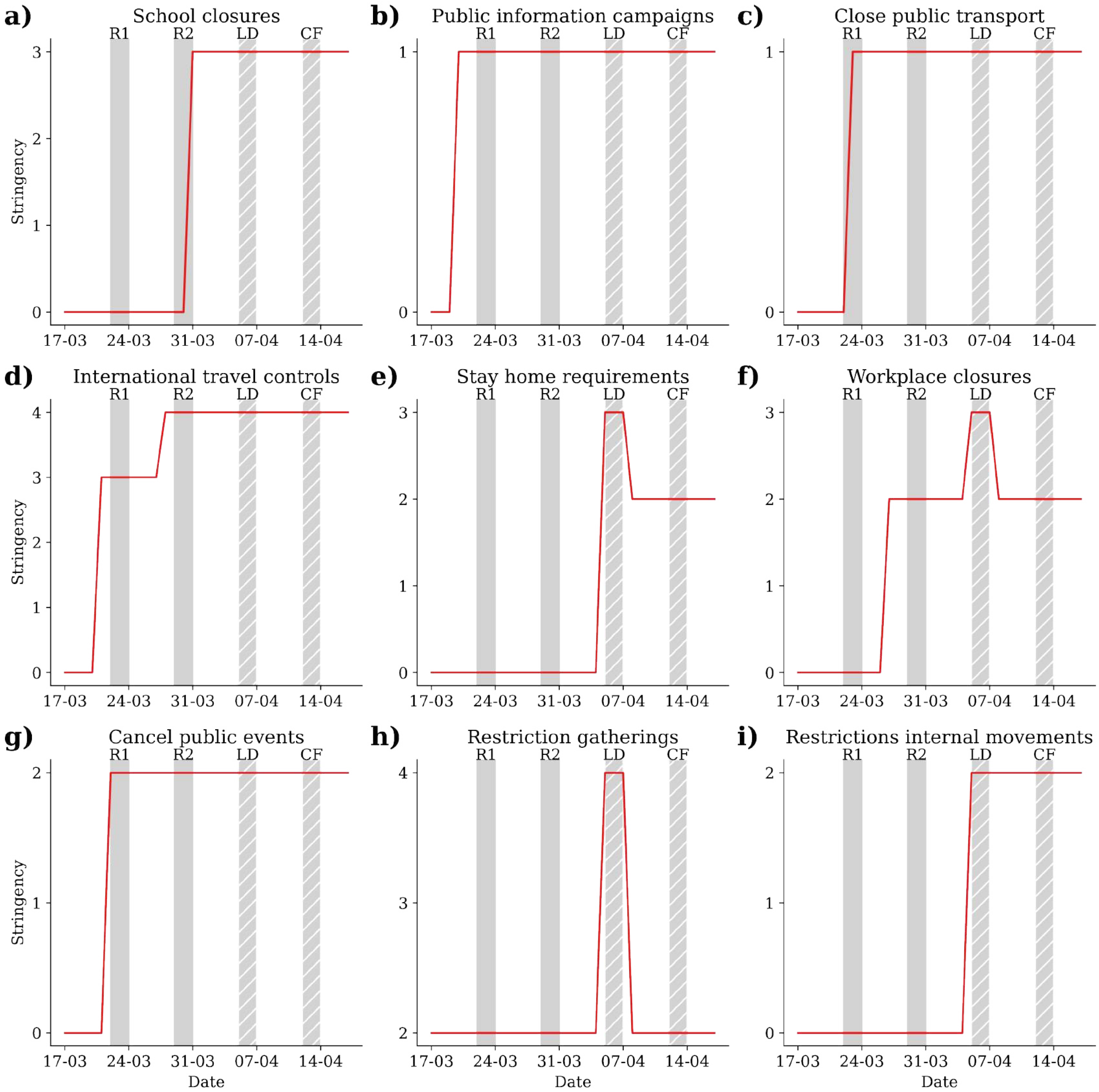

Figure B.1: COVID-19 restriction policies in Sierra Leone. Timeline of 9 restriction measures put in place in Sierra Leone in response to the first wave of COVID-19. The meaning of each category can be found in the Tables section or the original paper [5].

workplace closures were already put in place during R2. The only restrictions that were put in place during R1 were the closure of public transport and the cancellation of public events, but it is unlikely that such measures had a strong effect on people's behavior.

To assess if mobile phone behavior was different before our reference periods we can analyze the results obtained by Ndubuisi-Obi et al. [383]. The authors also had access to CDRs data from Sierra Leone during the first wave of COVID-19 and used it to study compliance with mobility restrictions. However, their data spans a longer time window (February 2020 to May 2020). We can see from Fig. 3.3 in [383] that mobility patterns at the national level did not change significantly in the months before the lockdown, and actually until the very beginning of the lockdown. This demonstrates that our reference periods are valid since a separate analysis of the same type of data shows that the temporal patterns were not significantly different before our reference periods. Moreover, they also show a very sharp decrease in mobility during the lockdown, similar to what we show in Fig.1 of Chapter 3. Their results add validity to our observations.

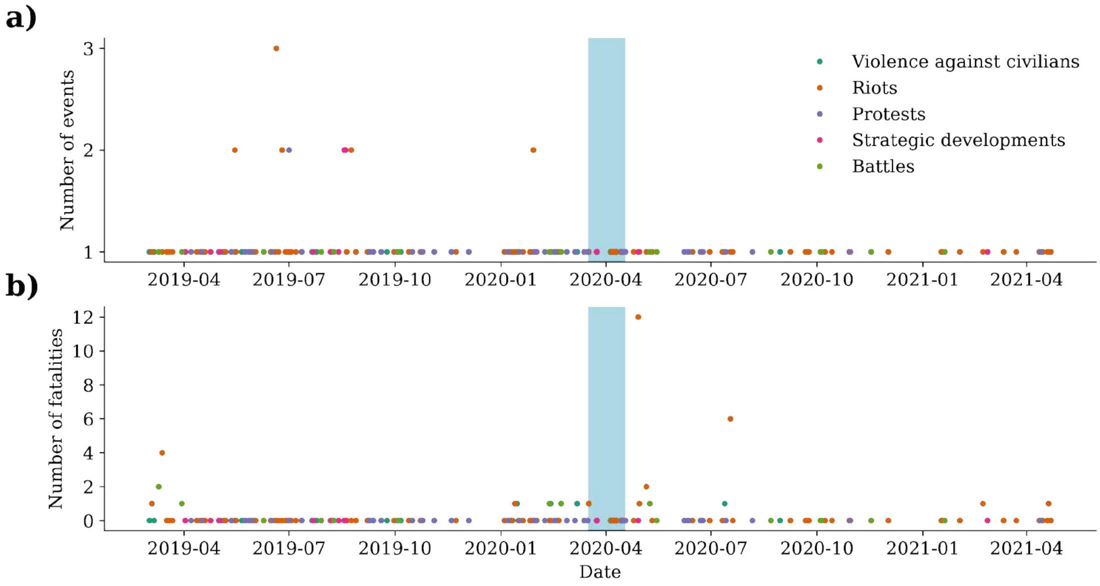

Figure B.2: Armed Conflict Location and Event Data. (a) Number of reported events from different categories. (b) Number of reported fatalities in the events from different categories. The observation period of our study (March 17 - April 17, 2020) is highlighted in light blue in both figures.

Finally, we check if there were notable conflict events that occurred in that period and that could have affected the segregation patterns by analyzing the publicly available Armed Conflict and Location Events Database (ACLED) [384]. In particular, we look at whether there were spikes of violence or protests in the same period of our observation. Fig. B.2 shows this data. We do not observe any particular increase in any type of events before, and during our observation period, so we can exclude that other events could have affected segregation. Also, looking specifically at our observation period (see Fig. B.3), we can see that there are only two reported events with zero fatalities during the lockdown, so we can also exclude that such events have been relevant for segregation.

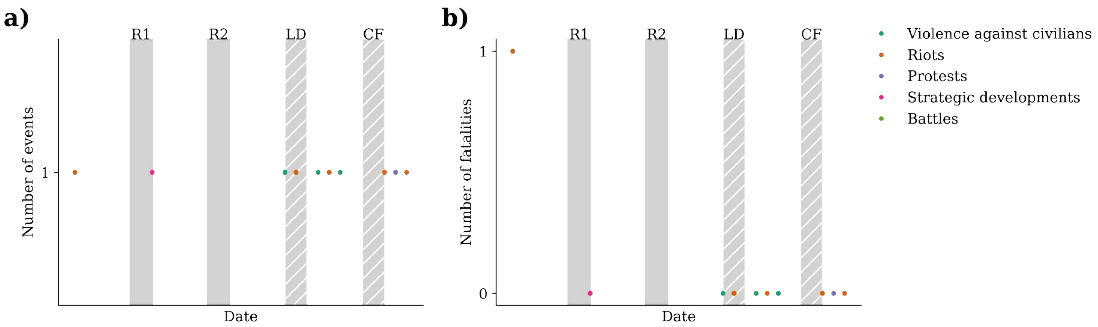

Figure B.3: Armed Conflict Location and Event Data (restricted). Same as Fig. B.2 but restricted to the observation period of our study. (a) Number of reported events from different categories. (b) Number of reported fatalities in the events from different categories. The time windows of interest (R1, R2, LD, CF) are highlighted.

The only festivity during our observation period is the Eastern Weekend, from April 10 (Good Friday) to April 13 (Eastern Monday). This festivity falls during our CF time window and might have had an impact on segregation that we can not disentangle from our observations. However, we think that the restriction policies put in place during the curfew played a major role in driving segregation dynamics. Also, the possible confounding effect of Eastern during the curfew is of relative interest to us as our main focus in this study is the sharp and short-term impact on segregation observed during the lockdown period.

From this analysis, we can conclude that we can consider R1 and R2 as valid reference periods (especially R1) and that we do not notice other relevant events that could have affected significantly segregation during our observation period and especially the shock that we observed during the lockdown.

B.3 Individual segregation

We computed the individual segregation index for each nodes in each network for each observation period (R1, R2, LD, and CF). Subsequently, we measured the $P(r_u(t))$ distribution of this index separately for each SE class and computed the median and standard deviation of this distribution for each period and SE class. This is summarized in Supp. Table B.10 for the $G_S(t)$ and $G_M(t)$ networks separately. For the interpretation of this table see Chapter 3.

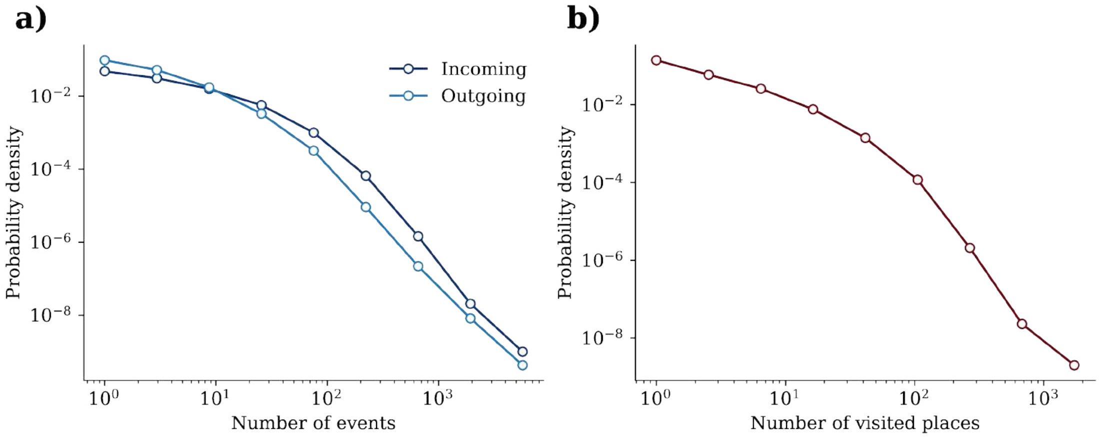

Figure B.4: Event distribution (a) Distribution of the number of incoming and outgoing communication events per user. (b) Distribution of number of visited places per user.

B.4 Segregation measured with entropy

Segregation is often measured in terms of diversity, through entropy-like metrics [92, 121, 295]. The principle behind these metrics is that the more entropic (i.e. diverse) the SES distribution of a given person's contacts, the less segregated the person is. In terms of mobility, the more entropic the SES distribution of places visited by a given person (or of people visiting the same places as a given person), the less segregated the person is. Despite being widely used metrics, we choose not to work with entropy-like metrics and to use assortativity for two main reasons:

- The homophily phenomenon: while entropy only considers the socioeconomic status (SES) distribution of an ego's neighbors without referring it to the ego's SES, assortativity explicitly measures the correlation between an ego's SES and its neighbors' SES. To make an example, if a node that belongs to the poorest class has connections only with the poorest class or only with the richest class, the entropy will be the same (zero). On the other hand, assortativity gives us two opposite outcomes (in this case, segregation and anti-segregation). We believe it is better to use assortativity in our case because it is more nuanced, and captures the homophily mechanisms behind segregation.

- Data sparsity: given the typical long-tail distribution of user's activity (see Fig. B.4), for most individuals, we only observe a few links in a single time window. As such, most people will be assigned low entropy values, if not exactly 0 (see Fig. B.5). Global assortativity is not affected by this issue because it measures one single correlation coefficient between all nodes' SES and their neighbors' SES. Moreover, individual-level segregation overcomes this issue, since it captures the correlation between an ego's network and its local network, by assigning exponentially decreasing weights to distant nodes.

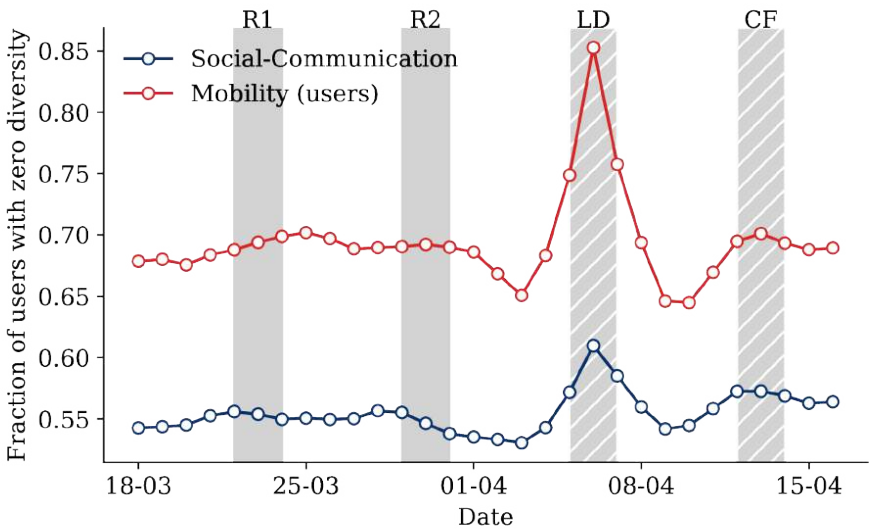

Figure B.5: Users with no diversity. Fraction of active users with $D_m(u) = 0$ (red) or $D_s(u) = 0$ (blue).

One advantage to working with entropy in mobility, however, is that we can compute individual values not only at the location level (the diversity of users visiting a location) but also at the user level (the diversity of places visited by a user).

In this section, we replicate the segregation analysis with an entropy-like measure of diversity and analyze explicitly the two problems mentioned above. Likewise the individual-level analysis, we assign each user and each location to one out of nine socioeconomic classes. Given the set of places visited by a user u, we compute the empirical probability that the user visits places of a given socioeconomic class $P_m(u;c)$, with $c \in \{1, ..., 9\}$, by normalizing the frequency of visits to places of each socioeconomic class. We then define the user mobility diversity $D_m(u)$ as:

$$D_m(u) = -\sum_{c=1}^{9} P_m(u;c) \log P_m(u;c)$$

Similarly, given the set of users visiting a place p, we compute the empirical probability that the place is visited by users from a given socioeconomic class $P_m(p;c)$, with

$c \in \{1, ..., 9\}$, by normalizing the frequency of visits from users of each socioeconomic class. We then define the place mobility diversity $D_m(p)$ as:

$$D_m(p) = -\sum_{c=1}^{9} P_m(p;c) \log P_m(p;c)$$

Finally, given the set of users with whom a user u communicates, we compute the empirical probability that the user communicates with users from a given socioeconomic class $P_s(u;c)$, with $c \in \{1, ..., 9\}$, by normalizing the frequency of communication with users from each socioeconomic class. We then define the user social communication diversity $D_s(u)$ as:

$$D_s(u) = -\sum_{c=1}^{9} P_s(u;c) \log P_s(u;c)$$

Since we have 9 socioeconomic classes, the three metrics are all bounded between 0 (connections to a single socioeconomic class) and $\log 9 \sim 2.20$ (equal frequency of connections with all socioeconomic classes). The temporal evolution of the mean values of $D_m(u)$, $D_m(p)$ and $D_s(u)$ is shown in Fig. B.6.

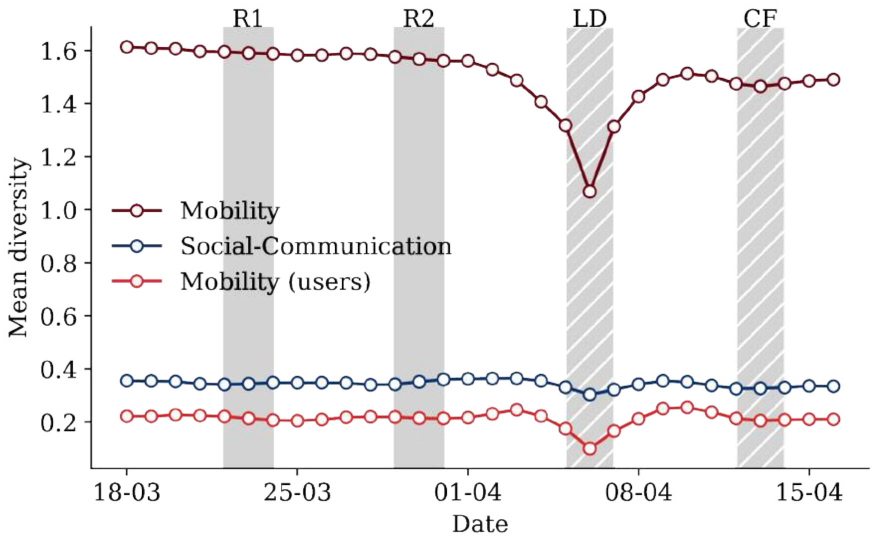

Figure B.6: Mean diversity. Dynamics of mean diversity for mobility in terms of locations (dark red) and users (light red) and for social communication (blue).

Regarding mobility, we can see that places' diversity is overall significantly higher than users' diversity. The main reason behind this difference is data sparsity since single users have much fewer connections than single locations (the same number of user-to-location links is shared among 505,676 nodes on the user side and 405 nodes on the location side). Therefore, for many users $D_m(u) = 0$ simply because they have very few recorded events. We can also see that both diversities strongly decrease during lockdown, which is the analogous observation we do with assortativity in Fig. 1 in Chapter 3 (decreasing diversity is equivalent to increasing assortativity in terms of segregation).

Regarding social communication, the diversity measure allows us to do a direct comparison with mobility at the user level. We can see that on average the set of of contacts of a user is more diverse than the set of visited places. This is likely due to the constraints of physical distance, which are clearly stronger in mobility than in communication (it is easier to communicate with someone living far away than to visit a faraway location, and at the same time faraway places are more diverse in terms of RWI). During lockdown, in Fig. 1 in Chapter 3, we observe a decreasing assortativity, which would correspond to an increasing diversity. However, we can see that diversity slightly decreases during lockdown. The reasons behind this discrepancy are the two issues with entropy-like metrics that we mentioned initially, namely data sparsity and the inability to capture homophily. To demonstrate it, for each time step, we analyze only the set of users with $D_s(u) = 0$, which are always the majority due to data sparsity (see Fig. B.5). At a given time step, these users communicate exclusively with a single socioeconomic class. However, entropy alone is not able to tell us if this single socioeconomic class is similar to the user's socioeconomic class and if this changes with time. We can measure this effect if we group users with $D_s(u) = 0$ by their socioeconomic class and look at the distribution of their aggregate interactions. For example, we take all users with $D_s(u) = 0$ at a given time belonging to class 1 (the poorest). We look at the socioeconomic class distribution of their aggregate interactions (which indicates how many of these users interacted exclusively with class 1, with class 2, ...). Temporal changes of this distribution indicate some homophily change that entropy can not capture because users are anyway assigned with $D_s(u) = 0$. The two distributions for R1 and LD are shown in Fig. B.8(a). We can see from the figure that during LD more users are communicating exclusively with richer (i.e. more distant) socioeconomic classes than in R1. To measure this effect we calculate the entropy of these distributions for every day. From Fig. B.8(b) we can see that during LD the entropy is higher than during reference periods, and this observation holds for every class (see the right panels in Fig. B.8). This indicates that in every socioeconomic class during LD, more users are communicating exclusively (i.e. with $D_s(u) = 0$) with more distant socioeconomic classes than during reference periods. This means in turn that there is a decrease in homophily during lockdown that entropy-like metrics like diversity are not capturing, and given the predominant number of users with $D_s(u) = 0$ (see Fig. B.5) this difference is determinant of the discrepancy that we observe between assortativity and diversity in the social communication network.

Regarding mobility, the same analysis can be applied to users with $D_m(u) = 0$ and it is shown in Fig. B.7. However, in this case, the effect goes in the same direction as the mean diversity, which means that in mobility there is an even stronger increase in homophily and segregation than what is captured by entropy-like metrics.

B.5 Effects of local spatial correlations

As discussed in Chapter 3, in the analysis of segregation dynamics we removed all home-to-home events from both $G_M(t)$ and $G_S(t)$ networks. Here we check that the main results of the segregation analysis are not altered by these removals. In Fig. B.9 we show the analogous panels of Fig. 1 in Chapter 3, without removing home-to-home events. We can see that the presence of home-to-home events induces the sharply diagonal shape of the distributions in panel (c) of Fig. B.9, with a network assortativity coefficient ρ close to 1, visible in panel (e). However, we can see that the main result claimed in Chapter 3 (the opposite direction of the network assortativity change of $G_M(t)$ and $G_S(t)$ during the lockdown) is clearly visible also in Fig. B.9, and hence not determined by the presence/absence of home-to-home events.

