Structure and Dynamics of Mobility Segregation and Spatial Diversity in Cities

By

Rafiazka Millanida Hilman

Submitted to
Central European University
Department of Network and Data Science

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

Supervisor: Márton Karsai

Vienna, Austria

2023

Declaration

I, the undersigned, Rafiazka Millanida Hilman, candidate for the degree of Doctor of Philosophy in Network Science at the Central European University Department of Network and Data Science, declare herewith that the present thesis is exclusively my own original 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, 31 May 2023

Rafiazka Millanida Hilman

Abstract

City is a living laboratory for more than half of population across the globe. Conceptualisation of city is not limited to the design of physical structure but rather far reaching to the dynamics of each element that makes it alive respectively people, places, and interactions. People visit various locations on daily basis to run their errands. Places occupy particular area given their functions. Individual preference regarding places they would like to visit differ from one to another and to the extent dictated by socioeconomic background. Considering such multifaceted setting, this thesis aims to analyse the structure and dynamics of mobility segregation and spatial diversity in cities through the lenses of network science and complex system with computational approach. At first, we lay integrated theoretical foundation in portraying city, complex system, and urban mixing. It is followed by the investigation on mixing patterns of urban mobility, supported by data driven study in the twenty largest cities of the United States. We show that stratification exists in visit preference, presaging the 'upwards bias' and closely related to segregation at large. Furthermore, we focus on analysing the impact of COVID-19 on the configuration of mobility patterns in a number of main capital cities namely New York, London, Jakarta and Bogota. The findings suggest that dynamical segregation phenomena exist which inevitably worsened during lockdown and sustained into long term effects as residual segregation. In the last study, we formulate artefact of measurement for economic diversity in urban context. It results in the notion of dual centrality, highlighting the importance of economic diversity in determining the location centrality. Finally, we present the overall contributions of this thesis and draw the line for both empirical implementations and future extensions.

Voor De Gordel Van Smaragd,
een land ver hier vandaan, maar nooit vergeten.
Tiada lain tiada bukan, namun ianya yang senantiasa
menjadi motivasi tak terperi.

To every single bit that implacably makes life worth a struggle, *La lutte continue*.
To inter generations that relentlessly bring colours to life, *Ik ben gezegend.*
To each day that continuously makes life alive, *Lehre mich in der Stille*.
To rooted cosmopolitanism, *Di mana bumi dipijak, di situ langit dijunjung*.

Acknowledgements

"*Kadadosan ing alam dunya menika*
sampun gadhah wekdalipun piyambak-piyambak
ingkang sampun ginaris dening Gusti."
-A Portrait of Javanese: *Aja Nggege Mangsa-*

"*Fela lao, lila se honyoli, ruku se sodabi.*"
-A Reflection of Tidorenese: *Borero Gosimo*

A long journey with non-linear paths and myriad externalities. I could not better frame the entire doctoral phase but the aforementioned. Life will never be the same. The process takes unorthodox toll to the final gate under relocation and pandemics, between epics and relics. Without indebted supports from the network, I just stand in the periphery as an emulation of isolated note. My utmost gratitude is to Márton Karsai, whom supervise my research and scientific endeavours during PhD duration. I thank my collaborators, Gerardo Iñiguez (Tampere University), Manuel Garcia-Herranz (UNICEF Office of Innovation), Vedran Sekara (IT University of Copenhagen), and Clémentine Cottineau (Delft University of Technology) for their incredible contributions in our joint works.

I thank my friends and colleagues with whom I shared PhD life at the Department of Network and Data Science, among others Luis, Orsi, Rebeka, Tamer, Teo, and Hao. To the faculties and coordinators, as well as postdocs and visitors, Lisette, Iacopo, and Greg to mention a few, I depart with my highest appreciation.

A sparking enthusiasm for the PhD chapter could be traced to the day in Rotterdam when I worked with Rommert Dekker (Erasmus University Rotterdam) for my master program. I would like to convey my appreciation for his invaluable guidance and support to this time. An equivalence expression is addressed to Wihana Kirana Jaya (Gadjah Mada University) and Tri Widodo (Gadjah Mada University), the main figures during my undergraduate days who seed the interest in academic research. The dynamics of research and the strategies to cope with challenges along my PhD track is inseparable from mentors: Júlia Koltai (Eötvös Loránd University), Silvia Fierăscu (West University of Timișoara), and Hiroki Sayama (Binghamton University, State University of New York). I could not sufficiently give credit to their encouragement and support.

It is not solely about the people, but also institutions that I encounter to complete the PhD configuration: Central European University (Budapest-Vienna), Complexity Science Hub (Vienna), Alan Turing Institute (London), and Network Science Institute (Boston). The role of scientific community is also evident: Network Science Society, Complex System Society, Women in Network Science, Women in Data Science, Women in Big Data, and Vienna Data Science Group. Many thanks for all supports and inspirations.

The role of social network remains crucial. In Hungary, let me thank Fovam Ter Community: Simbah Permadi, Om Arsena, Tante Effy, Tante Arni, and Mbak Agatha. The same recognition is for Hirtenberg Community in Austria: Chandra, Michael, and Richard. Last but not least, it is *Österreichisch-Indonesischen Gesellschaft*. Thank you for being there for me when needed.

A supportive family that deliberates me to shed my own light for my preferred path makes me even more thankful than ever. To my parents, Hilman and Indrina, my siblings, Ersa and Zela, your incredible encouragements are highly appreciated. To my family, relatives, and friends in Indonesia, the Netherlands, and elsewhere across the globe, thank you for the moments we cherished together in the lows and the highs. To the late generation that has brought colourful life in raising me, Eyang Soegiri, Eyang Tien, Eyang Prawirokartono, Opa Haroen, and Oma Aisjah, thank you for everything and so long. This thesis is dedicated to every single page in the book of life. *Ik ben gezegend.*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Cities are an immense laboratory of
trial and error, failure and success,
in city building and city design."

"This is what a city is,
bits and pieces that supplement each other
and support each other."
-Jane Jacobs-

1.1 City, complex system, and urban mixing

What makes a city a city? Once we hear about New York, our visual memory throws the mesmerising Liberty Statue, the hustle and bustle in financial district of Wall Street, or the masterpiece of landscape architecture in Central Park. Travelling to London, Buckingham Palace symbolises the epicentre of largest monarchy in the world, located steps away from Big Ben by the side of Thames River. A trip to South America leads us to Bogota with the historical charm of La Candelaria or immersing street art culture. Crossing the equatorial line, we embrace the warm smile of Jakarta with its remarkable melting pot in between skyscrapers, newly developed metro line and monorail which are integrated with wider transportation corridor, or indulging cuisines.

All in one, New York, London, Bogota, and Jakarta carry a capacity as an organic identity to deal with the needs and demands of its inhabitants. Urban form and urban function are inseparable when talking about cities as a whole, even the distinction is context dependent [110]. In general, the description about physical characteristics and shape of a city belongs to urban form, ranging from buildings, streets, to elements that make up the urban space [48]. It is also interchangeably with urban morphology, emphasising the geometry layout of a city [19].

In conjunction with physical aspect, urban function embodies human activities that are occurring within its spatial boundary, namely residential, productive, social, commuting, recreational, and administrational purposes [48; 204; 205]. Considering intertwined relationship between urban form and urban function mutually shapes urban dynamics through symbiotic interaction of infrastructures, people, and activities [22], Arcaute and Ramasco (2022) further argue that *"many of the spatial correlations of the different processes taking place in cities, are tightly related to the spatial distribution of functions and transport, which are both closely linked to the morphology of cities"* [10].

1.1.1 On city

The most common impression we have in mind could be a high density area filled with lively wide range of activities where segmented attributes among individuals such as economic status, social well-being, cultural background, and political leanings are quite apparent. Another constructed perception about city refers to the socioeconomic distance between the rich and the poor, the high rise buildings and the slump dwellers, or the well-connected side and the patchy infrastructure with rather limited transportation accessibility. On top of that, the way urban qualities are weighted also matters, for instance how safe the area is, how functional the urban form is, and how attractive the urban design is.

Unfortunately, negative associations particularly inequality [55], traffic congestion [42], criminal misconduct [196], and health issues [31], often appears while mentioning city. Cities are under pressure during the fight against diseases. The outbreak of COVID-19 poses a significant challenge to cities and their residents, affluent and lower socioeconomic status alike. Its impact and the measures use to stop the virus's spread disproportionately impacted the poor, marginalised, and vulnerable, exposing criticality in cities' economic structures and readiness for such external shock, particularly the public health quality and service delivery capacity.

The pandemic commencing from outside the system can be catastrophic to socioeconomic configuration and the related aspects and perpetuate individual mobility that is already constrained by socioeconomic stratification [58; 39; 112]. For some people, their capacities to adjust preferences and way of living in response to disruption are limited by their well-being. People from lower income class might have narrow flexibility and keep doing the usual means during lockdown period as seen by regularity in mobility since their works demand on-site physical presence. Interestingly, regardless aforementioned drawbacks, city still captivates people to move in and be a part of elements that make a city a city.

Generic term of city is moulded into two types of definition, one is derived by quantitative threshold and another is classified by qualitative category. An administrative unit is coined a city if population size reaches 5,000 inhabitants, with exceptions in Japan and China where higher threshold is imposed respectively 50,000 and 100,000 residents [57]. Alternatively, city is attached to the presence of amenities with certain standards, including first-rate hotels, international airport, premium hospital, and leading governmental institutions. Another qualitative attribution depends on typical economic activities, mainly non-agricultural sectors [64]. Policy stipulation might also reinstate the dimension that counts for city according to political evaluations and redistribution of resources. In addition, there has been several attempts to refine the definition of a city, shifting from spatial-centric to human activity-centric. Mumford [132] incorporates the notion of cities as agglomeration by formulating the presence of points where power and culture that belong to community are concentrated. Sjoberg [180] makes a distinction between cities and agricultural areas based on attribution to size, occupation, and literacy. Toynbee [187] identifies the economic functionality attached to cities as clusters with densely developed population where supply of goods and services takes place.

1.1.2 On complex systems

Conceptualisation of city surpasses physical boundaries. The entanglement between various elements that makes up a city namely human, space, and interactions, signifies its dynamic and complex nature [14; 54]. Human and space could be studied in isolated manner, for instance solely looking at the organisation of mobility flux or agglomeration process through time. Nonetheless, the interactions between the self-standing pieces remains unobserved. It comes to the realisation that towards comprehensive understanding about city, one should have understanding more than the behaviour of each element [8]. Those interacting entities generate emergent properties that are beyond the summation over each property separately [140; 15], making urban life more than the overall additive relationship between urban form as seen in buildings and infrastructure, and urban function as reflected in business fluctuation along with daily commuting flows [89].

For that reason, city stands out as an exemplary complex system and could be properly analysed by using complexity theories and methods. Jacobs envisages that a city is supposed to accommodate *"bits and pieces that supplement each other and support each other"* [89]. She addressed the problem of *organised complexity* arising from the condition in which *"several dozen quantities are all varying simultaneously and in sub-*

tly connected ways". Elements in the city concurrently affect and drive each other and mostly result in circular causality, envisaging the nature of connections and interdependencies [23]. Correspondingly, the problems of organised complexity lies in the nature of connections and interdependencies through social, economic, political, and any other means to fulfil the necessary condition for well-functioning life in cities where opportunities and development are visible for everyone.

A number of more recent studies argue the necessity of applying complex system framework in the science of cities. The notion of "inherent spatiality of complexity science" [147] reflects explicit interactions between system components. Further identifications of the specific properties in urban complex system result in emergence (the creation of new property driven by interactions among elements), nonergodicity (dynamically changing process over time), phase transition (cutoff point for which element convergence changes the system), and universality (persistent characterisation across times and spatial scales) [63]. The later is especially conscientiously reflected as *"cities are seen around the world and through history, suggesting 'universal' reasons for their existence"* [13]. Universality is also linked to urban scaling. Theoretical exponent of 5/6 for density scaling (population area) is validated in various empirical analysis covering ancient and cities in modern times, denoting higher density by 12% in a city of double the size as to compare to two cities of half the size [146].

There are two approaches employed in investigating simultaneous dynamic processes in urban complex system respectively urban dynamics research and human dynamics research [176]. While urban dynamics research puts forward the evolution of an urban area and attempts to trace growth, change, and decline of land use, human dynamics research explicitly underlines the role of human activities at individual level, followed by interactions among themselves that induces flows and patterns in an urban area. Therefore, the object of interest slightly differs between the two.

Urban dynamics research emphasises on spatial emergence of city and the subsequent interactions between cities due to enlargement of city scale. In contrast, human dynamics research perceives increasing interconnections between cities as outcomes of higher intensity of human flows beyond a single urban boundary. Nevertheless, both share mutual recognition towards the usefulness of spatiotemporal approach to better understand the tangible physical space and intangible interactional space of city.

To refine the resolution of urban complex system, elements that contribute to the structure and dynamics of city could be represented as network. Human dynamics defines the configuration of mobility network where people's home and places could be mapped as nodes, while links constitutes visiting patterns. In addition, transportation network is an aggregation over distinctive public transit modes (e.g.: bus, tram, and

metro) in which nodes are stations/stops and links are connections stated in the travel route. Street network is also present in cities with nodes stand for intersection and links describe streets.

1.1.3 On urban mixing

This foundational question emerges as an retrospective notion in dealing with thousands of years span human civilisation. The agglomeration of city usually absorbs its closest neighbouring space that altogether creates an urban area. Expansion of space sends signal to the inhabitants and even to the outsiders that the economic prospect is larger, attracting more people to come and inducing wave of urbanisation. Urban territory has been remarkably expanded in the modern day, serving as the economic backbone of most countries globally.

Within a century, rapid transformation of settlement has been more concentrated in city, giving a surge from 13% to 56% of the world population today [198]. Urbanisation will likely be accelerated in the coming years as an estimated 2.5 billion more people move into cities by 2050, making 7 of 10 people will live in cities. Urban sprawl puts inevitable pressure on the land use patterns with new urban built-up will acquire additional 1.2 million km^2 coverage in 2030. This increased urbanisation is a key driver of economic dynamism and social development, but it can also create enormous social challenges.

The promise that city brings is not always adequately delivered. The exclusion of the poor from the city's socioeconomic fabric, the upswing of social and economic inequalities, and the risks of epidemics crisis that recently absorbed our utmost attention are all concerns of urbanised populations. As the division becomes more visible, inequality and segregation are manifested in many aspects of life. Consequently, potential encounters between people coming from different background is constrained by limited shared physical space, pulling urban mixing process to bare minimum. Tackling these challenges is of paramount importance to fulfil the economic and social promises that cities hold, keeping them from becoming a landscape of segregation in various aspects of life such as mobility and residence.

On this ground, urban mixing evolves as a pivotal issue to carefully deal with. Individual mobility is dictated by one's meaningful places like home and work locations, or frequently visited places for shopping, children activities, and leisure. The choice of these places is strongly determined by the financial capacities, education, race and social network of people [149; 200]. The preference on residential areas and employment type emulate mobility pattern in order to meet daily errands. An interplay between

inequality and the way people organise their mobility in urban space is inevitable.

In retrospect, urban mixing of people with different backgrounds is far from homogeneous and depends on a variety of factors including geographic constrains and socioeconomic status. This leads to patterns of segregation in mobility and biased mixing patterns of the population of a city that might be observable with appropriate data. Moreover, a recent agent-based modelling simulation develops an extension of the Schelling segregation model and finds that the types and locations of venues affect segregation tendency in cities [178].

1.2 Aims and structure of the thesis

This thesis is dedicated to scientific queries evolving around socioeconomic mixing process in cities through the lenses of human dynamics namely mobility network in urban complex system. In the beginning, we provide conceptual framework of complex system to accommodate a rigorous interplay between elements that make up a city: mobility of people, spatial distribution of places, socioeconomic stratification of both people and places, as well as the interaction between people and places. Social dynamics, including urban mixing process, take place as a product of interconnections among those elements. We propose the application of spatiotemporal approach in order to capture its dynamic nature.

In the following chapters, the discussions are specified into two parts. At individual level, mobility pattern is observed to determine the initial indication of segregation, or commonly known as assortative mixing, given the presence of socioeconomic inequality. At aggregate level, residential area becomes focus of interest because it simultaneously reflects spatial embeddedness in individual preference dictated by prominent socioeconomic stratification.

Throughout six chapters, we come up with systematic review of state-of-the-art on city as a complex system where urban mixing process is closely related to the sociodemographic characteristics of area as seen in Chapter 2.

In Chapter 3 we investigate mixing patterns of mobility in the twenty largest cities of the United States. Our methodology integrates Home Detection Algorithms (HDAs), frequency visit modelled as a weighted bipartite network, and residential segregation estimate. We find strong signs of stratification indicating that people mostly visit places in their own socioeconomic class, occasionally visiting locations from higher classes. The intensity of this 'upwards bias' increases with socioeconomic status and correlates with standard measures of racial residential segregation. Our results indicate an even stronger socioeconomic segregation in individual mobility than one would expect from

system-level distributions, shedding further light on uneven mobility mixing patterns in cities. This chapter is presented as a stand alone publication in EPJ Data Science [85].

In Chapter 4 we present our study on the changing residual isolation and segregation patterns in mobility patterns in response to COVID-19 for large urban areas such as New York, London, Jakarta and Bogota. Sliding window algorithm is implemented to refine dynamically shifting assortativity in mobility that reflects the segregation level at a given time and complimented by entropy based measurement on individual trajectory. In our findings, increasing segregation in mobility overlaps with lockdown period and residual segregation in most cities suggests the emergence of long-term effects of pandemic interventions on socioeconomic mixing. Our results highlight population level dynamical segregation phenomena observed at the individual level, that provides important conclusions for better policy design with more equal consequences among people from all socioeconomic classes.

In Chapter 5, we focus on the effects of size, scale and location on the spatial agglomeration of economic diversity/complexity in New York City. Diversity in space is measured by fitting Leibovici entropy at a number of grid sizes, in comparison to our random models based on spatial randomisation procedure namely Poisson point process. We introduce the use of bipartite projection on sectoral linkage to show the centrality of area where economic diversity exists and a bipartite projection on spatial linkage to reveal the structure of sectoral proximity that drives economic diversity in city.

The final chapter of this thesis summarise the overall contributions of the presented PhD projects and explores further scientific directions that could be taken into account.

Chapter 2

Literature review and basic concepts

This chapter is dedicated to the state-of-the-arts of complex system approach to urban human dynamic process, connecting urban dynamics and human dynamics. As previously proposed in Chapter 1, urban dynamics research brings in the importance of tractable use of space across time and its implication on human activities as seen in the process of urban sprawl driven by changing land use. This also includes the inter-connectivity induced by various types of flows, among others flow of people and flow of resources, between cities or areas within a city. Alternatively, human dynamics emphasises on dynamics at both individual and group level behaviours that shapes the foundation of human society in space and time, for instance daily commuting pattern. As a complex system (Section 2.1), city accommodates urban human dynamic process through mixing patterns and spatial embededdness (Section 2.3) for which spatiotemporal pattern serves as an unifying framework.

2.1 Complexity in urban system

Since the publication of *The Death and Life of Great American Cities* by Jacobs [89], the abstraction of city as composite elements that are connected by interactions has significantly developed into the integrated notion of complex system. It is the conception of interactions that brings the centrality of urban process after dominant tradition of urban structure to the forefront discussion about cities.

Now, the main question is what does it mean with complex system and how does it relate to the conceptualisation of cities. In broader perspective, Newman [140] defines complex system as *"a system composed of many interacting parts, often called agents, which displays collective behaviour that does not follow trivially from the behaviours of the individual parts"* in which he explicitly mentions human society in the exemplary

note. Human society is built on top of three aspects namely the physical structure of society found in urban planning, the social structure of society represented by social networks, and societal discrepancies reflected in social hierarchical structure.

Batty [19; 14; 15; 16] is widely regarded as the main figure in translating and applying complex system approach in urban design and science of cities. In the early days, Batty redirects research in urban planning towards fractal geometry paradigm [19], resulting in refined simulation and visualisation of its physical form such as networks, hierarchies, and spatial distribution of population. Noticeably, distinct fractal structure is revealed in the sense that urban function and process are self-similar across levels or scale where hierarchy exists in the shape of nested structure, from neighbourhood to city, dictated by rank-size mechanism.

What emerges from his body of work is well identified elements in city or urban context that could be directly linked to the generic characterisation of complex system: scaling, interaction, networks, dynamics, and morphologies [17; 18]. Scaling pinpoints underlines a distinct feature that a city has proportional adjustment as a function of size to maintain a number of essential functions. Adjustments might appear in shape whenever there is change in size, known as allometry. Cities occupy space and separate analysis namely between cities (interurban) and within cities (intraurban). As cities differ in size due to agglomeration economies, it generates the distribution of city size where small cities outnumber big cities due to asymmetric conditionality. Firstly, to be a big city, one should pass the transformation from small a city. Secondly, a small city usually starts as a nested structure in the hinterlands of a big city. Interaction connects diverse individuals in cities via various processes, for instance production chains, trade links, and exchange of ideas, as well as bridging different areas with dedicated land use which creates economic and functional linkages, such as physical movement of people from residential area to business district. Networks constitute formal abstraction of relationships (links) between components/agents (nodes) within a system by using graph representation. It is useful in modelling urban infrastructure and adaptive in dynamic setting in which simulations are deployed to study the changes in topology and flows over time. Dynamics provides explanations regarding structural change resonating from multi-dimensional processes along temporal dimension, respectively human factors, economic exchange, and natural spatial boundary. Morphologies conceive the configurations of urban structure as a product of formation and transformation of urban forms (e.g.: streets, public spaces, and buildings). Sufficient condition for a city to reach a level of self-organisation is credited to its morphology, along with connectivity.

In a similar fashion, Battencourt [23] suggests that to be perceived as complex (adaptive) systems, cities should exhibit five properties respectively heterogeneity, inter-

connectivity, scaling, circular causality, and evolution. Firstly, the term of heterogeneity refers to the fact that major cities are extremely diversified in many aspects, including occupation type, income level, ethnic background, and socioeconomic status. Secondly, inter-connectivity highlights the fact that everything is inextricably linked in cities, for example the discourse on economic development is bounded by physical spaces and the availability of services which in turn is connected to budget capacity of individual and municipality. Thirdly, scaling, usually weighted by the number of population, determines the character of cities. A scale-dependent problem illustrates increasing returns to socioeconomic activity due to higher intensity of infrastructure use and higher economic productivity in bigger cities. Fourthly, circular causality is the situation in which cause and effect are embodied as in dualism, such as whether building infrastructure makes cities richer or being rich allows cities to build infrastructures. Last, evolution envisions capability of cities and the subsystems (e.g.: people and business) to adapt to gradual changes in uncertain environments, as seen in the economic cycle of boom (expansion) and bust (depression), or any other external shocks.

Complexity theory applied to the urban is perceived as appropriate approach to deal with many urban processes. Johnson [90] incorporates the reference of multi-level dynamics between subsystems (e.g.: transportation, retail, finance, and political), therefore every city should be treated as *systems of systems of systems*. Although these subsystems possesses own self-organised subsystems, clear boundary could be barely defined due to close knitted connectivities among them. More importantly, given the fact that every subsystem impact on are impacted on by the others, he asserts that hypernetworks should be at disposal in observing dynamic multilevel systems. It is in line with previous concept of *space of flows* from Castells [38] in which network diversities enhance interdependencies in and among cities. This particular term steps on the dynamics of urban process where urban space assumes a role as a nexus of flows, simultaneously mediating people, capital, goods, and information. Subsequently, Pflieger and Rozenblat [153] stress on the importance of gaining better understanding on how urban networks function and interact by focusing on specific features embedded in urban space namely social, economic, political, infrastructural, and technical.

We conceptualise the aforementioned aspects into a comprehensive research agenda and contextualise empirical investigation under data-driven approach. The discussion on socioeconomic stratification is provided in Subsection 2.1.3 with focuses on inequality and segregation in city (Subsection 2.3.1), followed by further elaboration on urban mobility in Subsection 2.1.1 and a closer attention given to the changing patterns of mobility during pandemics (Subsection 2.3.2). Insights on physical structure of city renowned as urban morphology is supplemented in Subsection 2.1.2 with an exemplary

case circulating around spatial diversity (Subsection 2.3.3).

2.1.1 Urban mobility

The structure and dynamics of cities are governed by mobility of people, most prominently daily commuting patterns. People move from one locations to a bunch different points sparsely located throughout space to run their errands, such as going to work, buying groceries, eating out at restaurants, visiting parks, before returning to their residences. Urban mobility patterns could be extracted by using a network approach [138] or an Origin–Destination (OD) matrix approach [53].

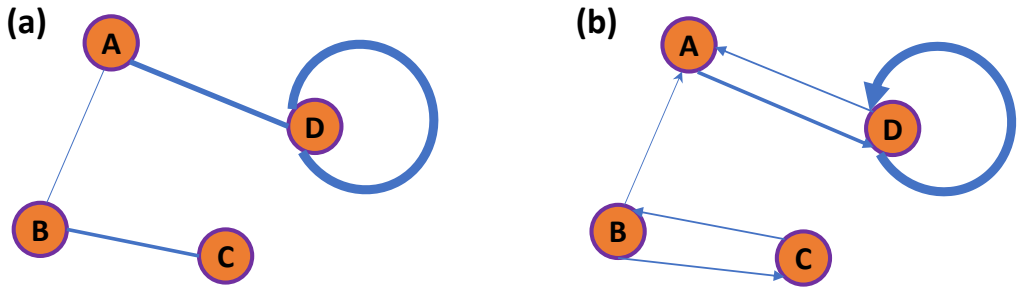

Figure 2.1: Mobility network construction. (a) Weighted undirected mobility network and (b) weighted directed mobility network, both having self-loops.

We formally express mobility network $G_s = (N, E)$ as a weighted network. Nodes $n \in N$ are spatial units, for example neighbourhoods, and edges are flows of people with volume (e.g.: number of people) w among these neighbourhoods $e \in E$. If we disregard the direction of mobility flows, only taking the aggregated number of people moving between a pair of nodes, then the network could be specified as weighted undirected network (Fig. 2.1a) in which $e_{i,j} = e_{j,i}$ and $E = \{(i,j) \in N \times N \wedge w(i,j) \neq 0\}$. Otherwise, the network is constructed as weighted directed network, with lines pointing from a node to another (Fig. 2.1b) and coding the direction and volume of movement between connected nodes for which $e_{i,j} \neq e_{j,i}$ where $E = \{(i,j)|(j,i) \in N \times N \wedge w(x,y) \neq 0\}$. In both networks above, we have self-loops or self-edges, representing the coincidence between origin node and destination node $i = j \in N$. In this case, self-loops captures intra-mobility, flows within the same neighbourhood.

Bipartite networks (or two-mode networks) could be a better fit if we distinguish between locations (nodes) (e.g.: separately identifying people's home and places they visit), instead of at spatial unit. We define $G_b = (U, P, E)$ as a weighted bipartite network. It connects people u in the set of node $u \in U$ and visited places p for set of node $p \in P$ if visit $e_{u,p} \in E$ exists. Frequency of visit is counted as edge weights $w_{u,p}$.

The second approach, OD matrix $A_{i,j}$, constituting aggregated flows of individuals from an area to various others. The origin of trip is denoted by the row i for set of origin area $i \in I$ and the destination is kept by the column j for set of destination area $j \in J$ (with $I \equiv J$). Intra-mobility is observable for $i = j$, in addition to inter-mobility for which $i \neq j$ exists. Spatial resolution of areas in cities can be any administrative units, such as census tract, zip code, or neighbourhood. Each matrix element $a \in A$ contains number of people/trip (traffic counts) between a pair of areas $g_{i,j}$ such that:

$$A_{i,j} = \sum_{i,j} g_{i,j}, i, j \in I, \tag{2.1}$$

The value of $A_{i,j}$ is an integer due to its nature as a count data. If there is no flow between a pair if areas, the value is coded as 0.

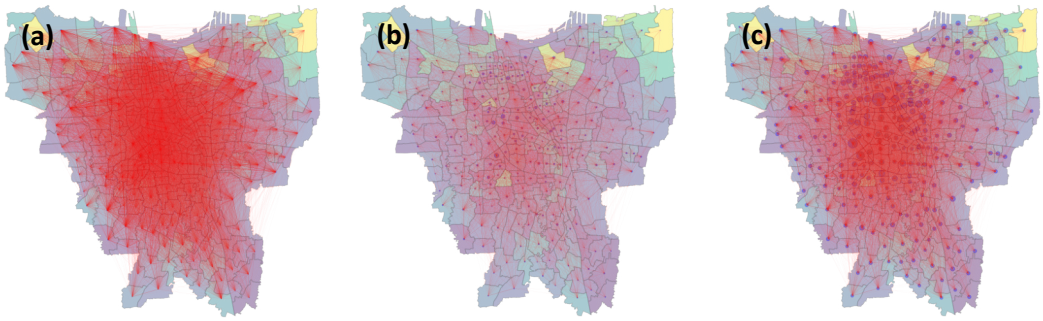

Figure 2.2: Urban mobility in Jakarta at the administrative level 4 *(kelurahan)* boundaries. Three mobility snapshots are presented along the course of COVID-19 pandemic period: Before Lockdown (Fig. 2.2a), Lockdown I (Fig. 2.2b), and Reopening I (Fig. 2.2c). Node is the centroid of *kelurahan*, weighted by number of infections. The thickness of edge is proportional to the flow it represents. Lighter colour of boundaries denotes lower socioeconomic status.

Fig. 2.2 shows the changing mobility network structure of urban mobility in Jakarta due to COVID-19 resurgence. Before lockdown policy was implemented (Fig. 2.2a), mobility of people were spread out across areas in the city but significantly dropped during the lockdown period (Fig. 2.2b). Decrease in mobility erodes the connectivity of areas located in the periphery, leaving away smaller proportion of flows in the central part of the city where main nation-wide administrative functions and business activities are located. As the mobility restrictions were uplifted to an extent (Fig. 2.2c), the mobility intensity is still lower than the baseline level.

2.1.2 Urban morphology

Talking about urban morphology, multifaceted definitions have been proposed with different stressing points. The most basic and generic definition is *"the study of urban form"* [47] and could be extended as *"the study of the physical (or built) fabric of urban form, and the people and processes shaping it"* [76]. This particular definition touches upon broad elements that build physical form of cities, respectively urban fabric. It is also interchangeably with urban tissue, constituting diverse components observable through different levels of resolution [101] namely streets, street blocks, plots and buildings.

In details, street system has two dimensions which are the open spaces for circulation (e.g.: roads, avenues, and boulevards) and the open spaces for permanence (e.g.: squares and gardens). There is a demarcation line drawn by street along different street blocks, the territory that belongs to private or semi-public use on one hand and public space where everyone crosses the path on the other hand. The later signifies the role of city in facilitating social contact between people with diverse background, transforming potential encounter into physical interaction in social terms. Consequently, the decision made of how street system being designed (e.g.: the space for pedestrians vs the space for vehicles) affect the dynamics of social mixing process in the city.