Also, in Fig. B.10 we show the analogous of the bottom panels (D-F) of Fig. 2 in Chapter 3 for the social communication network $G_S(t)$, without the removal of communication events between people with the same inferred home location. For this result, we

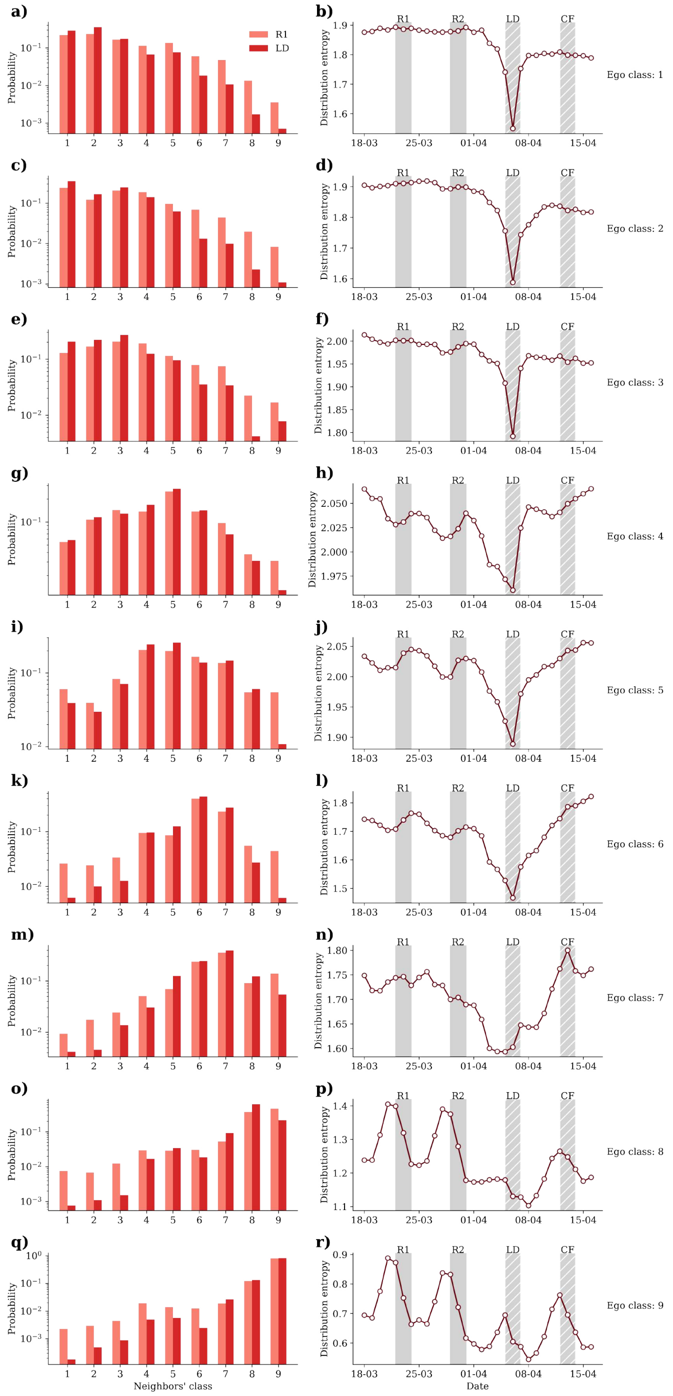

Figure B.7: Homophily changes not captured by diversity in the mobility network. (Left figures) Comparisons between the socioeconomic distribution of the aggregate interactions of users with $D_m(u) = 0$ during R1 and LD, for all classes from class 1 (a) to class 9 (q). (Right figures) Dynamics of the entropy of the socioeconomic distribution of the aggregate interactions of users with $D_m(u) = 0$, for all classes from class 1 (b) to class 9 (r).

can not reproduce the results in Chapter 3 for the mobility network $G_M(t)$ without the removal of home-to-home travels as the local assortativity index $r_u(t)$ is defined only for a network with no self-loops. Since nodes are locations in the mobility network $G_M(t)$, home-to-home links are trips that start and end in the same node (hence self-loops).

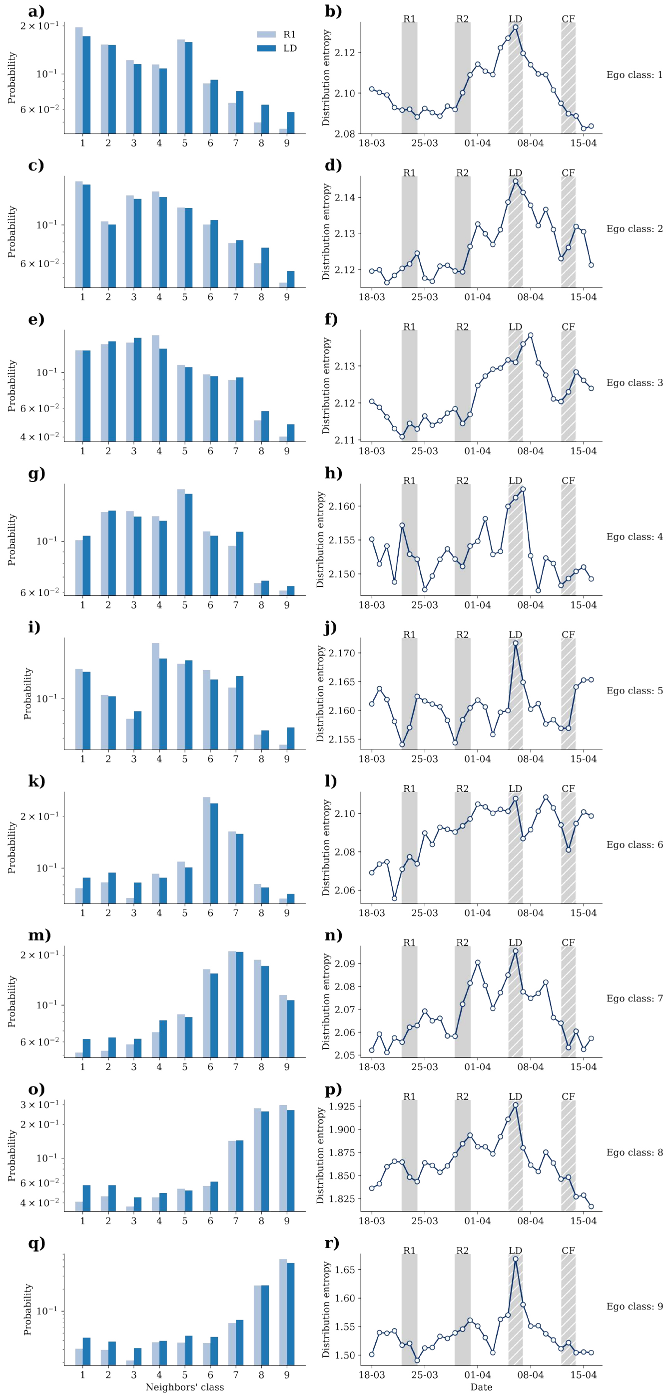

Figure B.8: Homophily changes not captured by diversity in the social communication network. (Left figures) Comparisons between the socioeconomic distribution of the aggregate interactions of users with $D_s(u) = 0$ during R1 and LD, for all classes from class 1 (a) to class 9 (q). (Right figures) Dynamics of the entropy of the socioeconomic distribution of the aggregate interactions of users with $D_s(u) = 0$, for all classes from class 1 (b) to class 9 (r).

On the other hand, in the social communication network $G_S(t)$ home-to-home links are communication events between different nodes that live in the same place (hence are not self-loops). We can see from panels (a), (b), and (c) in Fig. B.10 that the results change only slightly from the bottom panels of Fig. 2 in Chapter 3. The main message

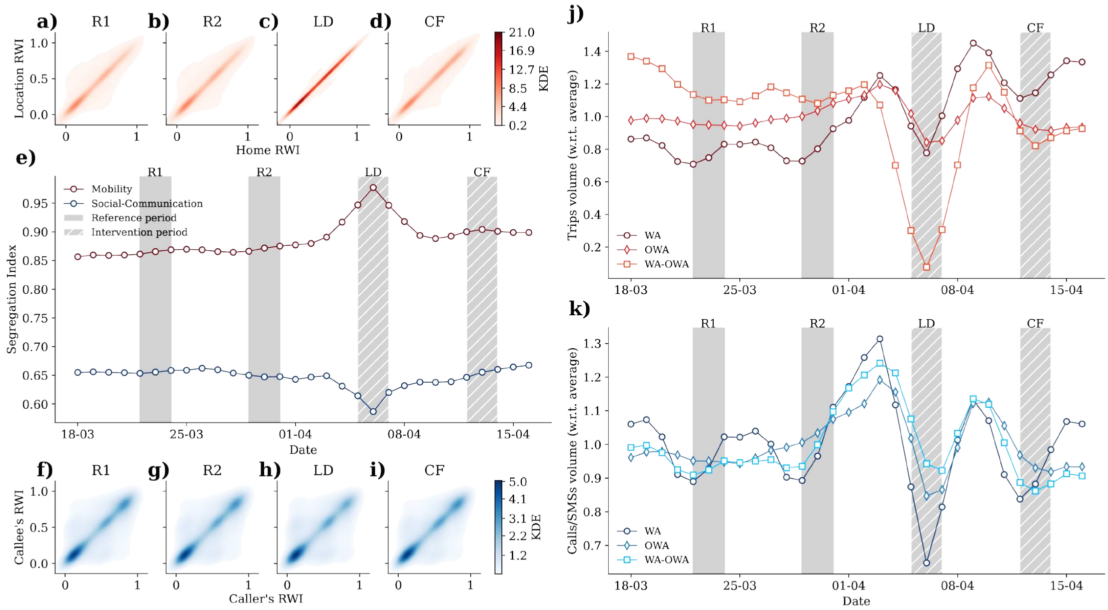

Figure B.9: Spatial effects on global segregation. Same calculations of Fig. 1 in Chapter 3, here made without removing links between nodes located at the same location (respectively, movements to the home locations and calls/SMSs to users with the same home location as the caller). (a-d) SE assortativity matrices (shown as the kernel density of the joint probability of RWIs) of the mobility network during the two reference (R1 and R2), lockdown (LD), and curfew (CF) periods. (e) The dynamics of the ρ SE assortativity index computed for the mobility (red) and social communication (blue) networks. (f-i) Same as (a-d) but for the social communication network. (j) Relative number of travels within WA, OWA, and between the areas WA-OWA (also accounting for OWA-WA trips). (k) Number of communication events between people living in WA and OWA, or between the two geographic areas. All curves are normalized by their average computed over the full data period. For calculations on panels (e), (j), and (k), we used 3-day symmetric rolling time windows with a 1-d shift to obtain aggregated networks around the middle day at time t of the actual window.

remains the same, with a clear dependence on the SES of the relative shift of the distribution $P(r_u(t))$ during intervention periods (panel c). The majority of classes become less segregated than during reference periods (the poorer the class, the higher the segregation decrease) and only the richer classes become more segregated, leaving the global segregation to decrease just like in Fig. 3E in Chapter 3.

Rolling time window

As explained in the caption of Fig. 1 in Chapter 3 for all calculations we use a 3-d rolling time window with 1-d shift. To make an example, if $t =$ April 10, $G_M(t)$ or $G_S(t)$ are the aggregate networks obtained from all the movements or communication events (resp.) recorded between April 9 and April 11. We make this choice to smooth the time series and to reflect the 3-d nature of the lockdown implemented by the Government of Sierra Leone during April 5-7. In this way, there is one point in the time series (April 6th) that incorporates all and only the interactions recorded during the lockdown.

However, the rolling time window can hide or smooth weekly patterns. We show the "raw" time series (without a rolling time window) in Fig. B.11: in this case, every point refers to the segregation index computed from all the events recorded within a

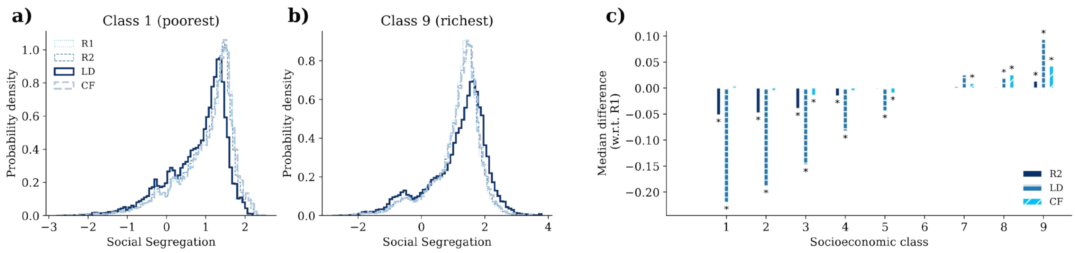

Figure B.10: Spatial effects on individual segregation in the social communication network. Same calculations of the bottom panels of Fig. 2 in Chapter 3, here made without removing links between nodes located at the same location (calls/SMSs to users with the same home location as the caller). The $P(r_u(t))$ individual assortativity index distributions for the poorest (class 1 in a) and the richest (class 9 in b) SE classes for the two reference periods (R1 and R2, thin dashed lines), the lockdown (LD, solid line), and curfew (CF, dashed thick line). Panel (c) depicts the pairwise differences of median assortativity values of $P(r_u(t))$ for each of the nine SE groups in the social communication network. Differences are calculated pairwise between R1 and the R2, LD, and CF periods. The asterisk symbols over the bars (when bars are positive, otherwise under them) in panels (c) indicate statistically significant differences computed with the one-tailed Mann-Withney U-test (with p-value < 0.01)

given day. We can clearly see that our main finding is not an effect of the rolling time window. Indeed, during the lockdown days, both curves show the same trends that we find in Fig. 1 in Chapter 3, with segregation increasing significantly in the mobility network and decreasing significantly in the communication network, with respect to reference periods. However, we can also clearly see that the rolling time window is hiding some weekly patterns related to work and school routines (working days in Sierra Leone are from Monday to Friday). From the mobility curve in Fig. B.11 we can see that during reference times mobility segregation is higher during weekdays than during weekends. The same can be said, to a lesser extent, for communication. We can link this observation to the fact that people during the weekend have normally more free time to explore different places and to communicate with a different set of people than during work days. Also, we can distinguish Sunday during the lockdown in the communication curve much more than in reference periods, while we can not in the mobility curve. This means that while the weekly pattern of social mixing is flattened in the physical space by stay-at-home requirements, it is amplified in the communication space.

B.6 Effects of professional activities

As we discussed in Chapter 3, one of the possible confounding effects that can induce the reduction of mobility and social communication activities during the lockdown period is the lack of professional communications due to interrupted businesses and closed offices. Here we investigate this factor by separating the mobility and call activities in our data for office hours, between 9 AM and 7 PM, and out-of-office hours (7 PM-9 AM) and recompute the daily segregation indices for the networks constructed from events of mobility and communication falling within these periods, with the same resolution we used in Chapter 3 (three-day time window with daily shift).

We can see from Fig. B.12 that the effects of professional activities are not the driver of the segregation changes during the lockdown, neither for the mobility nor for the social communication network. Indeed, separating the activity during office hours from the one during out-of-office hours, we do not find significant differences in the

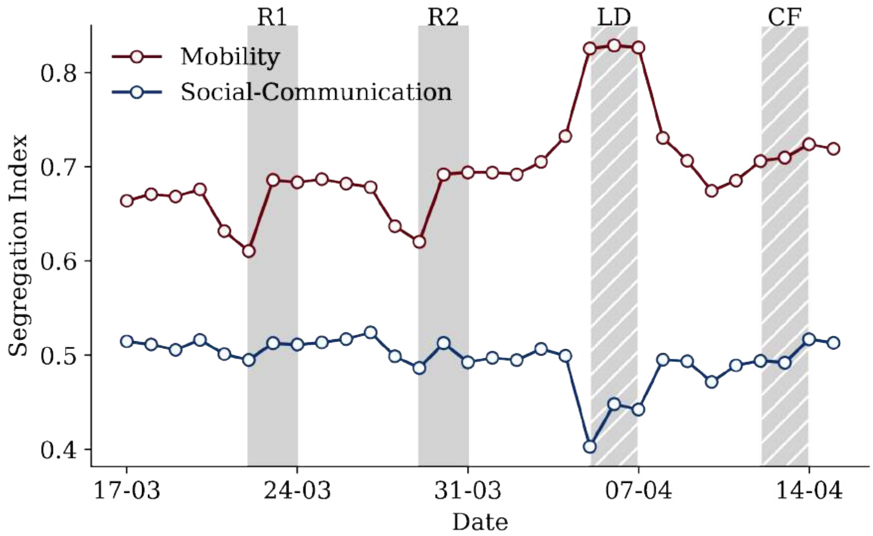

Figure B.11: Raw segregation time series. Segregation dynamics for the mobility (red) and social communication (blue) networks, with no rolling time window (every point refers to all events recorded within a given day).

segregation dynamics. The only relevant difference is between the two curves of social communication segregation, where segregation during working hours is systematically higher than during the remaining times. Nevertheless, a very similar segregation decrease appears during the lockdown in both curves, proofing that these reorganization patterns were not due to the interruption of professional communications.

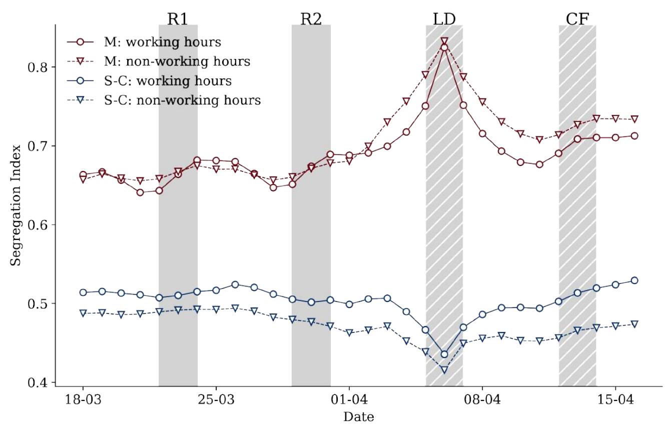

Figure B.12: Effects of professional activities. Segregation curves for mobility (red) and social communication (blue) networks, obtained by explicitly separating the activities during office hours (9 AM - 7 PM, solid lines) from out-of-office hours (dashed lines).

B.7 Effects of spatial and network correlations

In Chapter 3, we have shown signs of SE segregation both in mobility and in the social network. However, also simple random network models might produce positive segregation values, thus making our observations ambiguous from the point of status homophily. Indeed, segregated configurations might result from the convolution of multiple factors,

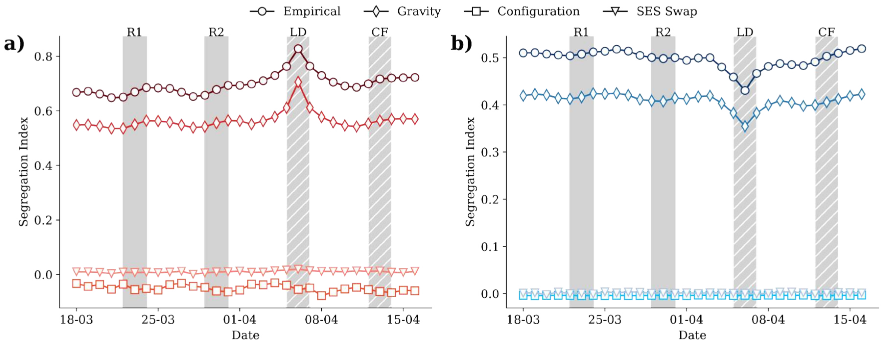

Figure B.13: Reference models. Evolution of segregation levels obtained from the data (darkest curves with circles, the same as in Fig. 1E in Chapter 3) and from reference models (lighter curves) in the (a) mobility and (b) social communication networks. As reference models the gravity model (diamond curve), the configuration model (square curve), and SES label swapped reference model (triangle curve) were considered.

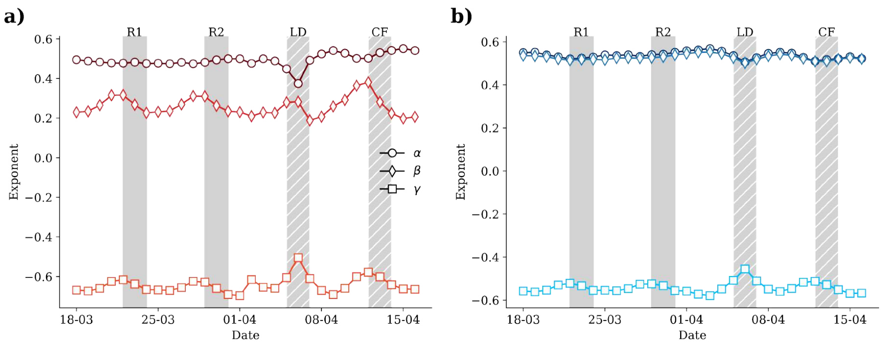

Figure B.14: Gravity models The three exponents (alpha, beta, gamma) of the gravity model, fitted with an OLS linear model, in the (a) mobility and (b) social communication network.

among which status homophily provides only one explanation. To assess the significance of status homophily in the assortative network formation, we identify three main confounding factors and measure separately their contributions to the observed segregation levels.

Physical distance

It is known that distance has a strong determining effect on spatial network formation [385]. On one hand, nearby places might host populations with similar SE profiles. On the other hand, nearby places are likely to have more mobility or even communication connections among each other. The convolution of these two effects could explain the observed segregation patterns. To measure the impact of physical distance, we consider a gravity model [385], where the number of connections W_{ij} between places i and j is only determined by the number of people living in the two places (N_i and N_j) and the distance d_{ij} between them, according to the law:

$$W_{ij} = C \frac{N_i^{\alpha} N_j^{\beta}}{d_{ij}^{\gamma}} \quad \text{(B.1)}$$

The three exponents α, β, and γ, and the constant C are fitted at every time t from the data with an ordinary least square (OLS) regression. The values of the exponents can be seen in Fig. B.14. As evident from Fig. B.13, assortative indices computed from the gravity model structures are relatively high but do not fully reproduce the level of segregation observed in the real network. Consequently, distance effects contribute to the emergence of the observed segregation but they do not fully explain them.

However, we can see that gravity model curves have similar temporal tendencies as the empirical ones, which is due to overfitting. Indeed, we fit a gravity model per day, instead of fitting a model for the full time-period. We say the model is overfit because we are fitting the three exponents of the gravity model at every timestep (i.e. every day), and predicting the weights of the same set of links on which the model was trained. The three exponents are supposed to be universal, but as seen in Fig. B.14 they fluctuate over time and significantly change during the lockdown. The purpose of this analysis is to show that even the most overfitted gravity-like model, despite having strong explanatory power, is not able to reproduce the same level of segregation, implying that there is also a social preference mechanism that comes into play.

SE network correlations

Next, we consider the impact of SE status and network correlations on the observed assortativity correlations to see whether they are contributing at all to the emergent network segregation patterns. To test the impact of the RWI distribution, we randomly swap the RWI labels among nodes (being people in $G_s(t)$ or places in $G_m(t)$), keeping the overall distributions and the network structures fixed [386]. In our implementation we perform 100 random swap iterations and consider the mean assortativity value, for every day, resulting from such iterations. This procedure destroys the network-SES correlations and results in assortativity indices close to zero (see Fig. B.13). Consequently, SE correlations are important in the emergence of network segregation patterns, as their removal leads to the vanishing of the observed patterns.

Network structure correlations

Finally, our last goal is to verify how much degree heterogeneities are important for the emergence of the segregation patterns in the networks. We test these effects by using configuration network models where we swap the ending nodes of a pair of edges to remove any structural correlations from the network [387]. Note that this method keeps intact the degrees (number of connections) of nodes and the degree-SES label correlations. Also in this procedure we perform 100 random swap iterations and consider the mean assortativity value, for every day, resulting from such iterations. Results are shown in Fig. B.13, where the assortativity indices measured in the configuration networks appear around zero, indicating that the removal of structural correlations completely destroys network segregation, thus the degree distribution and degree-label correlations do not contribute at all to the original observations neither in case of the mobility or the social communication network.

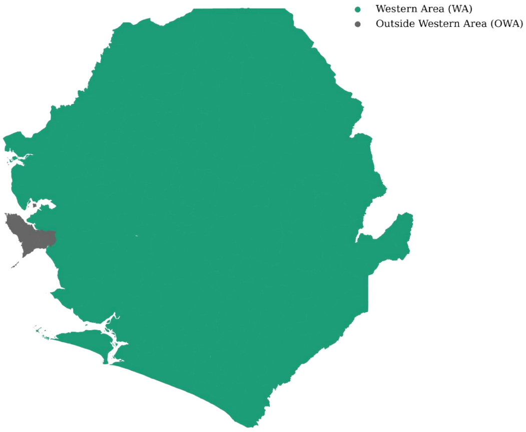

Figure B.15: Spatial division. Western Area (WA) in grey and Outside Western Area (OWA) in green.

Table B.1: School closures

0	No measures
1	Recommend closing
2	Require closing (only some levels or categories, e.g. just high school, or just public schools)
3	Require closing all levels

Table B.2: Workplace closures

0	No measures
1	Recommend closing (or work from home)
2	Require closing (or work from home) for some sectors or categories of workers
3	Require closing (or work from home) all but essential workplaces (e.g. grocery stores, doctors)

Table B.3: Cancel public events

0	No measures
1	Recommend cancelling
2	Require cancelling

Table B.4: Restrictions on gatherings

0	No restrictions
1	Restrictions on very large gatherings (the limit is above 1,000 people)
2	Restrictions on gatherings between 100-1,000 people
3	Restrictions on gatherings between 10-100 people
4	Restrictions on gatherings of less than 10 people

Table B.5: Close public transport

0	No measures
1	Recommend closing (or significantly reduce volume/route/means of transport available)
2	Require closing (or prohibit most citizens from using it)

Table B.6: Public information campaigns

0	No COVID-19 public information campaign
1	public officials urging caution about COVID-19
2	coordinated public information campaign (e.g. across traditional and social media)

Table B.7: Stay at home

0	No measures
1	recommend not leaving house
2	require not leaving house with exceptions for daily exercise, grocery shopping, and 'essential' trips
3	Require not leaving house with minimal exceptions (e.g. allowed to leave only once every few days, o

Table B.8: Restrictions on internal movement

0	No measures
1	Recommend movement restriction
2	Restrict movement

Table B.9: International travel controls

0	No measures
1	Screening
2	Quarantine arrivals from high-risk regions
3	Ban on high-risk regions
4	Total border closure

Table B.10: Median values and standard deviations of individual assortativity index distributions. Values are computed for nodes (locations in the mobility network or people in the social network) from the nine SE classes during the two reference periods (R1 and R2) and the intervention periods (LD and CF). Standard deviation values are shown in parentheses.

G_M	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
R1	0.41 (0.16)	0.33 (0.14)	0.24 (0.11)	0.16 (0.1)	0.1 (0.04)	0.09 (0.01)	0.08 (0.07)	0.2 (0.11)	0.34 (0.11)
R2	0.4 (0.16)	0.33 (0.14)	0.24 (0.12)	0.15 (0.1)	0.1 (0.04)	0.1 (0.01)	0.09 (0.07)	0.21 (0.1)	0.32 (0.11)
LD	1.28 (0.32)	1.1 (0.3)	0.92 (0.29)	0.62 (0.33)	0.48 (0.24)	0.31 (0.08)	0.28 (0.06)	0.46 (0.11)	0.62 (0.11)
CF	0.59 (0.24)	0.47 (0.19)	0.4 (0.17)	0.24 (0.16)	0.15 (0.08)	0.13 (0.02)	0.12 (0.05)	0.23 (0.07)	0.34 (0.09)
G_S									
R1	1.24 (0.71)	0.97 (0.62)	0.78 (0.52)	0.39 (0.32)	0.17 (0.18)	0.12 (0.17)	0.18 (0.34)	0.86 (0.61)	1.28 (0.75)
R2	1.18 (0.71)	0.92 (0.6)	0.75 (0.51)	0.38 (0.31)	0.16 (0.17)	0.12 (0.18)	0.18 (0.34)	0.87 (0.62)	1.3 (0.77)
LD	1.02 (0.73)	0.78 (0.62)	0.64 (0.52)	0.31 (0.3)	0.12 (0.17)	0.12 (0.21)	0.21 (0.4)	0.88 (0.73)	1.38 (0.93)
CF	1.24 (0.72)	0.96 (0.62)	0.77 (0.52)	0.39 (0.32)	0.16 (0.17)	0.12 (0.18)	0.19 (0.35)	0.89 (0.63)	1.32 (0.78)

Table B.11: List of p-values (order of magnitude) obtained from the statistical comparison of distributions (shown in Fig. 2 in Chapter 3) with the one-tailed Mann-Withney U-test.

G_M	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
R2-R1	10^{-1}	10^{-1}	10^{-1}	10^{-1}	10^{-1}	10^{-1}	10^{-1}	10^{-1}	10^{-1}
LD-R1	10^{-20}	10^{-21}	10^{-19}	10^{-9}	10^{-9}	10^{-10}	10^{-8}	10^{-10}	10^{-10}
CF-R1	10^{-7}	10^{-7}	10^{-7}	10^{-3}	10^{-3}	10^{-8}	10^{-2}	10^{-1}	10^{-1}
G_S									
R2-R1	10^{-15}	10^{-15}	10^{-11}	10^{-6}	10^{-4}	10^{-1}	10^{-1}	10^{-1}	10^{-4}
LD-R1	10^{-224}	10^{-210}	10^{-169}	10^{-137}	10^{-205}	10^{-101}	10^{-1}	10^{-13}	10^{-77}
CF-R1	10^{-1}	10^{-3}	10^{-4}	10^{-2}	10^{-4}	10^{-5}	10^{-1}	10^{-5}	10^{-16}

Appendix C

Deviations from universality in human mobility modeling

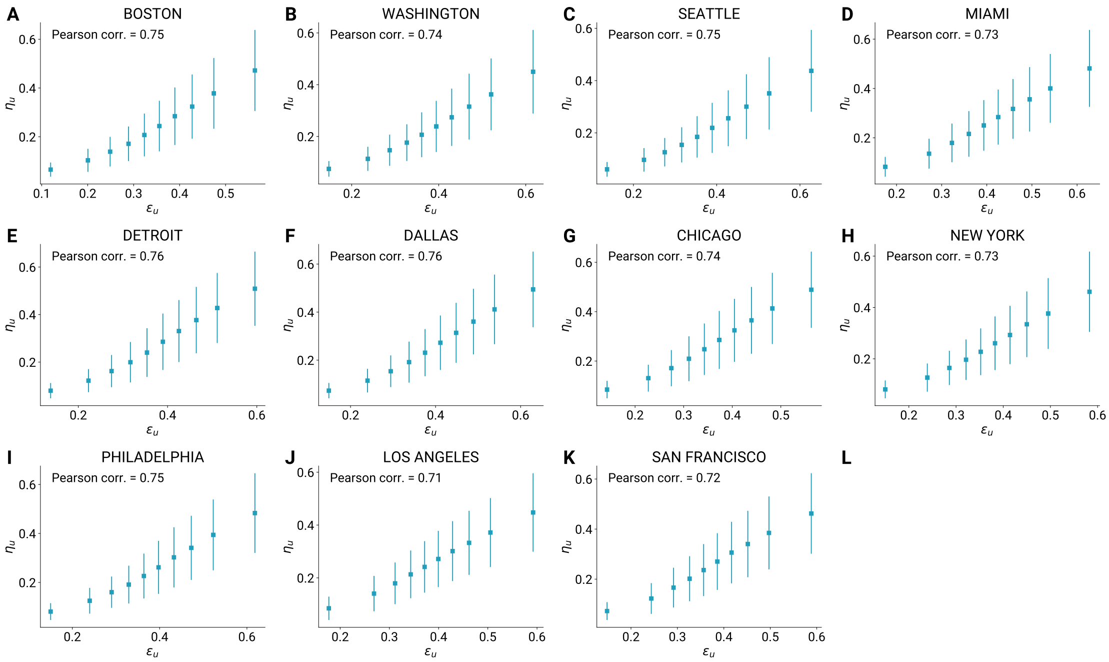

Figure C.1: The relation between ϵ_u and η_u in all CBSAs.

C.1 Deviations

As mentioned in Chapter 4, the two deviation metrics ϵ_u and η_u are highly correlated, despite measuring aspects of the EPR model apparently uncorrelated. Indeed, ϵ_u measures how good the EPR model is in predicting in how much time the next exploration step is going to happen. On the other hand, η_u measures how "far" is the final visitation frequency distribution from the expected $\langle f_k \rangle \sim k^{-\gamma}$. The relation between the two variables is shown in Fig.C.1, for all CBSAs. The Pearson correlation coefficients go from 0.71 in Los Angeles to 0.76 in Dallas and Detroit, indicating that if a user's exploration dynamics is not well described by the EPR model, then also its visitation frequency is likely not well described.

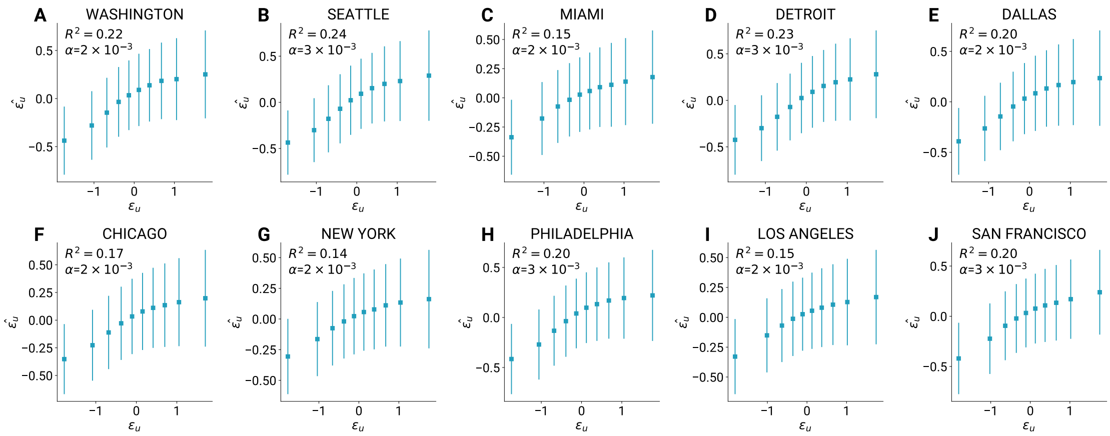

Figure C.2: Results of the LASSO regression for the other CBSAs (dependent variable: ϵ_u): true ϵ_u vs predicted $\hat{\epsilon_u}$. R^2 is the coefficient of determination while α is the regularization parameter.

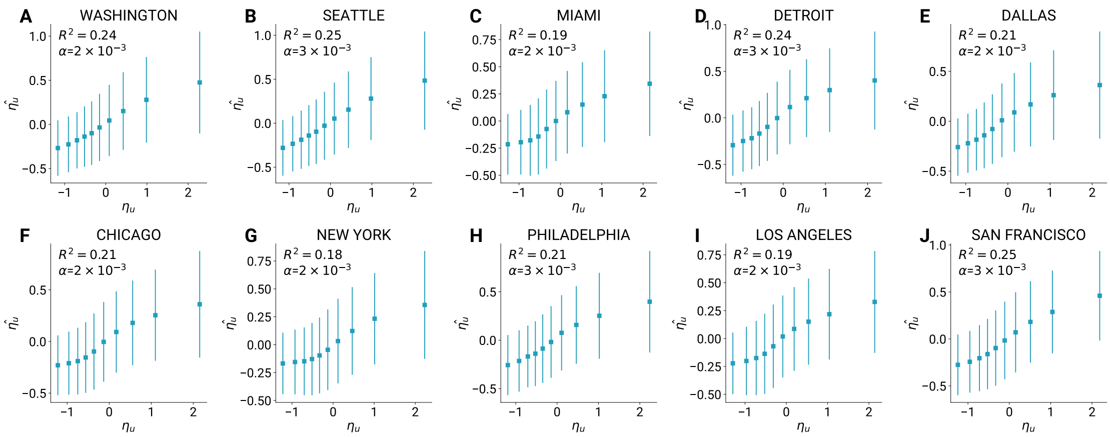

Figure C.3: Results of the LASSO regression for the other CBSAs (dependent variable: η_u): true η_u vs predicted $\hat{\eta_u}$. R^2 is the coefficient of determination while α is the regularization parameter.

C.2 Regression models

As mentioned in Chapter 4, both ϵ_u and η_u can be partially predicted from sociodemographic and life habits features, indicating that the EPR model is biased towards certain groups of people. We perform a LASSO regression for both deviation variables, estimating the regularization parameter α through cross-validation (see the Methods section in Chapter 4 for details), to quantify this bias in every CBSA. The regression results for ϵ_u are shown in Fig. C.2, where we show the relation between the actual ϵ_u and the predicted $\hat{\epsilon_u}$. The coefficient of variations goes from $R^2 = 0.14$ for New York to $R^2 = 0.24$ for Seattle. The results for η_u, on the other hand, are shown in Fig. C.3, where we show the relation between the actual η_u and the predicted $\hat{\eta_u}$. The coefficient of variations in this case goes from $R^2 = 0.18$ for New York to $R^2 = 0.25$ for Seattle and San Francisco. The results show the bias towards sociodemographic and life habit variables is robust and consistent across all CBSAs.

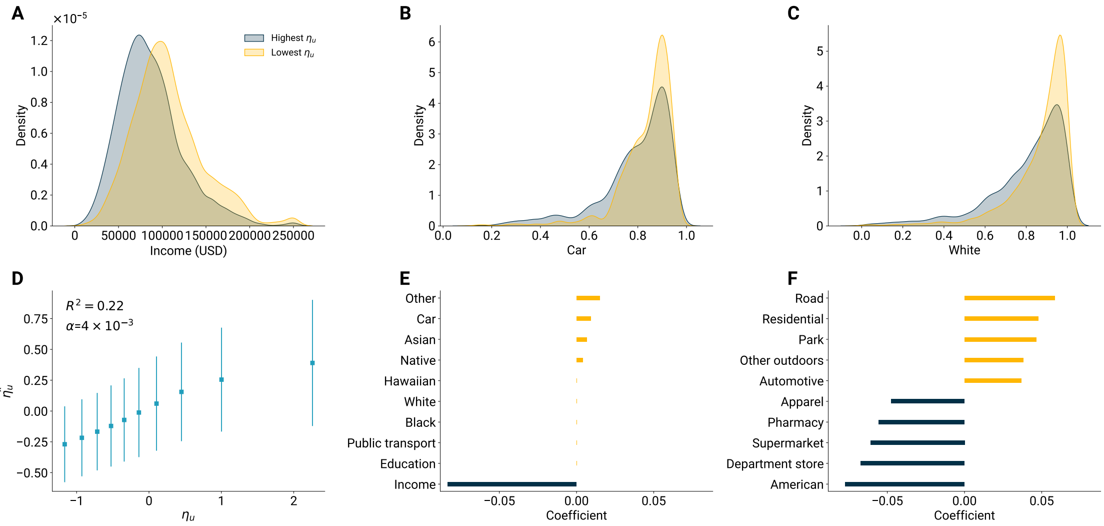

Figure C.4: A) The income distribution of users in the highest (blue) and lowest (yellow) 10% quantile of η_u. B) Same as A) for the use of car. C) Same as A) for the probability of being white. D) Results of the LASSO regression: true η_u vs predicted $\hat{\eta_u}$. R^2 is the coefficient of determination while α is the regularization parameter. E) Coefficients of the socioeconomic features. F) Coefficients of the life habit features (only highest and lowest 5 shown).

C.3 Results for η_u in Boston

The analogous results of Fig.2 in Chapter 4 for η_u are shown here in Fig.C.4. As we can see in the figure, results are consistent with ϵ_u, with the only significant exception being the use of car, which is negatively associated with ϵ_u and positively with η_u.

C.4 Feature importance

In Chapter 4, we measure the bias of the EPR model towards sociodemographic and life habit characteristics in Boston through the coefficients of the LASSO regressions. Here we analyze the same results for all the other CBSAs.

In Fig.C.5 we show the coefficients of the regressions on ϵ_u. As mentioned in Chapter 4, the results are consistent across CBSAs. First, regarding the sociodemographic features (first and third columns in the figure), income is always the most important predictor, with a negative coefficient: the lower the income, the higher ϵ_u. Moreover, also the use of cars has a consistent negative impact on ϵ_u, except for Los Angeles, Philadelphia, and Detroit, where the coefficient is almost zero. Complementary to the use of cars, the use of public transport is either positively associated with ϵ_u or irrelevant. Education is not always significant, but when it is the coefficient is negative. Finally, the role of race is not uniform: a negative coefficient is observed for the probability of being white in New York, Los Angeles, Seattle, and San Francisco, while it is observed for the probability of being black in Washington, Los Angeles, San Francisco, and Chicago. On the other hand, the probability of being Asian is often positively associated with ϵ_u, though with a small coefficient, e.g. in Washington, Dallas, Chicago, and Philadelphia.

Regarding the life habit features (second and fourth columns in the figure), we also find some similar results as in Chapter 4 for the other CBSAs. Indeed, among the categories of places with the highest positive coefficients, we find mostly routine places like roads, factories, buildings, and offices. On the other hand, among places that are

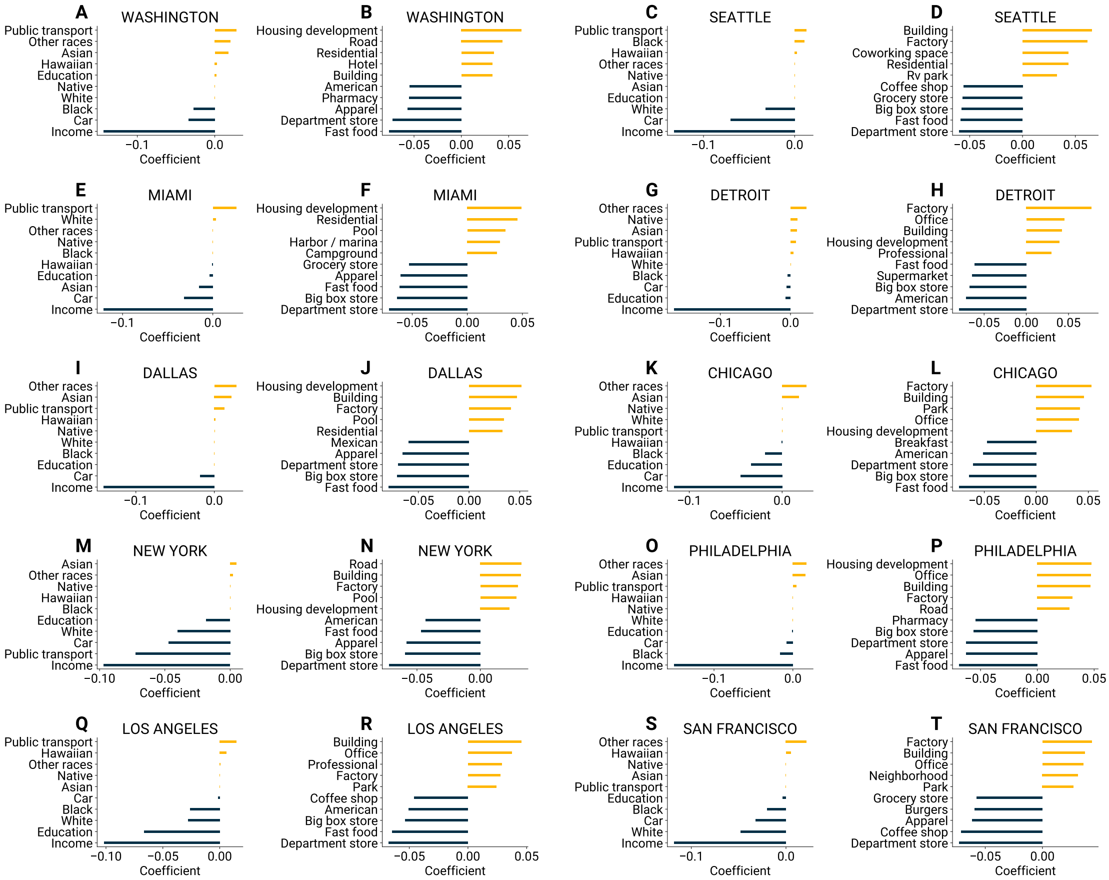

Figure C.5: First and third column: Coefficients of the socioeconomic features for the other CBSAs (dependent variable: ϵ_u). Second and fourth column: Coefficients of the life habit features (only highest and lowest 5 shown) for the other CBSAs (dependent variable: ϵ_u.)

negatively associated with ϵ_u, we find also many restaurants, like fast food and American, and shops, like supermarkets, grocery stores, department stores, and apparel shops.