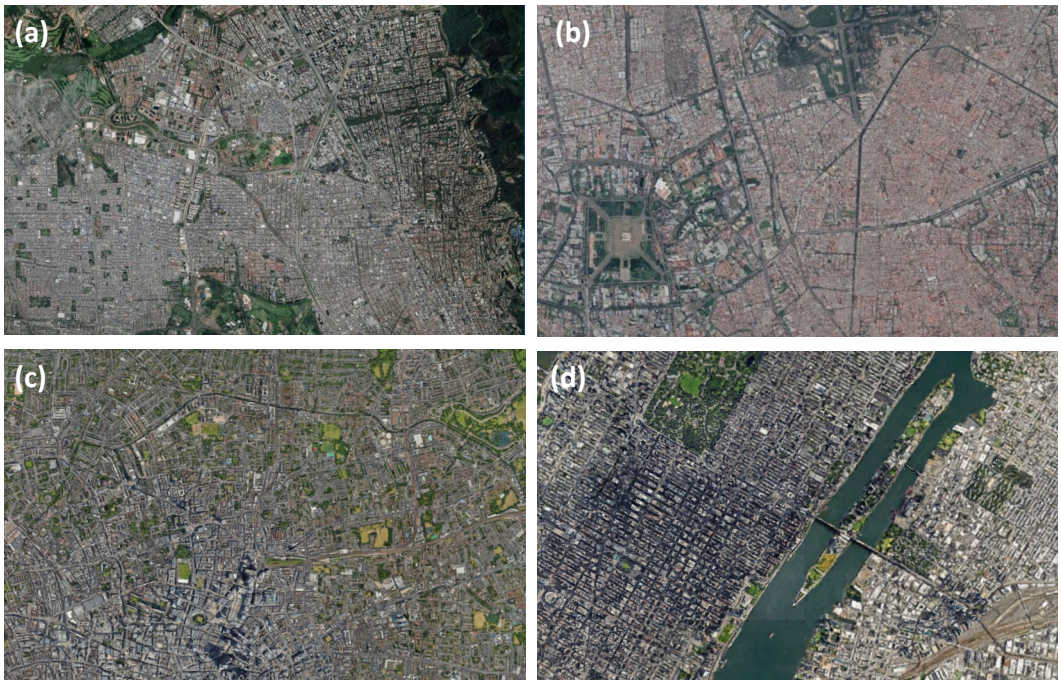

Figure 2.3: Urban tissues of cities at comparable scale (1:1000 m). We have 4 cities located in different continents: Bogota (Fig. 2.3a), Jakarta (Fig. 2.3b), London (Fig. 2.3c), and New York (Fig. 2.3d).

Physical layout of cities approximately at the same scale might look different. Fig. 2.3 visualises urban tissues where open space is favourable in the periphery of Bogota (Fig. 2.3a) and Jakarta with its extremely dense built environment with small street blocks interrupted by a number of main roads (Fig. 2.3b). Meanwhile, compact urban design is fairly recognisable in London (Fig. 2.3c) and New York (Fig. 2.3d) with noticeable presence of body of water and regular pattern of streets and of buildings alignment.

2.1.3 Urban socioeconomic patterns

Elements that compose urban structure are endowed with urban functions, altogether designate the emergence of cities. Stratification comes up as a consequence of unequal spatial distribution of embedded attributes, from the conventional measure of income and education [5; 106], to the immaterial features such as quality and happiness of life [174; 11]. On the pervasive scale, socioeconomic stratification might result in socioeconomic inequality. In the manifestation of socioeconomic inequality, elements impacting human activity in numerous sectors, such as opportunities, resources, and power, are unequally allocated [173].

Inequality is a multidimensional issue that is not only rooted in income disparity but also reflected in many aspects of life including occupation and housing [95; 182]. Attachment to spatial delineation, in contrast to randomly distributed, makes inequality even more complex to deal with, therefore applying multidimensional analysis of space offers better understanding regarding the socioeconomic structure of space and reproduction mechanism of embedded inequality [108; 141].

Inferring socioeconomic attributes from digital trace data such as social media and mobile phone communication could be taken as a strategy to capture population-level distribution of selected attribute. Moreover, it allows us to generate a relative socioeconomic position of an individual based on three factors namely cultural, social, and material capital, a concept coined as socioeconomic status (SES) in the literature [91]. It is manifested as statistical indicators comprising education, occupation, and income in census or administrative record data. SES is also a substitute for socioeconomic class defined by Ritzer (2013) [162] as attributes used to rank people within economic stratification system. Consequently, socioeconomic hierarchy is clearly seen in a way that the stratification in the modern society can be captured [170].

Given the multifaceted nature of socioeconomic measures, we attempt to reveal the most influential attribute that shapes stratification and amplifies inequality. Feature extraction technique namely Principal Component Analysis (PCA) is applied to identify

the most relevant socioeconomic features by narrowing down the range of variables that might portray inequality and segregation at best. It reveals that the largest loading belongs to income features as presented in Fig. 2.4 that in total count for 11 variables, comprising mean household income, mean earnings, median household income, median earnings for male full-time, per capita income, median earnings for female full-time, median earnings for workers, mean retirement income, mean social security income, mean supplemental security income, and mean cash public assistance income.

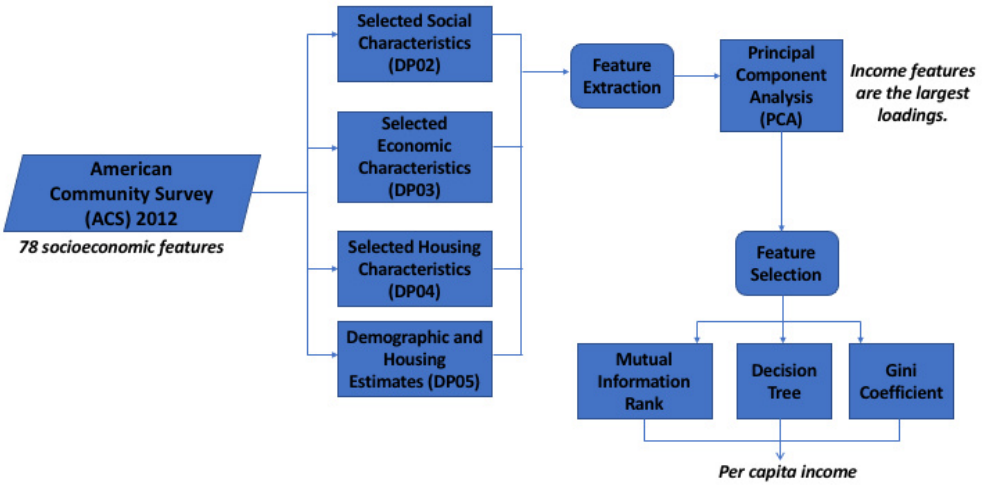

Figure 2.4: Feature collection, extraction, and selection. We collect 78 socioeconomic features from the ACS data. Based on PCA, the largest loadings correspond to income features (11 variables). Different feature selections (Mutual Information Rank, Decision Tree and Gini Coefficient) follow after. Per capita income stands out comparing to the rest of features.

The result from feature selection in Fig. 2.5 confirms the previous finding given by feature extraction. Mutual Information (MI) Rank suggests that income features have the highest contribution to mutual information of SES. In Decision Tree, per capita income remains the most important to SES. Gini Coefficient indicates that mean cash public assistance (0.46) has the highest sensitivity in capturing inequality, followed by per capita income (0.40). We choose per capita income over mean cash public assistance due to several reasons. Firstly, the coverage of mean cash public assistance is limited to particular segments in society. Secondly, it will be an issue when we are going further for a comparative study across countries since not every country has similar cash public assistance programs. Last, per capita income reflects various sources of income, including salary, social assistance, and wealth. Therefore, it is considered as a comprehensive measurement for average income and widely available.

Figure 2.5: Principal Component Analysis (PCA) and Mutual Info (MI) Rank. To determine the fittest feature in reflecting socioeconomic aspect, number of feature extraction and selection techniques are applied accordingly. (a) We take the first principal component which explains 93.5 % of the variability. (b) We estimate MI for discrete target variable (SES) to measure the dependency between variables. MI with higher value indicates higher dependency, thus, more explanatory contributions to SES are gained. PCA and MI Rank show corresponding results where variables associated with incomes appear on the top list.

2.2 Quantifying Inequalities and Socioeconomic Classes

Taking income level as a reference for socioeconomic status, for instance, gives us showcase of income concentration in the city. A common methodological practice is to run quantile-based discretisation [122] on income. It allows us to divide up the income data with given underlying distribution into equal sized bins. The percentiles rule is set to determine the bins (e.g.: 10 categories of socioeconomic status indicating quantile membership of each data point). For continuous income distribution X, the k^{th} q-quantile called x is defined as the value falls into a point where the cumulative distribution function intersects k/q such that:

$$Pr[X < x] \leq k/q, \tag{2.2}$$

which is equivalent to

$$Pr[X \geq x] \geq 1 - k/q, \tag{2.3}$$

and

$$Pr[X \leq x] \geq k/q. \tag{2.4}$$

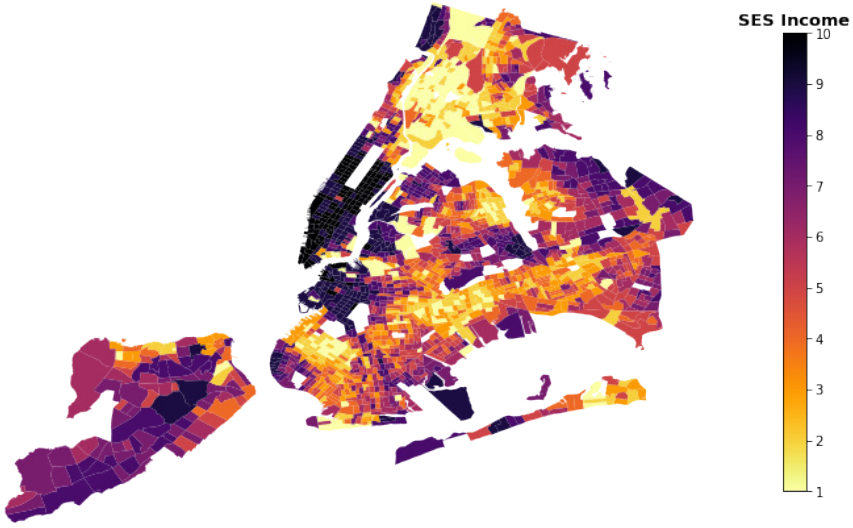

Figure 2.6: Socioeconomic Status Map of New York. Discretisation of income distribution generates 10 equally populated classes. Colour is sorted from light to dark shade, indicating ascending label from the poorest class (1) to richest class (10).

Fig. 2.6 visualises spatial distribution of diverse socioeconomic classes in New York. Areas in the surrounding downtown Manhattan, the home of main central business district, appears with darkest shade because it constitutes the highest income among others. In contrast, Bronx located on the upper side is dominate by yellow as income level of the residents are relatively low. It further implies the persistence of sociospatial inequalities that indirectly reflects the organisation of urban fabric.

2.3 Characterising mixing patterns and spatial embeddedness

Social fabric and inequality feedbacks are recursively produced and reproduced by mobilities [188; 94], bringing another layer of complexity in spatial embededdness

of urban mixing patterns. To capture multidimensionality of mixing patterns in urban landscape, we synthesis literature discussing the three main dimensions namely inequality and segregation (Subsection 2.3.1), mobility and pandemics (Subsection 2.3.2), and spatial economic diversity (Subsection 2.3.3).

2.3.1 Inequality and segregation in mobility

Earlier studies on human mobility present evidence of characteristic spatial scales [30; 73; 183; 4], as well as a correlation between human spatial behaviour and socioeconomic dynamics [119; 148; 27]. Rather than being homogeneously mixed, human mobility (as represented by daily individual trajectories throughout urban spaces) is strongly influenced by socioeconomic preferences. People sharing socioeconomic backgrounds are more likely to visit similar places within their class range and interact amongst themselves [26; 202; 192; 128], thus generating stratified mobility and social network patterns. In the presence of homophily mixing [124], spatial exploration is dictated by one's socioeconomic class, reducing the number of visits to locations with different economic status, and thus inducing highly predictable trajectories. However, when people aspire to diversify their experiences by, e.g., visiting lavish areas of the city, where they have never been able go before, the potential for an upwards bias in visiting patterns appears. Meanwhile, other has studied the effects of segregation of mixing in urban places using location data to explore exploration/exploitation behavioural patterns and their correlations with socioeconomic status [131].

In related studies, Dong and other [59] investigate segregation in economic and social interactions by using credit card transactions and Twitter data. They found that segregation increases with difference in socioeconomic status but is asymmetric for purchase activity. Meanwhile, neighbourhood isolation has been used [192] to observe travel patterns of individuals extracted from Twitter data. These findings show racial differences in the composition of the neighbourhoods visited. Black and Hispanic neighbourhoods, regardless of their socioeconomic status, are less exposed than white neighbourhoods. Moreover, white poor neighbourhoods are substantially isolated from non-poor white neighbourhoods. Morales et al. [128] aims to investigate polarisation in shopping, communication, and mobility reflected by online interaction in Twitter. It confirms the theoretical underpinning in which within-group homogenisation and between-group differentiation promote social fragmentation. They provide in-depth assessment on polarisation of conversations between neighbourhoods and show that the differentiation of online conversations reflects the distribution of wealth.

To quantify the degree of stratification in mobility network, we define a stratified

bipartite network $G = (U, P, E)$, where individual u is a node in set U, and place p belongs to set P (with $U \cap P = \emptyset$). People and places are connected by edges $e_{u,p} \in E$ with weights $w_{u,p}$ coding the number of times person u visited place p (see Fig. 3.1). Further, we stratify U into a set of socioeconomic classes indexed by values from C_U thus assigning a class membership $c_u = i \in C_U$ to each individual.

In the same way we define $c_p = j \in C_P$ classes for places. This network representation captures all information about the socioeconomically stratified visiting patterns of people to venues, coding their possible encounters and giving an aggregated description of the potential mixing patterns of people of different socioeconomic classes.

Based on the bipartite network representation we can measure the frequency at which people of a given class visit places in different classes. To summarise these visiting patterns we use stratification matrices [109; 85]. An *empirical stratification matrix* gives the probability that a person $u \in U$ from a given socioeconomic class $c_u = i \in C_U$ visits a place $p \in P$ belonging to a class $c_p = j \in C_P$. More formally:

$$M_{i,j} = \frac{\sum_{U,c_u=i} \sum_{P,c_p=j} w_{u,p}}{\sum_{j \in C_P} \sum_{U,c_u=i} \sum_{P,c_p=j} w_{u,p}}, \tag{2.5}$$

where the numerator counts the number of times people from class i visit places of class j, and the denominator normalises this frequency matrix column-wise to obtain a visiting probability distribution for each individual class $i \in C_U$.

As we have discussed, signatures of segregation can be associated to strong diagonal elements in these matrices, indicating that people of a given SES are the most likely to visit places associated with the same or similar SES, as compared to random visiting patterns. To quantify the strength of diagonal concentration of visiting probabilities, we measure the diagonality index of the normalised stratification matrices [25], which is similar to the assortativity coefficient used by others [59; 139]. It is defined as the Pearson correlation coefficient of matrix entries as

$$r = \frac{\sum_{i,j} ijN_{i,j} - \sum_{i,j} iN_{i,j} \sum_{i,j} jN_{i,j}}{\sqrt{\sum_{i,j} i^2 N_{i,j} - \left(\sum_{i,j} iN_{i,j}\right)^2} \sqrt{\sum_{i,j} j^2 N_{i,j} - \left(\sum_{i,j} jN_{i,j}\right)^2}}. \tag{2.6}$$

Here $i \in c_u$ indicates the socioeconomic class of individuals and $j \in c_p$ is the same for places. The diagonality index takes values between -1 and 1. In case it is 1, it indicates perfect assortative mixing corresponding to a fully stratified matrix with non-zero elements in its diagonals and zero anywhere else. Cities with large r values are characterised by visiting patterns of people who are strictly bounded to places associated to their own socioeconomic class. On the contrary, if r takes smaller than zero values (in

extremity $r = -1$), it indicates dis-assortative connections between people and places of different socioeconomic status. This corresponds to mobility mixing patterns where people prefer to visit places of different SES rather than places from their own class. In case $r = 0$, the normalised stratification matrix is flat indicating no choice preferences of people to visit places with particular SES.

Heterogeneity in urban setting could be traced by looking at two indivisible aspects that shape the dynamics of city, accordingly mobility trajectory of people and spatial distribution of places. We measure urban heterogeneity by computing mobility entropy and spatial entropy on the basis of generic Shannon's formula [175]. In the context of mobility entropy, entropy could be employed to quantify predictability of a visiting pattern. Generally, higher entropy is in line with lower predictability, eliciting the more heterogeneous preference of places to visit in the individual trajectory. At first, we define (*spatial mobility entropy* $Hm(X)$) as:

$$Hm(X) = -\sum_{x \in X} p_{(x)} \log_2 p_{(x)} = E[-\log p_{(X)}], \tag{2.7}$$

where X is a discrete random variable representing geographic location for which variable's possible values $x \in X$ denote sequence of POI locations visited by people.

We replicate above formulation to measure (*ses mobility entropy* $Hs(X)$) such that:

$$Hs(X) = -\sum_{x \in X} p_{(x)} \log_2 p_{(x)} = E[-\log p_{(X)}], \tag{2.8}$$

where X is replaced by a discrete random variable substituting SES of POI for which variable's possible values $x \in X$ denote sequence of socioeconomic classes of POI found in individual trajectory.

The value is normalised for each period, therefore the maximum value 1 and minimum value 0 is comparable across temporal snapshots. Upper bound value $Hm(X) = 1$ implies the sporadic visit to heterogeneous POI locations, while lower bound value $Hm(X) = 0$ indicates homogeneous visit pattern to rather limited POI locations but mostly with high return rate. In parallel, $Hs(X) = 1$ (heterogeneous SES POI) shows visit to places located in various socioeconomic classes and $Hs(X) = 0$ signifies visit pattern characterised by strictly preferred socioeconomic class (homogeneous SES POI).

In the context of spatial entropy, we take a step ahead in response to the need of suitable entropy measure in describing the spatial organisation, Leibovici (2009) [107] attempts to properly account for space dimension by introducing an additional univariate categorical variable Z containing every single realisation of coupled X in space,

for example the co-occurrences between 2 sectors (xi, xi') with preserved order such that $i \neq i'$ due to spatial dependence in co-occurrence configuration. It is explicitly expressed as

$$H_L(Z) = -\sum_{r=1}^{R} p_{(z_r|L_d)} \log_2 p_{(z_r|L_d)}, \tag{2.9}$$

where $p_{(z_r|L_d)}$ represents probability of distance d among occurrences which is conditional on cumulative distribution $d^* \leq d \in L_d$. This computation is iterated over spatial units under study.

In addition, those aforementioned stratification dynamics that takes place in space could be further signified by measuring how pervasive the process diffuses across spatial area. One way to capture it is through computation of spatial autocorrelation, a tendency of clustering formation between closely located neighbouring areas with similar attribute (e.g.: income or socioeconomic status). It could be quantified by fitting the Moran's I statistic given as [44]:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{2.10}$$

where w_{ij} is the spatial weight between locations i and j, and summing all w_{ij} results in S_0 specified as:

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}, \tag{2.11}$$

The I values take range from -1 to 1, where spatial dispersion happens at -1, spatial randomness holds at 0, and spatial clustering occurs at 1.

The Moran's I statistic could be transformed from global measure capturing general tendency of clustering pattern to relative measure locating a smaller spatial unit (e.g.: census tract) under observed larger delineation (e.g.: city), whether it is aligned with global pattern and similar to its neighbouring areas. For that purpose, we provide Local Form of Moran's I [9], an integral part of spatial measures under Local Indicators of Spatial Association (LISA), is computed to locate area where similar level of economic diversity is concentrated. It is defined as

$$I_i = z_i \sum_{j} w_{ij} z_j, \tag{2.12}$$

where z_i and z_j contain deviations of value in respectively spatial unit i and its neighbouring area j while w_{ij} is the weights in row- standardised with summation over neighbouring values $j \in J_i$. The visualisation is designated as Local Cluster Map.

2.3.2 Effects of external shocks on mobility segregation

In urban mobility network, social stratification in conjunction with unequal access to transport infrastructures brings social exclusion [116; 156] and social segregation [202; 126]. External shock such as COVID-19 outbreak commencing from outside the system can be catastrophic to socioeconomic configuration and the related aspects and perpetuate individual mobility that is already constrained by socioeconomic stratification [58; 39; 112]. For some people, their capacities to adjust preferences and way of living in response to disruption are limited by their well-being. People from lower income class might have narrow flexibility and keep doing the usual means during lockdown period as seen by regularity in mobility since their works demand on-site physical presence. After all, the impact of the pandemic on mobility segregation specifically, would likely be multifaceted and would depend on a variety of factors, including the specific characteristics of the people affected, the severity and duration of the pandemic, and the specific policies and responses implemented.

Existing literature suggests that higher income is associated with larger mobility reduction, while mobility inflexibility and less social distancing are observable among low-income, raising disparity in mobility [195; 60; 125]. In a finer grain resolution, Chang and other [40] decouple the differential spread of COVID-19 from racial composition and median income to study the disparities in mobility. They find that disadvantaged racial and socioeconomic groups, especially among the bottom income decile, for income are exposed to higher infection risk, confirming previous studies [150; 43]. One mechanism to explain this recurring pattern is their inability to reduce mobility in the first phase of outbreak (March–May 2020). In Chicago, for instance, they visited 27% more places than people from higher income. Moreover, transmission rates at those places are higher because of smaller in size with higher visitor density (59% more hourly visitors per square foot and 17% longer duration) as to compare to places visited by the richer group. Therefore, a larger burden of infection belongs to disadvantaged groups.

2.3.3 Spatial economic diversity

Literature in spatial economics, economic geography, and regional science widely discuss the significance of industrial clusters and reckon the contribution of economic

diversity in shaping competitiveness of a city. Malmberg and Maskell [117] underline that in economic process, the notion of proximity and place plays an important role. In this work, they specify that the backbone of economic process is knowledge creation or learning perspective in which spatial proximity facilitates inter-firm innovation. Hence, economic activities tend to form spatial agglomeration due to the presence of informational and knowledge exchange such as flow of labour [179], relationships between supplier and customer [157], and knowledge spillover [103] where geographic proximity facilitates these exchanges [159], pointing out the relevance of urban morphological structure at various spatial scale ranging from a street or block in a city to even larger geographical coverage.

The emergence of industrial clusters is also linked to the rise of regional specialisation that results in characterisation of heterogeneity of spatial concentration and motivated by the law of *increasing returns to scale* [102], the condition in which additional use of inputs in production generates increases in output by a larger proportion. Within a finite spatial limit, congregation of firms accessing nearby production factors including labours could accrue higher net gain as transaction costs coming from transportation and communication spending are compensated by reduced distance, therefore, *localised increasing returns to scale* gives a sound argument for the development of industrial clusters [104]. The term of *localisation economies* incentives positive externalities in a particular industry or the sector concerned [118] is broaden by recognising the dimension of spatial scales, known as *urbanisation economies* [87], making positive externalities available to various industries and all sectors. Mapping out the general agreements in the literature, it has been widely perceived that industry agglomeration and spatial clustering are related to cost sharing in infrastructure provision, the availability of a skilled labour, efficiency in transaction activities, and knowledge spillovers where taking them into account leads to firm learning and innovation [117].

Chapter 3

Socioeconomic biases in urban mixing patterns

Urban areas serve as melting pots of people with diverse socioeconomic backgrounds, who may not only be segregated but have characteristic mobility patterns in the city. While mobility is driven by individual needs and preferences, the specific choice of venues to visit is usually constrained by the socioeconomic status of people. The complex interplay between people and places they visit, given their personal attributes and homophily leaning, is a key mechanism behind the emergence of socioeconomic stratification patterns ultimately leading to urban segregation at large. Here we investigate mixing patterns of mobility in the twenty largest cities of the United States by coupling individual check-in data from the social location platform Foursquare with census information from the American Community Survey. We find strong signs of stratification indicating that people mostly visit places in their own socioeconomic class, occasionally visiting locations from higher classes. The intensity of this 'upwards bias' increases with socioeconomic status and correlates with standard measures of racial residential segregation. Our results suggest an even stronger socioeconomic segregation in individual mobility than one would expect from system-level distributions, shedding further light on uneven mobility mixing patterns in cities.

3.1 Introduction

Patterns of socioeconomic inequality can be found everywhere in a modern city. Large variations in earned income leading to uneven access to services, healthcare and education [164; 2], as well as spatial and housing segregation [184; 86], are just two of the most drastic examples of socioeconomic disparity. Less studied is the

segregation related to mobility mixing, where people from different socioeconomic classes encounter each other less often than what is potentially allowed by the city fabric [59; 130; 137].

Big data presents a unique opportunity to analyse the role of human mobility in segregation, from the level of individuals to the scale of societies. Digital data tracing human movements in cities ranges from mobile call detail records (CDRs) [73; 183] and GPS trajectories [186; 72; 3], to location-sharing services (LSS) and check-in sequences on social media platforms [80; 199; 92]. The analysis of these data sources, providing anonymised individual trajectories with unprecedented spatiotemporal resolution, has proven essential for our growing understanding of the underlying mechanisms of human mobility [30; 12; 193; 4], and the associated ability to predict future trajectories [20; 45]. It also offers the possibility to engage in a more comprehensive and nuanced exploration of urban socioeconomic segregation, by combining high-dimensional mobility data with information on the socioeconomic traits of individuals [114; 109; 59].

Homophily mixing is not the only mechanism influencing mobility patterns. The variability of socioeconomic traits such as ethnic group, education level, occupation sector, etc. also constrains the possibility of movement in urban spaces via residential segregation [184; 86; 56; 32], where people with similar backgrounds live next to each other and form fragmented areas in the city. Given the potentially complex interplay between human mobility and socioeconomic stratification, it is worth asking whether the presence of biased mobility across tracts of some socioeconomic trait is associated with lower residential segregation. This is particularly relevant given the number of studies reporting mobility as a key pillar in diminishing segregated spaces among people from diverse groups in society [172; 197; 65; 66]. People show heterogeneity in many aspects, including their mobility characteristics and socioeconomic capacities, which shape their patterns of movement across urban space.

Build on these results, our study is dedicated to reveal the role of visit preference in mobility bias caused across socioeconomic status at the individual and class levels. We specifically analyse the extent to which mobility may contribute to the emergence of socioeconomic stratification, ultimately leading to urban segregation at large. The scope of our discussion is concentrating on visit preference to detect generic patterns where 'upwards bias' increases with socioeconomic status. Additional investigation on ethnic isolation aims to instantiate the entanglement between characteristics and mobility patterns with other socioeconomic features, such as ethnic residential distribution.

In this study, we emphasise the need to query the extent to which behavioural segregation (bias in mobility) is related to residential segregation. We take a step forward in the current analysis of segregation in mobility by asking the following question:

How do socioeconomic attributes and geographic constraints affect the spatiotemporal process of individuals moving in urban spaces? To answer this question, we analyse individual check-in trajectories in the twenty largest cities of the United States, coupled with detailed socioeconomic maps indicating the economic status of people and places they visit. After a short data description, in the following we will introduce stratification matrix measures and individual- and class-level mobility bias scores to quantify patterns of mobility segregation, visiting biases, and their variation across cities with wide-ranging socioeconomic and ethnics segregation profiles. We base our analysis on observational behavioural data, which may not be fully representative for the observed populations. To address this shortcoming we carry out a careful analysis about the biases and confounding effects characterising the analysed data set. While the results of this analysis are reported in the discussion and the Appendix of this paper, they confirm the robustness of upward biased visiting patterns of people to places in cities with various socioeconomic stratification profiles.

3.2 Data and research methods

3.2.1 Data description

In order to simultaneously capture the mobility patterns and socioeconomic status of people, we concentrate on two independent sources (mobility and socioeconomic data, described below) and combine them using spatial information.

Mobility data: To construct individual mobility trajectories, we analyse a large, open Foursquare dataset [201], which records how people move from one place to another. Data comes as a sequence of user check-ins to places, or points of interest (POIs), thus providing information on mobility trajectories of individuals and visiting frequencies of places. This dataset is not collected directly through the Foursquare open API, but from Foursquare check-ins via Twitter. The crawling method corresponds to 18 months (549 days) of observations between April 2012 and September 2013 for users with Foursquare-tagged tweets. Using this mobility data, constituted by roughly 26,502 people with nearly 1,830,276 check-ins, we concentrate only on active users (who checked-in from at least two different places during the observation period).

Socioeconomic data: To estimate the socioeconomic status of people and places, we rely on the 2012 American Community Survey (ACS) [33] (recorded in the year matching the closest to the Foursquare observation period). After identifying the corresponding ACS census tract where a user's home location lies, we associate the socioeconomic indicators of this location to the individual. In order to estimate the economic status of

a place, we follow a similar strategy and assign local socioeconomic status indicators to POIs based on their locations.

By using the average per capita income as the socioeconomic indicator of active users living in a given tract, we sort them in an ascending order. To group them into distinct socioeconomic classes, we then segment this sorted list into 10 equally populated groups with people of the lowest income in class 1 and highest income in class 10. By means of this procedure we assign a socioeconomic class c_U to each user. In identical fashion, each venue is assigned a value c_P.

3.2.2 Pipeline description

Focusing on the 20 largest metropolitan areas in the US, we also infer the home locations of 26,502 users following a conventional pipeline of conditions [123] [for a detailed description of the method and a statistical summary see Section 7.1.1- 7.1.2]. The Foursquare dataset is not a uniform sample of the population, and as such, it may introduce bias in our analysis of mobility patterns. However, we expect that aggregation and averaging, as well as the length of the observation period (beyond yearly seasonality), decreases this potential for bias. In any case, in the Supplementary Material we estimate discrepancies between Foursquare data and the real population via a bootstrapping analysis (Section 7.1.1), a Kruskal-Wallis H test (Section 7.1.14), and a Dunn's test (Section 7.1.15).

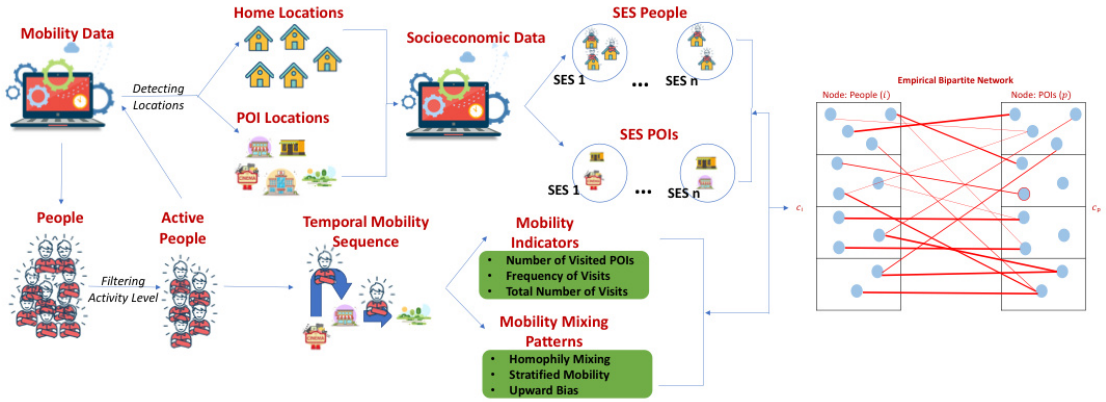

Figure 3.1: Mobility and socioeconomic data combination pipeline. (left) Overview of data sources, data processing pipelines and data combination steps to obtain data for the analysis of socioeconomic segregation in spatiotemporal urban mobility. **(right)** As a result we obtain a bipartite network, with nodes classified into two sets comprising individuals u and POIs p. Each node in both types is labelled by a socioeconomic indicator (c_U and c_P) assigned via our location-based method on the census tract level. Weighted edges between individuals and POIs indicate the frequency of visits of a given user to a given place.

Although the socioeconomic status of venues could arguably be better estimated from their pricing, this information is at present not available to us. Thus, we assume that the socioeconomic status (SES) of people living at a location is well correlated with the pricing of venues in the same neighbourhood and offered services around (for a summary of our data construction pipeline see Fig. 3.1).

In order to obtain a proper representation of socioeconomic status in the context of segregation, we consider 78 features from the ACS data. Although such a large number of dimensions in principle provides a rich way of quantifying the socioeconomic status of locations (and people living there), it turns out these variables have high redundancy. We perform a principal component analysis to identify the most relevant ones and find that income features (11 variables) have the largest loading, accounting for most of the socioeconomic variance between places. After implementing three different techniques (mutual information rank [99], decision tree [79], and Gini coefficient [160]), per capita income consistently stands out as the best indicator of individual SES: It accounts for the largest variance and it correlates strongly with other income variables such as earning/wage, wealth, and supplementary source of income (for more details on this analysis see Section 2.1.3).