Regarding the coefficients of the regressions on η_u, shown in Fig.C.6, the only notable difference with the ones for ϵ_u is the coefficient of the use of cars, which is positive and significant in most places, like we observed for Boston in Chapter 4.

C.5 Life habits categories

From Fig.C.7 to Fig. C.19, we show the full list of coefficients of the regression on ϵ_u (panel A) and η_u (panel B) for Boston grouped by macro category (as mentioned in the main text, all categories have been manually grouped in 13 macro categories in [121]), that have an absolute value higher than 10^{-3} for both regressions. In both panels of all figures, coefficients are sorted based on the values of the regression on ϵ_u. Notably, the sign of coefficients is always the same for the two regressions, except only for Cafeteria in Fig. C.9, which has a positive relation with ϵ_u and a negative one with η_u. Moreover, in addition to the sign, coefficients also share similar absolute values. In other words, categories that are important for predicting ϵ_u are likely to be important for predicting η_u, with the same direction.

As mentioned in Chapter 4 and in the previous section, categories that are most positively associated with deviations belong to the "City / Outdoors", "Entartainment",

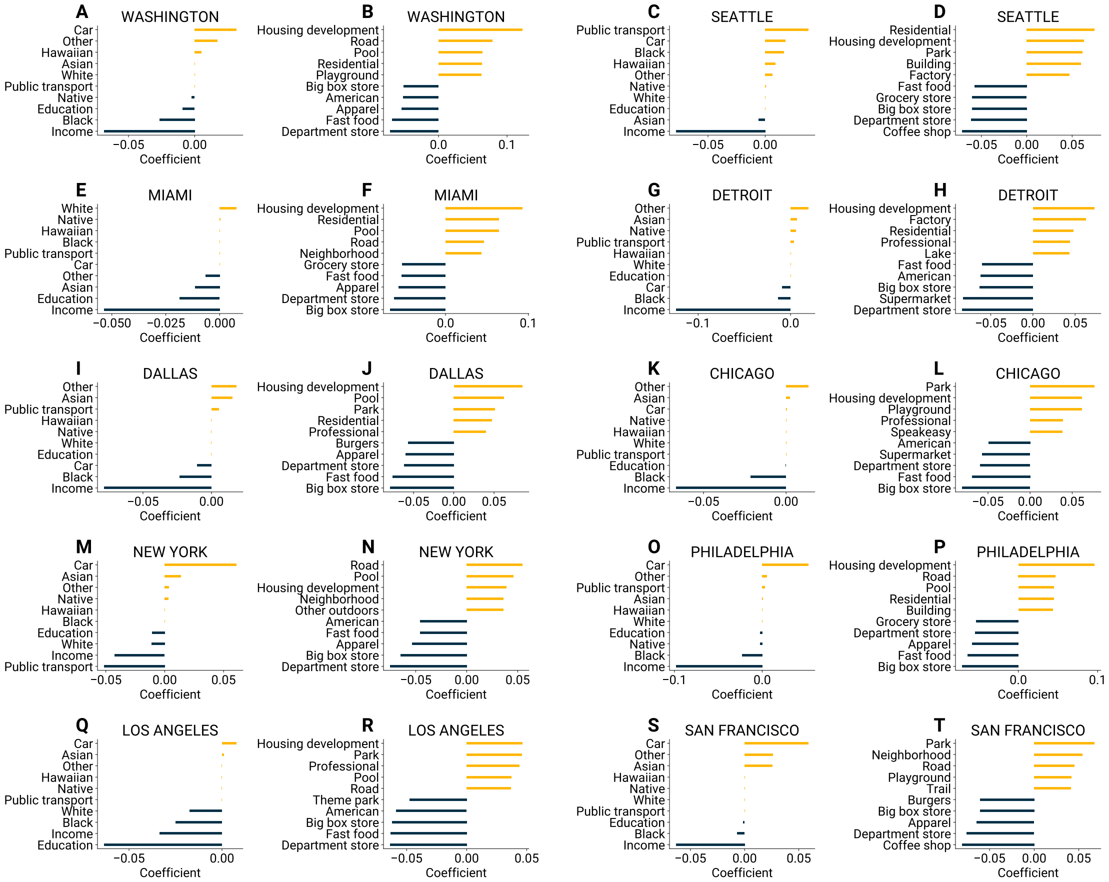

Figure C.6: First and third column: Coefficients of the socioeconomic features for the other CBSAs (dependent variable: η_u). Second and fourth column: Coefficients of the life habit features (only highest and lowest 5 shown) for the other CBSAs (dependent variable: η_u.)

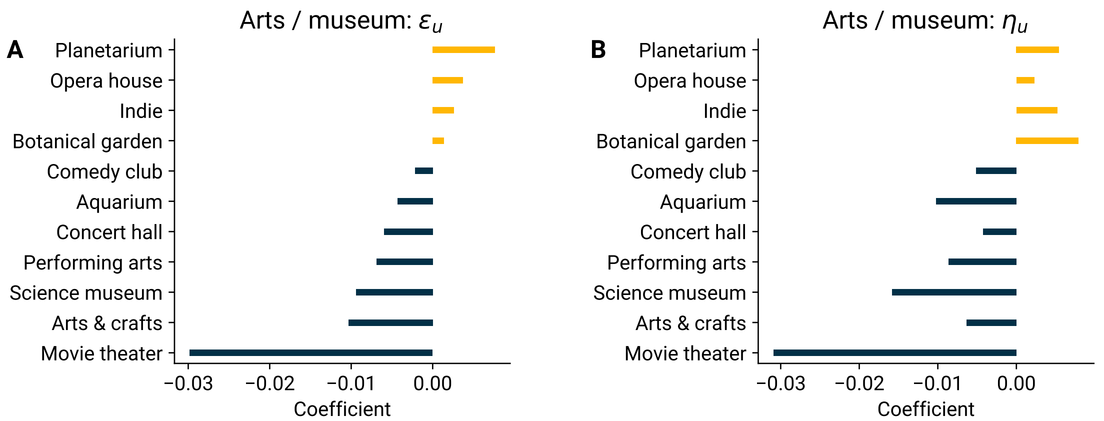

Figure C.7: A) Coefficients of habit features grouped in the "Arts / Museum" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

and "Work" macro groups. On the other hand, categories that are most negatively associated with deviations belong to the "Food" and "Shopping" macro groups.

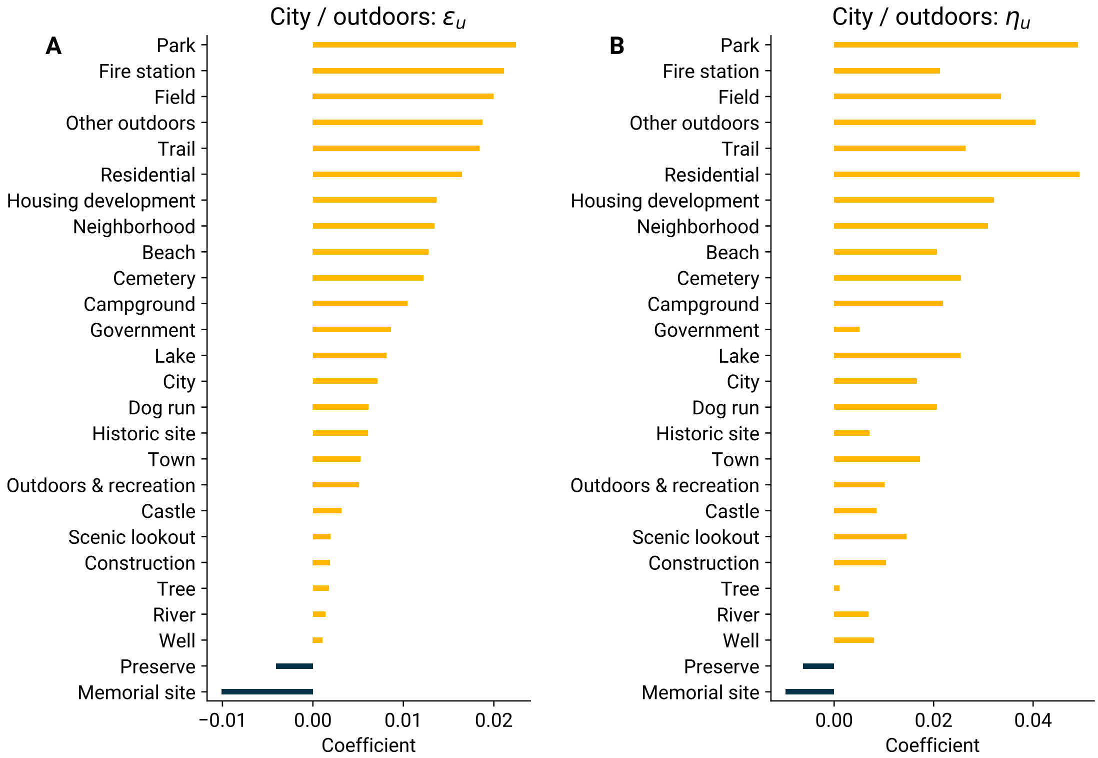

Figure C.8: A) Coefficients of habit features grouped in the "City / Outdoors" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

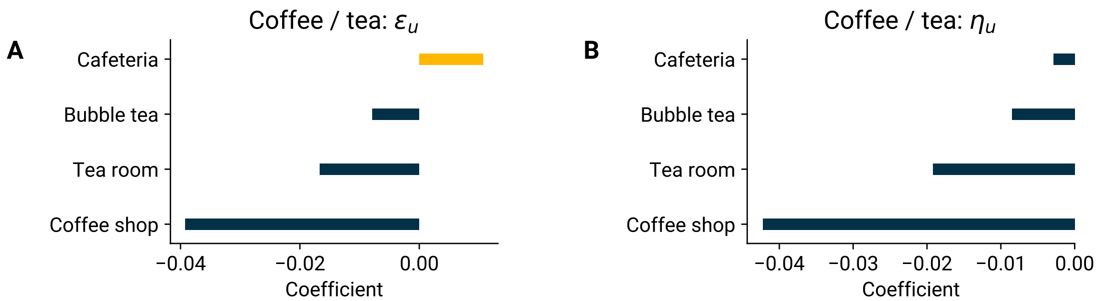

Figure C.9: A) Coefficients of habit features grouped in the "Coffee / Tea" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

C.6 Stochastic deviations

The stochastic deviations that we generate from simulations in the orange distributions in panels C-F of Fig.1 in Chapter 4 are generated using the individual parameters taken from the data, to make the results directly comparable with the empirical deviations, shown in the light blue distributions in the same figure. To generate a stochastic deviation ϵ_u for a user u, we run a simulation of the EPR model for a fictitious user with the same visitation tendency ρ_u and the same number of distinct places S_u of the real user. On the other hand, to generate a stochastic deviation η_u, we run another simulation, us-

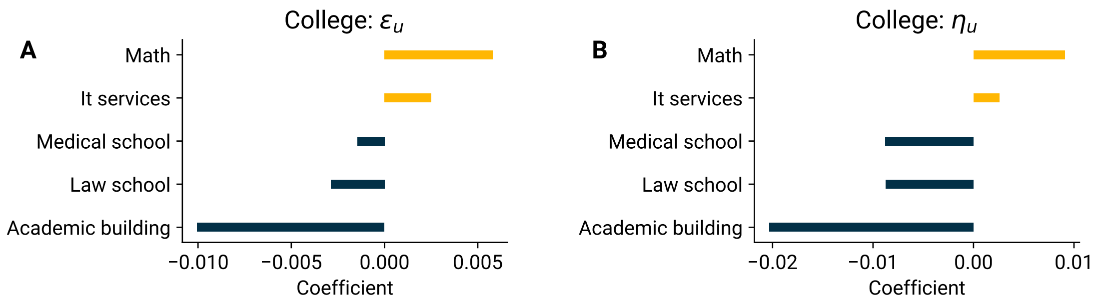

Figure C.10: A) Coefficients of habit features grouped in the "College" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

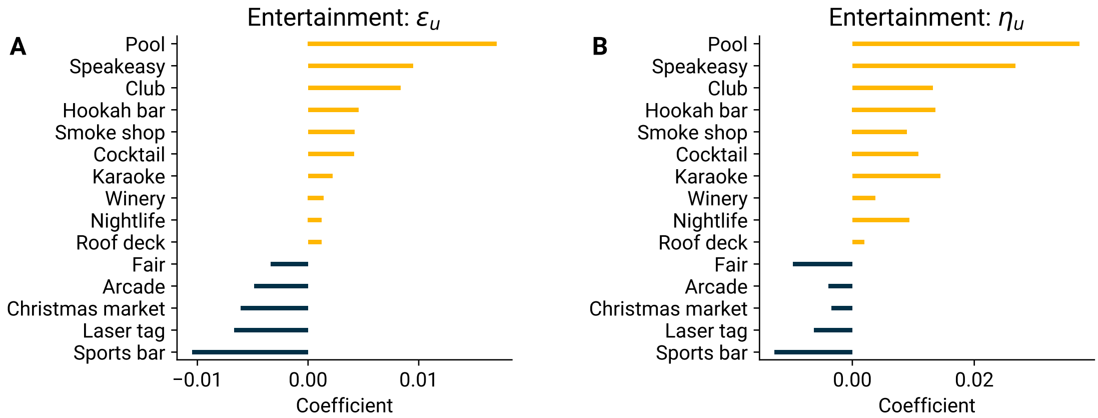

Figure C.11: A) Coefficients of habit features grouped in the "Entertainment" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

ing the same ρ_u and number of steps N_u, then we consider only the visitation frequency of the first K_u distinct places (where also in this case ρ_u, N_u and K_u are taken from the data. The details on the simulations can be found in the following paragraphs, with the only difference that instead of having a distribution of parameters taken from the data of individual users, we fix all of them and tune only ρ_u, because we are specifically interested in measuring the role of this parameter and to do this we need to control for the others.

As mentioned in Chapter 4, the stochastic errors are related to the exploration tendency, encoded in the parameter ρ_u. However, the stochastic deviations that we generate in the orange distributions in panels C-F of Fig.1 in Chapter 4, as we mentioned above, do not only depend on ρ_u but also on other individual parameters like the number of steps N_u and the number of distinct places S_u. To demonstrate the dependence on ρ_u, then, we run other stochastic simulations of the EPR models where we control for the other parameters, such that the final results can be directly associated with ρ_u. We run such controlled experiments for 20 different values of ρ_u, keeping all the other parameters fixed.

For ϵ_u, we consider simulations for $S_u = 100$ distinct places. As explained in the Methods section in Chapter 4, in the EPR model the inter-event time is drawn from the

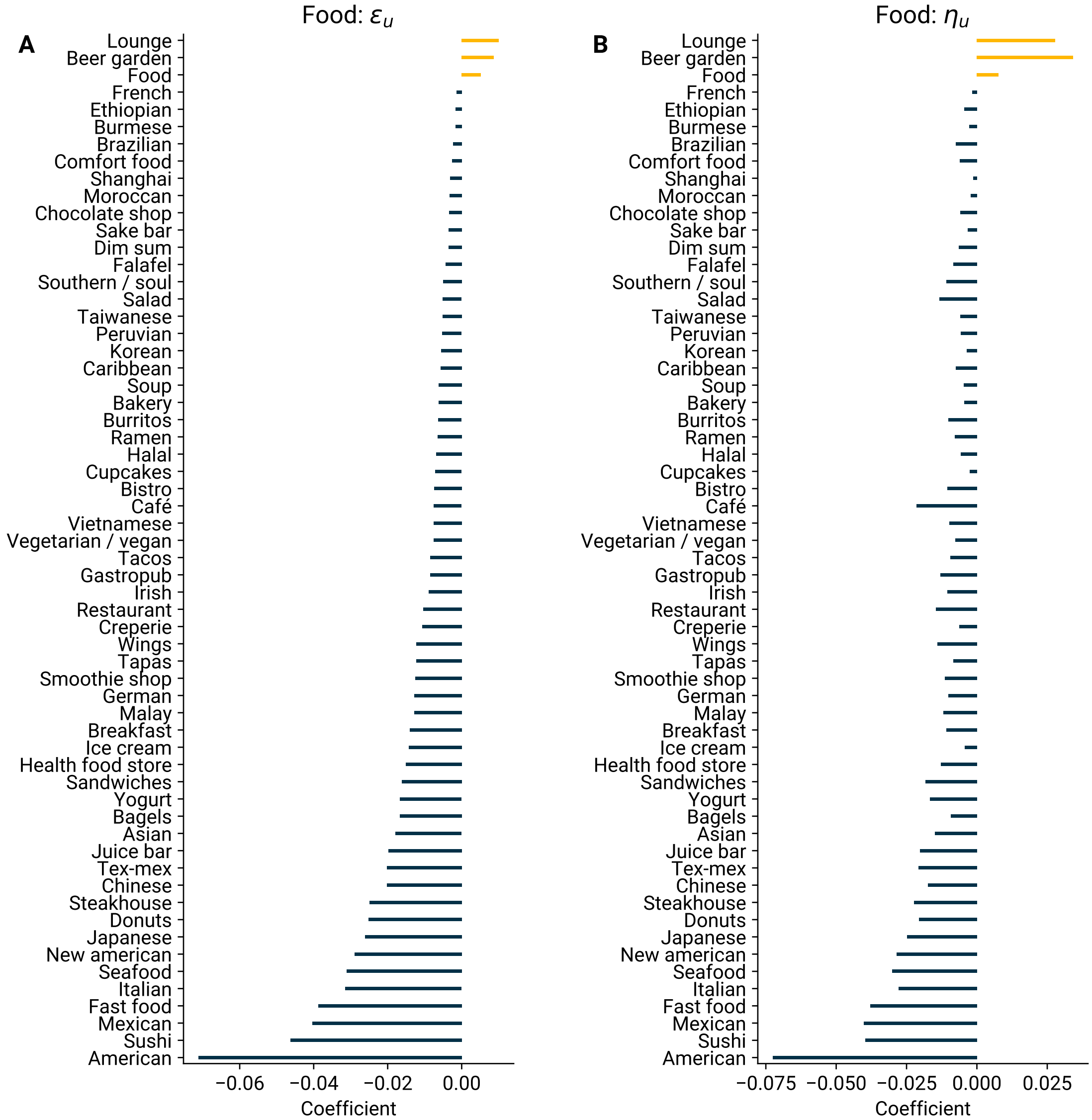

Figure C.12: A) Coefficients of habit features grouped in the "Food" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

distribution $P(\tau_{u,S} = T) = (1 - P_u(S))^{T-1} P_u(S)$, where $P_u(S) = \rho_u S^{-\gamma}$ is the user's probability of exploration. In every simulation, we compute the simulation's ϵ_u as:

$$\epsilon_u = \frac{1}{S_u} \sum_{S=1}^{S_u} \frac{|\tau_{u,S} - \langle \tau_{u,S} \rangle|}{|\tau_{u,S}| + |\langle \tau_{u,S} \rangle|}$$

where $\tau_{u,S}$ is the random value drawn from $P(\tau_{u,S} = T)$ and $\langle \tau_{u,S} \rangle = 1/P_u(S)$ is the expected value. We run 100 simulations for each value of ρ_u and compute ϵ_u for every simulation. In other words, we consider an ensemble of 100 identical users generated by the EPR model. In panel A of Fig.C.20 we show the mean and standard deviation of the distribution of ϵ_u for every value of ρ_u. As can be seen in the figure, the values that we get in the simulations decrease linearly with ρ_u. The result of these simulations

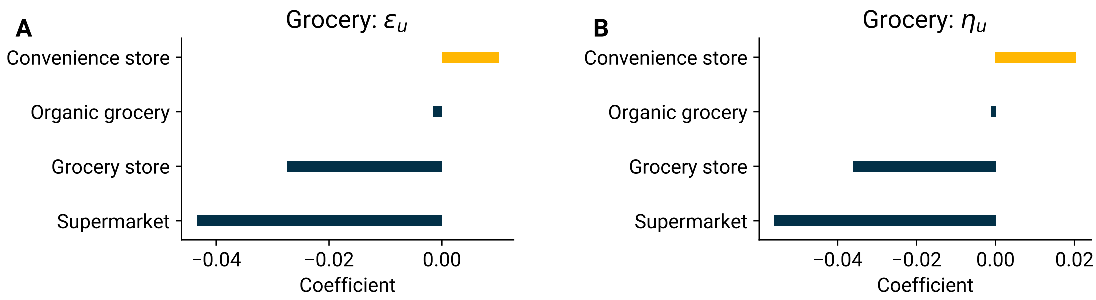

Figure C.13: A) Coefficients of habit features grouped in the "Grocery" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

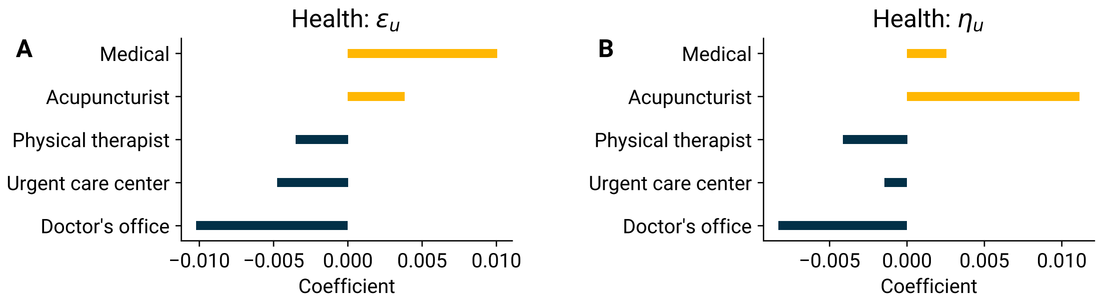

Figure C.14: A) Coefficients of habit features grouped in the "Health" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

indicates that the stochastic part of ϵ_u that we measure from data depends on ρ_u and decreases with it.

For η_u, we run simulations of 100 steps, where at each step the fictitious user explores or returns with a probability $P_u(S) = \rho_u S^{-\gamma}$. At the end of the simulation, we compute the visitation frequency distribution and compute the η_u as the KL-divergence with $f_{u,k} \sim k^{-\gamma-1}$:

$$\eta_u = \sum_{k=1}^{K_u} f_{u,k} \log \frac{f_{u,k}}{\langle f_{u,k} \rangle}$$

where K_u is the rank of the least visited among the locations visited more than once (as explained in Chapter 4, we don't consider locations visited only once to take out tail effects). Also in this case, we run 100 simulations for each value of ρ_u and compute η_u for every simulation. In panel B of Fig.C.20 we show the mean and standard deviation of the distribution of η_u for every value of ρ_u. Similarly to what we observe with ϵ_u, also η_u from simulations decreases with ρ_u, although with a sharper decrease for low values of ρ_u and a very smooth decrease for higher values. Also in this case, these simulations indicate that the stochastic part of η_u that we measure from data depends on ρ_u and decreases with it.

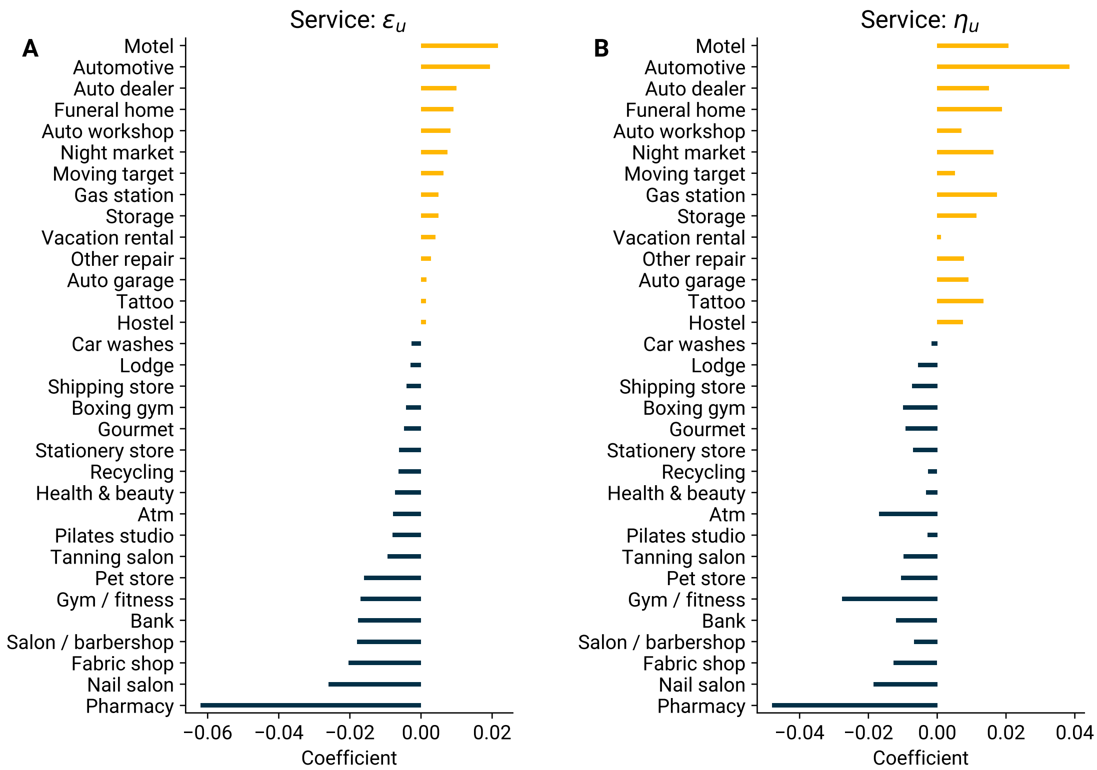

Figure C.15: A) Coefficients of habit features grouped in the "Service" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

C.7 Microscopic mechanisms

In Fig.C.21 we show the analogous results of Fig.3 A and B in Chapter 4 for all the other CBSAs. As we can see from the first and third columns of the figure, the positive association between exploration burstiness and ϵ_u, after controlling for the stochasticity through ρ_u, is visible and consistent across all CBSAs. On the other hand, from the second and fourth columns of the figure, we can see that also the role of the P.R. error in determining η_u is visible and consistent across all CBSAs, except for the users with an extremely high visitation tendency (similarly to what we have seen for Boston in Chapter 4).

C.8 Characterization of assumptions' violations

In Fig.C.22 we show the analogous results of Fig.3 C and D in Chapter 4 for all the other CBSAs. As can be seen in the figure, bursty trains are indeed characterized by the same specific categories in all CBSAs. Indeed, museums and art galleries, coffee shops, and shopping locations are visited significantly and consistently in all CBSAs more often during bursty exploration trains. Conversely and equally consistently in all CBSAs, workplaces and venues in the *City / Outdoor* category, such as parks, neighborhoods, playgrounds, and residential places, are rarely explored during bursty trains.

The same mirrored pattern as Boston in recency trains is observed in other CBSAs. In fact, when people repeatedly return to the same place, they tend to do so at routine

Figure C.16: A) Coefficients of habit features grouped in the "Shopping" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

and habitual locations, such as residential areas, workplaces, as well as transportation hubs and sports venues, and this is consistently true across CBSAs. In contrast and equally consistently, amusement places like coffee shops, restaurants, and shopping malls are not typically revisited continuously.

C.9 Spatial distribution

In Fig.4 of Chapter 4 2e can see a clear urban-rural pattern in all CBSAs. Indeed, users who are best represented by the EPR model, i.e. who are in the lowest 10% of the distribution of the combined deviation variable $\epsilon_u + \eta_u$, are mostly located far from the city center, in bigger and less urban census tracts. On the other hand, the most non-well-represented users, i.e. who are in the highest 10% of the distribution of the

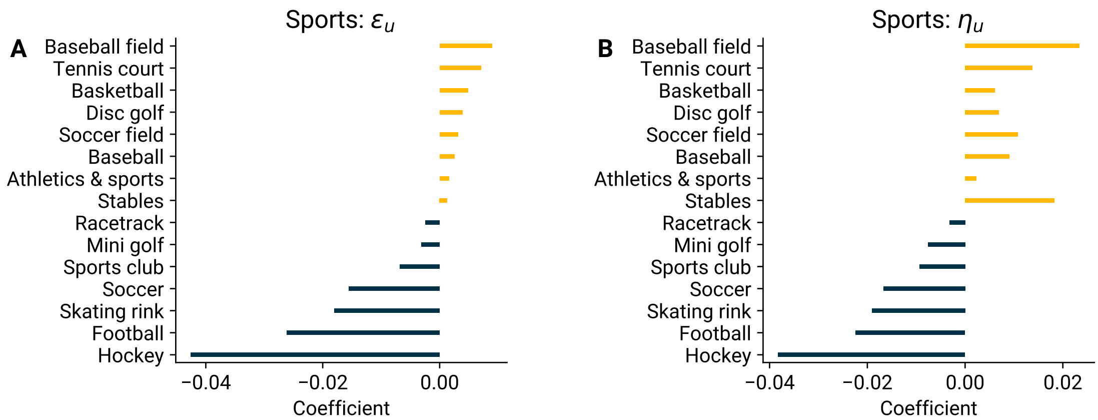

Figure C.17: A) Coefficients of habit features grouped in the "Sports" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

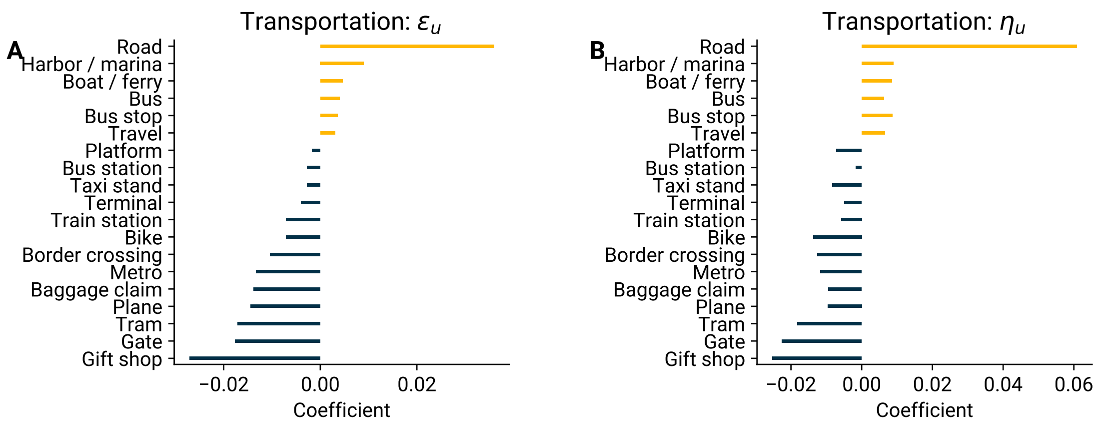

Figure C.18: A) Coefficients of habit features grouped in the "Transportation" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

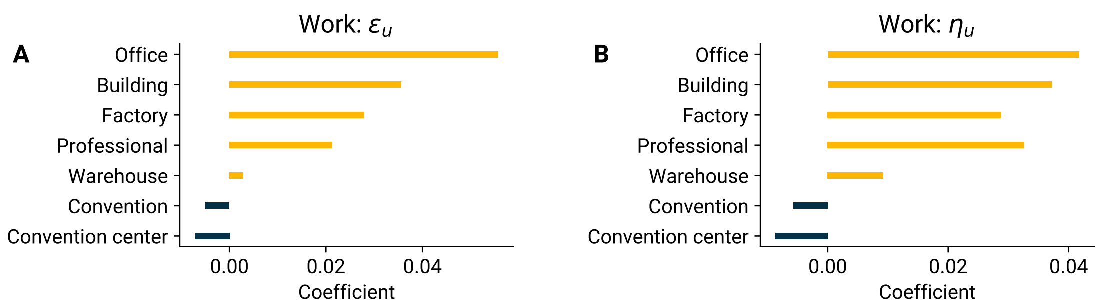

Figure C.19: A) Coefficients of habit features grouped in the "Work" taxonomy for the dependent variable: ϵ_u, with a coefficient greater than 10^{-3} in both regressions. A) Coefficients of the same features as panel A for the dependent variable η_u, with a coefficient greater than 10^{-3} in both regressions.

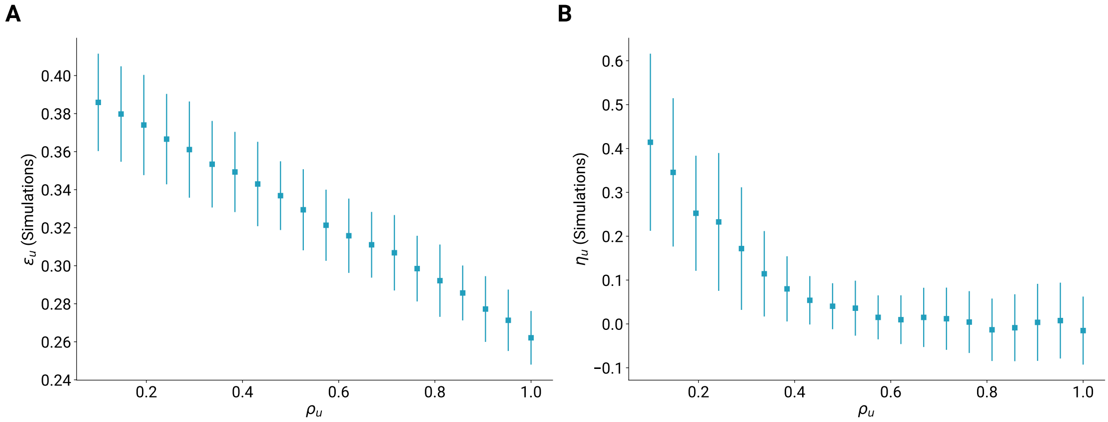

Figure C.20: A) Relation between ρ_u and ϵ_u in simulated experiments. B) Relation between ρ_u and η_u in simulated experiments.

BOSTON	-0.40	WASHINGTON	-0.32	SEATTLE	-0.29
MIAMI	-0.20	DETROIT	-0.20	DALLAS	-0.30
CHICAGO	-0.16	NEW YORK	-0.10	PHILADELPHIA	-0.33
LOS ANGELES	-0.22	SAN FRANCISCO	-0.22		

Table C.1: Pearson coefficients between the logarithm of tracts' population and the difference between the number of users in the highest and lowest 10% of ϵ_u and η_u.

combined deviation variable $\epsilon_u+\eta_u$, are mostly located closer to the city center, in smaller and more urban census tracts, though in the very central tracts, the two quantiles are mostly balanced. These visual results are confirmed by the more robust analysis shown in Fig.C.23, where we explicitly show the relation between a tract's population density (a proxy to its urbanization level) and the variable shown in the maps of Fig.4 in Chapter 4, i.e. the count difference between the two extreme quantiles. The relation is negative in all CBSAs, with Pearson correlation coefficients that go from -0.23 in New York and San Francisco to -0.37 in Dallas and Philadelphia and to -0.41 in Boston, as shown in the following Table C.1:

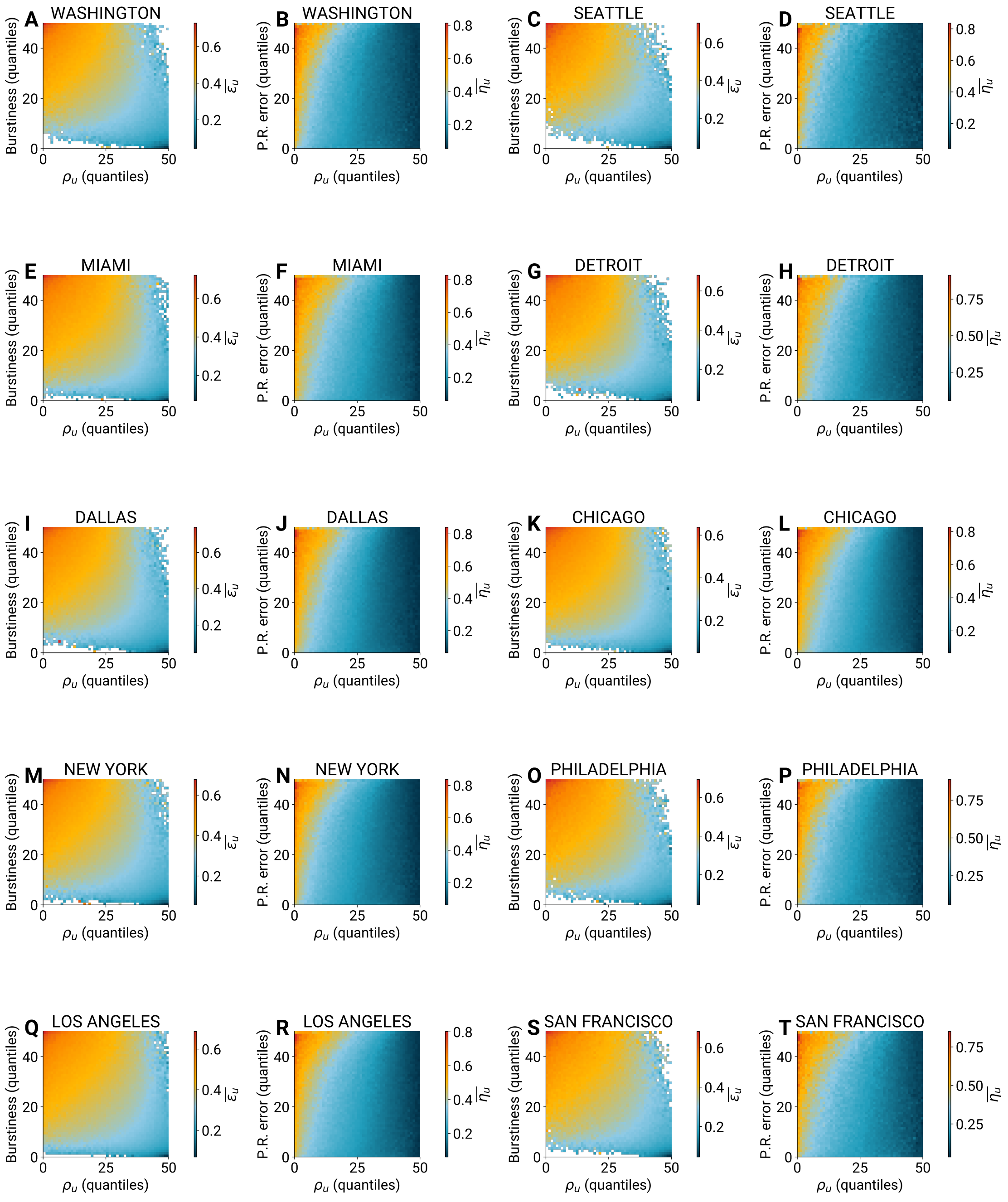

Figure C.21: First and third column: Average values of ϵ_u for users grouped in quantiles of ρ_u and burstiness, for the other CBSAs. Second and fourth column: Average values of η_u for users grouped in quantiles of ρ_u and P.R. error, for the other CBSAs.

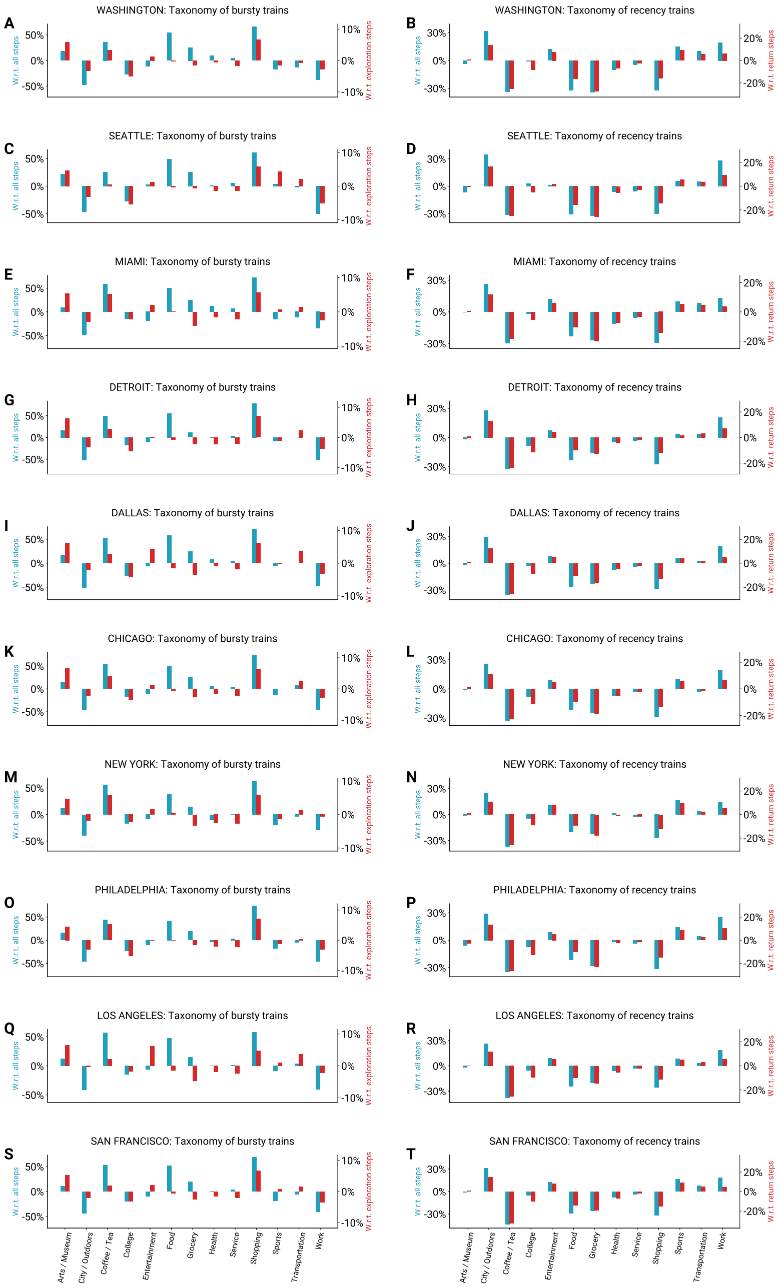

Figure C.22: Left column: Characterization of bursty trains in all CBSAs, i.e., sequences of consecutive exploration steps, in terms of relative visits to defined categories, compared to all visits (blue bars, left y-axis) and to exploration steps only (red bars, right y-axis). Right column: Characterization of recency trains in all CBSAs, i.e., sequences of consecutive visits to the same place, in terms of relative visits to defined categories, compared to all visits (blue bars, left y-axis) and to return steps only (red bars, right y-axis).

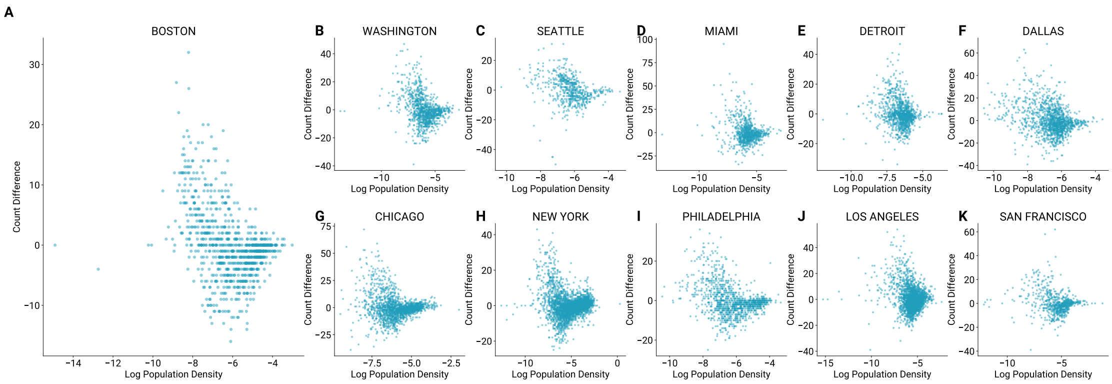

Figure C.23: Relation between the logarithm of tracts' population density and the difference between the number of users in the highest and lowest 10% of $\epsilon_u + \eta_u$, for the other CBSAs.

Bibliography

[1] Jacob Levy Abitbol, Márton Karsai, Jean-Philippe Magué, Jean-Pierre Chevrot, and Eric Fleury. Socioeconomic dependencies of linguistic patterns in twitter: A multivariate analysis. In *Proceedings of the 2018 World Wide Web Conference*, pages 1125–1134, 2018.