3.2.3 Matrix measures

To decide if these patterns appear as the consequence of population statistics or other confounding effects, we compare the matrix $M_{i,j}$ to a reference matrix, which measures similar stratification patterns in a system where visiting patterns appear uniformly at random with certain constraints. This *randomised stratification matrix* is defined through a random rewiring process of the bipartite network, while constraining the total number and frequency of visits of each individual (i.e. their activity and link weights), the class of individuals and places, but fully randomising links between individuals and visited places otherwise. The randomisation is performed by selecting randomly for each link of an individual u a place to visit from the set of places ever visited by their respected socioeconomic classes c_u, while keeping the link weight intact. This *in-class randomisation* allows us to compare an individual's behaviour to similar others, meanwhile distinguishing between socioeconomic classes, which potentially are characterised by very different visiting patterns.

After generating randomised bipartite networks via 100 independent realisations, we compute a similar column-wise normalised stratification matrix $R_{i,j}$, representing the probability of people from class c_i randomly visiting to places of class c_p. To finally obtain whether the empirical mobility patterns appear more frequently than by chance,

we compare the empirical and the preference-based null model matrices. We obtain a *normalised stratification matrix* $N_{i,j}$ by taking the element-by-element fraction of the empirical and random contact matrices as:

$$N_{i,j} = \frac{M_{i,j}}{R_{i,j}}. \tag{3.1}$$

For each element in the matrix $N_{i,j}$, the presence of $N_{i,j} > 1$ implies that higher probability of the visits made by individuals from class $i \in C_U$ to place of class $j \in C_P$ are found in the empirical observations compared to expected values in the random null model. Contrastingly, as $N_{i,j} < 1$ appears, smaller probability than expected by chance holds for corresponding visits. In broader picture, the domination of $N_{i,j} > 1$ values anywhere in $N_{i,j}$ indicates patterns of socioeconomic stratification, where people prefer to visit places of similar socioeconomic status as their own, rather than places, which are richer or poorer than them.

These normalised stratification matrices reveal further characters of possible biases of people in choosing places to visit, out of their own class. If in a city people exhibit upward visiting biases, thus they tend to choose more expensive places to visit when they step out of their own class, the upper diagonal matrix elements of $N_{i,j}$ would appear dominantly red. While, if the opposite is true, the lower diagonal elements would reflect similar but downward visiting biases. To simply quantify these patterns, we compute the average values $N_{i,j}$ elements of normalised stratification matrix of cities above, at, and under their diagonals. To measure the level of segregation in mobility mixing, we analyse the earlier introduced normalised stratification matrix $N_{i,j}$ for each city. The Pearson correlation coefficient of matrix entries r_N is imposed on $N_{i,j}$ as previously defined in Section 2.3.1.

3.2.4 Bias measures

We take a technical step ahead in order to adequately quantify this visiting bias that indicates deviations in mixing from the respected c_u socioeconomic class of an individual. We compute a single *empirical individual bias score* for each individual $u \in U$ as

$$B_u = \langle c_p \rangle_u - c_u, \tag{3.2}$$

where $\langle c_p \rangle_u = \frac{\sum_{p \in P} w_{u,p} \times c_p}{n_p^u}$ is the average socioeconomic status of places an individual u visited, defined as the fraction of the $\sum_{p \in P} w_{u,p} \times c_p$ sum of socioeconomic status of places in the trajectory of individual u and the $n_p^u = \sum_{p \in P} w_{u,p}$ number of times

individual u visited any places. An individual has upward visiting bias if her individual score B_u is positive, meaning that she tends to visit places located in more affluent areas than where she lives. Secondly, an individual with negative score value has downward visiting bias since places she usually visits are situated in lower socioeconomic class than her own. Otherwise, an individual does not have any indication of bias ($B_u = 0$) if she visits places within her own socioeconomic rank. A reference model for this measure can follow a similar logic as the in-class randomisation for the realisations of network reference models explained before. Given the individual trajectory resulted from the random visit generating process, we calculated a *randomised individual bias score* using the same formula as in Eq. 3.2.4. Note that in this measure boundary effects may appear, as people from the poorest class cannot exhibit downward bias, and similarly, the highest class cannot be upward biased. Individual bias scores can be fairly compared to null models, which retain these boundary effects. In-class randomisation fulfils this requirement, providing an average randomised bias score $\langle B^{rand} \rangle_u$ for each individual separately. Note, that the randomised individual bias scores take non-trivial values, different from zero, due to the individual variance of visiting frequencies of individuals to different places. These are represented by the weights $w_{u,p}$ in the bipartite network, which are preserved during the randomisation process.

The comparison of the empirical and in-class normalised individual bias scores can be best quantified by an individual bias z-score as

$$z_u^{B_u} = \frac{B_u - \langle B^{rand} \rangle_u}{\sigma_u^{B^{rand}}}, \tag{3.3}$$

where $\langle B^{rand} \rangle_u$ is the mean and $\sigma_u^{B^{rand}}$ is the standard deviation of the randomised individual bias scores across 100 independent realisations of the null model. The value of $Z_u^{B_u}$ reflects how much the individual bias deviates from the expected bias for an individual who chooses places to visit with the same frequency as before but selects them from a given set of places dictated by others within the same socioeconomic class.

The individual bias score B_u compares the average class of visited places of an individual to its own socioeconomic rank inferred from its home location. Meanwhile, its z-score $z_u^{B_u}$ indicates if this individual bias is weaker or stronger than expected from random behaviour. However, this measure is using the class label of the individual as a reference of comparison, and it says less about whether an individual visits higher or lower class places as compared to the random expected behaviour characterising other individuals in its own class. To directly measure this effect we introduce a class level z-score measure

$$z_u^{c_u} = \frac{\langle c_p \rangle_u - \langle c_p \rangle_{c_u}^{rand}}{\sigma_{c_u}^{rand}}, \tag{3.4}$$

where $\langle c_p \rangle_u$ is the average socioeconomic status of places individual u visited, and $\langle c_p \rangle_{c_u}^{rand}$ and $\sigma_{c_u}^{rand}$ are the average and standard deviation (respectively) of class of places that others from class c_u would visit if behave randomly. This reference measure, just like before, is generated by in-class shuffling to obtain null models over 100 realisations. The value of $z_u^{c_u}$ reflects directly how much the individual behaviour deviates from the expected level, when the individual could choose randomly places to visit from a given set dictated by others from the same socioeconomic class.

3.2.5 Residential segregation measure

To address the effects of residential segregation on mobility mixing, we took a similar path than others [100; 192] and considered the ethnic group distribution in a city as a proxy. Residential segregation is indicated by housing clustering tendency of individuals from the same ethnic group. This can be formally quantified by the so-called distance decay isolation [129], which measures the probability that a racial group minority interacts with members of their own group by considering the distance from the racial group minority's housing area. This is measured as:

$$Dp_{xx*} = \sum_{i=1}^{n} \left(\frac{x_i}{X} \sum_{j=1}^{n} \frac{k_{ij} x_j}{t_j} \right), \tag{3.5}$$

where x_i and x_j are the population sizes of a minority group in census tracts i and j (respectively), $X = \sum_i x_i$ is the total population of the minority group, and t_j is the total population of census tract j. The distance decay dimension is reflected by $k_{ij} = \frac{t_j^{-d_{ij}}}{\sum_{j=1}^{n} t_j^{-d_{ij}}}$, where d_{ij} is the distance between the centroids of census tracts i and j. Hence, higher index suggests higher probability of interaction with people from the same group, inferring isolation from the rest of population. In our case, we use a probabilistic individual profiling to identify the most likely socioeconomic profile of an individual based on the ethnic group with the highest proportion at the respected census tract where one lives. For instance, if an individual u lives at census tract i where the racial composition there is 60% white, 15% Hispanic, 10% black, and 5% Asian, this individual is considered as white. We consider different thresholds at first and we find out that considering a neighbourhood the ethnicity if such people consist of at least the 30% of the given tract is the optimum cut-off because it is the highest threshold with the lowest unidentified census tract ethnicity profiles.

3.3 Results

Our main scientific goal is to study socioeconomic segregation and biases in population mixing in cities by observing correlation patterns between the SES of people and places they visit. Using the collected data, this objective can be addressed by building a network of individuals visiting places.

3.3.1 Mobility stratification

Empirical stratification matrices $M_{i,j}$ are shown in Fig. 3.2 for selected cities (Houston, New York and San Diego). The dominant diagonal elements for Houston and San Diego indicate strongly stratified visiting patterns in these cities. People prefer to visit places of their own or similar socioeconomic class, rather than places from remote classes. Interestingly, for New York this pattern is less evident suggesting weaker socioeconomic preferences in visiting venues.

Each element in the normalised stratification matrices $N_{i,j}$, which are $N_{i,j} > 1$ (red bins in Fig. 3.2d-f) indicates that the visits made by individuals from class $i \in C_U$ to place of class $j \in C_P$ appeared with higher probability in the empirical observations than it was expected from the random null model. Otherwise, the blue blocks for $N_{i,j} < 1$ show that the corresponding visits appeared with a smaller probability than expected by chance. In cases red bins dominate the diagonal of the normalised matrix $N_{i,j}$, it indicates patterns of socioeconomic stratification, where people prefer to visit places of similar socioeconomic status as their own, rather than places, which are richer or poorer than them. This is the case of Houston and San Diego (see Fig. 3.2d and f respectively) and many other cities listed in Section 7.2.2. However, this character is less evident for New York (see Fig. 3.2e), where despite known strong residential segregation, the city fabric mitigates a more homogeneous mixing of people.

From Fig. 3.2g it is clear that, in all cities, diagonal elements dominantly concentrate visiting probabilities. However, in terms of off-diagonal averages, in most of the cities (like in Houston in Fig. 3.2d) the upper diagonal average takes a larger value as compared to the lower diagonal average, indicating present upward visiting biases in these metropolitan areas. Meanwhile, in some cases the contrary is true (like in San Diego in Fig. 3.2f) or in some cities these averages are very similar thus indicating no dominant upward or downward visiting biases, as in case of New York (see Fig. 3.2e and Fig. 3.2g).

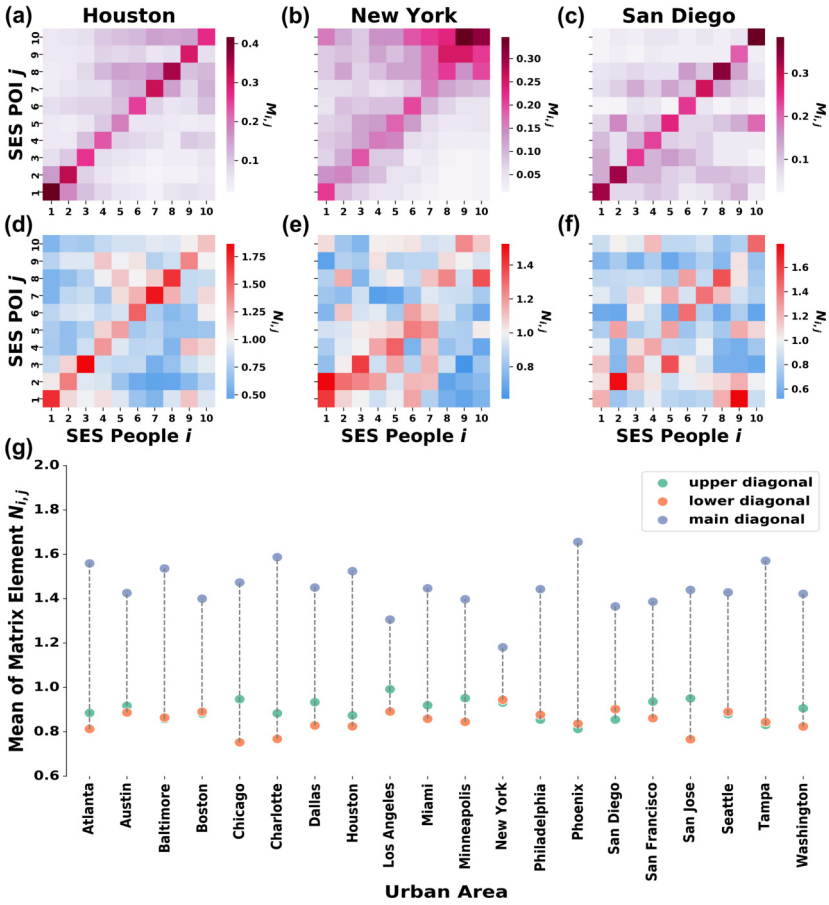

Figure 3.2: Socioeconomic stratification matrices. **(top)** Empirical stratification matrices $M_{i,j}$, showing the probabilities that individuals from a given class visit to places of different classes. The darker colour shades of bins represent larger visiting probability. Matrices of Houston (Fig. 3.2a), New York (Fig. 3.2b) and San Diego (Fig. 3.2c) all show strong stratification patterns, indicating that people tend to visit most likely places with similar status. The normalised stratification matrices $N_{i,j}$, defined as the fraction of the empirical and randomised stratification matrices. After normalisation, such stratification pattern becomes less evident for New York (Fig. 3.2e) and San Diego (Fig. 3.2f) but quite persistent in Houston (Fig. 3.2d). Similar matrices computed for other urban areas are available in Section 7.1.4. **(bottom)** Mean of matrix element $N_{i,j}$, computed separately for the upper, lower, and main diagonals. Among 20 urban areas, 12 of them (including Houston) have higher mean values for upper diagonal elements, indicating dominant upward visiting biases. In contrast, we see dominant downward visiting biases in San Diego, while mean values of upper and lower diagonal elements are almost indistinguishable in New York (respectively 0.932 and 0.945).

3.3.2 Mobility biases

The matrix measures presented in Fig. 3.2 reflect the coexistent socioeconomic configurations derived from visit trajectories. Firstly, the empirical stratification matrices

$M_{i,j}$ bring an initial indication of homophily mixing as seen in the dominant frequency visit within own class. Secondly, these results reveal the underlying inclination in visiting places situated in higher SES as depicted by the larger proportion of upper diagonal elements in the normalised stratification matrices $N_{i,j}$ in most of the cities. Taking these two configurations into account, it can be inferred that while individual mobility is dictated by the membership of socioeconomic class most of the time, the embedded motivation to visit upper class places is still present.

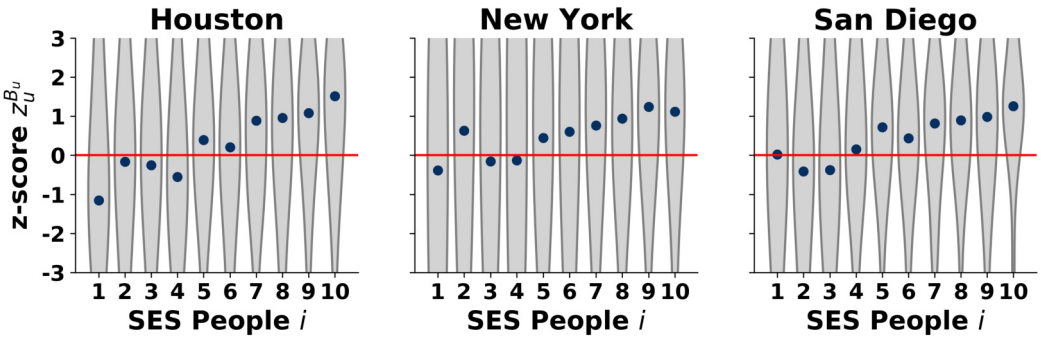

Figure 3.3: Individual Bias z-score $z_u^{B_u}$. Class level distributions and their median values are shown for each socioeconomic class in Houston (Fig. 3.3a), New York (Fig. 3.3b) and San Diego (Fig. 3.3c). The overall increasing trend of medians (blue dots) indicates that people from lower classes are less biased than expected, while the contrary is true for others from higher classes. Solid red line indicates the fully unbiased case. For results on other cities see Section 7.1.6 and Fig. 7.7.

The class distributions of individual z-scores together with their median values are shown in Fig. 3.3, where the unbiased level is assigned as a flat red line. These distributions appear broad for each class, indicating that actually people from any class exhibit upward or downward biases in terms of their visiting patterns to other socioeconomic classes. Interestingly, the median z-scores indicate an increasing trend in all the three depicted cities. The people from lower classes appear with slightly negative bias z-score, meaning they have a slightly weaker bias to visit places of different socioeconomic classes than expected from their random visiting patterns. In contrary, the middle and upper classes are evidently biased stronger than expected. This increasing trend of the median of the individual bias z-score with socioeconomic classes surprisingly characterises all the investigated cities as shown in Section 7.1.6 Fig. 7.7 in the SM.

Results in Fig. 3.4 show a different behaviour as compared to the individual bias scores. In case of New York (see Fig. 3.4b), the distributions of the class level bias z-scores indicate that, although the variation is large in each socioeconomic classes, the medians of these distributions are all slightly positive and independent of the socioe-

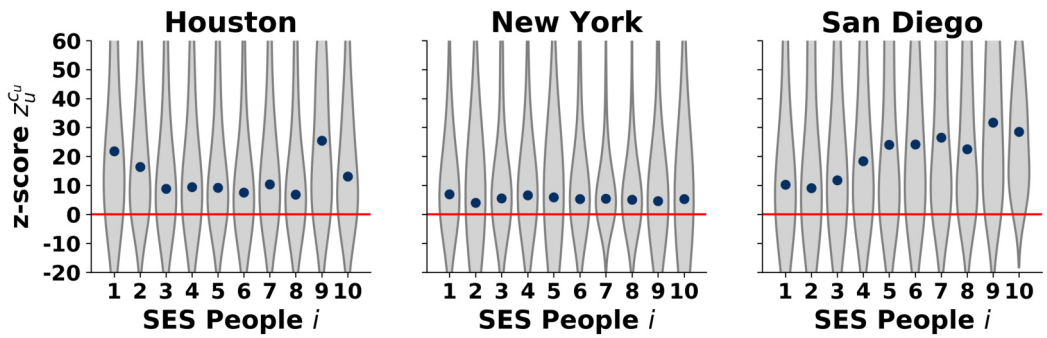

Figure 3.4: Class-level Bias z-score $z_u^{c_u}$. Distribution of class-level biased z-scores as the function of socioeconomic classes. Distributions are shown for each socioeconomic class with their median values as blue points for Houston (Fig. 3.4a), New York (Fig. 3.4b), and San Diego (Fig. 3.4c). Z-score values corresponding to unbiased cases are shown with red solid lines. Positive z-score values signal an upward visiting bias characterising each city. For results on other cities see Section 7.1.7 and Fig. 7.8 in the SM.

conomic class. This signals a weak upward bias in people's visiting patterns in New York as compared to the class behaviour that appears for each class. In other cities, we find several other bias patterns during our analysis (see Section 7.1.7 Fig. 7.8). In case of San Diego (in Fig. 3.4c) class-level biases are all positive and evidently increasing with the socioeconomic classes. This suggests that richer people in San Diego may visit even more affluent places, than one would expect from their random class behaviour. Somewhat the opposite trend can be observed for Houston (Fig. 3.4a), where although the class-level bias z-score is always positive and indicates upward bias for each class, it seems to follow an overall decreasing trend.

Visiting patterns measured by the class-level bias scores suggest that an upward socioeconomic bias characterise each cities we study. Although these measures incorporate the visiting frequency distribution of individuals, they do not show evidently that upward biases typically appear due to repeated visits to places with higher class scores, or due to several occasional visits to places out of ones socioeconomic classes. To answer this question, we recompute the median bias scores, excluding places which were visited less number of times by an individual than a given threshold. Results are depicted in Fig. 3.5 for each city. As expected, the median class-level bias score appears as a decreasing function of the frequency threshold in each city. This suggests that people visit more frequently places, which are closer in terms of socioeconomic status to their own class, while visit more affluent places occasionally only, that in turn causes upward bias patterns characterising their class. Beyond this general decreasing character, this function indicates large variance between different cities. For example, in case of San Diego (red line in Fig. 3.5), this curve starts from a high z-score value when

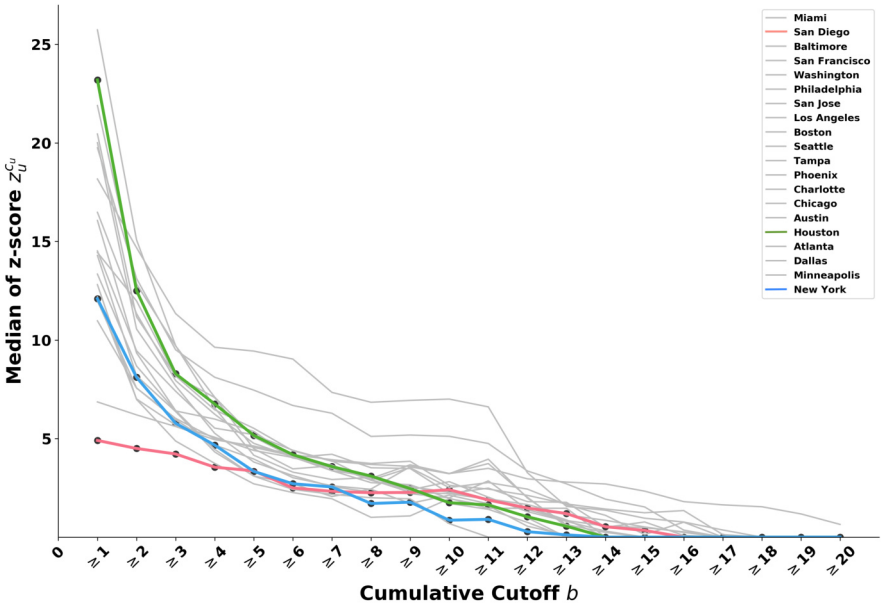

Figure 3.5: Sensitivity of class-level bias z-score z_u^{cu}. Lower bound cutoff is set as $b \geq 1$ to which we only take into account venues visited at least once. For each set of venues in individual trajectory cumulatively visited b times or higher, we measure class-level bias z-score z_u^{cu} and take the median values. Upper bound cutoff $b \geq 20$ is added to accommodate venues visited even more frequently. As b incrementally becomes larger, the medians are largely dropped closer to 0. It indicates that venues visited more frequently tends to be more homogeneous in term of mixing and closer to own socioeconomic status.

all visits are consider but decreases rapidly as repeated visits are taken into account. In case of New York this function starts from a relatively small z-score values and decrease linearly for larger threshold values. This suggests a different visiting behaviour where people typically visit places more than one time, but closer to their own socioeconomic class.

We prefer this particular measure over the one on individual bias, as our objective here is to reveal the source of upward bias, whether it is driven by repeated visits to places with higher class scores, or due to several occasional visits to places out of one's socioeconomic class. The class-level bias z-score serves this purpose as it already incorporates the visiting frequency distribution of individuals compared to their own class and gives positive z-score values.

3.3.3 Mobility mixing and segregated residences

While there is an expected relation between the mobility mixing patterns and residential segregation in a city, the combined investigation of these phenomena has not received much attention so far. Their relation is important however for several reasons.

For example, due to the multitude correlated socioeconomic factors it is likely that e.g. ethnicity, which strongly correlates with income status in US metropolitan areas, correlates also with residential segregation, as it has been shown in several studies [86; 1]. On the other hand, the daily mobility of people and their visiting patterns to different places are constrained also by these socioeconomic factors, thus they are likely to resemble similar segregation patterns. To investigate these correlations, we focus on different ethnic groups and the likelihood of their mixing in cities, which exhibits different level of mobility segregation patterns.

The mixing patterns in a city may not be only determined by the socioeconomic status of people but also by residential segregation. Residential segregation is strongly correlated with the ethnicity of people [121; 111; 1], which in turn, according to Wang and others [192], is an even stronger predictor of mobility mixing than socioeconomic status when it turns to black, Hispanic, and white poor and non-poor populations. This study found that the minority groups - despite their socioeconomic status - have lower exposure to richer or white neighbourhoods, comparing to poor white groups. The fact that they travel across similar distance and frequency to many places, does not change the persistent pattern of their isolation and segregation. Therefore, racial segregation emerges from a higher-order level, not limited to their residential neighbourhood but expanded to their mobility and potential contact.

Recalling the above mentioned diagonality index and individual bias, we take the average of z-score of each of these bias measures at the level of ethnic groups in every urban area and correlate them with their distance decay isolation value computed for the same ethnic group in the same city. By considering four ethnic groups (White, Hispanic, Black and Asian) we receive four data points for each cities as shown in Fig. 3.6. Although the total number of analysed individuals are not proportional to the total population of each city, the in-city fraction of different ethnic groups are similar to the census distributions.

There is a striking correlation emerging between the diagonality index (quantifying assortativity mobility mixing of each ethnic groups) and the distance decay isolation (measuring the isolation of different ethnic groups) with $R = 0.35$ ($p = 0.0$). Notably, almost all diagonality index measures appear with positive values suggesting assortative mixing for most ethnic groups, with a few exceptions. Further, the overall correlation suggests the intuitive picture that the stronger mobility mixing stratification patterns characterises a city (i.e. larger its corresponding diagonality index), the stronger isolation patterns emerge between its ethnic groups. In turn, it indicates that residential segregation (and thus physical proximity) play an important role in determining visiting and mixing patterns of people in a city. More interestingly, the de-coupled ethnic

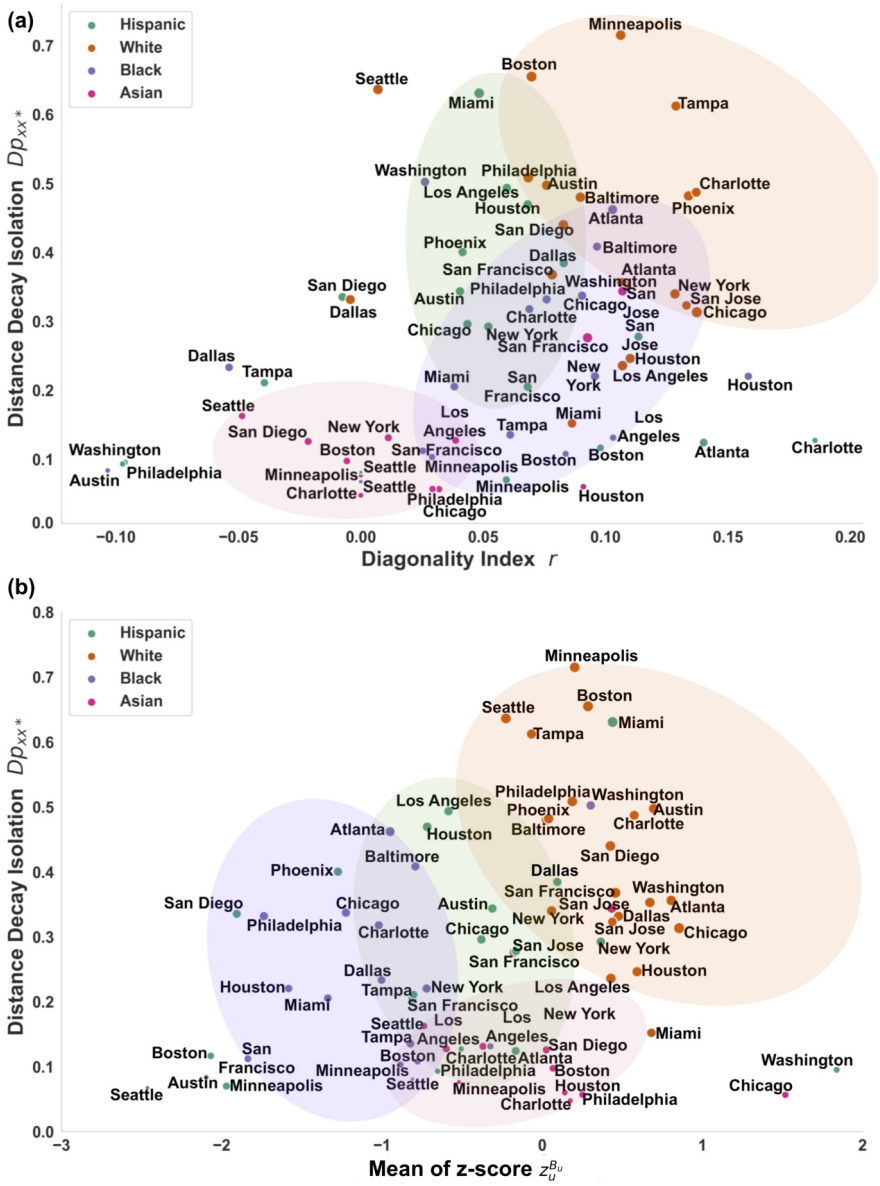

Figure 3.6: Segregation and bias measure correlations with isolation scores for different ethnic groups. Panel (a) depicts the correlation between the diagonality index r_N and distance decay isolation Dp_{xx*} while panel (b) show a similar correlation of the average individual bias z-score $z_u^{B_u}$. In each plot colours of symbols and blobs indicate ethnic groups of Hispanic (green), White (green), Black (purple) and Asian (red) people. The sizes of symbols are scaled with the size of these ethnic population identified in the Foursquare dataset in each city. Blobs with respected colour illustrate the cluster formation based on racial groupings. The shape is arbitrary, only to demonstrate the visibility of clusters.

groups for each city show an emerging clustering, which assigns the importance of racial differences in mobility segregation. From Fig. 3.6a it appears that people belonging to the white ethnic group (shown as orange points in Fig. 3.6a) appear to be the most isolated from the rest of the population (with the largest values of Dp_{xx*}), while

they appear with the strongest assortative mobility segregation patterns too (with the highest diagonality indices) consistently in several cities (to lead the eye we coloured this group as an orange blob in Fig. 3.6a). The contrary is true for the members of the Asian ethnic groups (indicated by red points and blob in Fig. 3.6a). In most cities they appear as the least isolated and the most dis-assortative (least stratified) ethnic group, thus mixing well with the rest of the population. In between these two groups, people from the Hispanic ethnic group (green points and blob) seem to be more isolated than people from the black ethnic group (purple points and blob) although they show comparable strength of segregation in mobility mixing, all weaker than white people.

Needless to say that the grouping patterns shown in Fig. 3.6a indicate overall trends only, while several exception exists for each ethnic group. For example, the Hispanic ethnic group of Charlotte appears with the strongest assortative pattern, although this group is not strongly isolated from the rest of the population. Or the black community of Austin appears with the lowest diagonality index, suggesting a strong dis-assortative mixing of these people with the rest of the population, while they also appear as one of the least isolated among any other communities.

A similar positive correlation appears in Fig. 3.6b with R=0.25 ($p = 0.04$) between the average individual bias z-score and distance decay isolation values over all the investigated ethnic groups and cities. Moreover, ethnic groups show certain clustering trends, which suggest ethnic trends in terms of visiting bias patterns. Interestingly, white ethnic groups (shown by orange points and blob), who we have already found the most isolated, show the strongest upward bias to visit more affluent places then their own socioeconomic class. As high SES classes are populated mostly by white people, this pattern derives from our earlier observations in Fig. 3.3, where we find upward bias to increase with the SES of people. Their high isolation score can be explained by their upward bias towards higher socioeconomic places, which are most likely to be visited by other white people. Meanwhile, it may also indicate that our data have an over-represented white population, as we find upward biases in all of the cities. Strikingly, other racial groups indicate negative visiting biases and lover level of isolation. This effect is the strongest for people from black racial groups all over the country, but also characterises Hispanic and Asian communities although they show more unbiased patterns, with average individual z-score values closer to 0.

Exceptions are again interesting. The Hispanic community of Washington appears as the most upward biased ethnic group, while the black ethnic group of Seattle sits on the other end of the spectrum and being the most downward biased minority among the analysed cities. Both of these communities appear with low level of isolation. Consequently, similar to the conclusion of Wang et al. [192], we observe that beyond socioe-

conomic status, ethnicity (strongly correlated with residential segregation) is another very important factor determining mixing patterns of people.