[2] Joshua Blumenstock, Gabriel Cadamuro, and Robert On. Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264):1073–1076, 2015.

[3] Jacob Levy Abitbol and Marton Karsai. Interpretable socioeconomic status inference from aerial imagery through urban patterns. *Nature Machine Intelligence*, 2(11):684–692, 2020.

[4] Sierra leone: High resolution population density maps + demographic estimates. https://data.humdata.org/dataset/highresolutionpopulationdensitymaps-sle, 2020. Accessed: 2024-09-26.

[5] Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, et al. A global panel database of pandemic policies (oxford covid-19 government response tracker). *Nature human behaviour*, 5(4):529–538, 2021.

[6] Bandy X Lee, Finn Kjaerulf, Shannon Turner, Larry Cohen, Peter D Donnelly, Robert Muggah, Rachel Davis, Anna Realini, Berit Kieselbach, Lori Snyder MacGregor, et al. Transforming our world: implementing the 2030 agenda through sustainable development goal indicators. *Journal of public health policy*, 37:13–31, 2016.

[7] Richard Wilkinson and Kate Pickett. The spirit level. *Why equality is better for everyone*, 2010.

[8] Joseph E Stiglitz. The price of inequality: How today's divided society endangers our future, 2012.

[9] Shigehiro Oishi, Selin Kesebir, and Ed Diener. Income inequality and happiness. *Psychological science*, 22(9):1095–1100, 2011.

[10] J Michael Oakes and Peter H Rossi. The measurement of ses in health research: current practice and steps toward a new approach. *Social science & medicine*, 56(4):769–784, 2003.

[11] Richard T Campbell and Robert Nash Parker. Substantive and statistical considerations in the interpretation of multiple measures of ses. *Social Forces*, 62(2):450–466, 1983.

[12] Nancy Krieger, David R Williams, and Nancy E Moss. Measuring social class in us public health research: concepts, methodologies, and guidelines. *Annual review of public health*, 18(1):341–378, 1997.

[13] August B Hollingshead. Commentary on the indiscriminate state of social class measurement. *Soc. F.*, 49:563, 1970.

[14] Cecil C North and Paul K Hatt. Jobs and occupations: A popular evaluation. *L. Wilson and WL Kolb (Eds.), Sociological Analysis. An Introductory Text and Case Book, New York (Harcourt, Brace and Co.) 1949, pp. 464-474.*, 1949.

[15] Charles B Nam and E Walter Terrie. Measurement of socioeconomic status from united states census data. In *Measures of socioeconomic status*, pages 29–42. Routledge, 2021.

[16] Charles B Nam and Mary G Powers. Variations in socioeconomic structure by race, residence, and the life cycle. *American Sociological Review*, 30(1):97–103, 1965.

[17] Bruna Galobardes, Mary Shaw, Debbie A Lawlor, John W Lynch, and George Davey Smith. Indicators of socioeconomic position (part 1). *Journal of Epidemiology & Community Health*, 60(1):7–12, 2006.

[18] James B Davies and Anthony F Shorrocks. The distribution of wealth. *Handbook of income distribution*, 1:605–675, 2000.

[19] Sudhir Anand and Paul Segal. The global distribution of income. In *Handbook of income distribution*, volume 2, pages 937–979. Elsevier, 2015.

[20] Nanak C Kakwani. *Income inequality and poverty*. World Bank New York, 1980.

[21] Thomas Piketty. *Capital in the twenty-first century*. Harvard University Press, 2014.

[22] Gabriel Zucman. Global wealth inequality. *Annual Review of Economics*, 11(1):109–138, 2019.

[23] Daniele Checchi. Inequality in incomes and access to education: a cross-country analysis (1960–95). *Labour*, 17(2):153–201, 2003.

[24] Keith M Lewin. Access to education in sub-saharan africa: patterns, problems and possibilities. *Comparative Education*, 45(2):151–174, 2009.

[25] Gary A Berg. *Low-income students and the perpetuation of inequality: Higher education in America*. Routledge, 2016.

[26] Jo-Ann S Hannah and Sharon E Kahn. The relationship of socioeconomic status and gender to the occupational choices of grade 12 students. *Journal of vocational behavior*, 34(2):161–178, 1989.

[27] Sebawit G Bishu and Mohamad G Alkadry. A systematic review of the gender pay gap and factors that predict it. *Administration & Society*, 49(1):65–104, 2017.

[28] Claudia Buchmann, Thomas A DiPrete, and Anne McDaniel. Gender inequalities in education. *Annu. Rev. Sociol*, 34(1):319–337, 2008.

[29] Gail E Thomas, Karl L Alexander, and Bruce K Eckland. Access to higher education: The importance of race, sex, social class, and academic credentials. *The School Review*, 87(2):133–156, 1979.

[30] Barbara F Reskin. Getting it right: Sex and race inequality in work organizations. *Annual Review of Sociology*, 26:707–709, 2000.

[31] Melvin Oliver and Thomas Shapiro. *Black wealth/white wealth: A new perspective on racial inequality*. Routledge, 2013.

[32] Nancy E Adler and Joan M Ostrove. Socioeconomic status and health: what we know and what we don't. *Annals of the New York academy of Sciences*, 896(1):3–15, 1999.

[33] John Mirowsky and Catherine E Ross. Socioeconomic status and subjective life expectancy. *Social Psychology Quarterly*, pages 133–151, 2000.

[34] David A Swanson, Mary A McGehee, and Nazrul Hoque. Socio-economic status and life expectancy in the united states, 1970–1990. *Population Review*, 48(1), 2009.

[35] Raj Chetty, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. The association between income and life expectancy in the united states, 2001-2014. *Jama*, 315(16):1750–1766, 2016.

[36] Michael G Marmot, Martin J Shipley, and Geoffrey Rose. Inequalities in death—specific explanations of a general pattern? *The Lancet*, 323(8384):1003–1006, 1984.

[37] Nancy E Adler, Thomas Boyce, Margaret A Chesney, Sheldon Cohen, Susan Folkman, Robert L Kahn, and S Leonard Syme. Socioeconomic status and health: the challenge of the gradient. *American psychologist*, 49(1):15, 1994.

[38] Yan Yu and David R Williams. Socioeconomic status and mental health. In *Handbook of the sociology of mental health*, pages 151–166. Springer, 1999.

[39] Franziska Reiss. Socioeconomic inequalities and mental health problems in children and adolescents: a systematic review. *Social science & medicine*, 90:24–31, 2013.

[40] Carles Muntaner, William W Eaton, Richard Miech, and Patricia O'campo. Socioeconomic position and major mental disorders. *Epidemiologic reviews*, 26(1):53–62, 2004.

[41] Gary W Evans and Elyse Kantrowitz. Socioeconomic status and health: the potential role of environmental risk exposure. *Annual review of public health*, 23(1):303–331, 2002.

[42] Elizabeth H Baker. Socioeconomic status, definition. *The Wiley Blackwell encyclopedia of health, illness, behavior, and society*, pages 2210–2214, 2014.

[43] Archana Singh-Manoux and Michael Marmot. Role of socialization in explaining social inequalities in health. *Social science & medicine*, 60(9):2129–2133, 2005.

[44] James S House. Understanding social factors and inequalities in health: 20th century progress and 21st century prospects. *Journal of health and social behavior*, pages 125–142, 2002.

[45] Luigi L Cavalli-Sforza and Francesco Cavalli-Sforza. The great human diasporas: a history of diversity and evolution. 1994.

[46] Mimi Sheller and John Urry. The new mobilities paradigm. *Environment and planning A*, 38(2):207–226, 2006.

[47] Robin IM Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavioral and brain sciences*, 16(4):681–694, 1993.

[48] Mark Granovetter. The strength of weak ties: A network theory revisited. *Sociological theory*, pages 201–233, 1983.

[49] Pierre Bourdieu. The forms of capital. In *The sociology of economic life*, pages 78–92. Routledge, 2018.

[50] Manlio De Domenico, Antonio Lima, and Mirco Musolesi. Interdependence and predictability of human mobility and social interactions. *Pervasive and Mobile Computing*, 9(6):798–807, 2013.

[51] Jameson L Toole, Carlos Herrera-Yaqüe, Christian M Schneider, and Marta C González. Coupling human mobility and social ties. *Journal of The Royal Society Interface*, 12(105):20141128, 2015.

[52] Marta C González, Pedro G Lind, and Hans J Herrmann. System of mobile agents to model social networks. *Physical review letters*, 96(8):088702, 2006.

[53] Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090, 2011.

[54] Miguel Picornell, Tomás Ruiz, Maxime Lenormand, José J Ramasco, Thibaut Dubernet, and Enrique Frías-Martínez. Exploring the potential of phone call data to characterize the relationship between social network and travel behavior. *Transportation*, 42:647–668, 2015.

[55] Zexun Chen, Sean Kelty, Alexandre G Evsukoff, Brooke Foucault Welles, James Bagrow, Ronaldo Menezes, and Gourab Ghoshal. Contrasting social and non-social sources of predictability in human mobility. *Nature communications*, 13(1):1922, 2022.

[56] David J Crandall, Lars Backstrom, Dan Cosley, Siddharth Suri, Daniel Huttenlocher, and Jon Kleinberg. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences*, 107(52):22436–22441, 2010.

[57] Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1100–1108, 2011.

[58] Roel Jennissen. Causality chains in the international migration systems approach. *Population research and policy review*, 26(4):411–436, 2007.

[59] Ernest George Ravenstein. The laws of migration. *Journal of the royal statistical society*, 52(2):241–305, 1889.

[60] Samuel A Stouffer. Intervening opportunities: a theory relating mobility and distance. *American sociological review*, 5(6):845–867, 1940.

[61] George J Borjas. Economic theory and international migration. *International migration review*, 23(3):457–485, 1989.

[62] Douglas S Massey, Joaquin Arango, Graeme Hugo, Ali Kouaouci, and Adela Pellegrino. *Worlds in motion: understanding international migration at the end of the millennium: understanding international migration at the end of the millennium*. Clarendon Press, 1999.

[63] Oded Stark, J Edward Taylor, and Shlomo Yitzhaki. Migration, remittances and inequality: A sensitivity analysis using the extended gini index. *Journal of Development Economics*, 28(3):309–322, 1988.

[64] Adansi A Amankwaa. The world economic system and international migration in less developed countries: an ecological approach. *International migration (Geneva, Switzerland)*, 33(1):93–114, 1995.

[65] James T Fawcett. Networks, linkages, and migration systems. *International migration review*, 23(3):671–680, 1989.

[66] Christian Dustmann and Anna Okatenko. Out-migration, wealth constraints, and the quality of local amenities. *Journal of Development Economics*, 110:52–63, 2014.

[67] Silvia Pedraza. Women and migration: The social consequences of gender. *Annual review of sociology*, pages 303–325, 1991.

[68] Ingrid Esveldt, Isik Kulu-Glasgow, Jeanette Schoorl, and Hanna Van Solinge. Migratiemotieven, migratienetwerken en partnerkeuze van turken en marokkanen in nederland. *Migration motives, migration networks and partner choice of Turks and Moroccans in the Netherlands). The Hague, NIDI. NIDI-report*, (43), 1995.

[69] Yu Xu, Andrey Belyi, Marceau Marsan, and Song Gao. Human mobility and urban structure in cities around the world. *Nature Communications*, 9(1):1–10, 2018.

[70] Olle Järv, Kerli Müürisepp, Rein Ahas, Ben Derudder, and Frank Witlox. Ethnic differences in activity spaces as a characteristic of segregation: A study based on mobile phone usage in tallinn, estonia. *Urban Studies*, 52(14):2680–2698, 2015.

[71] Siiri Silm and Rein Ahas. Ethnic differences in activity spaces: A study of out-of-home nonemployment activities with mobile phone data. *Annals of the Association of American Geographers*, 104(3):542–559, 2014.

[72] Yang Xu, Shih-Lung Shaw, Ziliang Zhao, Ling Yin, Zhixiang Fang, and Qingquan Li. Understanding aggregate human mobility patterns using passive mobile phone location data: A home-based approach. *Transportation*, 42:625–646, 2015.

[73] Qunying Huang and David WS Wong. Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? *International Journal of Geographical Information Science*, 30(9):1873–1898, 2016.

[74] Luca Pappalardo, Dino Pedreschi, Zbigniew Smoreda, and Fosca Giannotti. Using big data to study the link between human mobility and socio-economic development. In *2015 IEEE international conference on big data (big data)*, pages 871–878. IEEE, 2015.

[75] Luca Pappalardo, Maarten Vanhoof, Lorenzo Gabrielli, Zbigniew Smoreda, Dino Pedreschi, and Fosca Giannotti. An analytical framework to nowcast well-being using mobile phone data. *International Journal of Data Science and Analytics*, 2:75–92, 2016.

[76] Laura Lotero, Rafael G Hurtado, Luis Mario Floría, and Jesús Gómez-Gardeñes. Rich do not rise early: spatio-temporal patterns in the mobility networks of different socio-economic classes. *Royal Society open science*, 3(10):150654, 2016.

[77] Vanessa Frias-Martinez and Jesus Virseda. On the relationship between socio-economic factors and cell phone usage. In *Proceedings of the fifth international conference on information and communication technologies and development*, pages 76–84, 2012.

[78] Vanessa Frias-Martinez, Jesus Virseda, and Enrique Frias-Martinez. Socio-economic levels and human mobility. In *Qual meets quant workshop-QMQ*, pages 1–6. Citeseer, 2010.

[79] Narisra Limtanakool, Martin Dijst, and Tim Schwanen. The influence of socioeconomic characteristics, land use and travel time considerations on mode choice for medium-and longer-distance trips. *Journal of transport geography*, 14(5):327–341, 2006.

[80] Mei-Po Kwan. Gender, the home-work link, and space-time patterns of nonemployment activities. *Economic geography*, 75(4):370–394, 1999.

[81] Mariana Macedo, Laura Lotero, Alessio Cardillo, Ronaldo Menezes, and Hugo Barbosa. Differences in the spatial landscape of urban mobility: gender and socioeconomic perspectives. *Plos one*, 17(3):e0260874, 2022.

[82] Susan Hanson and Perry Hanson. Gender and urban activity patterns in uppsala, sweden. *Geographical Review*, pages 291–299, 1980.

[83] Susan Hanson. The determinants of daily travel-activity patterns: relative location and sociodemographic factors. *Urban Geography*, 3(3):179–202, 1982.

[84] Alice Battiston, Ludovico Napoli, Paolo Bajardi, André Panisson, Alan Perotti, Michael Szell, and Rossano Schifanella. Revealing the determinants of gender inequality in urban cycling with large-scale data. *EPJ Data Science*, 12(1):9, 2023.

[85] Abdullah Almaatouq, Francisco Prieto-Castrillo, and Alex Pentland. Mobile communication signatures of unemployment. In *Social Informatics: 8th International Conference, SocInfo 2016, Bellevue, WA, USA, November 11-14, 2016, Proceedings, Part I 8*, pages 407–418. Springer, 2016.

[86] Corrine W Ruktanonchai, Shengjie Lai, Chigozie E Utazi, Alex D Cunningham, Patrycja Koper, Grant E Rogers, Nick W Ruktanonchai, Adam Sadilek, Dorothea Woods, Andrew J Tatem, et al. Practical geospatial and sociodemographic predictors of human mobility. *Scientific reports*, 11(1):15389, 2021.

[87] Alejandro Llorente, Manuel Garcia-Herranz, Manuel Cebrian, and Esteban Moro. Social media fingerprints of unemployment. *PloS one*, 10(5):e0128692, 2015.

[88] Hugo Barbosa, Surendra Hazarie, Brian Dickinson, Aleix Bassolas, Adam Frank, Henry Kautz, Adam Sadilek, José J Ramasco, and Gourab Ghoshal. Uncovering the socioeconomic facets of human mobility. *Scientific reports*, 11(1):8616, 2021.

[89] Dong Li and Jiming Liu. Uncovering the relationship between point-of-interests-related human mobility and socioeconomic status. *Telematics and Informatics*, 39:49–63, 2019.

[90] Mark Granovetter. The impact of social structure on economic outcomes. In *The sociology of economic life*, pages 46–61. Routledge, 2018.

[91] Shaojun Luo, Flaviano Morone, Carlos Sarraute, Matías Travizano, and Hernán A Makse. Inferring personal economic status from social network location. *Nature Communications*, 8(1):15227, 2017.

[92] Nathan Eagle, Michael Macy, and Rob Claxton. Network diversity and economic development. *Science*, 328(5981):1029–1031, 2010.

[93] Jonathan Donner. The use of mobile phones by microentrepreneurs in kigali, rwanda: Changes to social and business networks. *Information Technologies & International Development*, 3(2):pp–3, 2006.

[94] Ronald S Burt et al. The social capital of structural holes. *The new economic sociology: Developments in an emerging field*, 148(90):122, 2002.

[95] Ronald S Burt. The network structure of social capital. *Research in organizational behavior*, 22:345–423, 2000.

[96] Nan Lin and Mary Dumin. Access to occupations through social ties. *Social networks*, 8(4):365–385, 1986.

[97] Scott E Seibert, Maria L Kraimer, and Robert C Liden. A social capital theory of career success. *Academy of management journal*, 44(2):219–237, 2001.

[98] Joel M Podolny and James N Baron. Resources and relationships: Social networks and mobility in the workplace. *American sociological review*, pages 673–693, 1997.

[99] Marc-David L Seidel, Jeffrey T Polzer, and Katherine J Stewart. Friends in high places: The effects of social networks on discrimination in salary negotiations. *Administrative Science Quarterly*, 45(1):1–24, 2000.

[100] James D Montgomery. Social networks and labor-market outcomes: Toward an economic analysis. *The American economic review*, 81(5):1408–1418, 1991.

[101] Yanjie Bian. Bringing strong ties back in: Indirect ties, network bridges, and job searches in china. *American sociological review*, pages 366–385, 1997.

[102] Bernd Wegener. Job mobility and social ties: Social resources, prior job, and status attainment. *American Sociological Review*, pages 60–71, 1991.

[103] Daniel J Brass. Being in the right place: A structural analysis of individual influence in an organization. *Administrative science quarterly*, pages 518–539, 1984.

[104] Daniel J Brass. Men's and women's networks: A study of interaction patterns and influence in an organization. *Academy of Management journal*, 28(2):327–343, 1985.

[105] Raj Chetty, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, et al. Social capital i: measurement and associations with economic mobility. *Nature*, 608(7921):108–121, 2022.

[106] Raj Chetty, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, et al. Social capital ii: determinants of economic connectedness. *Nature*, 608(7921):122–134, 2022.

[107] Martti Siisiainen. Two concepts of social capital: Bourdieu vs. putnam. *International journal of contemporary sociology*, 40(2):183–204, 2003.

[108] James S Coleman. Social capital in the creation of human capital. *American journal of sociology*, 94:S95–S120, 1988.

[109] Matthew O Jackson. A typology of social capital and associated network measures. *Social choice and welfare*, 54(2):311–336, 2020.

[110] Finn Hedefalk and Martin Dribe. The social context of nearest neighbors shapes educational attainment regardless of class origin. *Proceedings of the National Academy of Sciences*, 117(26):14918–14925, 2020.

[111] John A List, Fatemeh Momeni, and Yves Zenou. The social side of early human capital formation: Using a field experiment to estimate the causal impact of neighborhoods. Technical report, National Bureau of Economic Research, 2020.

[112] Bruce Sacerdote. Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*, volume 3, pages 249–277. Elsevier, 2011.

[113] Peter V Marsden and Jeanne S Hurlbert. Social resources and mobility outcomes: A replication and extension. *Social forces*, 66(4):1038–1059, 1988.

[114] Nan Dirk De Graaf and Hendrik Derk Flap. "with a little help from my friends": Social resources as an explanation of occupational status and income in west germany, the netherlands, and the united states. *Social forces*, 67(2):452–472, 1988.

[115] Ron Laschever. The doughboys network: Social interactions and the employment of world war i veterans. *Available at SSRN 1205543*, 2013.

[116] Lori A Beaman. Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the us. *The Review of Economic Studies*, 79(1):128–161, 2012.

[117] Michael Bailey, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, 126(6):2224–2276, 2018.

[118] Melanie Mitchell. *Complexity: A guided tour.* Oxford University Press, 2009.

[119] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[120] Barbara S Lawrence and Neha Parikh Shah. Homophily: Measures and meaning. *Academy of Management Annals*, 14(2):513–597, 2020.

[121] Esteban Moro, Dan Calacci, Xiaowen Dong, and Alex Pentland. Mobility patterns are associated with experienced income segregation in large US cities. *Nature communications*, 12(1):1–10, 2021.

[122] Wesley Shrum, Neil H Cheek Jr, and Saundra MacD. Friendship in school: Gender and racial homophily. *Sociology of Education*, pages 227–239, 1988.

[123] Peter V Marsden. Homogeneity in confiding relations. *Social networks*, 10(1):57–76, 1988.

[124] Lois M Verbrugge. The structure of adult friendship choices. *Social forces*, 56(2):576–597, 1977.

[125] Claude S Fischer, Robert Jackson, C Ann Stueve, Kathleen Gerson, and Lynne M Jones. Networks and places. 1977.

[126] Peter V Marsden. Core discussion networks of americans. *American sociological review*, pages 122–131, 1987.

[127] Karen E Campbell. Gender differences in job-related networks. *Work and occupations*, 15(2):179–200, 1988.

[128] Hugh Louch. Personal network integration: transitivity and homophily in strong-tie relations. *Social networks*, 22(1):45–64, 2000.

[129] Erik Olin Wright. *Class counts: Comparative studies in class analysis.* Cambridge university press, 1997.

[130] Edward O Laumann, Peter V Marsden, and Joseph Galaskiewicz. Community-elite influence structures: Extension of a network approach. *American journal of sociology*, 83(3):594–631, 1977.

[131] J Miller McPherson and Lynn Smith-Lovin. Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *American sociological review*, pages 370–379, 1987.

[132] Maureen T Hallinan and Richard A Williams. Interracial friendship choices in secondary schools. *American Sociological Review*, pages 67–78, 1989.

[133] Ronald S Burt. Decay functions. *Social networks*, 22(1):1–28, 2000.

[134] Matthijs Kalmijn. Intermarriage and homogamy: Causes, patterns, trends. *Annual review of sociology*, 24(1):395–421, 1998.