3.4 Discussion and conclusions

Mobility patterns are strongly determined not only by the fabric of a city but also by the socioeconomic structure of the population living there. This leads to biased mixing and segregation in mobility, which can be observed as stratification patterns in choices to visit places. We have addressed this complex phenomenon via a mobility analysis of people living in the 20 largest cities in the US, and aimed to quantify segregation patterns in mobility capturing their visit patterns to places of interest. We systematically found upward-biased mobility in all cities, with some variance across metropolitan areas. In one extreme, people living in New York do not exhibit dramatic stratification in their visit patterns but visit places in all kinds of locations, rich or poor, independently of their own socioeconomic status. Meanwhile, in Houston and San Diego people are more stratified and visit places of their own socioeconomic class, and show an upward bias towards richer places to visit. We found that this upward bias, which characterises most cities analysed, is usually induced by single visits of individuals to affluent places, while most visits correspond to their own socioeconomic class. We also revealed distinct patterns of individual mobility in terms of stratified correlations between the bias magnitude and residential segregation based on spatial distribution of racial groups in urban areas. Visual representations of ethnic clusters indicate overall trends of behaviour characterising most studied cities, where segregated mobility is bounded together with residential segregation and broadly contributes to the portrayal of inequality.

It should be taken into account that data for a given socioeconomic class in the population might not be comparable across cities due to sampling in the data collection process. Particularly, Foursquare data over-represents wealthy classes when compared to the underlying population. To understand better the fluctuations of the distribution of SES due to the representativeness of the used dataset, we designed a bootstrapping method (see Section 7.1.1). Bootstrapping results suggest that the SES distribution of Foursquare users is sufficiently similar to the SES distribution of the real population.

Multiple sources of data containing digital traces of human movements with higher resolution, such as mobile phone call records and GPS trajectories, may improve the robustness of findings presented in this paper. Methodological improvements to infer individual attributes (like racial group membership) provide a direction of future research. Moreover, algorithms for probabilistic individual profiling could be developed

by using machine learning techniques such as Random Forests and Support Vector Machines in the presence of ground truth information from alternative data sources.

One potential confounding factor of the emergent stratification patterns reported here is distance, as people visit places closer to their home more frequently, thus inducing similar correlation motifs. To check the robustness of our methods in investigating segregation in mobility and biased visiting patterns and the magnitude of such distance effect, we recomputed our results on out-of-class data after excluding own census tract visits for each individual's trajectory. Even with this constraint, SES plays a considerable role in shaping mobility (comparative observations for all cities can be found in Section 7.1.4-7.1.5 and Section 7.1.8-7.1.9, along with Section 7.1.12 Fig. 7.13 in the case of Houston, New York, and San Diego, in contrast to Fig.3.2 above). On the ground of visiting biases, there are some variations among cities regarding individual bias. The earlier notion of upward visiting bias is also very much present in the case of out-of-class measurements since z-score values are all positive above the red median unbiased line (complete plots are available in Section 7.1.6-7.1.7 and Section 7.1.10-7.1.11, while a deeper exploration for Houston, New York, and San Diego is available in Section 7.1.12 Fig. 7.14 and Fig. 7.15). Therefore, there are no conflicting results from our methodology even after controlling for this confounding factor. Enforcing out-of-class treatment is reasonable in this context because our study aims to analyse and quantify mixing patterns and not yet look for causal links or underlying reasons of their emergence.

Segregation is not an exclusive phenomenon to the quasi-static configuration of housing settlement, but also exists in more dynamic settings such as mobility. Questions about the conceptual relations between segregated mobility and segregated residence stand still in the literature, yet relatively untapped, while scientific investigations should follow this line of inquiry. We take a step forward through empirical data-driven analysis and yield an interaction effect between both types of segregation. Individual attributes (such as racial groups) partly explain the emergence of distinct clusters, beyond income levels. Our findings also highlight the notion that inequality is multidimensional in nature. A comprehensive policy design to address this issue should entail the wider possibility of individual movement across the urban landscape to accommodate larger socioeconomic heterophily and further interaction between socioeconomic classes.

Chapter 4

Mobility segregation dynamics during pandemic interventions

COVID-19 outbreak presents as the embodiment of external shock that suddenly disturbs the typical mobility pattern in urban areas. Everyone faces the need to reorganise their daily errands throughout space, but the magnitude of changes might not be identical for everyone across socioeconomic classes. This given condition raises concern on the emergence of further detrimental effect it potentially brings to inequality as the pandemic prolongs. In this study, we examine how COVID-19 outbreak and the restrictions that follow induce variability of mobility adjustment in terms of preferential mixing in socioeconomic and physical space. Anonymised and privacy-preserved mobility data in global cities namely Bogota, Jakarta, London, and New York, coupled with socioeconomic reference from Central Bureau of Statistics, are used to capture inequality in mobility among classes along temporal transitions, passing through before, during, and reopening phases. We find that the first lockdown induced considerable increase in mobility segregation, but the attempt to loosening mobility restriction did not necessarily diminish isolation within own neighbourhood, indicating that recovery is not fully made. Persistently lower heterogeneity in both locations and socioeconomic status of places visited by individual is also visible. Finally, we address the question, which interventions contributed the strongest to the change of segregation in different intervention periods. Our results highlight population level dynamical segregation phenomena observed at the individual level, that provides important conclusions for better policy design with more equal consequences among people from all socioeconomic classes.

4.1 Introduction

Inequality has always been a prominent feature of today's society. Inadequate distribution and access of resources, among others, stand as a preliminary setting. Untangled paths to income [113], education [181], and employment [37] are moulded into behavioural preference in daily life. Van Hamm et.al. [191] argue that rising inequality becomes more prevalent anywhere across the globe nowadays and it largely contributes to socioeconomic segregation especially in cities where heterogeneity in terms of income, education, and employment is the highest. The convergence of global trends is due to the fact that cities in lower-income countries have higher levels of inequality and segregation, but inequality and segregation are growing faster in high-income countries. It reveals the complex entanglement carried out by inequality because eventually socioeconomic segregation is none other than a spatial footprint of inequality itself.

Socioeconomic segregation is not the only factor that is linked to inequality. Beyond that, there are numbers of segregation type, namely residential segregation and employment segregation, to mention a few. Residential segregation is manifested as separation of different groups of people into different neighbourhoods within a city. At the same time, income could be a driving force behind the emergence of segregation in housing. Selective residential mobility flows, for instance, happen due to the quality of neighbourhoods moving farther away from each other and result in the highly segmented profile on the two extreme: downgraded neighbourhoods where people with low income live and upgraded neighbourhoods where high income people stay [133; 185]. Therefore, housing plays an intermediary role in reproducing inequality through the coupling effect between income inequality and residential segregation [185]. On the other spectrum, employment segregation entails changing occupational structure that is mainly characterised by professionalisation and socioeconomic upgrading. Growing proportion of high-income segment among workforce nonetheless increases demand for residential units located in inner city neighbourhoods, motivated by the centrality and accessibility of urban living [35; 190].

The preference on residential areas and employment type emulate mobility pattern in order to meet daily errands. An interplay between inequality and the way people organise their mobility in urban space is inevitable. In line with Urry [188], Olvera et al. [143] define inequality in mobility as behavioural differences in the level of transport use due to differences in the distribution of monetary ownership such as income or wealth. Furthermore, they find that car ownership is a strong determinant to mobility pattern and residential locations and diminishes potential interaction with people with heterogeneous backgrounds (as to compared with shared space in public transporta-

tion). Entitlement to more affluent socioeconomic status allows people to buy car and live in the outskirt of city since distance is no longer an issue and trip could be made promptly with higher degree of flexibility. As a result, segregation pattern comes out as an entanglement between inequality and mobility.

To clarify the conceptual interconnectedness between patterns of inequality and mobilities, Ohnmacht et al. [142] propose three separated mechanisms namely: inequality shapes mobility, mobility shapes inequality, and reciprocal linkage between inequality and mobility. The first mechanism works through social strata dictated by monetary ownership. Individual mobility, therefore, is determined by the ability to move in time and space in the presence of mobility capital (e.g: economic capacity). Unequal access to mobility capital ultimately marks different behavioural characteristics in mobility. The second mechanism emphasises on structural societal changes that require higher mobility intensity. Working households (e.g: mobility for work purpose), spatial dispersion in social networks (e.g: mobility for visiting friend) and perception as well as preference towards transportation mode (e.g: bus vs car). Consequently, unmet needs in mobility could induce deeper stratification that later translated into inequality. Other than that, inequality and mobility could be simultaneously interacting, pinned up by Kaufmann et al. [94] as motility. This terminology refers to the existence of capacity to move, including relational linkage to inequality and mobility needs. To an extent, motility may produce spatialisation even spatial isolation as unfulfilled mobility spatially bounds individual exploration in space. Residential relocation, for instance, is driven by motives to gain social and economic benefits due to better job prospects.

In urban mobility network, social stratification in conjunction with unequal access to transport infrastructures brings social exclusion [116; 156] and social segregation [202; 126]. External shocks such as COVID-19 outbreak commencing from outside the system can be catastrophic to socioeconomic configuration and the related aspects and perpetuate individual mobility that is already constrained by socioeconomic stratification [58; 39; 112]. For some people, their capacities to adjust preferences and way of living in response to disruption are limited by their well-being. People from lower income class might have narrow flexibility and keep doing the usual means during lockdown period as seen by regularity in mobility since their works demand on-site physical presence. After all, the impact of the pandemic on mobility segregation specifically, would likely be multifaceted and would depend on a variety of factors, including the specific characteristics of the people affected, the severity and duration of the pandemic, and the specific policies and responses implemented. Existing literature suggests that higher income is associated with larger mobility reduction, while mobility inflexibility and less social distancing are observable among low-income, raising disparity in

mobility [195; 60; 125].

The fact that social fabric and inequality feedbacks are recursively produced and reproduced by mobilities [188; 94] justifies our approach using mobility as an operational concept to analyse socioeconomic stratification and spatial isolation brought by the pandemic. This research investigates the impact of COVID-19 outbreak and the non-pharmaceuticals interventions (NPI) that later follows in some urban areas across the globe namely Bogota, Jakarta, London, and New York. Our ultimate goal is to analyse the size of changing residual isolation and segregation pattern in mobility due to external shock. It is also in our interest to observe whether such phenomenon is temporary caused by timely restrictions such as lockdown or rather become a long term effects.

To test this, firstly, we capture the changing segregation pattern by quantifying mobility stratification in every sequence of pandemic periods. Secondly, we empirically points out behavioural effects of spatial and socioeconomic exploration in mobility by computing entropy measures derived from spatial and socioeconomic property of visited places. Moreover, we identify types of intervention contributing to aforementioned behavioural effects and their impacts on mobility segregation. Interestingly, these procedures lead us to the still presence of residual effect of shock even after the removal of mobility restrictions.

4.2 Data and research methods

4.2.1 Data description

The first part of data collection is dedicated to mobility data from CUEBIQ dataset [50]. The dataset contains geolocation of places visited by anonymous smartphone users along with timestamps. Time period starts from January 2020 with last day of observation that varies between cities. Given the time gap among them, we assure that the time window still adequately covers temporal transition during pandemics namely before lockdown, lockdown, and reopening as presented in Supplementary Material (SM) Section 7.2.1. In total, we successfully identify 597,000 home locations of people with different sample sizes between cities. Home location is defined as the most frequent location visited by each individual during the night time (between 9PM to 6AM). Details of dataset coverage and home inference algorithm is specified respectively in Section 4.2.1 and Section 4.2.2. To check the general reproducibility of mobility pattern in New York, we also use SafeGraph dataset [93] which is available at coarser resolution (census tract level) and longer temporal coverage (until May 2021).

We overlay socioeconomic layer on top of the existing mobility layer. Income related features are fitted for this purpose. In Bogota, multidimensional poverty index [51] at urban section developed by Colombian bureau of statistics (DANE) becomes the basis for socioeconomic status computation. It captures quite comprehensive dimension of individual well-being: health, education, utilities and housing, employment, as well as childhood and youth condition. A simpler version of poverty index called poverty rate [28] is used in Jakarta at village-level resolution, taking the proportion of people living below particular amount of average monthly income. Meanwhile, socioeconomic configuration of London and New York is plotted respectively based on total annual income recorded by Office for National Statistics (ONS) [144] in 2015 at middle layer super output area (MSOA) level and per capita income in 2018 at census tract level taken from American Community Survey (ACS) [34]. We group the people by income distribution in the dataset into 10 equally populated groups from the lowest SES/poorest (1) to highest SES/riches (10). It should be taken into account that the direct comparison between cities could not be fully established because of diverse characterisation by nonidentical SES indicators. Therefore, comparison across period in the same city is more visible to derive in this context.

To synchronise the movement along mobility points and to derive observable structural break in mobility pattern induced by the epidemiological outbreak and policies coming after, we refer to the stringency index on Oxford COVID-19 Government Response Tracker (OxCGRT) dataset [77]. It contains a broad-spectrum of COVID-19 related policy responses taken daily by the government worldwide and measures the strength of strictness-degree of such mobility restriction. Stringency index consists of containment and closure policies (closings of schools and universities/C1; closings of workplaces/C2; cancelling public events/C3; limits on gatherings/C4; closing of public transport/C5; orders to stay-at-home/C6; restrictions on movement between cities/regions/C7; and restrictions on international travel/C8) complimented with health system policy (presence of public information campaigns/H1). We validate this with actual implementation at city level to ensure policy alignment between national and local government.

Timestamp for each recorded movement opens up possibility to identify residual effect of periodical intervention, such as the diversity of non-pharmaceutical intervention (NPI) taken by the government during the COVID-19 pandemics as reflected in the changing mobility patterns. On top of that, the degree of socioeconomic leaning in mobility pattern is also discernible, making it visible to measure the difference in mobility.

4.2.2 Pipeline description

Human motion exhibits a generic and reproducible pattern formulated as a function of temporal and spatial regularity [73]. The distribution of individual trajectories is spatially skewed to a few highly frequented locations with home or work taking largest proportion. With this significant regularity in mind, we build an integrated pipeline to refine individual mobility pattern. In our pipeline, data that have been processed will be used to construct *mobility network*.

We follow procedure described in Section 7.1.2 to infer home location and construct an algorithm to detect POI (non-home) locations as seen in Fig. 4.1. Our methodology combines the spatial and temporal attributes such as frequency of visit, time window of visit, as well as duration of stay at given locations. We take a further step to infer socioeconomic status for each people (based on home location and POI) by performing spatial projection and merge it with demographic data (average income) from bureau of statistics.

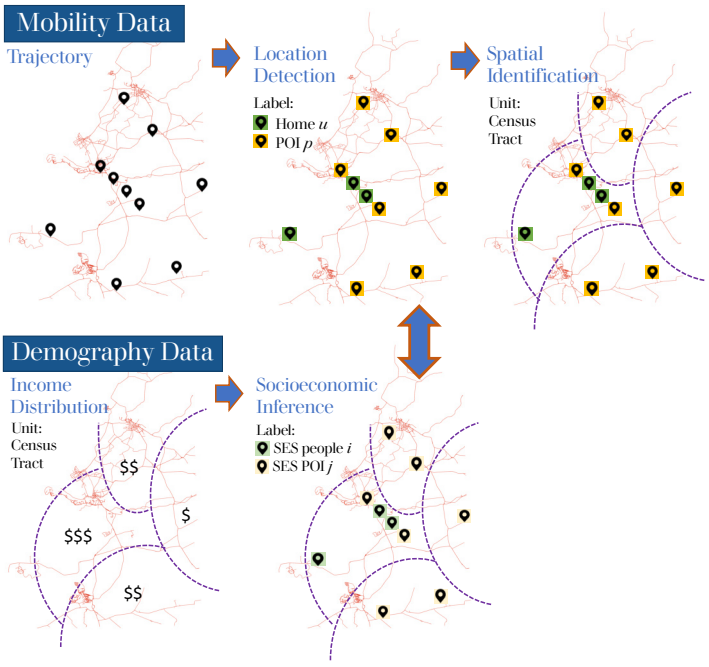

Figure 4.1: Inference Algorithm. Mobility data contains information regarding whereabouts of people namely geographic locations and timestamp (trajectory). Demographic data covers average income of given spatial unit (eg: census tract). We build an algorithm to separate home u and POI locations p and identify the inferred income based on its spatial delineation. Discretisation on distribution of inferred income results in two separated SES label: SES People i and SES POI j.

POI Location: Apart from home, human individual activities evolve around other

areas for some reasons, including work. Trip between home and work location dominates daily mobility, while visits to other locations are broadly distributed with short inter-event times [171]. We set criteria for POI location as place other than home where people with identified home locations are present during weekdays from 9AM until 3PM. Afterwards, the rest of locations that do not fall into either home or work category are labelled as others.

Socioeconomic Status (SES): We assign SES label to every individual and and POI based on socioeconomic data previously discusses in Section 4.2.1. The first step to SES people is to identify socioeconomic feature of area where they live (home location). Similarly, SES POI is inferred by mapping out the area where points (work and other locations) are spatially positioned. We sorted the values by ascending order and split them into equally populated bins of 10 SES labels, making SES 1 to be the poorest and SES 10 to be the richest.

4.2.3 Mobility matrix

In Section 2.3.1, we mention the basic formulation of stratification used in the *mobility stratification matrix* $M_{i,j}$ is based on *mobility network* $G = (U, P, E)$. Given a pair of *mobility stratification matrix* $M_{i,j}$ in two consecutive periods, we initialise *mobility adjustment matrix* $S_{i,j}$ is set up where the matrix element b_{ij} entails the difference in proportion of frequency visits. More formally:

$$S_{i,j} = M_{i,j}^{t1} - M_{i,j}^{t2}, \tag{4.1}$$

where $t1$ denotes the initial period and $t2$ is the succeeding rolling period. For instance, if we have 3 periods namely Before Lockdown (BL), Lockdown (L1) and Reopening (R1), we could generate three $S_{i,j}$ respectively:

$$S_{i,j}^{BL-L1} = M_{i,j}^{BL} - M_{i,j}^{L1}, \tag{4.2}$$

$$S_{i,j}^{L1-R1} = M_{i,j}^{L1} - M_{i,j}^{R1}, \tag{4.3}$$

$$S_{i,j}^{BL-R1} = M_{i,j}^{BL} - M_{i,j}^{R1}, \tag{4.4}$$

while $S_{i,j}^{BL-R1}$ shows the difference between period before enforcement of lockdown and reopening (removal some mobility restrictions in the post-lockdown).

Moreover, the degree of socioeconomic isolation is presented by assortativity mea-

sure of mobility stratification matrix r. Values closer to 1 signal the higher concentration of visiting venues within own socioeconomic range, while lower cutoff values at -1 reveals the tendency of visiting places outside own class. If the value is equal to 0, the dispersion in visiting pattern throughout classes without any structural choice preference regarding socioeconomic status of places.

4.2.4 Ranking of non-pharmaceutical intervention

To rule out the effectiveness of each type of restrictions, we initiate univariate linear regression model. There are 9 restrictions k listed as NPI respectively closings of schools and universities (C1), closings of workplaces (C2), cancelling public events (C3), limits on gatherings (C4), closing of public transport (C5), orders to stay-at-home (C6), restrictions on movement between cities/regions (C7), restrictions on international travel(C8) and presence of public information campaigns (H1). Stringency value S for every restriction in each temporal snapshots is obtained from OxCGRT dataset and to be used as independent variable. The dependent variable is two types of mobility entropy, being computed separately: geographic space-based $Hm(X)$ and socioeconomic space-based $Hs(X)$. If a city passes three temporal snapshots t (eg: BL-L1-R1), we will build upon 3 periods-by-9 restrictions that results in 27 univariate linear regression models.

To further understand the impact magnitude of a single restriction $k \in K$ at period $t \in T$, we fit the data to this form:

$$Hm(X)^t \sim S_k^t, \tag{4.5}$$

and

$$Hs(X)^t \sim S_k^t. \tag{4.6}$$

Given the configuration, each model gives $R-squared$ that is used as basis of ranking after sorting from the highest value. As a statistical measure indicating goodness-of-fit, $R - squared$ shows the extent variation of a dependent variable (spatial or ses mobility entropy) could be explained by the independent variable (a single restriction). Therefore, we base our framework on this approach.

4.3 Results

Organisation of daily mobility pattern is affected by dynamics of socioeconomic fabric. Spatial distribution of commercial areas, residential units, workplaces, and schools, among others, encourages people to move from a point to many different places across urban landscape. Built up on the notion of unequal distribution at individual level, mobility is also engendered and reinforced by inequality [142]. The presence of individual preferences over socioeconomic characteristics of places could be further signified at the class level by taking the visit ratio of people coming from particular class to places distributed in various other classes.

We set our scientific agenda to study the impact of COVID-19 outbreaks, including imposed restrictions that come after, on the way people make adjustments in their daily mobilities. We introduce two conceptual frameworks respectively *induced assortativity* and *residual isolation*. To operationalise the aforementioned concepts, we take the strategy previously proposed [109; 85] to indicate the existence of socioeconomic assortativity in visiting patterns.

4.3.1 Changing segregation pattern

Assortativity may lead to segregation in mobility where people of different socioeconomic characters meet less likely than with own counterparts in the same socioeconomic level. We take the first step to capture stratification tendency by transforming *mobility network* into *mobility stratification matrix* $M_{i,j}$ as defined earlier in Section 2.3.1. As a result, mobility stratification in each period is summarised in a single matrix. To standardise the assortativity measure for the sake of comparability and reproducibility, we compute correlation coefficient r between $i \in c_u$ and $j \in c_p$ as motivated by assortativity coefficient used in number of works [139; 59; 25]. Values closer to 1 signal the higher concentration of visiting venues within own socioeconomic range (assortative mobility), while 0 pinpoints the dispersion in visiting pattern throughout classes (non-assortative mobility). Otherwise, lower values with cutoff -1 indicates the tendency to visit places outside own socioeconomic class (disassortative mobility). Complete technical note on transformation technique and assortativity computation is discussed in Section 4.2.3.

Fig. 4.2 provides snapshots of mobility stratification pattern in London, starting from before lockdown and followed by the interchangeable periods between series of lockdown and reopening. We construct two different matrices. Firstly, Fig. 4.2a contains all locations in the trajectories, regardless identified label as either home or

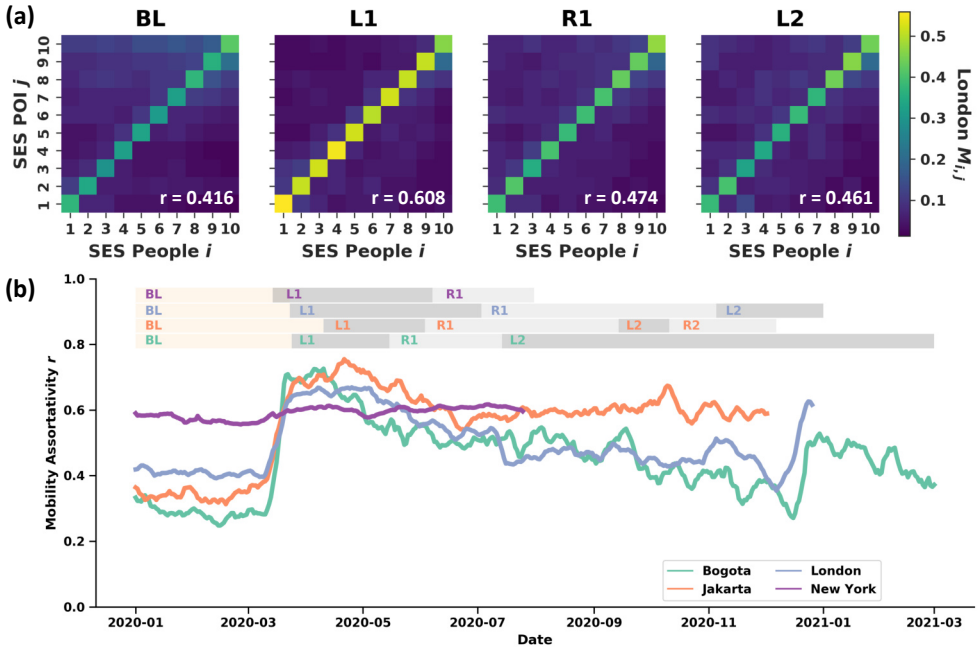

Figure 4.2: Mobility stratification matrix $M_{i,j}$. The structure of empirical socioeconomic stratification in London is visualised in a matrix form composing visit probabilities of individuals in each class to places located in various other classes. Fig. 4.2a reveals that larger visit proportion happens in a bin with lighter colour grades along diagonal elements across periods: Before Lockdown (BL), Lockdown (L1/L2), and Reopening (R). The strength of assortative mixing is quantified by a correlation coefficient between i and j denoted as r. We find stronger diagonal concentration during lockdown, denoting considerable visits to locations within own SES. Therefore, enforcing lockdown levels up assortative mixing. This is considered as a change in mobility preference due to NPI. Fig. 4.2b is constructed by implementing sliding window algorithm. For every 1 week window with 1 day slide interval, a mobility matrix is generated with computed r. Increasing r overlaps with lockdown period. Colour shades of line and block denotes city.

non-home areas. Inference method for location detection and labelling (home and non-home) is provided in Section 4.2.2. In Fig. 4.2a, x-axis represents socioeconomic classes of people i while y-axis denotes socioeconomic classes of places j. As people move, we calculate the frequency visit for each pair of classes (people-place), proportional to total visits made by everyone who belongs to $c_u = i$ (column-wise normalisation). Colour shades differ the visit magnitude where it becomes lighter as visit proportion gets larger.

Taking a closer look at Fig. 4.2a, we find noticeable main diagonal elements of the matrix where the colour gradient is away brighter than the the rest. The dominance of diagonal trace reveals that people usually visit places within their own class, independent of their socioeconomic rank. This generic pattern is consistent from time to time against the level of mobility restriction. More prominently, there is solid cycli-

cal pattern showing assortativity measure r as a response function towards strictness of such restrictions (NPI). The implementation of lockdown (L1 and L2), for example, harnesses mobility at large and encourages people to visit POI within own socioeconomic spectrum. Therefore, correlation coefficient r is on the peak during the first lockdown (L1). Once ease of mobility is reintroduced during reopening (R1), visiting more places is possible again. Chance for higher socioeconomic mixing in mobility is opened, resulting in lower r. However, it does not retrieve back to the original level before lockdown. We recognise this phenomena as induced assortativity. Similar matrices computed for other urban areas are presented in Section 7.2.2.

To refine the computational robustness, we look at more granular temporal length by introducing sliding window approach in constructing mobility stratification matrix (Fig. 4.2b). For every 2 weeks window with 1 day slide interval, we create a matrix and measure its r. This steps explodes the number of generated matrices, showing how segregation is changing as a function of t. It is also found anywhere across the hemisphere but quite unshifted in New York. Mobility assortativity in New York relatively stable across the time without any significant temporal cycle. Outlier pattern in New York could be addressed to the imbalance and asymmetric mobility between five boroughs within its territory: Manhattan, Brooklyn, Queens, Bronx, and Staten Island. In related studies, Rajput and other [161] state that stay-at-home orders implemented in the midst of COVID-19 outbreak disturbed 80% typical daily movement within city in New York from as early as the second half of March 2020. Recalling that Manhattan is the epicentrum of human dynamics where various mobility motifs and activities occur, we observe the case on Manhattan separately in Section 4.3.3 to clarify the upsurge in assortativity during lockdown that already found in other cities.

4.3.2 Residual isolation

The next logical question is regarding the persistence of the segregation pattern that being discussed so far. To refine the observation, we isolate home location effect on visiting pattern by removing own home location from mobility trajectory of each individual (see Section 7.2.3). Assortative mixing is consistently pronounced regardless types of policy imposed on mobility restriction, for instance lockdown and reopening. Moreover, it validates the finding as the revolving cycle persists even after we exclude own home location from mobility trajectory of each individual.

To further refine the observation related to changing segregation pattern, we motivate this study to measure the presence of *residual isolation*. The ultimate recovery is expected when mobility pattern and assortative mixing during the reopening stage are

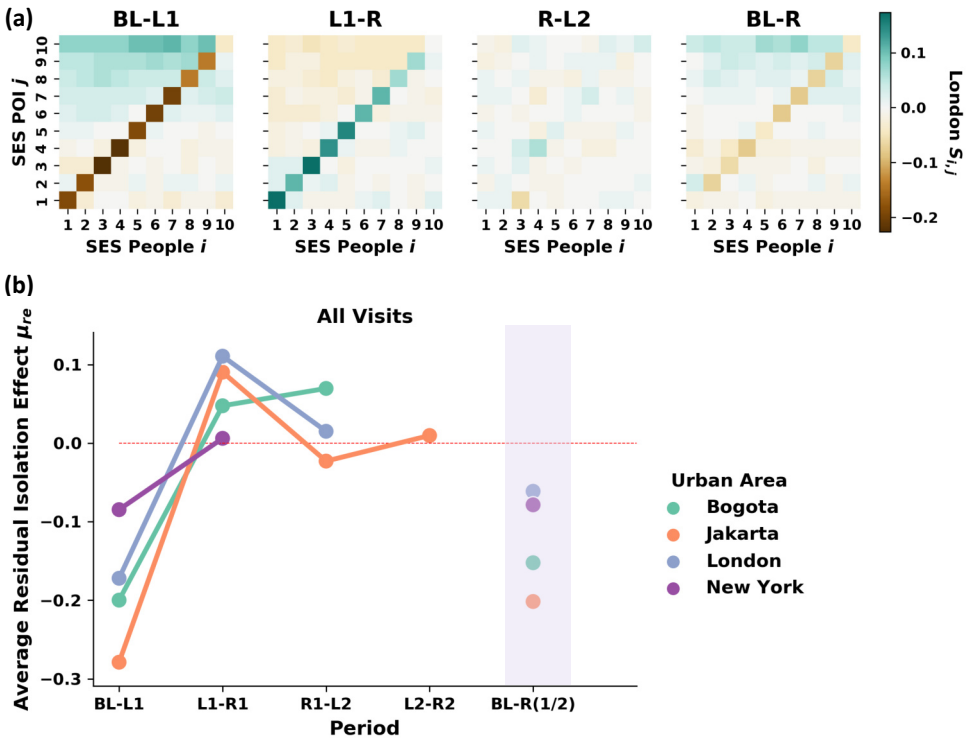

Figure 4.3: Mobility adjustment matrix $S_{i,j}$. It shows the ratio in stratified mobility pattern between a period during the pandemic namely Lockdown (L1/L2) and Reopening (R) as to compare to Before Lockdown (BL). Green shades indicate more visits made before the enforcement of lockdown, white blocks constitute equal visits, otherwise brown blocks appear. Therefore, we observe contrast proportion on the upper diagonal elements in London as visits to these places touch the lowest level in L1 relative to BL, burst in R1 and drop in L2 (Fig. 4.3a). Residual isolation effects as measured by average value of main diagonal trace in each matrix μ_{re}. Comparative measure across cities in terms of average residual isolation effect μ_{re} is provided in Fig. 4.3b. Purple block shows the difference between before lockdown baseline and reopening stage.

on the same level as before lockdown. If such conditions hold, sudden changes triggered by external shock namely COVID-19 outbreak might only carry short-temporal effect and people don't have any barrier to return to the normal pre-pandemics configuration. *Mobility adjustment matrix* $S_{i,j}$ is set up where the matrix element b_{ij} entails the difference in proportion of frequency visits between a pair of consecutive periods as seen in Fig. 4.3. In London, there are 4 pairwise comparisons. If the visits are more prominent in the second period, the bins have negative value with brown shades. Otherwise, visits mostly happen during the first period coded in green. If the bins turn white, visit proportion remains the same in those two periods.

Fig. 4.3a reveals the difference between before lockdown and the first lockdown, inferring that the first lockdown is the most stringent among others. It tells us that the

induced assortativity develops into isolation. In the extreme degree, individuals during lockdown restrict their preference to be present in the areas within own socioeconomic boundary around 20% more than they used to be. As the reopening is imposed after the first lockdown, the pattern is reversed. The difference between reopening and the second lockdown is very subtle. Interestingly, the reopening is not necessarily able to restore the typical configuration to before lockdown. We still see brownish gradient along main diagonal traces, revealing the existence of residual isolation effect.

Quantitative measure of residual isolation μ_{re} is provided by computing average value of matrix diagonal elements in Fig. 4.3b. It confirms the visual assessment in Fig. 4.3a. Initially, if we include all locations in the trajectories (Fig. 4.3b: All Visits), shifting to the first lockdown (BL-R1) results in the lowest μ_{re} because visits in the first lockdown is highly concentrated within own socioeconomic class. In this case, Jakarta exhibits highest average residual effect where people tend to spend almost more than 30% frequent activities in the class they belong to. It is followed by Bogota at about 20%, slightly lower in London, and nearly 10% in New York. However, the reopening (compared to before lockdown/BL-R) does not directly bring μ_{re} equal to zero in cities we observe, indicating the prevalent residual isolation. The order of residual isolation level changes a bit as New York precedes London. Weaker average residual isolation is found after removing local visits (see Section 7.2.3) and pushes μ_{re} closely distributed around zero.

4.3.3 Manhattan effect

New York is made up of five boroughs respectively Manhattan, Brooklyn, Queens, Bronx, and Staten Island. Among others, Manhattan is the centre of human activity agglomeration. Manhattan as a borough with the highest economic pull-factors in New York is massively affected, because mobility disruption hit not only movement of people inside borough, but also inter-borough movement that usually found in commuting pattern to workplace. People who reside in Brooklyn and Queens, for example, stop commuting to Manhattan as many of them switched to working from home practice. It is also reflected in lower use of public transportation and level of road traffic.

Segregation pattern changes as a response to mobility restriction imposed due to the pandemic. In Fig. 4.2b, we see that the mobility assortativity r in New York is relatively flat as to compare to other cities such as Bogota, Jakarta, and London, but a more substantial mechanism at work that shapes urban human dynamics might contribute as well. In this section we take two strategies to disentangle spatial scale. At first, we focus in the area of Manhattan where activities and mobilities are heavily concentrated.

Later on, we analyse mobilities in each borough that together unite as New York (intra-mobility), followed by mobilities between a pair of boroughs (inter-mobility).

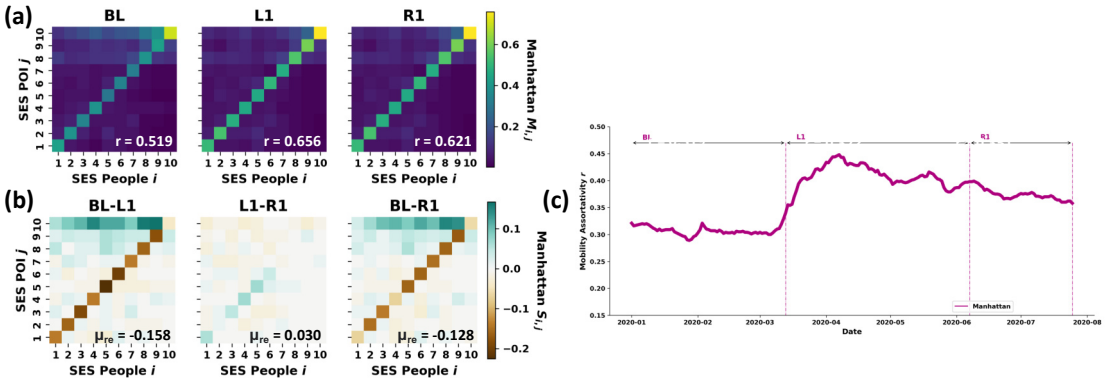

Figure 4.4: Mobility stratification matrix $M_{i,j}$ and adjustment matrix $S_{i,j}$ for Manhattan. We impose additional layer of filtering in New York by only looking at the locations within Manhattan boundary. Assortative mixing touches the highest level during lockdown ($r = 0.656$) as shown in Fig. 4.4a. Similar computation based on sliding window algorithm at higher temporal resolution (1 week window with 1 day slide interval) in Fig. 4.4c is also in concordance. After reopening, average residual isolation effect μ_{re} is still 12.8% higher as to compare to before lockdown period Fig. 4.4b.

Mobility stratification in Manhattan is visualised as matrix in Fig. 4.4a and a sliding time window plot in Fig. 4.4b. Measures taken during lockdown affect individual preference regarding their mobility. Inevitably, there is an increase in visits to places within own socioeconomic range. Reopening happen at some points, however nothing such fully recovery exists. Taking a pair of matrices in two consecutive periods, we have another form of matrix to show mobility adjustment as seen in Section 4.4b. We still find that the average value of diagonal elements is 12% higher than before lockdown (see right matrix). After all, residual isolation effect remains prominent in Manhattan.

Computations for mobility in New York based on CUEBIQ dataset (Fig. 4.5a) are reproduced for SafaGraph dataset (Fig. 4.5b). The two comes in conformity in terms of the proportion of mobility category in which individual flows within a single borough (intra-mobility) surpasses the fluxes across different territories (inter-mobility). The first is presented in Fig. 4.5c (CUEBIQ) and Fig. 4.5d (SafeGraph). A striking mirroring degree of assortativity in mobility r within Manhattan is seen, ranging from 0.6 before the implementation of lockdown to 0.8 in the aftermath. While the value of r is slightly different in Bronx (light green), Brooklyn (orange), Queens (purple), and Staten Island (pink), the pattern stays the same: increasing segregation since the lockdown period. One reason behind is that once people stay at residential area, they are bounded not only by spatial scale, but also socioeconomic homogeneity in the sur-

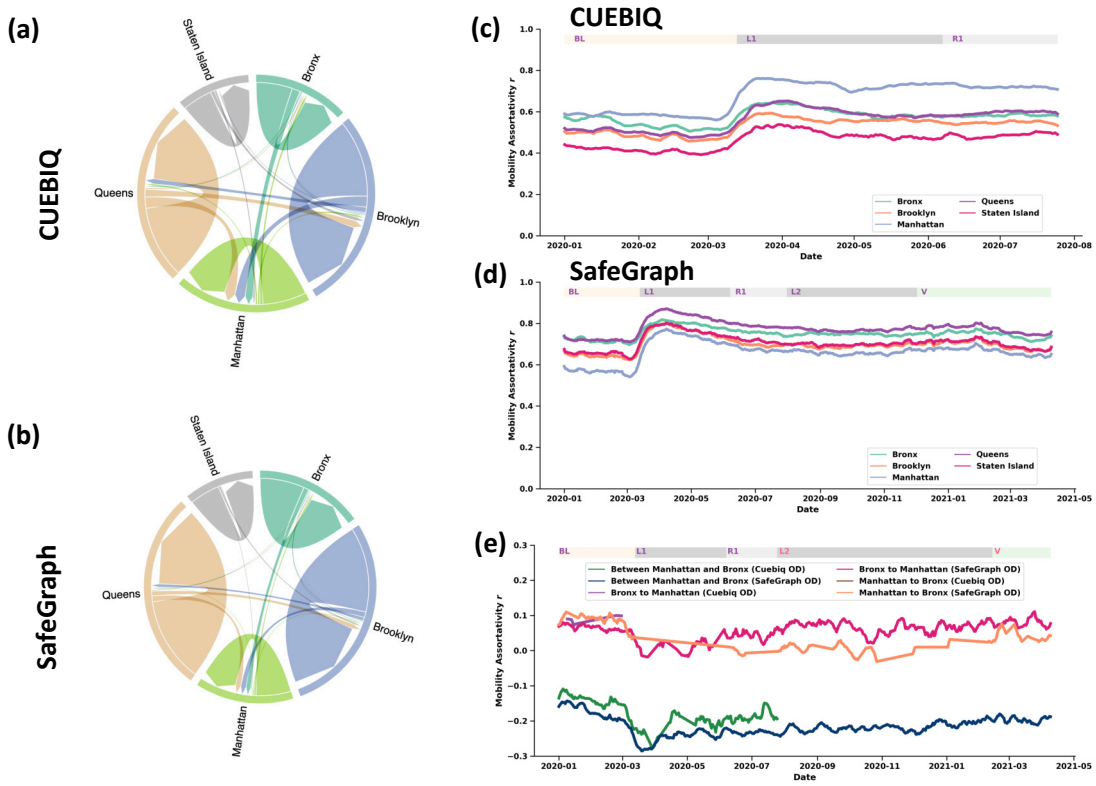

Figure 4.5: Trip composition and mobility assortativity r by category. Intra-mobility (mobility within borough) dominates trip proportion in both CUEBIQ (Fig. 4.5a) and SafeGraph dataset (Fig. 4.5b). Mobility assortativity is computed at census tract level based on OD matrix, showing similar pattern for intra-mobility mixing namely increasing segregation in the two datasets (Fig. 4.5c-d). Interestingly, segregation in inter-mobility (mobility between borough) tends to be lower instead, for instance in mobility flow between Manhattan and Bronx (Fig. 4.5e).

rounding neighbourhoods.

On contrary, individual flows across boroughs (inter-mobility) exhibits decreasing segregation as shown in Fig. 4.5e in the case of mobility flux between Manhattan and Bronx. As a undirected mobility network, mobility recorded in CUEBIQ dataset (dark green) and SafeGraph dataset (dark blue) indicate the emergence of disassortative mixing with value lower than 0, implying that people abruptly visit places differ from own socioeconomic status whenever they need to step out territory/borough where they reside due to multiple mobility reasons (e.g.: work or school).

4.3.4 Restriction and behavioural effects

Pandemic brings another complexity in the way people move from one location to numerous others across space. During the COVID-19 outbreak, mobility is not merely driven by established personal preference but also supplementary necessity to align with

prescribed mobility restrictions. Long distance and intensive human movements are no longer possible to take deliberately in order to restrain the surging number in COVID-19 infections and spreads. This inter-twinned mobility motif draws our attention to translate its effect on the variability of location preference among individuals.

We look at heterogeneity of where-to-go decision from two different aspects: spatial and socioeconomic composition. Entropy is employed to measure the regularity of mobility traces in term of geolocation that resulted in *spatial mobility entropy* $Hm(X)$ and socioeconomic class of those locations that gives *SES mobility entropy* $Hs(X)$. Lower entropy corresponds to higher domination of particular locations/SES of locations in the visit pattern, signalling the extensive locational/socioeconomic isolation. Given that the measure is normalised by period, the upper cut-off is 1 (absolute heterogeneity) and the lower cut-off is 0 (absolute homogeneity). Formal formulation of entropy is available in Section 2.3.1.

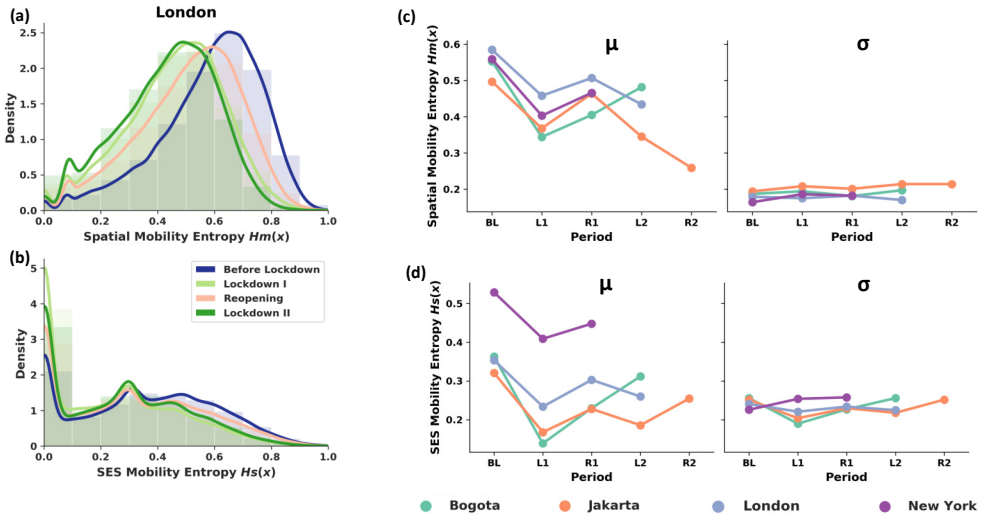

Figure 4.6: Spatial and SES mobility entropy. Spatial mobility entropy $Hm(X)$ (Fig. 4.6a) takes into account the heterogeneity of places in individual trajectory with value range from 0 (visiting same locations) to 1 (visiting various locations). SES mobility entropy $Hs(X)$ (Fig. 4.6b) takes similar computation after replacing set of locations with socioeconomic status of area where those places located implying visit variation between socioeconomic isolation (0) and socioeconomic diversity (1). In London, we observe less heterogeneity in both locations and socioeconomic status of places visited by individual during lockdown. Even after some relaxations are allowed, people do not experience mobility at pre-pandemics level. Similar observation also become evident in other cities globally (Fig. 4.6c).

In London, we deal with four phases of pandemic: Before Lockdown (BL), Lockdown I (L1), Reopening (R1), and Lockdown II (L2). Fig. 4.6a reveals the distribution of locational mixing degree in individual trajectory. Fig. 4.6b follows the similar way but rather emphasising on socioeconomic setting of those listed locations. In

both figures, skewness of the curve moves to the left (to the direction of zero) in the first lockdown (light green), so does in the second lockdown (dark green). It points out the tendency of upholding more homogeneous visiting pattern. In respect of spatial scale, urban explorability drops once policy limiting mobility flow implemented. Consequently, trajectory becomes more narrow (centred to smaller set of places) and localised (closer to where home is located). Similar pattern also holds with regard to socioeconomic range. As set of locations is shrunken by distance, in turn it becomes highly concentrated to particular socioeconomic level that reflects own well-being.

We check the shifting magnitude by computing average value (μ) and standard deviation (σ) of the two entropies. In Fig. 4.6c, the initial phase of lockdown (L1) characterises mobility pattern to be locationally more homogeneous since spatial mobility entropy $Hm(X)$ is lower than before lockdown period (BL). Spatial shrinkage largely happened in Bogota during L1, reaching the lowest record of average value at 0.35. Jakarta comes after at 0.37, while New York and London are somewhere between 0.4 and 0.5. The reopening phase that follows (R1) does not bounce the variability of locational and socioeconomic preference back to original level before lockdown even though it goes to recovery direction. Compared to spatial mobility entropy, ses mobility entropy $Hs(X)$ in Fig. 4.6d receives grave repercussions caused by the outbreak even more as μ ranges from about 0.5 to lower values. During L1, People in Bogota and Jakarta experience deeper socioeconomic isolation as $Hs(X)$ falls below 0.2. London is close to 0.35 while New York is around the borderline of 0.4.

4.3.5 Mobility intervention

To this point, we have revealed the embeddedness of residual isolation effect of shock even after mobility restrictions gradually lifted. However, what kind of restriction significantly contributes to such configuration is still unknown. Data on NPI contains the strictness level of every single restriction (9 categories) over period of time, including closing of main venues such as school and workplace. We rank those restrictions listed as NPI by running linear regression where the dependent variable is entropy values and the independent variable is stringency of each restriction. The $R - squared$ is descendingly sorted and label the rank from 1 to 9 starting from the restriction with highest (1) R-squared to the lowest (9). On top of that, we are motivated to reveal the similarity of restriction profile. We compute Spearman's rank correlation on the restriction ranking by city $rk_{Hm(X)}$. It measures the strength of monotonic relationship on the ranked data. Possible coefficient values are between -1 (perfect negative monotonic relationship) and 1 (perfect positive monotonic relationship), while 0 denotes the

condition where no correlation found. Methodological justification for this approach is further explained in Section 4.2.4.

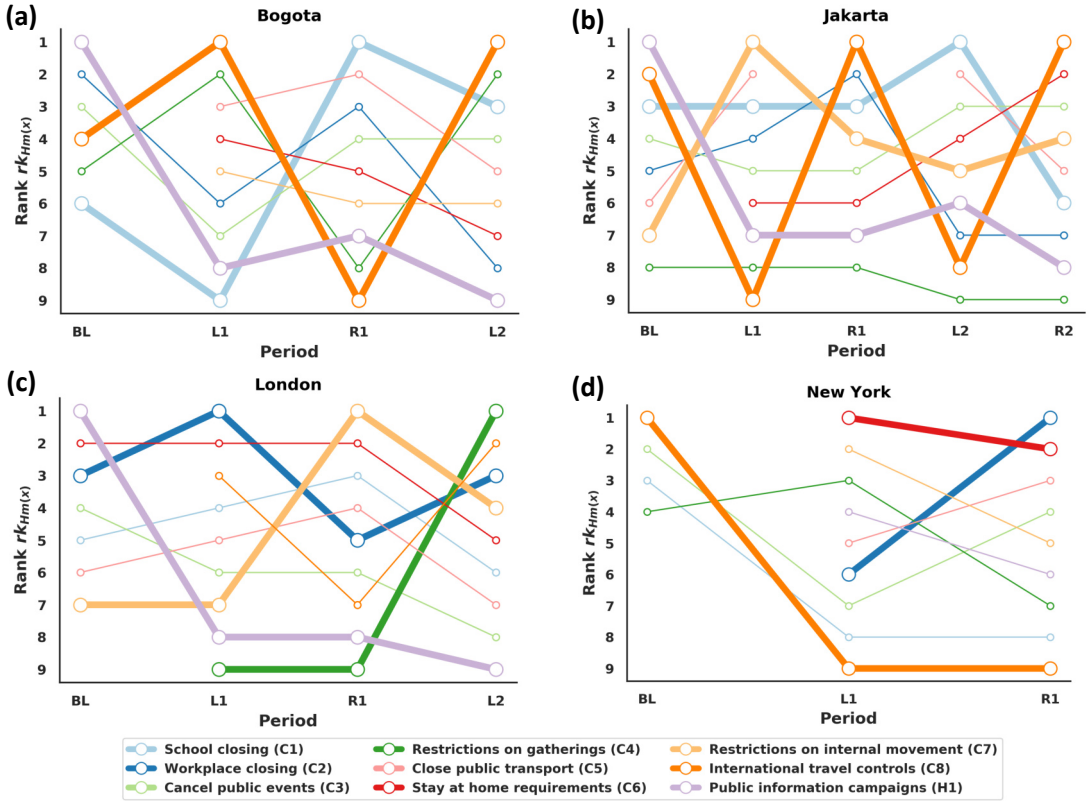

Figure 4.7: Spatial Constraint. The effectiveness of NPI in constraining spatial (Fig. 4.7a-d) exploration of individual is presented as rank based on significance of each type of restriction (C1-C8 and H1) in interfering spatial entropy $rk_{Hm(X)}$. Restrictions that come on the top rank is printed in bold. In all cities except New York, public information campaign (HI/light purple) is the most influential instrument before lockdown, but weakening later on.

Public information campaign (H1/light purple) is the most preponderant for everyone in Bogota, Jakarta, and London (Fig. 4.7a-c, top row), equally affects mobility in terms of locational diversity in mobility. This restriction gets less importance as time goes by. International travel controls (C8/dark oranges) predominantly dictates the trajectories in the first lockdown in Bogota (Fig. 4.7a) and in both reopening phases in Jakarta (Fig. 4.7b). Workplace closing (C2/dark blue) highly affects London in the first lockdown (Fig. 4.7c) but rather later to be effective in New York (Fig. 4.7d). Meanwhile in New York, stay at home requirements (C6/red) diminishes spatial exploration in the first lockdown.

Individual exploration occurs not only over physical space, but also beyond socioeconomic dimension. Therefore, enforcement of NPI (widely known as mobility re-

strictions) also reduces socioeconomic diversity of visiting places. The computational results on the effectiveness of NPI in restricting mobility in term of socioeconomic exploration $rk_{Hs(X)}$ is reported in Section 7.2.8.

4.4 Discussion and conclusions

Mobility network as a system is dictated by personal preferences over location, including home and workplace area as well as the availability of supporting amenities and infrastructures like school, park, shop, and public transportation. Given the fact that personal preferences are bounded by own socioeconomic background, social stratification characterises mobility at large. The strict preference towards such boundary may lead to social exclusion [116; 156] and social segregation [202; 126].

The presence of sudden change coming from outside of the system, for instance the burst of COVID-19 outbreak, definitely reshapes typical pattern of mobility network. To combat the spread of infection, governments worldwide take into account the non-pharmaceuticals interventions (NPI) with basic principle of controlling crowds. Social distancing stands on the frontier and later followed by various other measures such as cancellation of large gathering, closure of school and workplace, and stay-at-home order. In response to imposed regulation, people reorganise their mobility motif and reduce the activities outside home. Essential daily trips grocery or pharmacy cut short and become localised, closer to the neighbourhood where they live. Consequently, changing travel behaviour is visible, especially in terms of destinations and distances.

Earlier study [105] uses mobility data to show the effectiveness of aforementioned interventions in dealing with the surge of COVID-19 cases. However, the understanding related to the extent those mobility restrictions bring additional strength of assortative mixing in urban encounter that further stratifies the society remains untapped. Disadvantages such as isolation and segregation caused by reduced spatial scale and possibility of wider urban encounter and exploration may stimulate excess homophilic behaviour [192] at the cost of shrinking diverse encounters [167].

We took a step forward to analyse the impact of COVID-19 outbreak on structural preference reflected in mobility pattern by looking at the mobility dynamics in Bogota, Jakarta, London, and New York. In this study, we find that in-class visits dominate mobility pattern in every temporal snapshots, ranging from before lockdown, lockdown, to reopening (see Fig. 4.2a). Solid cyclical pattern of homophilic behaviour is also detected as the assortativity coefficient r remains highest during lockdown while the emergence of reopening does not directly brings the typical mobility mixing pattern to the original level before the enforcement of lockdown, indicating the existence of

residual isolation effect.

We further measured the degree of residual isolation by comparing stratification in mobility pattern between two consecutive periods (see Fig. 4.3a). It validates the presence of residual isolation effect where visits within own class during reopening is still higher than the usual rate. Another feature of isolation in mobility that has been presented in this study is the decreasing heterogeneity of where-to-go decision from two distinctive aspects: spatial and socioeconomic composition (see Fig. 4.6). Entropy measures reveal that visits becomes highly concentrated to particular locations and socioeconomic classes.

In literature, NPI is globally considered to have potent impact in reducing virus transmission brought by COVID-19 [98; 177]. Nevertheless which type of NPI does constrains mobility across time window is still open to investigation. We proposed ranking approach for each mobility restriction listed as NPI to examine its magnitude in intervening the diversity configuration of visiting pattern. In all cities we observed, the significance of public information campaign (H1) gains its highest importance in the early stage of pandemic but drops as other measures available on the table. Variability of ranking could be related to the structure of urban fabric in respected city as well as the level of socioeconomic well-being. The results of this study could be extended by mapping out the spatial distribution of urban forms in order to better understand better way to mitigate dissonance induced by residual segregation in the pandemic time, for example by alienating with functional mixing of urban amenities.

Apart from the computations demonstrated to this point, we realise that stronger evidence for residual isolation in the longer term could be presented if the access to more recent data is available. Our latest data only covers the initial period of reopening where NPI and the COVID-19 protocols is still at the frontier in controlling the outbreak. It solely depends on the behavioural conformity/attitude towards mask wearing and social distancing without any intervention from vaccination policy. The difference that the a 2-dose COVID-19 vaccination makes in the US is estimated reducing the overall attack rate 9.0% to to 4.6% [127]. Reduction in death rate at 69.3% is also quite remarkable, as well as decreasing adverse effects among ICU hospitalisations (65.6%) and non-ICU hospitalisations (63.56%). However, the considerable drop of epidemic spreading during this period is still under the condition of necessary NPI implementation. Once the attitude towards becomes NPI less adherent due to the massive vaccination and the arrival of booster program, we could further investigate whether mobility configuration of visit preference will be fully recovered and back to normal pre-pandemic level.

Another boundary that we would like to underline is the limitation in direct comparison between cities. This issue is raised up due to the difference variables we use

in defining SES, depending on the availability of data. In some cities, SES is constructed solely based on income. However, poverty index that captures aspect beyond income such as quality of life is preferred in other cities. If similar characterisation of SES could be made once data is accessible in the same spatial resolution (e.g: Relative Wealth Index for every 2 km sq), broader comparative interpretation could be brought in this study.

Chapter 5

Multiscale spatial economic diversity

Urban economic diversity is a complex phenomenon that entangles various aspects such as proximity and embededdness. Under the notion of agglomeration and spillover, cluster formation emerges as result. Apart from the multifaceted process, the understanding on economic diversity still faces the gap especially related to the relation between spatial scale and sector size that colloquially seeds proximity through colocation. In this study, our utmost motivation is to clarify the entanglement between the role of spatial proximity and spatial embeddedness in reproducing economic diversity across industrial sectors by presenting three mechanisms that affect economic diversity in New York with main focus in Manhattan namely size, spatial, and proximity. We also search for evidence of larger economic diversity in larger spatial unit by comparing computations over a number of spatial grid scales based on bipartite projection on spatial and sectoral linkage. The results show that sectors with larger number of firms or entities form close knitted connections in the network and largely contribute to the urban economic diversity.

5.1 Introduction

Global city emerges as a conceptual underpinning for a primary node in intersectoral economic activities at world wide scale. In the literature, a number of leading papers have explored the concept of global city and its interchangeable term called world city, including their role in shaping the economy. Friedmann (1986) [71] proposes the foundational hypothesis of world city by dismantling hierarchical organisation of city and identifying joint economic forces that form global network: investment flows, division of labour, and flow of people. While investment flows are widely supported by backbone of banking and financial sector, the existence of extended ser-

vices such as manufacturing and retail are not diminutive either. Dual side of labour market characterised by co-existing professionals and low-skilled workers brings monetary prospect to wider population, making flow of people at large inevitable. Sassen (1991) [168] rather observe the presence of global city as a spatial extension of global economy that incorporates various economic activities along production chain located in different places, implying *"advanced producer services which are the distinctive feature of contemporary world city formation"*. Consequently, the theorisation of the global city model consists of complex building blocks from central corporate functions, specialised service firm, growing agglomeration economies, highly specialised and networked services sector, expanding transnational urban system, increasing spatial and socioeconomic inequality, to rising informalisation of a range of economic activities [169].

Nonetheless, the position of New York in such configuration as the world's business hub is mainly driven by centrality of international financial and business centres located in the city with direct transactional connections to many other industries world wide. In 2022, with a market of 8.6 million potential clients and a metro area population of more than 25 million, the economic prospects in New York are appealing. On the supply side, the city has over 250,000 enterprises and 43 of the corporations in the S & P 500 Index have their headquarters here, the most of any city in which altogether contributing to the Gross Annual City Product of USD 678 billion [46]. Its robust economy and infrastructure, global workforce, on top of cross-industry innovation attracts many other sectors and people to be connected, creating multiplier for already highly established diversity level in the economic sectors. However, looking at finer spatial scale where New York as a city consists of 5 boroughs namely Bronx, Brooklyn, Manhattan, Staten Island, and Queens, activities and sectors apparently are not distributed homogeneously in space. Main central business districts (CBD) respectively Midtown, Midtown South, and Downtown are located in Manhattan.

Literature in spatial economics, economic geography, and regional science widely discuss the significance of industrial clusters and reckon the contribution of economic diversity in shaping competitiveness of a city. Malmberg and Maskell (2002) [117] underline that in economic process, the notion of proximity and place plays an important role. Economic activities tend to form spatial agglomeration due to the presence of informational and knowledge exchange such as flow of labour (Simpson, 1992) [179], relationships between supplier and customer (Porter, 1990) [157], and knowledge spillover (Krugman, 1992) [103] where geographical proximity facilitates these exchanges (Potter and Watts, 2011) [159]. The emergence of industrial clusters is also linked to the rise of regional specialisation that results in characterisation of hetero-

geneity of spatial concentration and motivated by the law of *increasing returns to scale* (Krugman, 1991) [102]. Nevertheless, the context of space becomes an integral part while discussing embeddedness, shaping the understanding towards inherently spatial framework (Martin, 1994) [120]. More precisely, spatial embededdness counts on *"who is embedded in what" and "what is so spatial about it"* [155; 83]. The first remark is linked to emulation of local institutional fabrics as suggested by MacLeod (1997) [115] and the second is articulated by Porter (1998)) [158] via the conceptualisation of cluster as an embodiment of geographical proximity in the firm networks. In addition, Halinen and Tornroos (1998) [78] land on the lenses prescribed by location theory by looking at the preference of businesses over locations in running their activities. Although these studies closes the gap on the absence of spatial scale in the Granovetter's preliminary work [75], the methodological toolbox purposely designed to capture the scale effect aspect in spatial embeddedness remains missing from the literature.

Within a finite spatial limit, congregation of firms accessing nearby production factors including labours could accrue higher net gain as transaction costs coming from transportation and communication spending are compensated by reduced distance, therefore, *localised increasing returns to scale* gives a sound argument for the development of industrial clusters (Krugman, 1993) [104]. According to Nefke and Henning (2013) [135], employees and their skills are listed as the main important resource that a firm has because skills as endowment factors enable people to transform resources into end products and bring relatedness among various industries via knowledge creation. In this context, skill intermediates connectivity due to switching employment driven by *revealed ability of skilled employees to move* principle. It is quantified as *Revealed Skill Relatedness (RSR)* and suggests that labour flows dictates industrial linkages, making the existing nested hierarchy-based industrial classification system less adequate in contrast to proposed web-based industrial skill connections.

Observing industrial relatedness breaks another layer of complexity, taken into account its presence in different phase of economic process. In general, industrial relatedness could be measured by a broad spectrum of approaches. The first is constructed based on similarity in output, widely known as the standard industry classification system [7]. It has a nested structure that implies higher relatedness among industries within same 4-digit class than 2-digit class, for instance. The second highlights the importance of shared inputs and knowledge in production process. Breschi et al. (2003) [29] use patent portfolios to distinguish one industry to another while Farjoun (1994) [67] takes occupational profile as an instrument in measuring such relatedness. The third is derived from industrial portfolios where economies of scope comes out as a further implication of resource co-utilisation at various levels, including firm level [84] and country level

[136]. Neffke et al. (2008) [136] and Frenken et al. (2007) [70] come into an agreement that externalities caused by industrial relatedness are superior than externalities solely raising from localisation and diversity.

Relatedness and embededdness are utterly two sides of the same coin. Relatedness draws connectivity among sectors while embededdness underline the localisation aspect of relatedness. Moreover, industrial relatedness on one hand brings sectoral proximity given sizeable shared resources. Locational embededdness on the other hand reinstates the notion of spatial proximity because of being closely located next to each other in the same area. The term of *localisation economies* incentives positive externalities in a particular industry or the sector concerned (Marshall, 1920) [118] is broaden by recognising the dimension of spatial scales, known as *urbanisation economies* (Jacobs, 1960) [87], making positive externalities available to various industries and all sectors. In details, Goya (2022) makes a distinction between Marshallian and Jacobian externalities [74]. According to Marshallian externalities, benefits due to spatial agglomeration is considered to spread out across firms in the same industry because they share the pool of labours, the chain of suppliers, and the accumulation of specialised knowledge. On the other spectrum, Jacobian externalities suggests that the presence of diverse industries induces interactions among firms and triggers innovation, therefore positive externalities simultaneously affect multiple firms and industries. Consequently, size of city matters recalling that a larger area tends to have more diverse amenities and greater capacity to accumulate input-output linkages. However, the relation between spatial scale and sector size that colloquially seeds proximity through colocation is left to the lack of elaboration. The availability of free and updatable place of interest (POI) data open opportunity to establish integrated observation under study.

Our objective in this research is therefore to unravel the role of spatial proximity and spatial embeddedness in reproducing economic diversity across industrial sectors. We are interested in addressing methodological issues and proposing an artefact of measurement that allows us to disentangle the interdependence between geographical space and sectoral diversity. In our perspective, tools in network science could be adopted to enrich our analysis, especially in dealing with the complex interconnected phenomenon that drives economic diversity such as the size problem, the spatial scale problem, and the proximity problem.

Jacobs (1961) pinpoints a higher dependency of smaller enterprises on city as a representation of consolidated market where varied material supplies and skills are abundant in a concentrated area and directly connected to other sector of the economy, allowing them to further contribute to the diversity at large. Accordingly, the size of each sector has an influence on how it contributes to the estimation of its effect on the overall

diversity level due to the aforementioned mechanism such as the effect of proximity between sectors as seen in the concept of relatedness. It motivates us to take an initial step in formulating the first research question (RQ1): *What are the effects of size of sector relating to inducing economic diversity?*.