[135] Jeffrey Pfeffer. A resource dependence perspective on intercorporate relations. *Intercorporate relations: The structural analysis of business*, 1(1):25–55, 1987.

[136] Bruce H Mayhew, J Miller McPherson, Thomas Rotolo, and Lynn Smith-Lovin. Sex and race homogeneity in naturally occurring groups. *Social Forces*, 74(1):15–52, 1995.

[137] CS Fischer. *To dwell among friends: Personal networks in town and city*. The University of Chicago Press, 1982.

[138] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[139] Fariba Karimi, Mathieu Génois, Claudia Wagner, Philipp Singer, and Markus Strohmaier. Homophily influences ranking of minorities in social networks. *Scientific reports*, 8(1):11077, 2018.

[140] Aili Asikainen, Gerardo Iñiguez, Javier Ureña-Carrión, Kimmo Kaski, and Mikko Kivelä. Cumulative effects of triadic closure and homophily in social networks. *Science Advances*, 6(19):eaax7310, 2020.

[141] Scott L Feld. The focused organization of social ties. *American journal of sociology*, 86(5):1015–1035, 1981.

[142] Scott L Feld. The structured use of personal associates. *Social Forces*, 62(3):640–652, 1984.

[143] Robin IM Dunbar. Neocortex size as a constraint on group size in primates. *Journal of human evolution*, 22(6):469–493, 1992.

[144] Jari Saramäki, Elizabeth A Leicht, Eduardo López, Sam GB Roberts, Felix Reed-Tsochas, and Robin IM Dunbar. Persistence of social signatures in human communication. *Proceedings of the National Academy of Sciences*, 111(3):942–947, 2014.

[145] George Kingsley Zipf. The p 1 p 2/d hypothesis: on the intercity movement of persons. *American sociological review*, 11(6):677–686, 1946.

[146] Herbert J Gans. People and plans: Essays on urban problems and solutions. *(No Title)*, 1968.

[147] Barry Wellman. Are personal communities local? a dumptarian reconsideration. *Social networks*, 18(4):347–354, 1996.

[148] Camille Zubrinsky Charles. The dynamics of racial residential segregation. *Annual review of sociology*, 29(1):167–207, 2003.

[149] Douglas S Massey and Nancy A Denton. The dimensions of residential segregation. *Social forces*, 67(2):281–315, 1988.

[150] Stanley Lieberson. *A piece of the pie: Blacks and white immigrants since 1880*. Univ of California Press, 2023.

[151] Eun Lee, Fariba Karimi, Claudia Wagner, Hang-Hyun Jo, Markus Strohmaier, and Mirta Galesic. Homophily and minority-group size explain perception biases in social networks. *Nature human behaviour*, 3(10):1078–1087, 2019.

[152] Sina Sajjadi, Samuel Martin-Gutierrez, and Fariba Karimi. Unveiling homophily beyond the pool of opportunities. *arXiv preprint arXiv:2401.13642*, 2024.

[153] Philip S Chodrow. Structure and information in spatial segregation. *Proceedings of the National Academy of Sciences*, 114(44):11591–11596, 2017.

[154] Rémi Louf and Marc Barthelemy. Patterns of residential segregation. *PloS one*, 11(6):e0157476, 2016.

[155] Sandro Sousa and Vincenzo Nicosia. Quantifying ethnic segregation in cities through random walks. *Nature Communications*, 13(1):5809, 2022.

[156] Douglas S Massey and Nancy A Denton. Trends in the residential segregation of blacks, hispanics, and asians: 1970-1980. *American sociological review*, pages 802–825, 1987.

[157] Andrew J Stier, Sina Sajjadi, Fariba Karimi, Luís MA Bettencourt, and Marc G Berman. Implicit racial biases are lower in more populous more diverse and less segregated us cities. *Nature Communications*, 15(1):961, 2024.

[158] Sean F Reardon and Kendra Bischoff. Income inequality and income segregation. *American journal of sociology*, 116(4):1092–1153, 2011.

[159] Richard Florida and Charlotta Mellander. The geography of inequality: Difference and determinants of wage and income inequality across us metros. *Regional Studies*, 50(1):79–92, 2016.

[160] Tiit Tammaru, Szymon Marcińczak, Marten Van Ham, and Sako Musterd. *Socio-economic segregation in European capital cities.* Routledge Abingdon, 2014.

[161] Paul Taylor and Richard Allan Fry. *The rise of residential segregation by income.* Citeseer, 2012.

[162] Sako Musterd, Szymon Marcińczak, Maarten Van Ham, and Tiit Tammaru. Socioeconomic segregation in european capital cities. Increasing separation between poor and rich. *Urban Geography*, 38(7):1062–1083, 2017.

[163] Edward Glaeser and Jacob Vigdor. *The end of the segregated century: Racial separation in America's neighborhoods, 1890-2010.* Manhattan Institute for Policy Research New York, 2012.

[164] Glenn Firebaugh and Chad R Farrell. Still large, but narrowing: the sizable decline in racial neighborhood inequality in metropolitan america, 1980–2010. *Demography*, 53(1):139–164, 2016.

[165] John Iceland, Daniel H Weinberg, and Erika Steinmetz. *Racial and ethnic residential segregation in the United States 1980-2000*, volume 8. Bureau of Census, 2002.

[166] Raj Chetty, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The quarterly journal of economics*, 129(4):1553–1623, 2014.

[167] Raj Chetty, Nathaniel Hendren, and Lawrence F Katz. The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902, 2016.

[168] David R Williams and Chiquita Collins. Racial residential segregation: a fundamental cause of racial disparities in health. *Public health reports*, 2001.

[169] Lincoln Quillian. Does segregation create winners and losers? residential segregation and inequality in educational attainment. *Social Problems*, 61(3):402–426, 2014.

[170] Alan J Abramson, Mitchell S Tobin, and Matthew R VanderGoot. The changing geography of metropolitan opportunity: The segregation of the poor in us metropolitan areas, 1970 to 1990. *Housing Policy Debate*, 6(1):45–72, 1995.

[171] Geoffrey T Wodtke, David J Harding, and Felix Elwert. Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high school graduation. *American sociological review*, 76(5):713–736, 2011.

[172] Russell W Rumberger and Gregory J Palardy. Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers college record*, 107(9):1999–2045, 2005.

[173] Dolores Acevedo-Garcia and Kimberly A Lochner. Residential segregation and health. *Neighborhoods and health*, pages 265–87, 2003.

[174] Yoo Min Park and Mei-Po Kwan. Beyond residential segregation: A spatiotemporal approach to examining multi-contextual segregation. *Computers, Environment and Urban Systems*, 71:98–108, 2018.

[175] David WS Wong and Shih-Lung Shaw. Measuring segregation: An activity space approach. *Journal of geographical systems*, 13:127–145, 2011.

[176] Teodoro Dannemann, Boris Sotomayor-Gómez, and Horacio Samaniego. The time geography of segregation during working hours. *Royal Society open science*, 5(10):180749, 2018.

[177] Xue Zhang, Jue Wang, Mei-Po Kwan, and Yanwei Chai. Reside nearby, behave apart? activity-space-based segregation among residents of various types of housing in beijing, china. *Cities*, 88:166–180, 2019.

[178] Nandini Iyer, Ronaldo Menezes, and Hugo Barbosa. Transport and mobility segregation in urban spaces. *International Journal of Population Data Science*, 8(3), 2024.

[179] John Östh, Ian Shuttleworth, and Thomas Niedomysl. Spatial and temporal patterns of economic segregation in sweden's metropolitan areas: A mobility approach. *Environment and Planning A: Economy and Space*, 50(4):809–825, 2018.

[180] Donald R Davis, Jonathan I Dingel, Joan Monras, and Eduardo Morales. How segregated is urban consumption? *Journal of Political Economy*, 127(4):1684–1738, 2019.

[181] Mariano G Beiró, Loreto Bravo, Diego Caro, Ciro Cattuto, Leo Ferres, and Eduardo Graells-Garrido. Shopping mall attraction and social mixing at a city scale. *EPJ Data Science*, 7(1):1–21, 2018.

[182] Susan Athey, Billy Ferguson, Matthew Gentzkow, and Tobias Schmidt. Estimating experienced racial segregation in us cities using large-scale gps data. *Proceedings of the National Academy of Sciences*, 118(46):e2026160118, 2021.

[183] Taylor Shelton, Ate Poorthuis, and Matthew Zook. Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and urban planning*, 142:198–211, 2015.

[184] Christopher R Browning, Catherine A Calder, Lauren J Krivo, Anna L Smith, and Bethany Boettner. Socioeconomic segregation of activity spaces in urban neighborhoods: Does shared residence mean shared routines? *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 3(2):210–231, 2017.

[185] Xiaowen Dong, Alfredo J Morales, Eaman Jahani, Esteban Moro, Bruno Lepri, Burcin Bozkaya, Carlos Sarraute, Yaneer Bar-Yam, and Alex Pentland. Segregated interactions in urban and online space. *EPJ Data Science*, 9(1):20, 2020.

[186] Rafiazka Millanida Hilman, Gerardo Iñiguez, and Márton Karsai. Socioeconomic biases in urban mixing patterns of US metropolitan areas. *EPJ data science*, 11(1):32, 2022.

[187] Yannick Leo, Eric Fleury, J Ignacio Alvarez-Hamelin, Carlos Sarraute, and Márton Karsai. Socioeconomic correlations and stratification in social-communication networks. *Journal of The Royal Society Interface*, 13(125):20160598, 2016.

[188] Vinicius M Netto, João Vitor Meirelles, Maíra Pinheiro, and Henrique Lorea. A temporal geography of encounters. *CyberGeo: European Journal of Geography*, 2018.

[189] Qi Wang, Nolan Edward Phillips, Mario L Small, and Robert J Sampson. Urban mobility and neighborhood isolation in america's 50 largest cities. *Proceedings of the National Academy of Sciences*, 115(30):7735–7740, 2018.

[190] Yossef Arie and Gustavo S Mesch. Interethnic ties via mobile communications in homogeneous and ethnically mixed cities: A structural diversification approach. In *Communication and Information Technologies Annual*, pages 175–203. Emerald Group Publishing Limited, 2015.

[191] Lijun Sun, Kay W Axhausen, Der-Horng Lee, and Xianfeng Huang. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences*, 110(34):13774–13779, 2013.

[192] Siiri Silm and Rein Ahas. The temporal variation of ethnic segregation in a city: Evidence from a mobile phone use dataset. *Social Science Research*, 47:30–43, 2014.

[193] Guillaume Le Roux, Julie Vallée, and Hadrien Commenges. Social segregation around the clock in the paris region (france). *Journal of Transport Geography*, 59:134–145, 2017.

[194] Rahul Goel, Rajesh Sharma, and Anto Aasa. Understanding gender segregation through call data records: an estonian case study. *Plos one*, 16(3):e0248212, 2021.

[195] Bjørn-Atle Reme, Andreas Kotsadam, Johannes Bjelland, Pål Roe Sundsøy, and Jo Thori Lind. Quantifying social segregation in large-scale networks. *Scientific reports*, 12(1):6474, 2022.

[196] Bas Hofstra, Rense Corten, Frank Van Tubergen, and Nicole B Ellison. Sources of segregation in social networks: A novel approach using facebook. *American Sociological Review*, 82(3):625–656, 2017.

[197] Thomas A DiPrete, Andrew Gelman, Tyler McCormick, Julien Teitler, and Tian Zheng. Segregation in social networks based on acquaintanceship and trust. *American journal of sociology*, 116(4):1234–1283, 2011.

[198] Linton C Freeman. Segregation in social networks. *Sociological Methods & Research*, 6(4):411–429, 1978.

[199] Michał Bojanowski and Rense Corten. Measuring segregation in social networks. *Social networks*, 39:14–32, 2014.

[200] Yang Xu, Alexander Belyi, Paolo Santi, and Carlo Ratti. Quantifying segregation in an integrated urban physical-social space. *Journal of the Royal Society Interface*, 16(160):20190536, 2019.

[201] Eduardo Marques. Social networks, segregation and poverty in são paulo. *International Journal of Urban and Regional Research*, 36(5):958–979, 2012.

[202] Gergő Tóth, Johannes Wachs, Riccardo Di Clemente, Ákos Jakobi, Bence Ságvári, János Kertész, and Balázs Lengyel. Inequality is rising where social network segregation interacts with urban topology. *Nature communications*, 12(1):1143, 2021.

[203] Eszter Bokányi, Sándor Juhász, Márton Karsai, and Balázs Lengyel. Universal patterns of long-distance commuting and social assortativity in cities. *Scientific reports*, 11(1):1–10, 2021.

[204] Joshua Blumenstock and Lauren Fratamico. Social and spatial ethnic segregation: A framework for analyzing segregation with large-scale spatial network data. In *Proceedings of the 4th Annual Symposium on Computing for Development*, pages 1–10, 2013.

[205] Daniel Rhoads, Ivan Serrano, Javier Borge-Holthoefer, and Albert Solé-Ribalta. Measuring and mitigating behavioural segregation using call detail records. *EPJ Data Science*, 9(1):5, 2020.

[206] Yossef Arie and Gustavo S Mesch. The spatial and social network dimensions of mobile communication: a test of the social stratification and social diversification hypotheses. *Communication Research*, 43(5):713–734, 2016.

[207] Antonio Luca Alfeo, Mario GCA Cimino, Bruno Lepri, Alexander Sandy Pentland, and Gigliola Vaglini. Assessing refugees' integration via spatio-temporal similarities of mobility and calling behaviors. *IEEE Transactions on Computational Social Systems*, 6(4):726–738, 2019.

[208] Felipe Montes, Roberto C Jimenez, and Jukka-Pekka Onnela. Connected but segregated: social networks in rural villages. *Journal of Complex Networks*, 6(5):693–705, 2018.

[209] Partha Pratim Adhikary, Pravat Kumar Shit, and Gouri Sankar Bhunia. Ntfps for socioeconomic security of rural households along the forest ecotone of paschim medinipur forest division, india. In *Forest resources resilience and conflicts*, pages 239–246. Elsevier, 2021.

[210] Robert M Groves. *Survey errors and survey costs*. John Wiley & Sons, 2005.

[211] Robert J Fisher. Social desirability bias and the validity of indirect questioning. *Journal of consumer research*, 20(2):303–315, 1993.

[212] Herminia Ibarra. Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Administrative science quarterly*, pages 422–447, 1992.

[213] Krista Casler, Lydia Bickel, and Elizabeth Hackett. Separate but equal? a comparison of participants and data gathered via amazon's mturk, social media, and face-to-face behavioral testing. *Computers in human behavior*, 29(6):2156–2160, 2013.

[214] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.

[215] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. Computational social science. *Science*, 323(5915):721–723, 2009.

[216] U.S. Census Bureau. 2016 American Community Survey 5-Year Data. https://www.census.gov/programs-surveys/acs, 2017. Accessed: 2024-09-26.

[217] ICF International. Demographic and Health Surveys (DHS). https://dhsprogram.com, 2024. Accessed: 2024-09-26.

[218] Thomas Hone, Jonathan Stokes, Anete Trajman, Valeria Saraceni, Claudia Medina Coeli, Davide Rasella, Betina Durovni, and Christopher Millett. Racial and socioeconomic disparities in multimorbidity and associated healthcare utilisation and outcomes in brazil: a cross-sectional analysis of three million individuals. *BMC Public Health*, 21:1–11, 2021.

[219] Leanne C Findlay, Elizabeth Beasley, Jungwee Park, Dafna E Kohen, Yann Algan, Frank Vitaro, and Richard E Tremblay. Longitudinal child data: what can be gained by linking administrative data and cohort data? *International Journal of Population Data Science*, 3(1), 2018.

[220] Neil T Coffee, Tony Lockwood, Peter Rossini, Theo Niyonsenga, and Stanley McGreal. Composition and context drivers of residential property location value as a socioeconomic status measure. *Environment and Planning B: Urban Analytics and City Science*, 47(5):790–807, 2020.

[221] Commuting Data Tables. https://www.census.gov/topics/employment/commuting/data/tables.html, 2020. Accessed: 2024-09-26.

[222] Origin-destination data explorer: Census 2021. https://www.insee.fr/fr/statistiques/2022117, 2009. Accessed: 2024-09-26.

[223] Origin-destination data explorer: Census 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestima 11-21, 2021. Accessed: 2024-09-26.

[224] Arline T Geronimus, John Bound, and Lisa J Neidert. On the validity of using census geocode characteristics to proxy individual socioeconomic characteristics. *Journal of the American Statistical Association*, 91(434):529–537, 1996.

[225] John Jerrim. Measuring socio-economic background using administrative data. what is the best proxy available? Technical report, Quantitative Social Science-UCL Social Research Institute, University . . . , 2020.

[226] Joshua Blumenstock, Ye Shen, and Nathan Eagle. A method for estimating the relationship between phone use and wealth. In *QualMeetsQuant workshop at the 4th international conference on information and communication technologies and development*, volume 13, pages 114–125, 2010.

[227] Yuxiao Dong, Yang Yang, Jie Tang, Yang Yang, and Nitesh V Chawla. Inferring user demographics and social strategies in mobile social networks. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 15–24, 2014.

[228] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 453:779–782, 2008.

[229] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[230] Shan Jiang, Joseph Ferreira, and Marta C Gonzalez. Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. *IEEE Transactions on Big Data*, 3(2):208–219, 2017.

[231] Márton Karsai, Hang-Hyun Jo, Kimmo Kaski, et al. *Bursty human dynamics*. Springer, 2018.

[232] Nathan Eagle, Alex Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, 106(36):15274–15278, 2009.

[233] Vincent D Blondel, Adeline Decuyper, and Gautier Krings. A survey of results on mobile phone datasets analysis. *EPJ data science*, 4:1–55, 2015.

[234] Renaud Lambiotte, Vincent D Blondel, Cristobald De Kerchove, Etienne Huens, Christophe Prieur, Zbigniew Smoreda, and Paul Van Dooren. Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387(21):5317–5325, 2008.

[235] Yves-Alexandre De Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3(1):1–5, 2013.

[236] Joshua E Blumenstock, Nathan Eagle, and Marcel Fafchamps. Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters. *Journal of Development Economics*, 120:157–181, 2016.

[237] Joshua Blumenstock and Nathan Eagle. Mobile divides: gender, socioeconomic status, and mobile phone use in rwanda. In *Proceedings of the 4th ACM/IEEE international conference on information and communication technologies and development*, pages 1–10, 2010.

[238] Frank Schlosser, Vedran Sekara, Dirk Brockmann, and Manuel Garcia-Herranz. Biases in human mobility data impact epidemic modeling. *arXiv preprint arXiv:2112.12521*, 2021.

[239] Huina Mao, Xin Shuai, Yong-Yeol Ahn, and Johan Bollen. Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to côte d'ivoire. *EPJ Data Science*, 4:1–16, 2015.

[240] Susan L Erikson. Cell phones≠ self and other problems with big data detection and containment during epidemics. *Medical Anthropology Quarterly*, 32(3):315–339, 2018.

[241] Tilottama Ghosh, Sharolyn J Anderson, Christopher D Elvidge, and Paul C Sutton. Using nighttime satellite imagery as a proxy measure of human well-being. *sustainability*, 5(12):4988–5019, 2013.

[242] Christopher D Elvidge, Kimberley E Baugh, Eric A Kihn, Herbert W Kroehl, Ethan R Davis, and Chris W Davis. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*, 18(6):1373–1379, 1997.

[243] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, 2016.

[244] Kumar Ayush, Burak Uzkent, Marshall Burke, David Lobell, and Stefano Ermon. Generating interpretable poverty maps using object detection in satellite images. *arXiv preprint arXiv:2002.01612*, 2020.

[245] Kumar Ayush, Burak Uzkent, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. Efficient poverty mapping from high resolution remote sensing images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12–20, 2021.

[246] Joshua Evan Blumenstock. Fighting poverty with data. *Science*, 353(6301):753–754, 2016.

[247] Jun Wang, Jian Gao, Jin-Hu Liu, Dan Yang, and Tao Zhou. Regional economic status inference from information flow and talent mobility. *Europhysics Letters*, 125(6):68002, 2019.

[248] Jacob Levy Abitbol, Márton Karsai, and Eric Fleury. Location, occupation, and semantics based socioeconomic status inference on twitter. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 1192–1199. IEEE, 2018.

[249] Nikolaos Aletras and Benjamin Paul Chamberlain. Predicting twitter user socioeconomic attributes with network and language information. In *Proceedings of the 29th on Hypertext and Social Media*, pages 20–24. 2018.

[250] Daniele Quercia, Jonathan Ellis, Licia Capra, and Jon Crowcroft. Tracking" gross community happiness" from tweets. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 965–968, 2012.

[251] Daniele Rama, Yelena Mejova, Michele Tizzoni, Kyriaki Kalimeri, and Ingmar Weber. Facebook ads as a demographic tool to measure the urban-rural divide. In *Proceedings of The Web Conference 2020*, pages 327–338, 2020.

[252] Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A tale of many cities: universal patterns in human urban mobility. *PloS one*, 7(5):e37027, 2012.

[253] Bartosz Hawelka, Izabela Sitko, Euro Beinat, Stanislav Sobolevsky, Pavlos Kazakopoulos, and Carlo Ratti. Geo-located twitter as proxy for global mobility patterns. *Cartography and geographic information science*, 41(3):260–271, 2014.

[254] Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna. Four degrees of separation. In *Proceedings of the 4th annual ACM Web science conference*, pages 33–42, 2012.

[255] Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

[256] Robin IM Dunbar. Do online social media cut through the constraints that limit the size of offline social networks? *Royal Society Open Science*, 3(1):150292, 2016.

[257] Xingjie Liu, Qi He, Yuanyuan Tian, Wang-Chien Lee, John McPherson, and Jiawei Han. Event-based social networks: linking the online and offline social worlds. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1032–1040, 2012.

[258] Jean-Pierre Eckmann, Elisha Moses, and Danilo Sergi. Entropy of dialogues creates coherent structures in e-mail traffic. *Proceedings of the National Academy of Sciences*, 101(40):14333–14337, 2004.

[259] Jitesh Shetty and Jafar Adibi. The enron email dataset database schema and brief statistical report. *Information sciences institute technical report, University of Southern California*, 4(1):120–128, 2004.

[260] Cuebiq Inc. Cuebiq data for mobility and location analytics. https://www.cuebiq.com. Accessed: 2023-09-24.