The existence of spatial embededdness is vastly examined within the literature, yet further investigation on the extent spatial scale matters in determining sectoral linkage among different business is still unavailable. As in Bergman and Feser (2020) [21], Van den Berg et al. (2001) [189] limit the observation to the imprecise dimension called local or regional if the emergence of industrial cluster, a network of co-located firms of a given industry, fills in a territorial boundary without specifying the spatial scale and shape. To tap the gap, we stand on the the importance of raising the following question (RQ2): *What are the effects of spatial and scale of sector relating to inducing economic diversity?*.

Last, our approach in response to proximity problem differs from the existing studies in which spatial proximity and sectoral proximity are treated separately. Spatial proximity could be measured based on metric distance or time distance (e.g.: travel time in the transportation networks subject to the availability of transportation modes) between two corresponding locations. On another spectrum, sectoral proximity is computed by taking industrial similarity (e.g.: 2-digit Standard Industrial Classification/SIC category) into account as seen in Bishop and Gripaios (2007) [24]. Instead of following the same construction, we combine the two and raise a question (RQ3): *To what extent does spatial proximity corresponds to co-location patterns between sectors?*.

5.2 Materials and methods

5.2.1 Data description

We decide to analyse New York as a case study in investigating multiscale spatial distribution of economic diversity because of two concerns. The first concern is related to the profile of New York as global city with its complex sectoral and spatial linkages that fits to our motivation in studying scale and shape of economic diversity. The second concern is due to the availability of POI data at a fine-grained resolution. These data are collected from SafeGraph Core Places [166] for 71,468 places across New York City in which 68,258 (96%) of them are located in Manhattan. Core Places dataset contains baseline information on business listing including business name, address, industry category, and geolocation (latitude and longitude).

Aggregated Sector	2 digit NAICS code	2 digit NAICS description
Agriculture	11 + 21	Agriculture, Forestry, Fishing and Hunting Mining, Quarrying, and Oil and Gas Extraction
Arts	71 + 72	Arts, Entertainment, and Recreation Accommodation and Food Services
Construction	23	Construction
Educational	61 + 62	Educational Services Health Care and Social Assistance
Finance	52 + 53	Finance and Insurance Real Estate and Rental and Leasing
Information	51	Information
Manufacturing	31 + 32 + 33	Manufacturing
Other	81	Other Services (except Public Administration)
Professional	54 + 55 + 56	Professional, Scientific, and Technical Services Management of Companies and Enterprises Administrative and Support and Waste Management and Remediation Services
Public	92	Public Administration
Retail	44 + 45	Retail Trade
Transportation	48 + 49 + 22	Transportation and Warehousing Utilities
Wholesale	42	Wholesale Trade

Table 5.1: Sector classification. We regroup economic sectors based on 2 digit NAICS into 13 aggregated sectors.

Industrial classification is taken from the North American Industry Classification System (NAICS) [134] that composes hierarchical classification system ranging from 2 to 6 digit code. Greater classification detail is found in more digits. At 2 digit level, it reflects the composition of economic sectors and we will group POI listed in SafeGraph Core Places dataset into 13 sectors based on aggregation of 2 digits NAICS as seen in Table 5.1.

5.2.2 Pipeline description

Fig.5.1a shows sectoral composition of firms in New York. Manhattan stands out in terms of number of firms especially in the 3 leading sectors respectively (Fig.5.1b). For the purpose of constructing random model as to compare with empirical data, there are 3 grid resolutions used in this study comprising 10km x 10km grid, 5km x 5km grid and 1km x 1km grid (Fig.5.1c). Due to the spatial concentration of different business sectors in this area, Manhattan is studied closely.

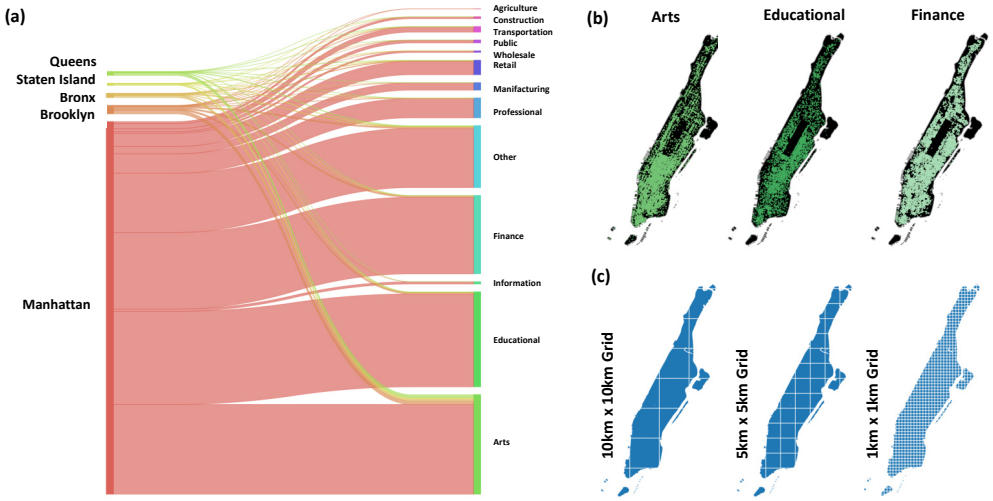

Figure 5.1: Sectoral composition, spatial distribution, and grid design. In New York, Manhattan remains the most densely populated by firms (Fig.5.1a). Within Manhattan, economic sectors falling under arts, educational, and finance have the highest spatial distribution (Fig.5.1b). We use 3 grid resolutions in the following computations namely 10km x 10km grid, 5km x 5km grid and 1km x 1km grid (Fig.5.1c).

5.2.3 Economic diversity

Jane Jacob's notion of urban diversity has been well established and widely accepted as a factor that veraciously contributes to economic prosperity and quality of life in cities. In her prolific work [88], the broad spectrum of definition on urban diversity touches upon the attribute of having urban areas with mixtures of land use and human activities that drives social and cultural dynamics to enhance inhabitants' lives. Additionally, she perceives that cities are *natural generators of diversity and prolific incubators of new enterprises and ideas of all kind*, assigning cities to a distinctive position as *natural economic homes of immense numbers and ranges of small enterprises*, to name a few are *supermarkets and standard movie houses plus delicatessens, Viennese bakeries, foreign groceries, and art movies*. Sectors like entertainment, retail trade, and cultural facilities gain advantage in urban geographical territory because there are enough people that could take part in the provision of service on supply side as well as sufficient demand for variety of attractions.

This particular condition shows the centrality of city in accommodating economic diversity where dependency on resources available in the city is translated into the diversity creation as more businesses open their doors. Therefore, a vibrant city is the one with ability to permit and stimulate the multiplier effect of diversity itself. In parallel, Foord (2013) expresses the requisite for which urban vitality could be achieved namely *the risky experimentation across co-located sectors in which hitherto unrelated knowl-*

edge and activities [68]. From an economic geography perspective, level of economic diversity is positively correlated with the size of city [61]. Henderson (1997) [82] draws a line between cities in the category of medium size (50,000–500,000 population) and large cities (over 500,000) and reveals that the later excels in sectors related to finance, insurance and real estate while the prior performs better in manufacturing. Regarding the scaling pattern of diversity, Henderson (1991) [81] provides evidence among cities under observation in which similarity in sectoral specialisation is in line with similarity in size. To measure spatial diversity given the distribution of diverse firms across industries in an area, we apply entropy approach especially the one that is designed for the purpose of spatial analysis such as Leibovici entropy.

To answer RQ1 (*What are the effects of size of sector relating to inducing economic diversity?*), we propose to measure economic diversity using the Leibovici entropy because it takes into account spatial mixing pattern. The Leibovici entropy consists in the relevance of spatial locations of occurrences, recognised as the primary characteristic of spatial data, and is different from the Shannon entropy (formally expressed in Section 2.3.1), used to measure economic diversity, for instance in Zachary and Dobson (2020) [203] and Palenzuela et al. (2022) [163], but subject to the limitation of discerning the role attached to space in measuring heterogeneity and dealing with different spatial distributions, for instance in the presence of spatial association and randomness [6].

Moreover, by running out-of-bag approach (removing one business sector at a time), we know which sector guarantees diversity in that particular area. We define change in Leibovici entropy for each sector as

$$\Delta H_L(Z) = \frac{H_L(Z) - H_L(Z)^*}{H_L(Z)^*} \times 100\%, \tag{5.1}$$

where $H_L(Z)$ is Leibovici entropy of respected area and $H_L(Z)^*$ constitutes the aforementioned value after removing the business sector of an interest. Presented in percentage, large negative value indicates the ability of a sector in considerably reducing spatial diversity in the local area, while large positive value contributes to higher diversity. Otherwise, no observable impact is brought by that particular sector in the case of zero value. We state our hypothesis that sector with larger number of firms within a spatial unit induces higher economic diversity in that respected area (H1).

5.2.4 Sectoral linkages and co-location

Diversity in economic sector is argued to promote Jacobian externalities as Frenken et al.(2004) [69], for example, argue that diversity magnifies spillover effects with

higher likelihood of agglomeration emergence when there are some linkages between sectors in terms of technology or knowledge. In the large cities, economic diversity tends to be higher with more complex linkages which later encourages firms to increase productivity level and results in positive growth of average earning for the labours. Within the theoretical framework, Duranton and Puga (2004) [62] distinguish a number of mechanisms for cities to strengthen their economic existence called as *micro-foundations of localised increasing returns* among co-located firms. The first makes use of sharing mechanism including production facilities, input suppliers, and risks. The second gives an importance to matching mechanism in which probability of matching and quality of matches are expected to improve and reduce the hold-up quantity. Moreover, there is learning mechanism that creates, distributes, and takes stock of knowledge in production process. We introduce the measurement strategy based on networks and bipartite projections of co-located firms to reveal the mechanism at work.

In the previous section, we mention that the literature conform the impact of size on the level of economic diversity and sectoral linkages in cities. A widely accepted assumption is the larger the city, the higher variability of firms and sectors as well as complexity of linkages. Nevertheless, the spatial dimension in which the two takes place is rarely formulated, making the size of spatial boundaries becomes fuzzy [152] and may lead to the modifiable areal unit problem (MAUP) problem. It arises from the changing aggregation of shape and size of areal units or spatial data, generating possible contradicting results in statistical analysis [145; 52]. Not only aggregation into fewer and larger spatial units diminishes variation of the data, but also the alternative combinations of spatial units at comparable scales. Consequently, both carries associated effects on the quantitative findings. To tackle MAUP issues in multiscale organisation of economic activities, we compare various spatial delineations specifically 10km x 10km, 5km x 5km, and 1km x 1km cell grid and investigate the extent robust results could be obtained.

To answer RQ2 (*What are the effects of spatial and scale of sector relating to inducing economic diversity?*), we vary the size and shape of geographical units in which to measure economic diversity. We use a spatial delineation of 10km x 10km, 5km x 5km, and 1km x 1km cell grid with a given area as previously shown in Fig. 5.1c and calculate the aforementioned entropy. Distance effect at multiscale dimension is now observable. This result is compared to the entropy measurement on random model based on 100 iterations of synthetics business location. We control the confounding effects in observing locational aspect by keeping the number of POIs and distribution across sector constant at system level (aggregate). There are two scenarios in the random model: (i) Complete Randomisation (no specific pattern of clustering nor dispersion) and (ii)

Clustered Randomisation (y-axis is divided into partitions that belong to specific sector.) The first scenario is realised by shuffling sector label for each POI. Meanwhile the second scenario is implemented by ordering a single sector into a unique partition and running Poisson point process [96] to create synthetic locations. Our hypothesis is that spatial embededdness shapes the spatial distribution of firms across sectors (H2) as suggested by Halinen and Tornroos (1998) [78] and Malmberg and Maskell (2002) [117]. Two-sided Kolmogorov–Smirnov Test [97] is employed to test H2. In this test, spatial embeddedness is signified if there is significant difference in $\Delta H_L(Z)$ between empirical and synthetic spatial distribution of POIs. The test statistic D^* evaluates the difference between the cumulative distribution functions (CDFs) of two spatial data distributions as specified in

$$D^* = \max \lim_{x}(|\hat{F}_1(x) - \hat{F}_2(x)|), \tag{5.2}$$

where $\hat{F}_1(x)$ denotes the ratio of values that belong to the first data vector x_1 (e.g.: empirical data) which are less than or equal to the range CDFs of data distribution from those two data vectors x. Similarly, $\hat{F}_2(x)$ stands for the second data vector x_2 (e.g.: synthetic data) ess than or equal to x. Given D^* is generated with corresponding $p - values$, therefore in the case of $p-value < \alpha = 5\%$, the null hypothesis stating that the two data distributions are identical is rejected, validating that location embeddedness matters in shaping economic diversity.

To answer RQ3 (*To what extent does spatial proximity corresponds to co-location patterns between sectors?*), we construct a bipartite projection on spatial and sectoral linkage. A distinct characteristics of bipartite network $G = (S, A, E)$ is the existence of two sets of nodes, for instance sector s in the set of node $s \in S$ and area/grid cell a for set of node $a \in A$. Edges $e_{s,a} \in E$ represent connections between nodes in different sets with multiple edges counted as edge weights $w_{s,a}$. In the first projection, the node is sector and the spatial linkage is taken into account by pairing nodes if that area/grid cell has those sectors co-located. Inversely, in the projection based on sectoral linkage, the node constitutes the centroid of each area/grid cell and an edge emerges connecting those nodes if sectors are co-located in both areas/grid cells. We compute betweenness centrality of each node as follows

$$b_v = \sum_{i,j} \frac{n^v_{i,j}}{n_{i,j}}, \tag{5.3}$$

where $n^v_{i,j}$ is the number of shortest paths spanning from i to j passing v and $n_{i,j}$ is the total number of shortest paths from i to j. Higher value of betweenness centrality of

node v indicates larger proportion of nodes in the networks connected via node v, contributing to implication of its strategic location in controlling, mediating and facilitating information flows passing between others. Reflecting on the pioneering work of Jacob (1960) [87] and Hidalgo et al. (2007) [84], we hypothesise that spatial and sectoral proximity simultaneously trigger the emergence of linkages which further stimulates economic diversity (H3).

5.3 Results

Do some sectors contribute more to economic diversity than others? The idea of diversity is rooted in Jacobian externalities where recombination of wide varieties of knowledge may stimulate novelty in ideas and application, a source of innovation potential [88]. Notably, the likelihood of sectoral contribution to diversity, mainly via joint innovation, is not uniformly distributed. Rutten et al. (2011) [165] and Chapain et al. (2010) [41] highlight the strategic position of creative industry in promoting diversity through novelty creation. The first paper comes up with an argument in which creative industry is able to set its leverage because of high growth rate, high specialisation degree, and supported by crosslinking nature with other sectors in the economy, signifying the importance of sectoral linkage or inter-industry effects. In the second paper, the postulate regarding colocation is used as a basis for boosting potential spillovers commencing from creative industry to any other industries. Creative spillovers are also found in various types such as product spillovers with the demand generation for complementary creative goods and network spillovers with the rise of tourism industry in the aftermath of creative activities.

5.3.1 Size effects

In this study, we investigate sectoral spatial configuration of economic activities. POIs are grouped into 13 sectors based on aggregation of 2 digits NAICS. Among others, arts (71+72), educational (61+62), and finance (52+53), make up the highest number. In details, arts comprises POI identified as arts, entertainment, recreation, accommodation, and food services, while educational represents educational services, health care, and social assistance. Moreover, finance is used to label commercial activities in finance, insurance, real, estate, rental, and leasing.

Fig. 5.2 shows the spatial distribution of business locations in Manhattan. For each sector, we use three spatial scaling to illustrate the impact of granularity on concentration in spatial distribution. At coarse grain level in 10km x 10km grid (first row), all of

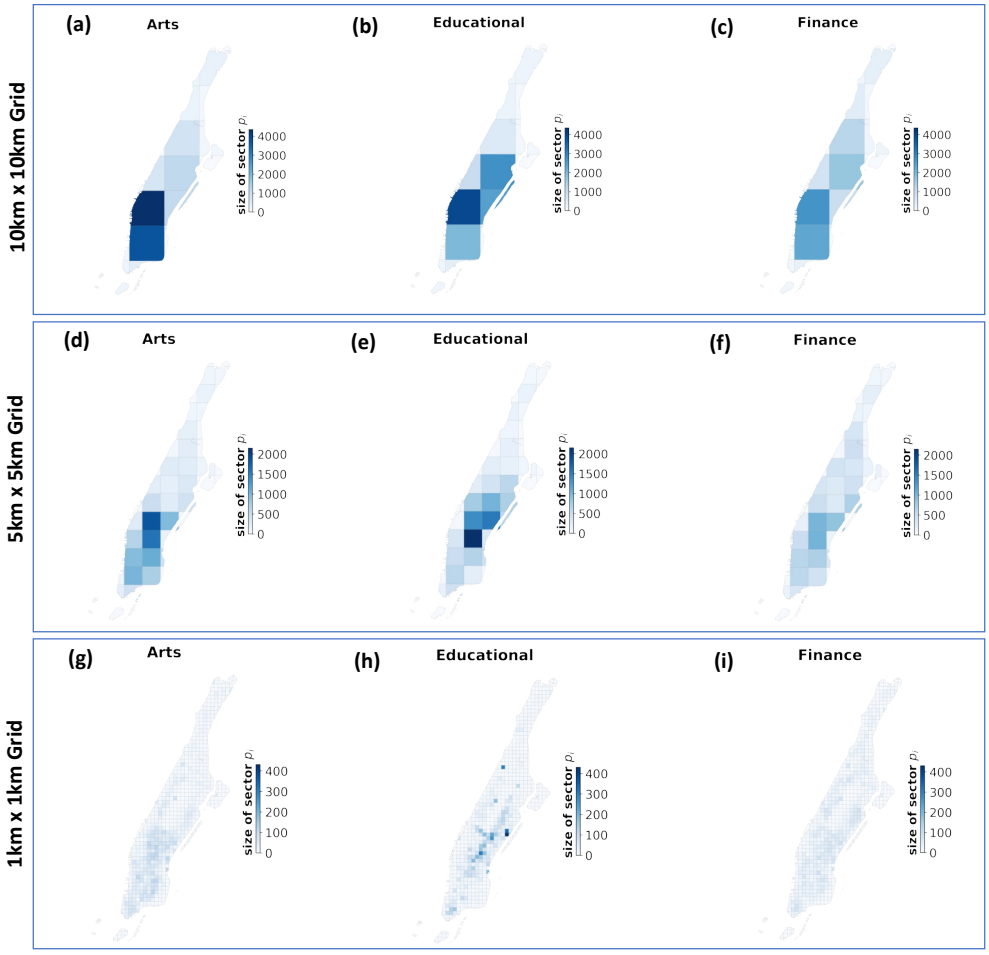

Figure 5.2: Spatial concentration of business in Manhattan by sector (left to right): arts, educational, and finance and by granularity (top to bottom): 10km x 10km grid, 5km x 5km grid and 1km x 1km grid.

three sectors mainly occupy the same grid cell/tile. As the spatial scale becomes finer at 5km x 5km, different sector is concentrated on different tiles. Higher resolution of grid at 1km x 1km even further indicates more visible cluster tendency by sector. Therefore, distribution of business exhibits spatial pattern that might differ across sector.

Lower Manhattan exhibits striking profile in which all sectors are highly concentrated. The density of POIs in this area is very high, with arts sector dominating the proportion of business along with educational and finance sector succeeding. Within the economically most active grid in 10km x10km resolution, arts sector alone opens the door for more than 4,000 business entities, while educational sector accommodates slightly lower number around 3000 and finance sector takes half the figure. Arts sector consists of any business falling into 'Arts, Entertainment, and Recreation' (71) and 'Accommodation and Food Services' (72). A primary reason for these listed activities to stand out is it functions as derived demand from what it is well known for, finance

sector, consisting of 'Finance and Insurance' (52) and 'Real Estate and Rental and Leasing' (53). For similar reason, it provides ancillary support to educational sector which composes 'Educational Services' (61) and 'Health Care and Social Assistance' (62).

As the area with immense level of diversity on daily life, Lower Manhattan is home for diverse attractions, raging from business deals, cultural epicentres, to governmental affairs. Business district is located in the heart of Lower Manhattan where the Financial District or Wall Street neighbourhood is spaced out by leading financial institutions such as Wall Street, the New York Stock Exchange, and the Federal Reserve Bank of New York. Cultural epicentres becomes a trademark especially in the area around Broadway. The presence of educational sector is highly visible along Park Avenue and its surrounding neighbourhoods where Columbia University, City College of NYC, Manhattan School of Music, and The Apollo Theatre can be found.

The importance of proper scaling is inevitable to better capture the sectoral distribution pattern. Sectors with intense spatial concentration like arts are better preserved over scale than the more ubiquitous sector like finance. Moreover, artefact of Central Park situated in the middle of Manhattan is visually present at higher scales in which 1km x 1km grid shows surpasses the others. This preliminary observation on the contextual entanglement between spatial scaling and size of sectors provides a foundation for further inference towards the level of diversity. It is in line with our first hypothesis stating that the sector size measured by the number of firms within itself largely contributes to the economic diversity in the delineated area.

5.3.2 Scale effects

Contribution of a sector to spatial diversity is measured by taking the percentage change between Leibovici entropy with all sectors and without sector in question as formally expressed in Eq. 5.2.3. We repeat this procedure for all sectors in every grid cell across grid resolutions. Higher percentage change implies larger contribution to spatial diversity in the given area. Leibovici entropy is preferred to measure spatial diversity due to its ability to capture not only variability of sectoral composition, but also the role of space in shaping spatial distribution of sector, for instance the degree of spatial association/randomness.

Fig. 5.3 presents the intertwining relationship between spatial diversity and size of sector. The x-axis points out the impact brought by the sector of interest to economic diversity in the given grid cell ($\Delta H_L(Z)$), showing the magnitude of contribution that a sector has. Positive value shows its ability to induce higher level of diversity, in contrast to negative value that reduces the diversity degree in respected area. In the case

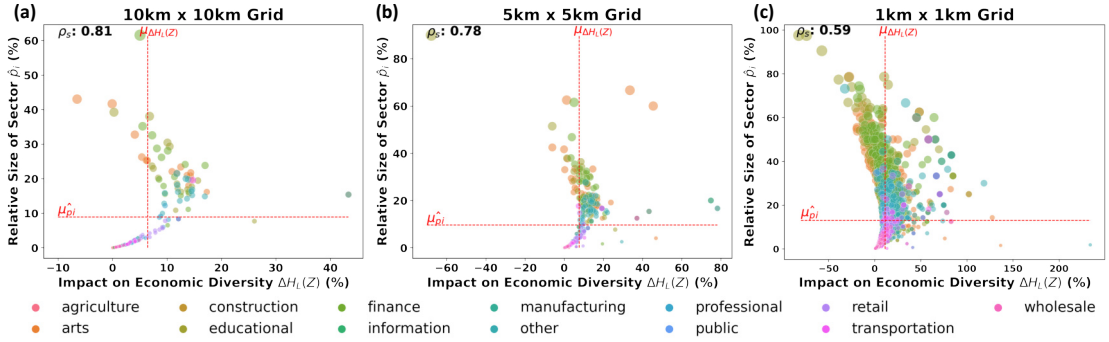

Figure 5.3: Sectoral entropy and size distribution. Various grid resolutions are presented: (a) 10km x 10km grid, (b) 5km x 5km grid, and (c) 1km x 1km grid. Dot size is normalised by number of POI per grid while colour denotes sector.

of zero value, no observable contribution coming from the sector is detected. On the y-axis, relative size of sector ($\hat{p}_i$) is plotted as percentage. The value is generated by dividing the number of POI per sector category by total number of POI in each grid cell. Maximum number of dot appears in every grid resolution equals to number of sector exist in every grid and multiplied by number of grid cell in that respected grid resolution. For instance, in 10km x 10km grid, there are 18 grid cells for which 13 sectors exist in total, resulting in 234 dots in the plot if all sectors are presents in each grid cells. For each grid resolution (Fig. 5.3a-c), vertical red dash is the average value of entropy change (denoted as the average impact to diversity) and horizontal red dash is the average value of relative sector size.

Cluster tendency is captured in the plot for all grid resolutions where transportation and wholesale sector are less affluent in bringing up the economic diversity to the local area and relatively small in size (less than 20% in proportion to sectors in the economy). Quite the opposite, arts, educational and finance sector are characterised by considerable role in shaping economic diversity, so do they have larger relative size. It is hypothesised earlier that larger sector has larger role in inducing economic diversity. This condition holds as Spearman correlation coefficient ρ ranges from 0.81 (10km x 10km grid) to 0.59 (1km x 1km grid), signifying high correlation between sectoral contribution to spatial diversity and size of sector.

5.3.3 Spatial effects

This section is dedicated to reveal the significance of scale-dependence spatial effects in shaping economic diversity. To operationalise this approach, we compare $\Delta H_L(Z)$, change in Leibovici entropy for each sector as an indicator for sectoral contribution to economic diversity, generated from empirical distribution of POI and ran-

dom model. The randomisation mechanism takes two procedures namely Complete Randomisation where no clustering tendency considered and Clustered Randomisation where each sector being located in a specific cluster. For each entropy computation, we iterate over spatial scale granularity respectively 10km x 10km, 5km x 5km, and 1km x 1km. The comparison is tested against statistical significance based on Kolmogorov-Smirnov Test introduced earlier in this paper (Section 5.2.1). This methodological construct allow us to further investigate whether higher economic diversity is expected at larger spatial scale.

The presence of difference in term of statistical significance indicates the existence of spatial embeddedness. As a concept, spatial embeddedness highlights the preference of businesses over locations [78] based on local institutional fabrics [115] that later incentifies the emergence of cluster [158]. Fig. 5.4 provides the result of Kolmogorov-Smirnov Test performed across spatial scales under two types of random models for that purpose. For both random models, the importance of spatial embeddedness becomes more evidence at higher spatial resolution (smaller grid size) because in Complete Randomisation, all sectors show statistically significant different impact to economic diversity comparing to empirical data in at 1km x 1km spatial resolution (Fig.5.4e). Similar pattern is also found in Clustered Randomisation for each sector except Agriculture (Fig.5.4f), considering diminutive size of agriculture being clustered together might not be distinctively observable. Moreover, Clustered Randomisation gives larger convergence to spatial embeddedness as replacing 10km x 10km grid with 5km x 5km grid results in more sectors showing different impact to economic diversity given the spatial distribution of POI (from 6 sectors to 10 sectors). In comparison, Complete Randomisation triples the number of sectors from 2 to 6 sectors and includes all sectors at the finest spatial resolution (1km x 1 km grid). Therefore, the Clustered Randomisation fits better as a random model in reference to real-life distribution in empirical setting and spatial and scale effect are more finely captured at higher resolution which in this case is 1km x 1 km grid.

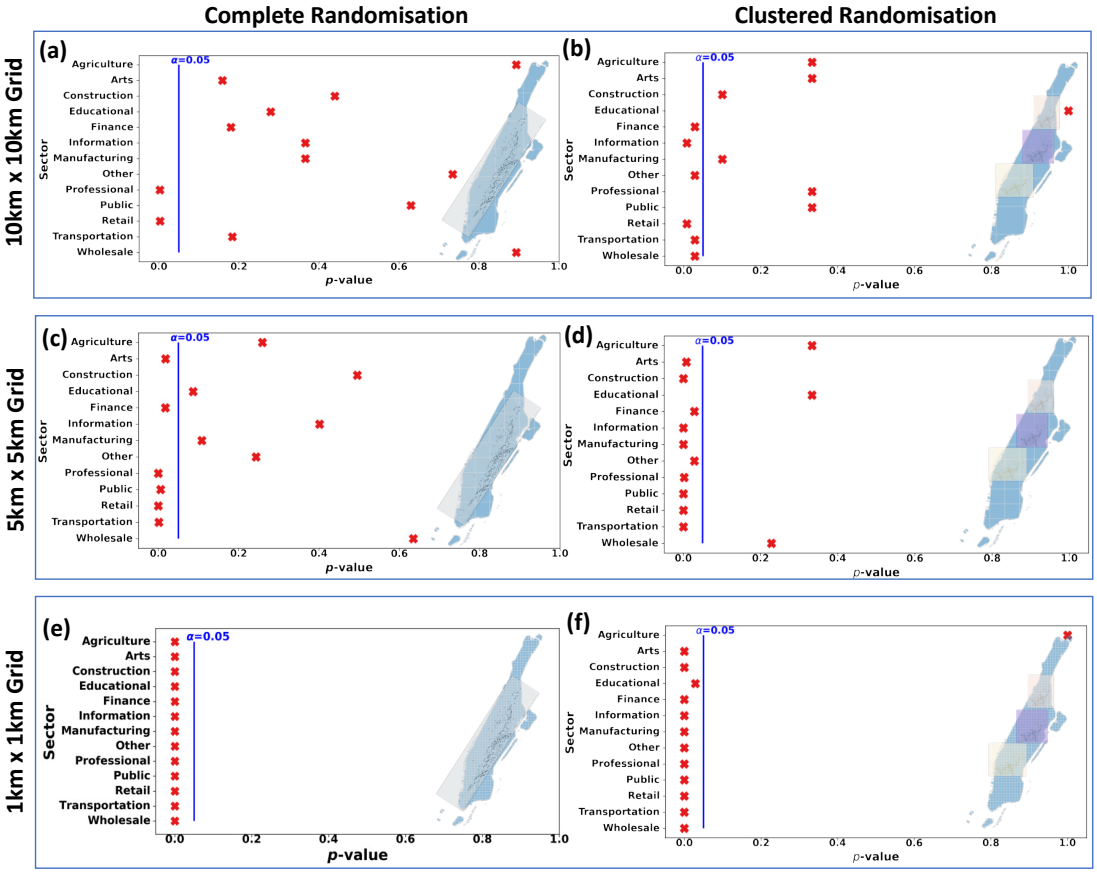

Figure 5.4: Kolmogorov-Smirnov Test for 10km x 10km grid, 5km x 5km grid and 1km x 1km grid. The squares on the map illustrates the way randomisation takes place. In the Complete Randomisation, there is only a single square exists, showing no sectoral clustering pattern in space. In contrast, Clustered Randomisation consists of multiple squares for which each is populated by POI within the same sector.

5.3.4 Sectoral proximity and colocation

To test whether spatial and sectoral proximity simultaneously trigger the emergence of linkages which further stimulates economic diversity, we analyse sectoral proximity and colocation derived from bipartite network. There are two bipartite projections we present in this section. First, bipartite projection on spatial linkage (Fig. 5.5) is constructed by pairing nodes if that area/grid cell has those sectors co-located. Second, bipartite projection on sectoral linkage (Fig. 5.6) is initialised by connecting the centroid of each area/grid cell if sectors are co-located in both areas/grid cells.