[261] SafeGraph Inc. Safegraph data. https://www.safegraph.com. Accessed: 2024-09-24.

[262] Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, and Song Chong. On the levy-walk nature of human mobility: Do humans walk like monkeys? 2008.

[263] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. Understanding mobility based on gps data. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 312–321, 2008.

[264] Luca Pappalardo, Salvatore Rinzivillo, Zehui Qu, Dino Pedreschi, and Fosca Giannotti. Understanding the patterns of car travel. *The European Physical Journal Special Topics*, 215:61–73, 2013.

[265] Where's George? https://www.wheresgeorge.com/, 2024. Accessed: 2024-09-26.

[266] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.

[267] Nathan Eagle and Alex Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10:255–268, 2006.

[268] Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. Dynamics of person-to-person interactions from distributed rfid sensor networks. *PloS one*, 5(7):e11596, 2010.

[269] Arkadiusz Stopczynski, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann. Measuring large-scale social networks with high resolution. *PloS one*, 9(4):e95978, 2014.

[270] Sicheng Dai, Hélène Bouchet, Márton Karsai, Jean-Pierre Chevrot, Eric Fleury, and Aurélie Nardy. Longitudinal data collection to follow social network and language development dynamics at preschool. *Scientific Data*, 9(1):777, 2022.

[271] Guanghua Chi, Han Fang, Sourav Chatterjee, and Joshua E Blumenstock. Microestimates of wealth for all low-and middle-income countries. *Proceedings of the National Academy of Sciences*, 119(3), 2022.

[272] Lisette Espín-Noboa, János Kertész, and Márton Karsai. Interpreting wealth distribution via poverty map inference using multimodal data. In *Proceedings of the ACM Web Conference 2023*, pages 4029–4040, 2023.

[273] Kamwoo Lee and Jeanine Braithwaite. High-resolution poverty maps in subsaharan africa. *World Development*, 159:106028, 2022.

[274] Juan Antonio Carrasco, Bernie Hogan, Barry Wellman, and Eric J Miller. Collecting social network data to study social activity-travel behavior: an egocentric approach. *Environment and Planning B: Planning and Design*, 35(6):961–980, 2008.

[275] Joshua E Blumenstock. Calling for better measurement: Estimating an individual's wealth and well-being from mobile phone transaction records. 2015.

[276] Yannick Leo, Márton Karsai, Carlos Sarraute, and Eric Fleury. Correlations of consumption patterns in social-economic networks. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 493–500. IEEE, 2016.

[277] Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. Where is this tweet from? inferring home locations of twitter users. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 6, pages 511–514, 2012.

[278] Tian-ran Hu, Jie-bo Luo, Henry Kautz, and Adam Sadilek. Home location inference from sparse and noisy data: models and applications. *Frontiers of Information Technology & Electronic Engineering*, 17:389–402, 2016.

[279] Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. Modelling the scaling properties of human mobility. *Nature physics*, 6(10):818–823, 2010.

[280] Albert Einstein. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der physik*, 4, 1905.

[281] Micheal F Shlesinger, George M Zaslavsky, and Uriel Frisch. Lévy flights and related topics in physics. 1995.

[282] Luca Pappalardo, Salvatore Rinzivillo, and Filippo Simini. Human mobility modelling: exploration and preferential return meet the gravity model. *Procedia Computer Science*, 83:934–939, 2016.

[283] Hugo Barbosa, Fernando B de Lima-Neto, Alexandre Evsukoff, and Ronaldo Menezes. The effect of recency to human mobility. *EPJ Data Science*, 4:1–14, 2015.

[284] Shan Jiang, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C González. The timegeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences*, 113(37):E5370–E5378, 2016.

[285] Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. Returners and explorers dichotomy in human mobility. *Nature communications*, 6(1):8166, 2015.

[286] Luca Pappalardo and Filippo Simini. Data-driven generation of spatio-temporal routines in human mobility. *Data Mining and Knowledge Discovery*, 32(3):787–829, 2018.

[287] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[288] P. Erdős and A. Rényi. On random graphs i. *Publ. math. debrecen*, 6(290-297):18, 1959.

[289] Sergio Currarini, Jesse Matheson, and Fernando Vega-Redondo. A simple model of homophily in social networks. *European Economic Review*, 90:18–39, 2016.

[290] Gautier Krings, Francesco Calabrese, Carlo Ratti, and Vincent D Blondel. Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(07):L07003, 2009.

[291] David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Sciences*, 102(33):11623–11628, 2005.

[292] Juan Antonio Carrasco, Eric J Miller, and Barry Wellman. How far and with whom do people socialize? empirical evidence about distance between social network members. *Transportation Research Record*, 2076(1):114–122, 2008.

[293] Mark EJ Newman. Assortative mixing in networks. *Physical review letters*, 89(20):208701, 2002.

[294] Leto Peel, Jean-Charles Delvenne, and Renaud Lambiotte. Multiscale mixing patterns in networks. *Proceedings of the National Academy of Sciences*, 115(16):4057–4062, 2018.

[295] Takahiro Yabe, Bernardo García Bulle Bueno, Xiaowen Dong, Alex Pentland, and Esteban Moro. Behavioral changes during the COVID-19 pandemic decreased income diversity of urban encounters. *Nature Communications*, 14(1):2310, 2023.

[296] Nandini Iyer, Ronaldo Menezes, and Hugo Barbosa. Network entropy as a measure of socioeconomic segregation in residential and employment landscapes. In *International Workshop on Complex Networks*, pages 26–37. Springer, 2023.

[297] Thomas C Schelling. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186, 1971.

[298] Thomas C Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.

[299] Daniele Gambetta, Giovanni Mauro, and Luca Pappalardo. Mobility constraints in segregation models. *Scientific Reports*, 13(1):12087, 2023.

[300] Alexander J Laurie and Narendra K Jaggi. Role of'vision'in neighbourhood racial segregation: a variant of the schelling segregation model. *Urban Studies*, 40(13):2687–2704, 2003.

[301] Alexander V Mantzaris. Incorporating a monetary variable into the schelling model addresses the issue of a decreasing entropy trace. *Scientific reports*, 10(1):17005, 2020.

[302] Egemen Sert, Yaneer Bar-Yam, and Alfredo J Morales. Segregation dynamics with reinforcement learning and agent based modeling. *Scientific reports*, 10(1):11771, 2020.

[303] H Randy Gimblett. *Integrating geographic information systems and agent-based modeling techniques for simulating social and ecological processes*. Oxford University Press, 2002.

[304] Itzhak Benenson, Itzhak Omer, and Erez Hatna. Entity-based modeling of urban residential dynamics: the case of yaffo, tel aviv. *Environment and Planning B: Planning and Design*, 29(4):491–512, 2002.

[305] Mark Fossett and David R Dietrich. Effects of city size, shape, and form, and neighborhood size and shape in agent-based models of residential segregation: Are schelling-style preference effects robust? *Environment and Planning B: Planning and Design*, 36(1):149–169, 2009.

[306] David Abella, Maxi San Miguel, and José J Ramasco. Aging effects in schelling segregation model. *Scientific Reports*, 12(1):19376, 2022.

[307] Junfu Zhang. Residential segregation in an all-integrationist world. *Journal of Economic Behavior & Organization*, 54(4):533–550, 2004.

[308] Giovanni Mauro and Luca Pappalardo. The role of relocation policies in urban segregation dynamics. In *EDBT/ICDT Workshops*, 2024.

[309] Daniel Silver, Ultan Byrne, and Patrick Adler. Venues and segregation: A revised schelling model. *PLoS one*, 16(1):e0242611, 2021.

[310] Mateo Neira, Carlos Molinero, Stephen Marshall, and Elsa Arcaute. Urban segregation on multilayered transport networks: a random walk approach. *Scientific Reports*, 14(1):8370, 2024.

[311] Adam Douglas Henry, Paweł Prałat, and Cun-Quan Zhang. Emergence of segregation in evolving social networks. *Proceedings of the National Academy of Sciences*, 108(21):8605–8610, 2011.

[312] Angelo Mele. A structural model of segregation in social networks. *Available at SSRN 2294957*, 2013.

[313] Giorgio Fagiolo, Marco Valente, and Nicolaas J Vriend. Segregation in networks. *Journal of economic behavior & organization*, 64(3-4):316–336, 2007.

[314] United States Census Bureau. Core-Based Statistical Areas. https://www.census.gov/topics/housing/housing-patterns/about/core-based-statisticalareas.html, 2020. Accessed: 2024-09-26.

[315] Ramaswamy Hariharan and Kentaro Toyama. Project lachesis: parsing and modeling location histories. In *International Conference on Geographic Information Science*, pages 106–124. Springer, 2004.

[316] Fang Zhao, Ajinkya Ghorpade, Francisco Câmara Pereira, Christopher Zegras, and Moshe Ben-Akiva. Stop detection in smartphone-based travel surveys. *Transportation research procedia*, 11:218–226, 2015.

[317] Ulf Aslak and Laura Alessandretti. Infostop: scalable stop-location detection in multi-user mobility data. *arXiv preprint arXiv:2003.14370*, 2020.

[318] Lauren Alexander, Shan Jiang, Mikel Murga, and Marta C González. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation research part c: emerging technologies*, 58:240–250, 2015.

[319] Cornelia Schneider, Simon Gröchenig, Verena Venek, Michael Leitner, and Siegfried Reich. A framework for evaluating stay detection approaches. *ISPRS International Journal of Geo-Information*, 6(10):315, 2017.

[320] Andrea Cuttone, Sune Lehmann, and Jakob Eg Larsen. Inferring human mobility from sparse low accuracy mobile sensing data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 995–1004, 2014.

[321] Cuebiq. Sensitive Points of Interest Policy. https://www.cuebiq.com/spoi-policy/. Accessed: 2024-12-26.

[322] Franz Aurenhammer and Rolf Klein. Voronoi diagrams. *Handbook of computational geometry*, 5(10):201–290, 2000.

[323] Maarten Vanhoof, Fernando Reis, Zbigniew Smoreda, and Thomas Ploetz. Detecting home locations from cdr data: introducing spatial uncertainty to the state-of-the-art. *arXiv preprint arXiv:1808.06398*, 2018.

[324] Reynolds Farley and William H Frey. Changes in the segregation of whites from blacks during the 1980s: Small steps toward a more integrated society. *American sociological review*, pages 23–45, 1994.

[325] Nicola Perra. Non-pharmaceutical interventions during the COVID-19 pandemic: A review. *Physics Reports*, 2021.

[326] Nicolò Gozzi, Michele Tizzoni, Matteo Chinazzi, Leo Ferres, Alessandro Vespignani, and Nicola Perra. Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. *Nature communications*, 12(1):1–9, 2021.

[327] Giovanni Bonaccorsi, Francesco Pierri, Matteo Cinelli, Andrea Flori, Alessandro Galeazzi, Francesco Porcelli, Ana Lucia Schmidt, Carlo Michele Valensise, Antonio Scala, Walter Quattrociocchi, et al. Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences*, 117(27):15530–15535, 2020.

[328] Giulia Pullano, Eugenio Valdano, Nicola Scarpa, Stefania Rubrichi, and Vittoria Colizza. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. *The Lancet Digital Health*, 2(12):e638–e649, 2020.

[329] Matteo Chinazzi, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489):395–400, 2020.

[330] Frank Schlosser, Benjamin F Maier, Olivia Jack, David Hinrichs, Adrian Zachariae, and Dirk Brockmann. COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proceedings of the National Academy of Sciences*, 117(52):32883–32890, 2020.

[331] Daniel Green and Erik Loualiche. State and local government employment in the COVID-19 crisis. *Journal of Public Economics*, 193:104321, 2021.

[332] Martina Luchetti, Ji Hyun Lee, Damaris Aschwanden, Amanda Sesker, Jason E Strickhouser, Antonio Terracciano, and Angelina R Sutin. The trajectory of loneliness in response to COVID-19. *American Psychologist*, 2020.

[333] Elisabet Alzueta, Paul Perrin, Fiona C Baker, Sendy Caffarra, Daniela Ramos-Usuga, Dilara Yuksel, and Juan Carlos Arango-Lasprilla. How the COVID-19 pandemic has changed our lives: A study of psychological correlates across 59 countries. *Journal of clinical psychology*, 77(3):556–570, 2021.

[334] Juan D Caicedo, Joan L Walker, and Marta C González. Influence of socioeconomic factors on transit demand during the covid-19 pandemic: A case study of bogotá's brt system. *Frontiers in Built Environment*, 7:642344, 2021.

[335] Marco Dueñas, Mercedes Campi, and Luis E Olmos. Changes in mobility and socioeconomic conditions during the COVID-19 outbreak. *Humanities and Social Sciences Communications*, 8(1):1–10, 2021.

[336] Clodomir Santana, Federico Botta, Hugo Barbosa, Filippo Privitera, Ronaldo Menezes, and Riccardo Di Clemente. Covid-19 is linked to changes in the time–space dimension of human mobility. *Nature Human Behaviour*, 7(10):1729–1739, 2023.

[337] Minha Lee, Jun Zhao, Qianqian Sun, Yixuan Pan, Weiyi Zhou, Chenfeng Xiong, and Lei Zhang. Human mobility trends during the early stage of the COVID-19 pandemic in the United States. *PLoS One*, 15(11):e0241468, 2020.

[338] Ruth F Hunter, Leandro Garcia, Thiago Herick de Sa, Belen Zapata-Diomedi, Christopher Millett, James Woodcock, Esteban Moro, et al. Effect of COVID-19 response policies on walking behavior in US cities. *Nature communications*, 12(1):1–9, 2021.

[339] World economic situation and prospects: April 2020 briefing, no. 136. https://www.un.org/development/desa/dpad/publication/world-economic-situation-and-prospects-april-2020-briefing-no-136/, 2020. Accessed: 2024-09-26.

[340] Gonzalo E Mena, Pamela P Martinez, Ayesha S Mahmud, Pablo A Marquet, Caroline O Buckee, and Mauricio Santillana. Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. *Science*, 372(6545):eabg5298, 2021.

[341] Brian L Levy, Karl Vachuska, SV Subramanian, and Robert J Sampson. Neighborhood socioeconomic inequality based on everyday mobility predicts COVID-19 infection in San Francisco, Seattle, and Wisconsin. *Science advances*, 8(7):eabl3825, 2022.

[342] Gini index - sierra leone. https://data.worldbank.org/indicator/SI.POV.GINI?locations=SL, 2018. Accessed: 2024-09-26.

[343] Samuel Engle, John Stromme, and Anson Zhou. Staying at home: mobility effects of COVID-19. *Available at SSRN 3565703*, 2020.

[344] Adrián Glodeanu, Pedro Gullón, and Usama Bilal. Social inequalities in mobility during and following the COVID-19 associated lockdown of the Madrid metropolitan area in Spain. *Health & Place*, 70:102580, 2021.

[345] Statistics Sierra Leone, Oxford Poverty and Human Development Initiative, and United Nations Development Programme. Sierra Leone multidimensional poverty index, 2019. Accessed: 2024-09-26.

[346] Lorenzo Lucchini, Ollin Langle-Chimal, Lorenzo Candeago, Lucio Melito, Alex Chunet, Aleister Montfort, Bruno Lepri, Nancy Lozano-Gracia, and Samuel P Fraiberger. Socioeconomic disparities in mobility behavior during the covid-19 pandemic in developing countries. *arXiv preprint arXiv:2305.06888*, 2023.

[347] Jaehee Park, Ming-Hsiang Tsou, Atsushi Nara, Somayeh Dodge, and Susan Cassels. Examining human mobility changes during covid-19 across socioeconomic groups: a comparative analysis of san diego county and new york city. *Computational Urban Science*, 4(1):21, 2024.

[348] Xiao Li, Xiao Huang, Dongying Li, and Yang Xu. Aggravated social segregation during the covid-19 pandemic: Evidence from crowdsourced mobility data in twelve most populated us metropolitan areas. *Sustainable cities and society*, 81:103869, 2022.

[349] Giovanni Bonaccorsi, Francesco Pierri, Francesco Scotti, Andrea Flori, Francesco Manaresi, Stefano Ceri, and Fabio Pammolli. Socioeconomic differences and persistent segregation of italian territories during covid-19 pandemic. *Scientific reports*, 11(1):21174, 2021.

[350] Alessandro Gabbiadini, Cristina Baldissarri, Federica Durante, Roberta Rosa Valtorta, Maria De Rosa, and Marcello Gallucci. Together apart: the mitigating role of digital communication technologies on negative affect during the covid-19 outbreak in italy. *Frontiers in psychology*, 11:554678, 2020.

[351] Ryuichi Kitamura, Cynthia Chen, Ram M Pendyala, and Ravi Narayanan. Microsimulation of daily activity-travel patterns for travel demand forecasting. *Transportation*, 27:25–51, 2000.

[352] Marc Barthelemy. *The structure and dynamics of cities.* Cambridge University Press, 2016.

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

[354] Kai Nagel and Maya Paczuski. Emergent traffic jams. *Physical Review E*, 51(4):2909, 1995.

[355] Serdar Çolak, Antonio Lima, and Marta C González. Understanding congested travel in urban areas. *Nature communications*, 7(1):10793, 2016.

[356] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the national academy of sciences*, 106(51):21484–21489, 2009.

[357] Michele Tizzoni, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Christian M Schneider, Vincent Blondel, Zbigniew Smoreda, Marta C González, and Vittoria Colizza. On the use of human mobility proxies for modeling epidemics. *PLoS computational biology*, 10(7):e1003716, 2014.

[358] Laura Alessandretti, Ulf Aslak, and Sune Lehmann. The scales of human mobility. *Nature*, 587(7834):402–407, 2020.

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

[360] Markus Schläpfer, Lei Dong, Kevin O'Keeffe, Paolo Santi, Michael Szell, Hadrien Salat, Samuel Anklesaria, Mohammad Vazifeh, Carlo Ratti, and Geoffrey B West. The universal visitation law of human mobility. *Nature*, 593(7860):522–527, 2021.

[361] Yanni Yang, Alex Pentland, and Esteban Moro. Identifying latent activity behaviors and lifestyles using mobility data to describe urban dynamics. *EPJ Data Science*, 12(1):15, 2023.

[362] Foursquare Venue Category Hierarchy. https://docs.foursquare.com/data-products/docs/categories, 2020. Accessed: 2024-09-26.

[363] Christian M Schneider, Vitaly Belik, Thomas Couronné, Zbigniew Smoreda, and Marta C González. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246, 2013.

[364] Eun-Kyeong Kim and Hang-Hyun Jo. Measuring burstiness for finite event sequences. *Physical Review E*, 94(3):032311, 2016.

[365] Ravi Maheswaran, Tim Pearson, Hannah Jordan, and David Black. Socioeconomic deprivation, travel distance, location of service, and uptake of breast cancer screening in north derbyshire, uk. *Journal of Epidemiology & Community Health*, 60(3):208–212, 2006.

[366] Elisa Omodei, Manuel Garcia-Herranz, Daniela Paolotti, and Michele Tizzoni. Complex systems for the most vulnerable. *Journal of Physics: Complexity*, 3(2):021001, 2022.

[367] Shikha Verma. Weapons of math destruction: how big data increases inequality and threatens democracy. *Vikalpa*, 44(2):97–98, 2019.

[368] Vedran Sekara, Márton Karsai, Esteban Moro, Dohyung Kim, Enrique Delamonica, Manuel Cebrian, Miguel Luengo-Oroz, Rebeca Moreno Jiménez, and Manuel Garcia-Herranz. Are machine learning technologies ready to be used for humanitarian work and development? *arXiv preprint arXiv:2307.01891*, 2023.

[369] Michael Muthukrishna, Adrian V Bell, Joseph Henrich, Cameron M Curtin, Alexander Gedranovich, Jason McInerney, and Braden Thue. Beyond western, educated, industrial, rich, and democratic (weird) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological science*, 31(6):678–701, 2020.

[370] Przemyslaw A Grabowicz, Jose J Ramasco, Bruno Gonçalves, and Víctor M Eguíluz. Entangling mobility and interactions in social media. *PloS one*, 9(3):e92196, 2014.

[371] Juan Antonio Carrasco and Eric J Miller. Exploring the propensity to perform social activities: a social network approach. *Transportation*, 33:463–480, 2006.

[372] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[373] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph (p*) models for social networks. *Social networks*, 29(2):173–191, 2007.

[374] Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(1):016107, 2011.

[375] Tiago P Peixoto. Disentangling homophily, community structure, and triadic closure in networks. *Physical Review X*, 12(1):011004, 2022.

[376] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.

[377] Nicola Perra, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani. Activity driven modeling of time varying networks. *Scientific reports*, 2(1):469, 2012.

[378] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014.

[379] James S Coleman. Relational analysis: The study of social organizations with survey methods. *Human organization*, 17(4):28–36, 1958.

[380] David Krackhardt and Robert N Stern. Informal networks and organizational crises: An experimental simulation. *Social psychology quarterly*, pages 123–140, 1988.

[381] Antonio Rodriguez-Moral and Marc Vorsatz. An overview of the measurement of segregation: classical approaches and social network analysis. *Complex networks and dynamics: Social and economic interactions*, pages 93–119, 2016.

[382] Coralio Ballester and Marc Vorsatz. Random walk-based segregation measures. *Review of Economics and Statistics*, 96(3):383–401, 2014.

[383] Innocent Ndubuisi-Obi, Ziyu Ran, Yanchao Li, Chenab Ahuja Navalkha, Sarah Williams, and Lily Tsai. Using public-private data to understand compliance with mobility restrictions in Sierra Leone. *Urban Informatics and Future Cities*, pages 33–49, 2021.

[384] Clionadh Raleigh, rew Linke, Håvard Hegre, and Joakim Karlsen. Introducing ACLED: An armed conflict location and event dataset. *Journal of peace research*, 47(5):651–660, 2010.

[385] Marc Barthélemy. Spatial networks. *Physics reports*, 499(1-3):1–101, 2011.

[386] Laetitia Gauvin, Mathieu Génois, Márton Karsai, Mikko Kivelä, Taro Takaguchi, Eugenio Valdano, and Christian L Vestergaard. Randomized reference models for temporal networks. *SIAM Review*, 64(4):763–830, 2022.

[387] Ron Milo, Nadav Kashtan, Shalev Itzkovitz, Mark EJ Newman, and Uri Alon. On the uniform generation of random graphs with prescribed degree sequences. *arXiv preprint cond-mat/0312028*, 2003.