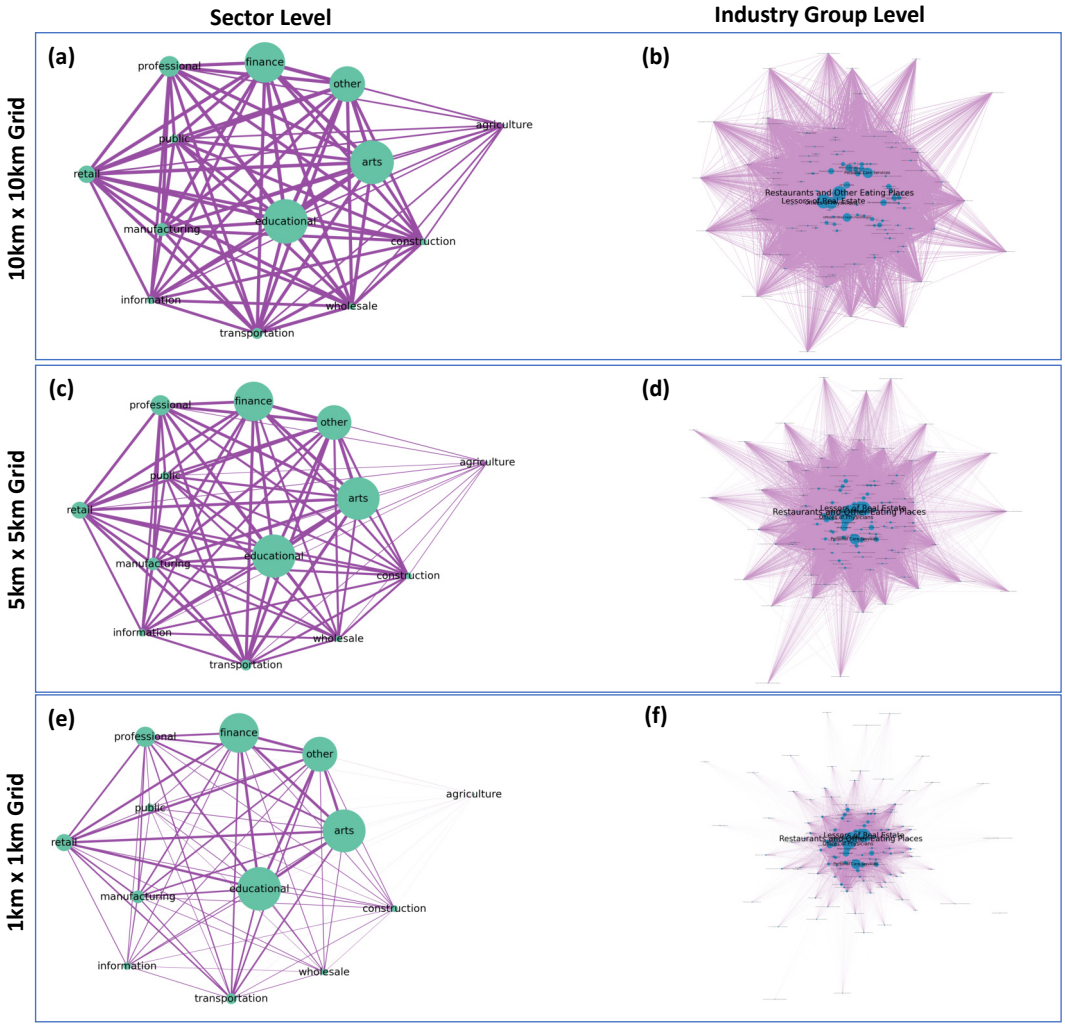

Figure 5.5: Bipartite projection on spatial linkage by grid resolution (row) and aggregation (column). Node is sector with size is comparable to the number of POI.

In Fig. 5.5, the size of each node is proportional to the number of POI. We observe robust structure regardless spatial grid resolution. There are strong closely connected main nodes at sector level (Fig.5.5a, c, and e): arts, educational, finance, and other. The

order of node size is also consistent at industry group level (Fig.5.5b, d, and f). It is in line with findings presented by Park et al. (2019) [151] where the leisure industry under aggregated 2 digit NAICS in our study (71/arts) is one of the most widely connected, exhibiting strong connections to many other sectors, including healthcare, education, art, media, and manufacturing. Moreover, it confirms the findings by Chapain et al. (2010) [41] and Rutten et al. (2011) [165] in which creative spillovers is the driving force behind diversification of activities and products in the larger spectrum of creative industry, including Arts (71), Educational (61), and Finance (53) that appears to be well connected and quite dominant in size.

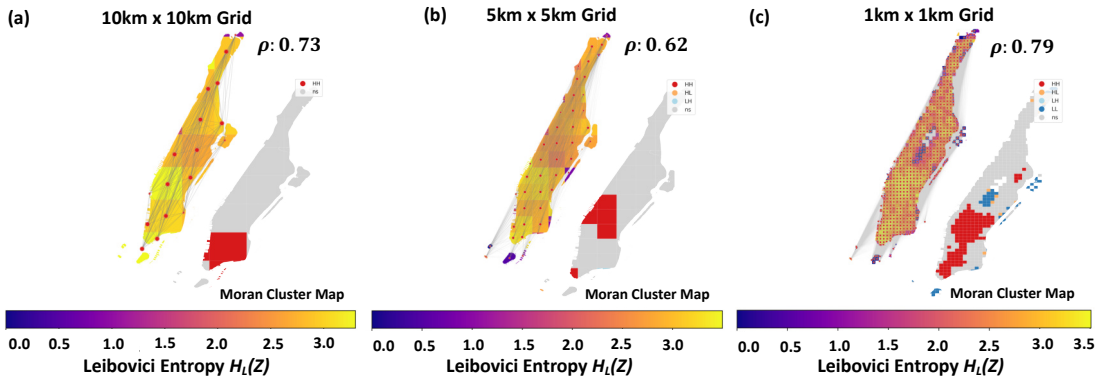

Figure 5.6: Bipartite projection on sectoral linkage by grid resolution: (a) 10km x 10km grid, (b) 5km x 5km grid, and (c) 1km x 1km grid. Node is the centroid of each grid cell.

As seen in Fig. 5.6, there is high correlation between Leibovici entropy (Eq. 2.3.1) and betweenness centrality (Eq. 5.2.4) of given grid. The area in the closely located to Lower Manhattan has the lightest colour, denoting the highest economic diversity, and remains consistent across spatial scale. Among others, the finest resolution at 1km x 1km spatial grid retains the highest correlation ρ equals to 0.79, implying that the conjunction between economic diversity and importance of location is seeded at localised boundary. This structure indicates the importance of economic diversity in determining the centrality of each area in mediating the business interconnection across sectors.

On the bottom right map in each grid resolution in Fig. 5.6, Moran Cluster Map is visualised based on computation of Local Form of Moran's I [9] in Eq.2.3.1. It delineates similar areas of particularly high/low economic diversity. Moreover, the appearance of read area captures the inclination toward cluster-like formation among areas with economic diversity higher than average (high-high/HH), meanwhile the dark blue represents cluster tendency among areas with lower than average economic diversity (lower-lower/LL). It is also possible that the surrounding areas instead show dissimilar value level, for instance light blue (low-high/LH) and orange (high-low/HL). Other-

wise, the space is filled by grey, exhibiting no statistically significant clustering pattern. Hence, it shows areas in which values are concentrated: Lower/Downtown Manhattan. It provides suggestive evidence about the processes that might be at work: economic diversity is optimised by the degree of urban process complexity (higher income area, higher similarity in term of socioeconomic profile) and the radius where similar level of economic diversity is concentrated (1 km for the smallest spatial component to 10 km for the largest spatial component).

5.4 Discussion and conclusions

The existing literature in spatial economics, economic geography, and regional science provide conceptual framework regarding the distribution of POI over space through location theory [78]. Spatial embededdness [120] take a steps forward to show that preference on location matters for business entities due to the presence of two types of positive externalities, the intra industry based known as Marshallian externalities [118] and the inter industry externalities recognised as Jacobian externalities [87]. In the following phase of urban dynamics, the later stimulates more businesses from diverse industries to come in urban territory, increasing the economic diversity level. Therefore, externalities are perceived as a pivotal means in facilitating spillovers that incentives cluster formation. Two causes of externalities are cited widely: industrial relatedness [136; 70] and spatial localisation [104; 158]. In this study, we label the first as sectoral proximity and the second as spatial proximity.

We identify the gap in this configuration especially on the the relation between spatial scale and sector size that colloquially seeds proximity through colocation. The objective we want to reach is to unravel the role of spatial proximity and spatial embeddedness in reproducing economic diversity across industrial sectors. Methodological building block and data driven computation is therefore necessary to bring more comprehensive analysis. Here we offer an artefact of measurement that allows us to disentangle the interdependence between geographical space and sectoral diversity through the lenses of network science by specifying the three forces behind urban economic diversity respectively the size problem (RQ1), the spatial scale problem (RQ2), and the proximity problem (RQ3).

Evidence from New York shows that Manhattan dominates the spatial distribution of POI. At sector level, arts, educational, and finance remain on the top three in business composition. On the size problem, our result suggests that larger sector has larger role in inducing economic diversity as shown by higher Spearman correlation coefficient ρ between relative size of sector ($\hat{p_i}$) and the impact brought by the sector of interest to economic diversity ($\Delta H_L(Z)$). Stronger entanglement is captured in larger area coverage noticeably 10km x 10km with ρ equals to 0.81, reiterating earlier proposition of diversity in bigger spatial unit [87; 61]. On the spatial scale problem, we find that the Clustered Randomisation at finer resolution (1km x 1km) gives a better fit to empirical distribution of POI, validating the presence of clusters in various sectors in the economy [115; 158; 78]. In addition, on the proximity problem, we show that the largest sectors namely arts, educational, and finance are closely connected in the network and remains robust across grid resolution and at different level of industrial aggregation tax-

onomy (sector level/2 digit NAICS as and industry group level/4 digit NAICS), in line with previous studies discussing the role of the aforementioned sectors in stimulating economic diversity [41; 165]. Furthermore, we reveal the dualism profile of spatial unit between being economically diverse and locationally important. A high correlation between Leibovici entropy and betweenness centrality (up to 0.79 in the finer scale of 1km x 1km) implies the importance of economic diversity in determining the centrality of each area in mediating the business interconnection across sectors as in the case of Lower Manhattan.

The results obtained from this research should be followed by studies that observe wider spatial coverage in the future. The limit in terms of spatial coverage that merely focuses in Manhattan could be tested further by replicating the artefact of measurement in other cities or urban area. Another direction to pursue is checking on smaller incremental scale limit (e.g.: incremental grid of 1km x 1km in a step) instead of only providing stand alone comparative scale setting like the one used in this study (e.g.: 10km x 10km, 5km x 5km, and 1km x 1km). The results of analyses of this kind will be better presented after its robustness could be extrapolated across space and scale in order to bring about integrated understanding on the complexity of urban economic diversity.

Chapter 6

Conclusion and future extensions

"Science must, over the next 50 years, learn to deal with these problems of organised complexity."

—Warren Weaver-

As a scientific work of its nature, this thesis is expected to contribute to the characterisation of city in accommodating socioeconomic mixing process. The analysis presented along the chapters fills the scientific queries related to human dynamics in urban setting by taking into account main spectrum, the spectrum drawn by perplexed and multifaceted elements that all together shape the social dynamics of a city: people visiting places from time to time, places located in various neighbourhood, people and places being socioeconomically stratified, as well as the interaction pattern commencing between people and places. One reflective thinking we formulate is based on the idea of complex system that flourished in the mid-century through the terms of *organised complexity* by Warren Weaver (1948) [194] and its presence in city as *"bits and pieces that supplement each other and support each other"* by Jane Jacob (1961) [89]. We leverage the analytical framework of network science with spatiotemporal approach due to the structural compatibility it offers in dealing with the phenomena under study, organised complexity in cities, respectively the elements in the system (nodes), the connections among elements (edges), and the way both shapes interaction dynamics.

6.1 Conclusion

In the initial chapter, this thesis sets out conceptual framework used in perceiving city as a complex system and urban mixing that takes place within its spatial boundary. Systematic review of state-of-the-art and mathematical representation of basic concepts

covering urban mobility network, urban morphology, and urban socioeconomic pattern are provided in Chapter 2.

In Chapter 3, we propose the integration Home Detection Algorithms (HDAs) into a data driven and network based computation to reveal mixing patterns of mobility. Case studies presented are the twenty largest cities of the United States in which they come to conformity regarding the emergence of strong signs of stratification in visiting preference. Interestingly, not only people are driven by "upward bias" to visit places with higher socioeconomic status whenever possible, but the degree of such bias is also positively correlated with own socioeconomic status and racial residential segregation. Consequently, uneven mobility mixing patterns become more prominent in cities. Therefore, policy aiming to improve social cohesion should take into account this persistent aspect by designing inclusive shared space for people across socioeconomic class to have potential encounters. Alternatively, promoting adequate living quality across neighbourhood to make it appealing for anyone to live side by side is also a direction to pursue.

In our study presented in Chapter 4, we show that segregation pattern in mobility is not a static object. External disturbance such as COVID-19 outbreak affects existing mobility configuration and contributes to the changing residual isolation. We approach the problem by implementing measures on dynamic assortativity in mobility network based on sliding time window, complimented with entropy on inseparable dimensions of individual trajectory: heterogeneity of locations and their respected socioeconomic status. We apply the methods in four metropolitan areas across hemisphere namely Bogota, Jakarta, London, and New York. In comparison among them, we observe that the degree of mobility restriction especially the highest during lockdown period induces increasing segregation in mobility which on the later stage seeds the emergence of long-term effects on socioeconomic mixing. This approach allows us to look at broader temporal spectrum on how epidemic affect segregation phenomena in mobility that turns to be country-dependent and socioeconomic-dependent, a feature that should be carefully considered in proposing better policy.

The final research part in Chapter 5 demonstrates that the heterogeneity in urban context is not limited to mobility aspect, but also manifested in spatial distribution of places of interest (POI). Performing entropy measure on space as suggested by spatial entropy, among others Leibovici entropy, serves as the first step to disentangle the configuration of urban economic diversity as a complex phenomenon. We introduce the use of scale dependent random models as a reference to validate empirical mechanism that colloquially seeds economic diversity constituting sectoral size, spatial embededdness, and sectoral-spatial proximity. Furthermore, we synthesise insight from complex sys-

tem and network science by projecting bipartite network on spatial and sectoral linkage. We show the result from New York with focus in Manhattan and the findings suggest that sectors with larger number of firms or entities form close knitted connections in the network and largely contribute to urban economic diversity. Furthermore, it implies the presence of dualism profile that a spatial unit has, between being economically diverse and locationally important as denoted by high correlation between the two. In retrospect, urban planner could translate this attribute in the agenda making of more equitable cities.

6.2 Future extension

Capturing socioeconomic mixing process in cities is the ultimate goal of this presented thesis. We offer methodological framework and empirical evidence to investigate structure and dynamics of embedded segregation which occurs not only in the preference over visiting places in mobility, but also the selection of locations in commercial activities. Through the perspective of mobility of people, the measures we propose suggest that segregation is not trivial. In cities, segregation is pervasive, reaching wide range of human dynamics where mobility presents as the intermediary which amplifies the existing residential segregation. This degree of segregation is even elevated when restriction imposed on mobility, producing higher isolation to the already stratified socioeconomic groups. Through the perspective of spatial distribution of places in commercial activities, similar logical consequence also appears. An area with more homogeneously segregated economic activity tend to be afar from gaining centrality. Afterall, these discussed findings fulfil the original hypothesis stated earlier in which segregation is a consequence of stratification and homophily leaning in preferences.

As an extension in the future scientific agenda, building probabilistic model to estimate the occurrence of segregation given mobility pattern and socioeconomic attribute could be considered. This direction is taken into account by fitting the Exploration and Preferential Return (EPR) model, a modelling approach to derive properties of individual human mobility by estimating the decision on either visiting new places or returning to usual places given related parameters namely distance, waiting time and the jump length. We expect to empirically determine the probability that two people may meet (e.g.: share a public space) by chance at any location in a city in terms of their socioeconomic status or any other attributes. As a result, more refined understanding on the intertwining dimensions in segregation or mixing process becomes highly visible.

In achieving them, there is a limitation regarding the resolution of data. The show-

cases we present in this thesis unravel the data driven approach and computational capabilities in handling heterogeneous sources of data in urban mobility and complex system. The use of Social network platform such as Foursquare, Open Street Map (OSM) and POI, as well as digital trace data for instance anonymous smartphone records that have been intensively used in this thesis could be enriched by data extracted from smart transportation card, public Wi-Fi signals, and Call Detail Records (CDRs)/ eXtended Detail Records (XDRs). The role of future technology let alone the forefront of super computing power is also incredibly important because it allows the deployment of more realistic analysis that is not only limited to static measure but also simulation of segregation dynamics at city level in the form of Agent Based Modelling (ABM) and Digital Twins. A successful analysis would increase the predictability of epidemic outbreaks and help answer several societal challenges such as efficient urban design and inequality reduction based on the main findings in this thesis namely the importance of mixing process in urban encounters from both the perspective of people (mobility and visiting preference) and places (location and land-use mix preference).

Chapter 7

Appendices

7.1 Socioeconomic biases in urban mixing patterns

7.1.1 Summary statistics

This study focuses on the presence of mixing patterns in the largest 20 urban areas in the US. We consider urban area as the level of analysis because this densely developed territory fits very well with our objective in capturing the interaction between peoples and places. While people are represented by Foursquare users (who are also Twitter users in this case due to the crawling technique used in data collection), places are functional point of interests (POI) in accordance with the Foursquare venue category hierarchy. After combining the selected geographic and cartographic information from the U.S. Census Bureau (American Community Survey/ACS 2012) and Foursquare check-ins data (April 2012 - September 2013), we infer a set of active users and urban land uses (e.g.: residential, commercial, and other nonresidential) along with their socioeconomic status (SES).

During the SES assignment procedure, first we identify the home location of Foursquare users and group them into distinct socioeconomic classes such that each class presents the same size. To signify the representativeness of Foursquare users, we create synthetic data based on the population size and income. We segment this sorted list into 10 equally populated groups of the lowest income in class 1 and highest income in class 10 as seen in Fig. 7.1. Meanwhile, to assign the SES of POIs in Fig. 7.2, we match each POI with the boundary income of users. Recalling that we segment users to equal SES, this SES segmentation is used for POIs. Consequently, POIs are not equally distributed in each city as it is less plausible in reality that the spatial distribution of POIs identically resembles residential distribution all over space in the city.

Fig. 7.3 shows the bootstrapping result as a strategy to deal with the data represen-

Urban Area	Number of population	Average income (US$)	Number of users	Average income of users (US$)	Number of POIs	Number of check-ins
Atlanta	1,515,880	32,154	1,597	40,880	12,729	117,403
Austin	964,309	32,508	1,101	32,127	7,654	59,698
Baltimore	1,558,743	30,275	1,006	33,762	9,697	66,848
Boston	2,090,520	39,810	1,653	49,789	13,548	103,395
Charlotte	864,236	31,520	614	41,298	5,579	42,730
Chicago	2,540,254	28,279	2,318	55,997	18,059	166,133
Dallas	1,368,607	26,425	884	43,966	6,160	49,683
Houston	2,027,223	27,149	1,063	45,351	9,967	67,999
Los Angeles	2,066,612	27,162	1,284	41,929	10,101	90,234
Miami	2,146,572	23,599	1,542	34,445	10,509	74,525
Mineapolis	2,074,556	34,483	1,109	35,230	12,784	97,669
New York	7,669,696	32,225	1,353	50,457	31,060	191,564
Philadelphia	2,430,063	28,960	1,369	37,958	12,445	100,518
Phoenix	1,806,374	23,744	969	28,537	8,746	67,922
San Diego	1,732,145	29,692	1,330	41,386	11,783	84,476
San Francisco	2,059,293	39,878	2,348	51,715	16,353	122,846
San Jose	1,446,254	38,089	762	42,280	6,918	41,023
Seattle	1,316,257	41,967	1,148	43,595	10,260	63,755
Tampa	1,120,985	27,904	765	30,888	7,506	55,255
Washington	600,566	45,237	2,287	58,226	17,662	166,627

Table 7.1: Summary statistics. It summarises size of dataset used in this study, including population information (eg: number and income) extracted from census as well as Foursquare users and POIs in the 20 Urban Areas. Number of population is the summation of real population in the census tract that appears in the Foursquare dataset. Number of users is counted from pool of users with identified home locations after running home location algorithm explained in Section 7.1.2. Number of check-ins is the total frequency of users tagging themselves at every listed POI in the urban area where they live.

tativeness issue in this study. As previously seen in Table 7.1, the ratio between the number of Foursquare users and population could be quite small in some cities. Therefore, it is necessary to estimate the potential fluctuations of the distribution of the SES of people due to the small sample of populations. In response to this issue, we design a bootstrapping method. We apply bootstrapping with replacement by taking the original Foursquare population size for each city to generate random samples of the SES of Foursquare users for 1000 iterations in each city. After obtaining a re-sampling distribution, we compute mean values of the SES of Foursquare users for each iteration and plot them as histograms to observe the bootstrapping distribution. The SES distribution of Foursquare users is similar to that of the underlying population, since the 95% confidence interval is quite narrow with boundary value less than 1 SES class.

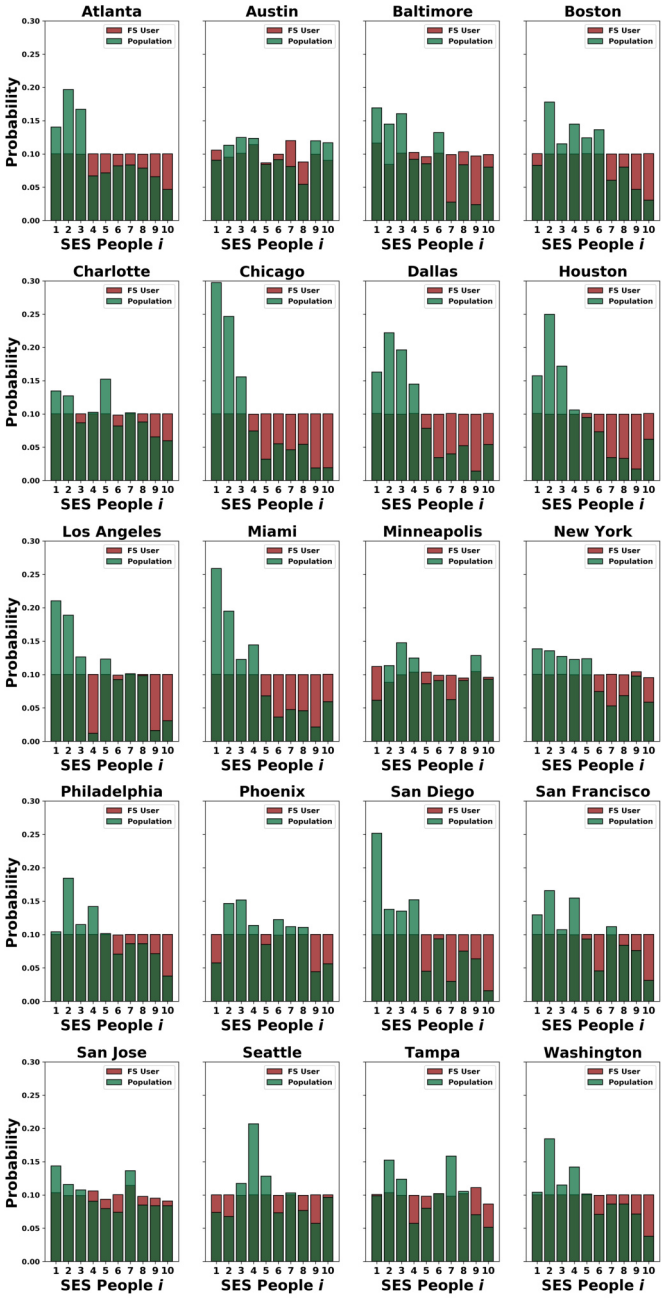

Figure 7.1: Data representativeness of Foursquare User. The representativeness of the user data is visualised as density distributions between the size of populations from census (green bars) and number of Foursquare users in each SES class (red bars). In general, it reveals over-represented rich people in some cities in the Foursquare data, as red bars are above green bars for high SES classes.

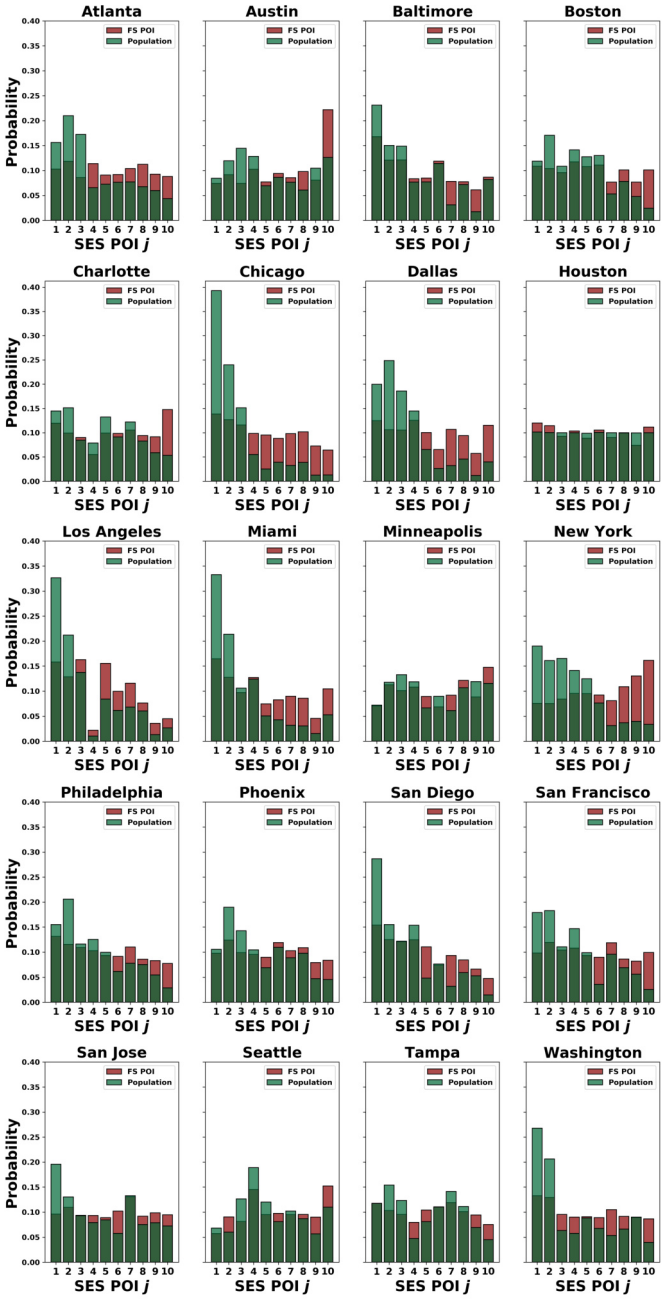

Figure 7.2: Data representativeness of Foursquare POI. The representativeness of the POI data is visualised as density distributions between the size of populations from census, and number of Foursquare POI in each SES class. Given the variability of SES distribution of POI, it's not always the case that POI are highly concentrated POI in richer areas in even though it is quite common.

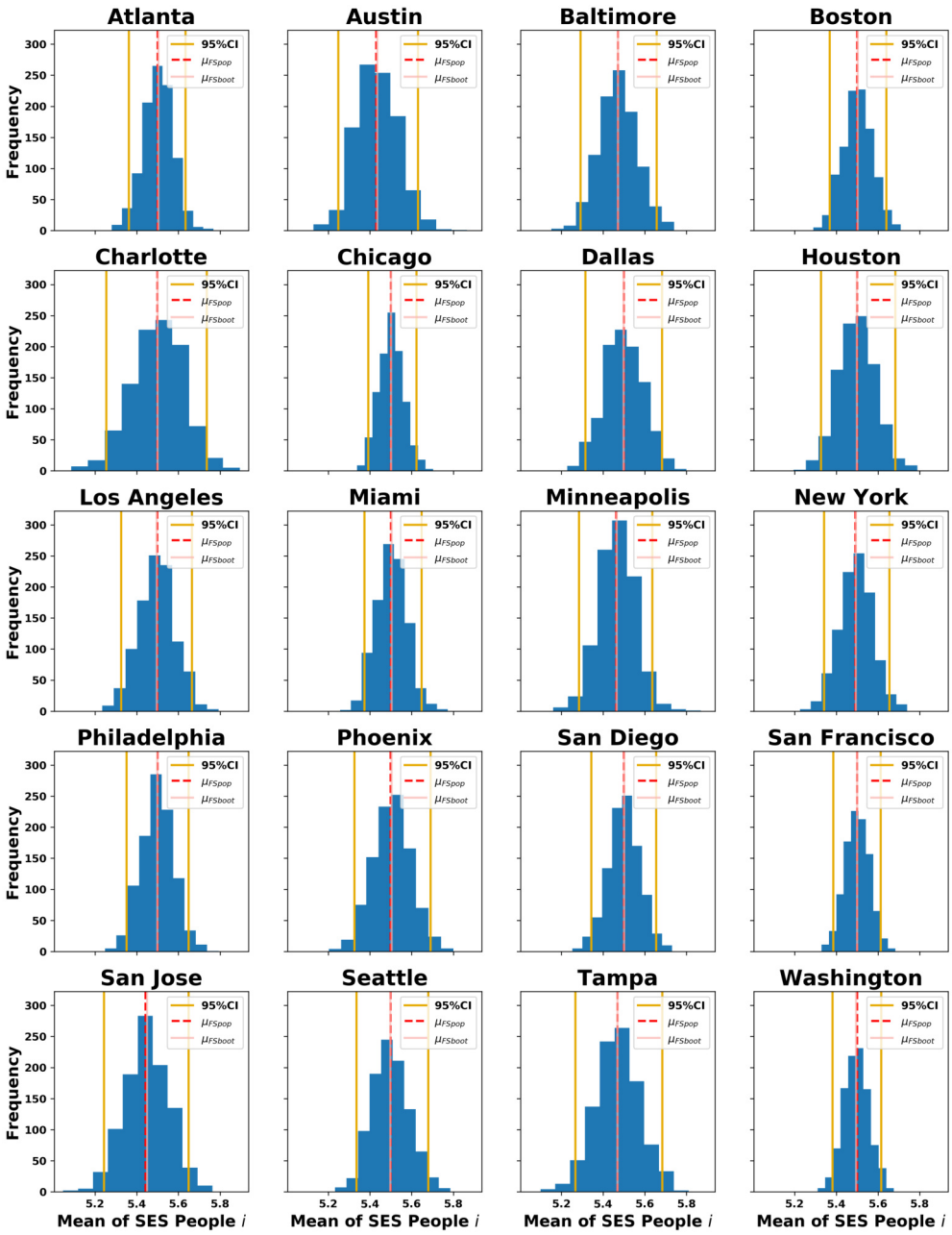

Figure 7.3: Bootstrapping. We perform bootstrapping to estimate potential fluctuations in the results due to limited sample size in the dataset. As the results show, the repeated samples have mean value (light pink line) in close approximation to the mean value of the Foursquare user population (red dash), while the yellow line represents 95% confidence interval. Therefore, the bootstrap method provides a solid ground to examine the proximity of the sample to the original distribution.

7.1.2 Home location inference

Detecting home location is a primary step in dealing with mobility data because spatial identifier serves as an intermediary instrument among heterogeneous source of data, including census data. Various decision rules have been developed to identify the whereabouts of people reside. In mobility literature, a single rule home detection algorithm is widely applied in both continuous (eg: global positioning system/GPS data) and non-continuous location traces (eg: call detailed record/CDR data). Home is defined as the location where highest proportion of activities occurs during night hours with variations regarding time window [36; 49; 154].

For each individual mobility trajectory, we identify which locations are categorised as home and venues, including visits to places located in their own SES tract. Home inference algorithm is constructed based on daily temporal windows. It consists of 8 time slots starting from midnight with 4-hour intervals. The main criterion of home location is consistent check-ins from 9PM to 6AM (time window 1, 2, and 8) at locations labelled by Foursquare as 'Assisted Living', 'Home (private)', 'Housing Development', 'Residential Building (Apartment / Condo)', and 'Trailer Park'. The rank of home location candidacy is sorted based on frequency of check-ins. We iterate the process for other venues to accommodate users that might live in the downtown, next to multi-purposes buildings attached to various urban functionalities (e.g.: a flat on the top of deli or store). This algorithm is designed with ability to properly identify the logical consequence of urban built environment where the human activities frequently take place at a location serving numerous settings as previously mentioned.

7.1.3 Distribution of Mobility Stratification

This section is dedicated to show the socioeconomic distribution of area where POIs are located and the socioeconomic segmentation of people visiting them. POIs are grouped based on a category taxonomy extracted from Foursquare API called category labels. The hierarchical taxonomy record allow us to characterise the place to different levels of granular categorisation. For this purpose, we opt for parent categories: 'Arts and Entertainment', 'Business and Professional Services', 'Community and Government', 'Dining and Drinking', 'Health and Medicine', 'Landmarks and Outdoors', 'Retail', 'Sports and Recreation', and 'Travel and Transportation'.

Fig. 7.4 shows that most of POI categories across cities have positive values, indicating assortativity and the presence of considerably high homophily mixing. Meanwhile, disassortative visiting patterns only characterise visits to POI categories in Miami ('Community and Government'), San Diego ('Health and Medicine'), and Tampa ('Health and Medicine'). Furthermore, we observe deviations of disaggregated r from aggregate r to both directions and none of single POI category consistently and uniformly dictates visit patterns in all cities.

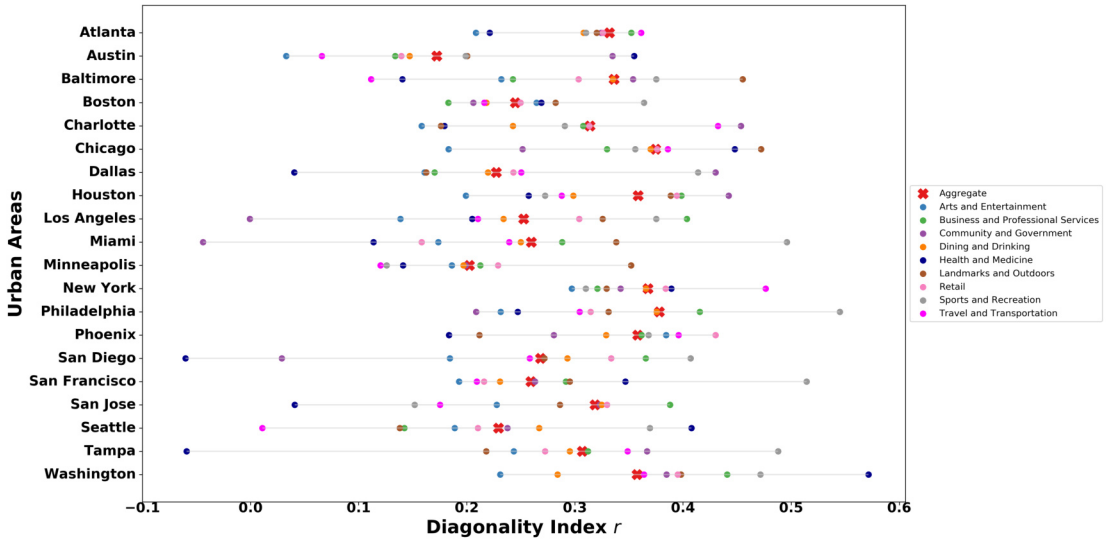

Figure 7.4: Diagonality Index of Arts and Entertainment POI. We further demonstrate this point by a comparative summary of the diagonality index r for all visits (aggregate, red cross) and visits to particular places (by poi categories, coloured points). Aggregate r takes positive values, ranging from about 0.2 to 0.4, while disaggregated r retains a broader distribution. The diagonality index r for 'Arts and Entertainment' (light blue dot) deviates from the aggregate r (red cross), implying that variability in visit patterns are not merely triggered by the visit patterns to particular POIs, for example in 'Arts and Entertainment'.

7.1.4 Empirical stratification matrices

The mixing pattern in mobility network conceptualised in Section3.3.1 measures the proportion of visits by people to places stratified by their respective SES. According to the empirical stratification matrices $M_{i,j}$ shown in Fig. 7.5, we find that individual mobility in some cities is less stratified than others, in which colour of bins fades away from diagonal elements such as in New York and Austin. In contrast, strong empirical homophily mixing appears in most cities, for instance in Phoenix, Philadelphia, and Tampa.

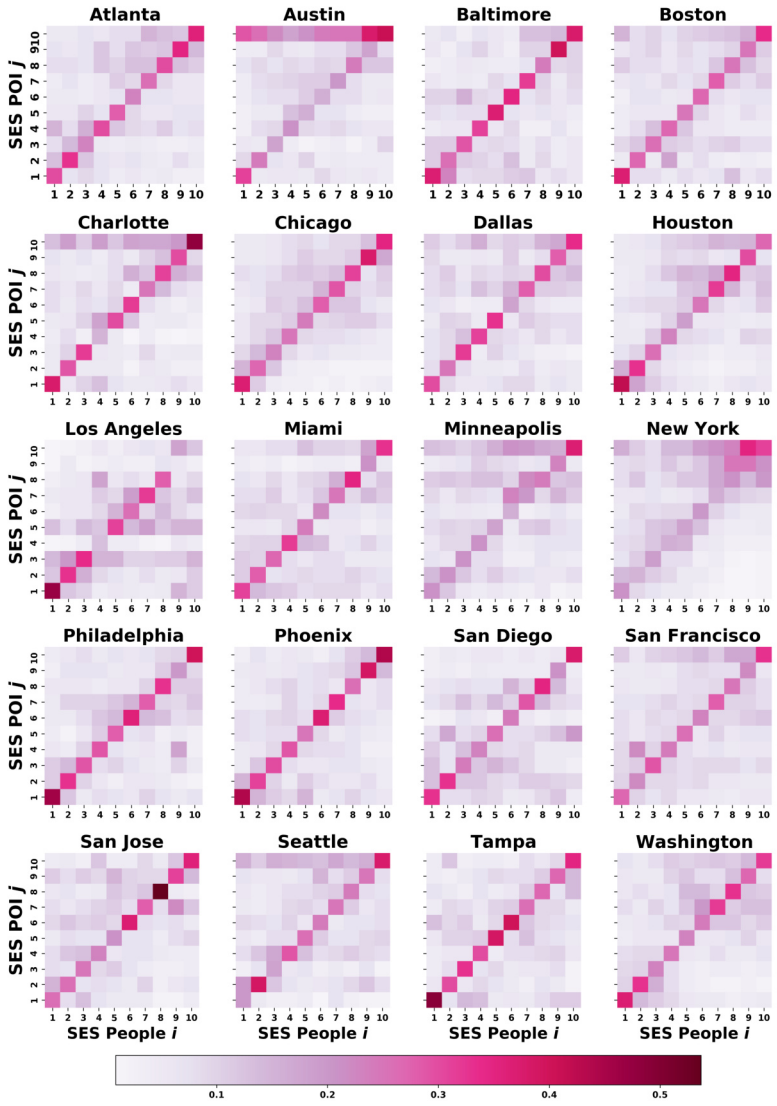

Figure 7.5: Empirical stratification matrices $M_{i,j}$. Each bin represents the visiting probabilities of individuals from a given class to places of different classes. The darker colour shades of bins represent larger visiting probability.

7.1.5 Normalised stratification matrices

We replicate methodological approach of the empirical stratification matrices $M_{i,j}$ used in Section 3.3.3 to construct normalised stratification matrices $N_{i,j}$. In Fig. 7.6, red colour amplifies the higher frequency visit than expected, while blue denotes fairly less visits. Furthermore, the presence of red blocks in the upper diagonal elements signifies the upwards bias where people are aspired to drop by at places with higher SES. The bold red gradient along the diagonal exhibits homophily mixing.

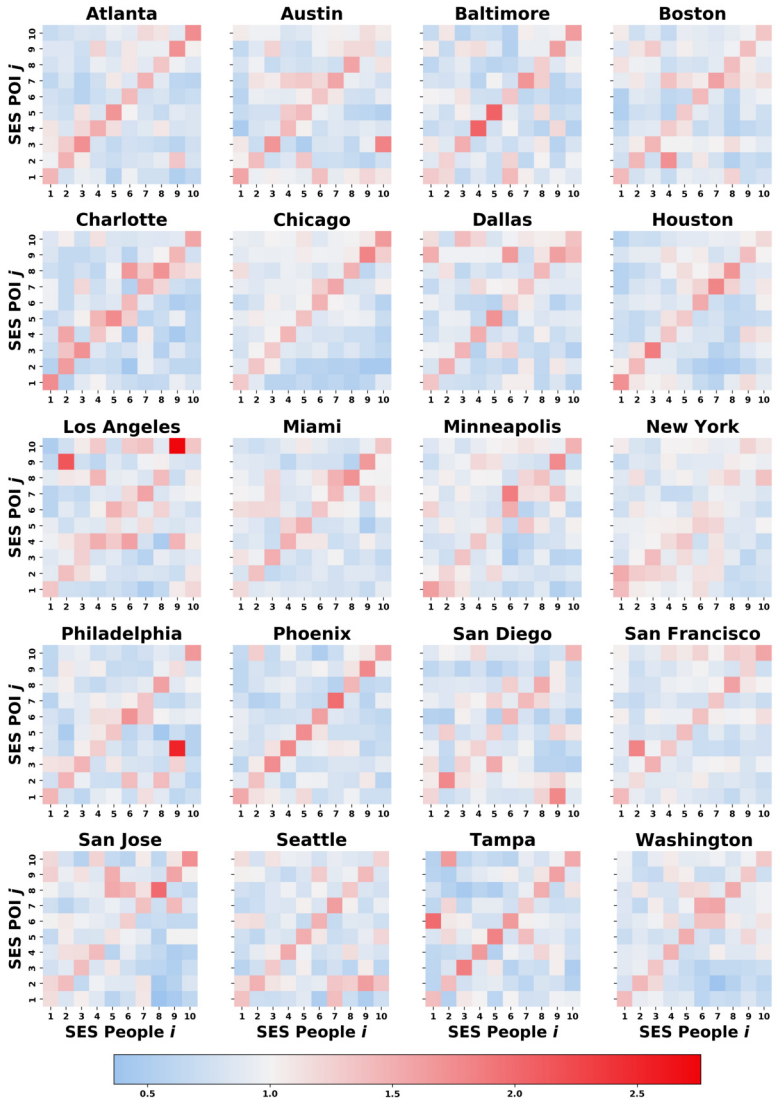

Figure 7.6: Normalised stratification matrices $N_{i,j}$. Each bin contains the fraction of the empirical and randomised stratification matrices. Colour gradients ranges from darker blue to darker red, taking the condition of lower to higher frequency visit than expected.

7.1.6 Individual bias z-score

The individual bias z-score $z_u^{B_u}$ (Fig. 7.7) measures how far empirical individual bias score deviates from the random mobility model. The technical formulation is explained in Section 2.3.1. Similar to typical individual mobility in other studies, individual mobility in our data is also characterised by large jumps between two types of people: a small proportion of people with large trajectories and a largely dominant type with small trajectories. Fig. 7.7 reveals the coexistence of upward and downward biases in terms of visiting patterns to other socioeconomic classes in any class.

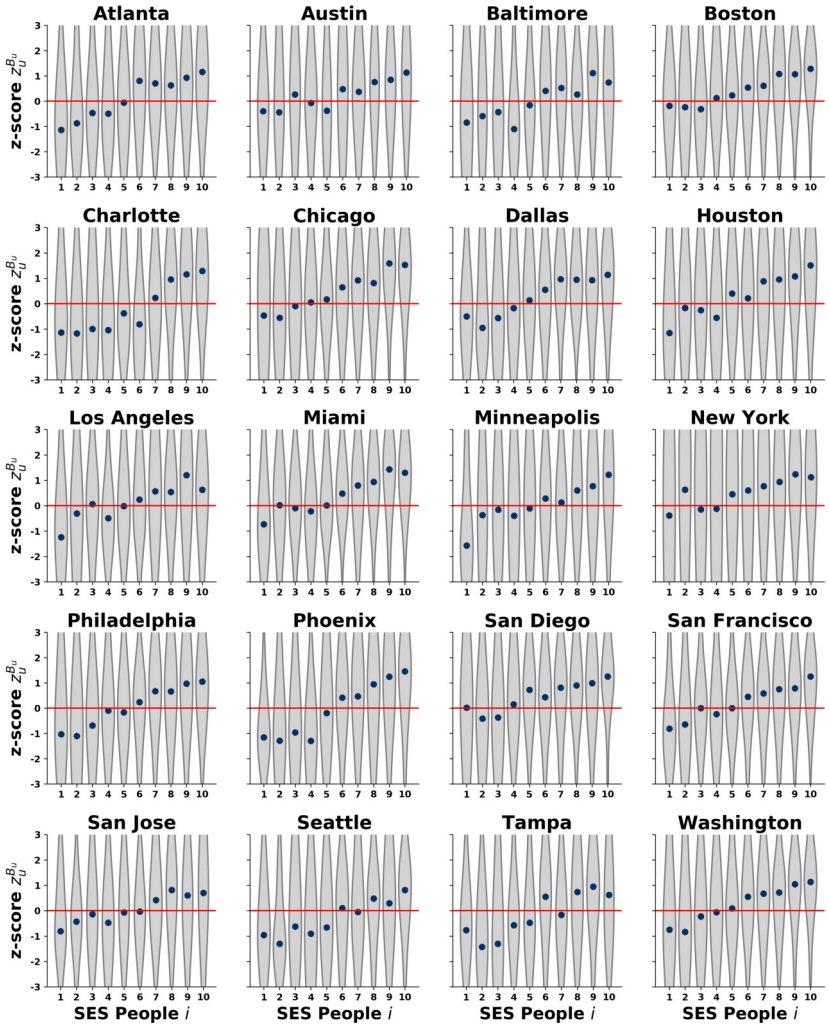

Figure 7.7: Individual Bias z-Score $z_u^{B_u}$. Class distributions of individual z-scores reflects how much the individual bias differs from the expected bias for an individual who chooses places to visit with the same frequency as before but selects them from a given set of places dictated by others within the same socioeconomic class.

7.1.7 Class-level bias z-score

The class-level bias z-score $z_u^{c_u}$ (Fig. 7.8) is formalised to quantify the difference between typical visited places in an individual trajectory from the trajectory in the random class average. The composition of visit patterns differs with respect to SES. Many places, usually within its own range, are visited many times along the trajectory. Meanwhile, some others are only visited less frequently. We dedicate Section 3.2 to formulate and discuss this measurement.

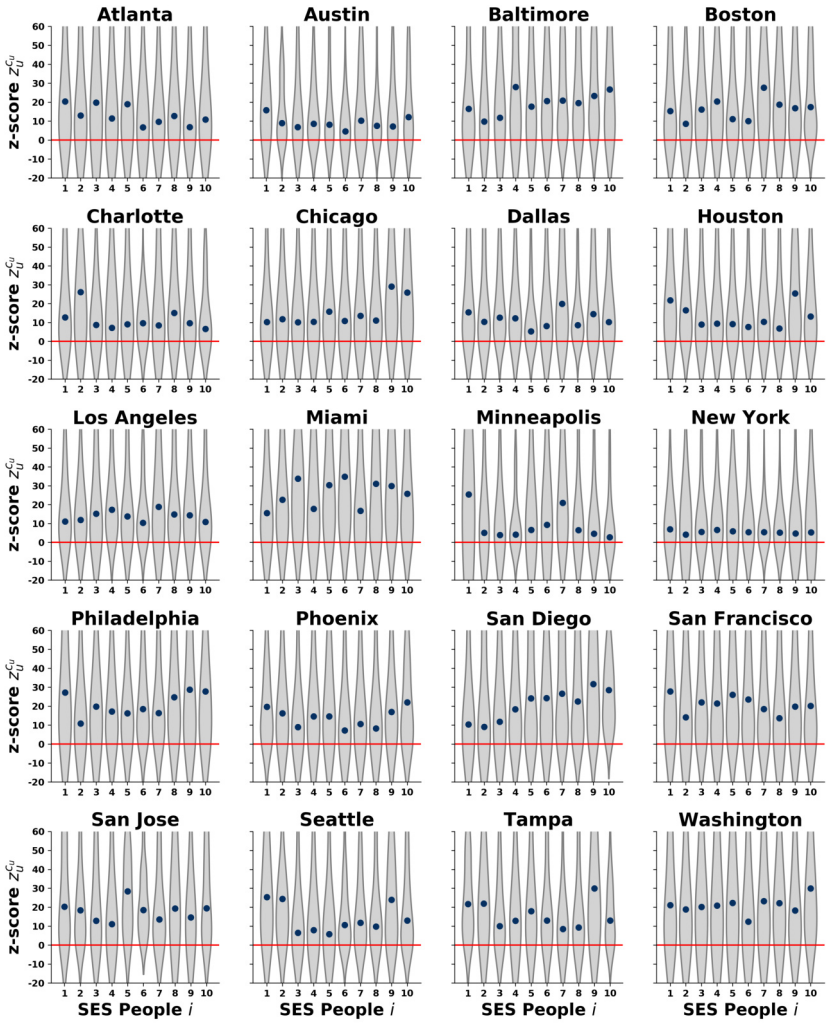

Figure 7.8: Class-level Bias z-score $z_u^{c_u}$. Distribution of class-level biased z-scores reflects directly how much the individual behaviour deviates from the expected level, when the individual could choose randomly places to visit from a given set dictated by others from the same socioeconomic class.

7.1.8 Out-of-class empirical stratification matrices

The out-of-class empirical stratification matrices $Mc_{i,j}$ shown in Fig. 7.9 take the same methodological approach as the empirical stratification matrices $M_{i,j}$ shown in Fig. 7.5. The two differ in term of the census tract scope, since the out-of-class measure excludes the own census tract in each individual trajectory. This step is taken into account to control distance effect that contributes to the homophily mixing condition. Even though we observe less diagonality here, some stratification patterns remain in several cities.

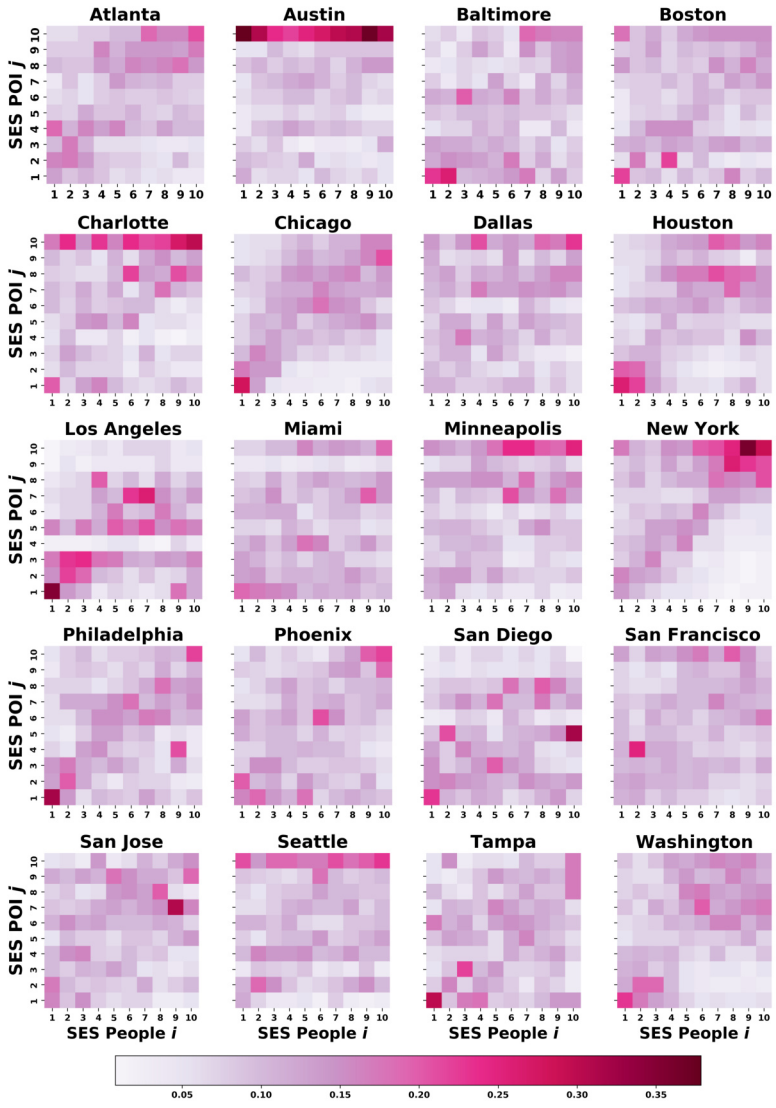

Figure 7.9: Out-of-class Empirical Stratification Matrices $Mc_{i,j}$. Probabilities that individuals from a given class visit to places of different classes are visualised with colour shades. The darker the bin, the higher visiting probability.

7.1.9 Out-of-class normalised stratification matrices

We construct out-of-class normalised stratification matrices $Nc_{i,j}$ Fig. 7.10) in order to examine the robustness of out-of-class empirical stratification matrices $Mc_{i,j}$ Fig. 7.9). A normalised version of out-of-class stratification matrices is generated to measure the difference in magnitude between the out-of-class empirical stratification matrices and random visit occurrence. We observe that an upward bias tendency (red gradient in upper diagonal matrices) is still considerably present in some cities and such distinct pattern does not emerge by chance.

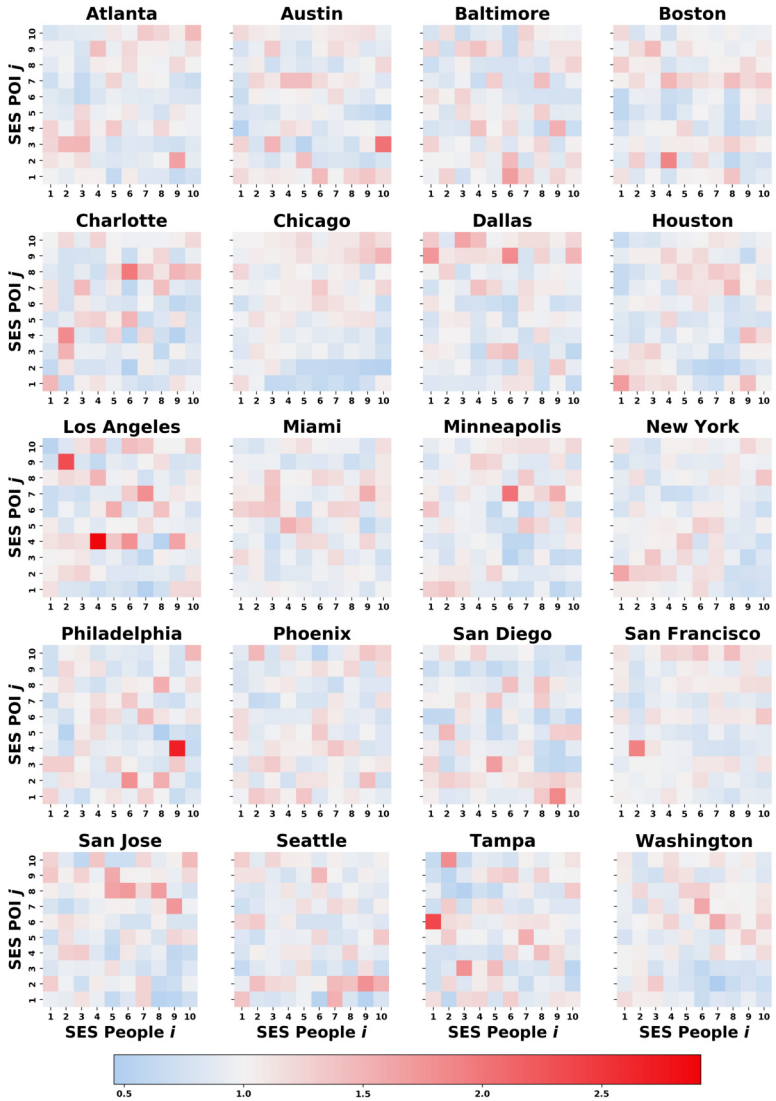

Figure 7.10: Out-of-class Normalised Stratification Matrices $Nc_{i,j}$. Defined as the fraction of the empirical and randomised stratification matrices without own census tract, out-of-class normalised stratification matrices reveals the visiting patterns of upward bias (red gradient), downward bias (blue gradient), and no bias (white) tendency across a SES pair of people and visited places in their trajectories.

7.1.10 Out-of-class individual bias z-score

The out-of-class individual bias z-score $zc_u^{B_u}$ depicted in Fig. 7.11 takes the same methodological approach as the individual bias z-score $z_u^{B_u}$ shown in Fig. 7.7. The two differs in term of census tract scope where out-of-class excludes the own census tract in each individual trajectory. This step is taken into account to control the distance effect that contributes to biases in visiting patterns, recalling that mobility mostly takes place in home census tract or nearby locations.

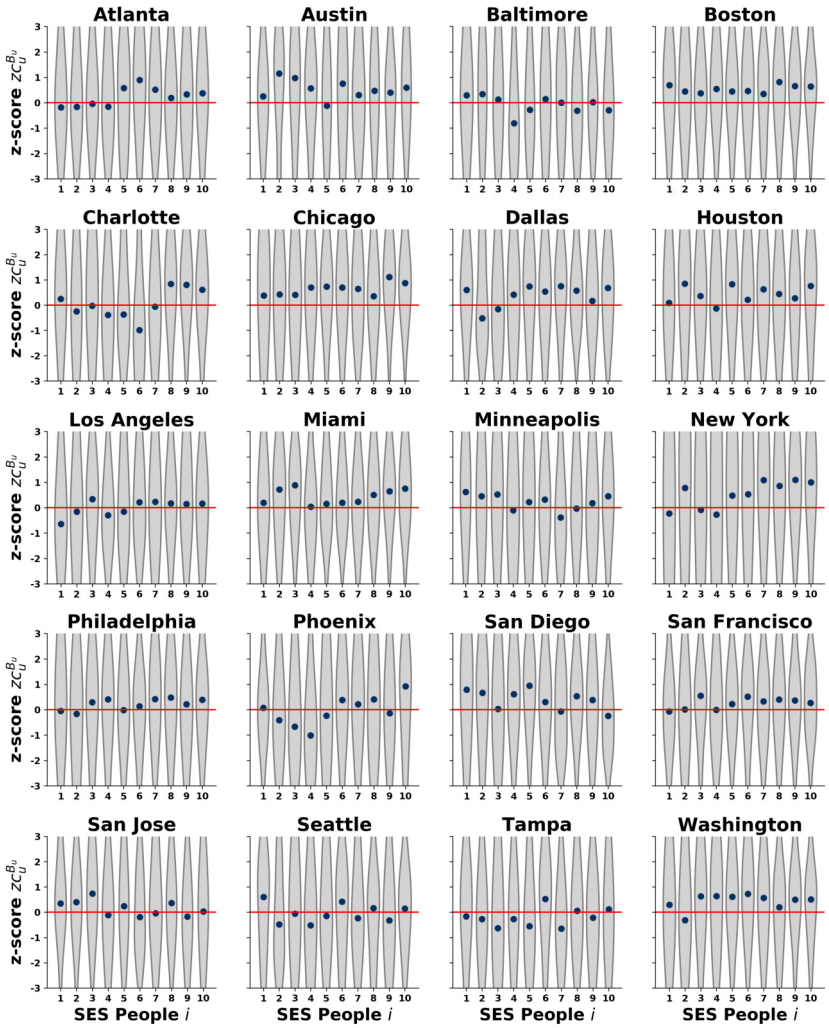

Figure 7.11: Out-of-class Individual Bias z-score $zc_u^{B_u}$. We show class level distributions and their median values (blue dots) for each socioeconomic class after removing own census tract. Unbiased condition is depicted by horizontal red line. Upper red line areas contain upward biases in terms of visiting patterns, in contrast to downward biases on the lower part.

7.1.11 Out-of-class class-level bias z-score

The out-of-class class-level bias z-score $zc_u^{c_u}$ (Fig. 7.12) indicates if this individual bias is weaker or stronger than expected from random behaviour, considering the absence of visit to places located in the own census tract. It reflects directly how much the individual behaviour deviates from the expected level, when the individual could choose randomly places to visit from a given set dictated by others from the same socioeconomic class.

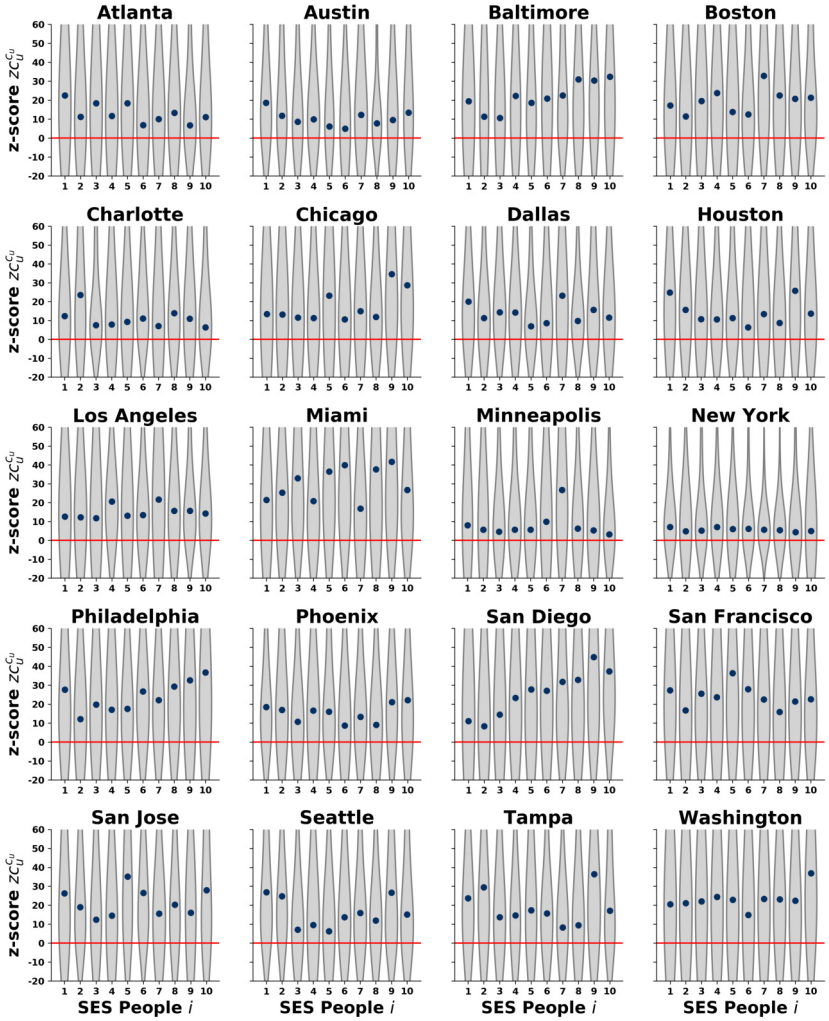

Figure 7.12: Out-of-class Class-level Bias z-Score $zc_u^{c_u}$. After removing own census tract, distributions are shown for each socioeconomic class with their median values as blue points. Above the red unbiased median line, an upward visiting bias holds, otherwise upward visiting bias is prevalent.

7.1.12 Out-of-class measures in Houston, New York, and San Francisco

In Section 7.1.8- 7.1.9 and Section 7.1.10- 7.1.11, we raise the issue of potential confounding factors driven by distance effects. It is important to realise that shortcomings may arise from this particular computation which disregards distance aspects. In dealing with that, we remove mobility data in own census tracts and recompute the stratification matrices as well as bias measures. In line with the discussion in the main text (Fig. 3.2), this section is dedicated to analyse the robustness and sensitivity of our homophily mixing and visiting bias measures in Houston, New York, and San Diego. We find that mobility is still stratified by SES as seen in Fig. 7.13a, 7.13b and 7.13c. Moreover, the presence of upward bias tendency is still visible in those cities by looking at Fig. 7.13d, 7.13e, and 7.13f. It is confirmed by the fact that mean of upper diagonal matrix elements surpass mean of lower diagonal matrix elements in 7.13g.

In a deeper analysis, we exploit the persistence of biased behaviour in mobility by employing procedural bias measures at individual and class level that have been introduced earlier in Section 7.1.8-7.1.9 and 7.1.10-7.1.11. In the case of the out-of-class individual bias z-score $zc_u^{B_u}$, visiting bias is less expected among lower class people in New York (Fig. 7.14b), while anyone in Houston (Fig. 7.14a) and San Diego (Fig. 7.14c) can be as much biased in their visiting patterns regardless their socioeconomic classes. Upward visiting bias is even stronger at class level in all three as none of z-score values fall below the red unbiased median line (Fig. 7.14a, Fig. 7.14b and Fig. 7.14c).

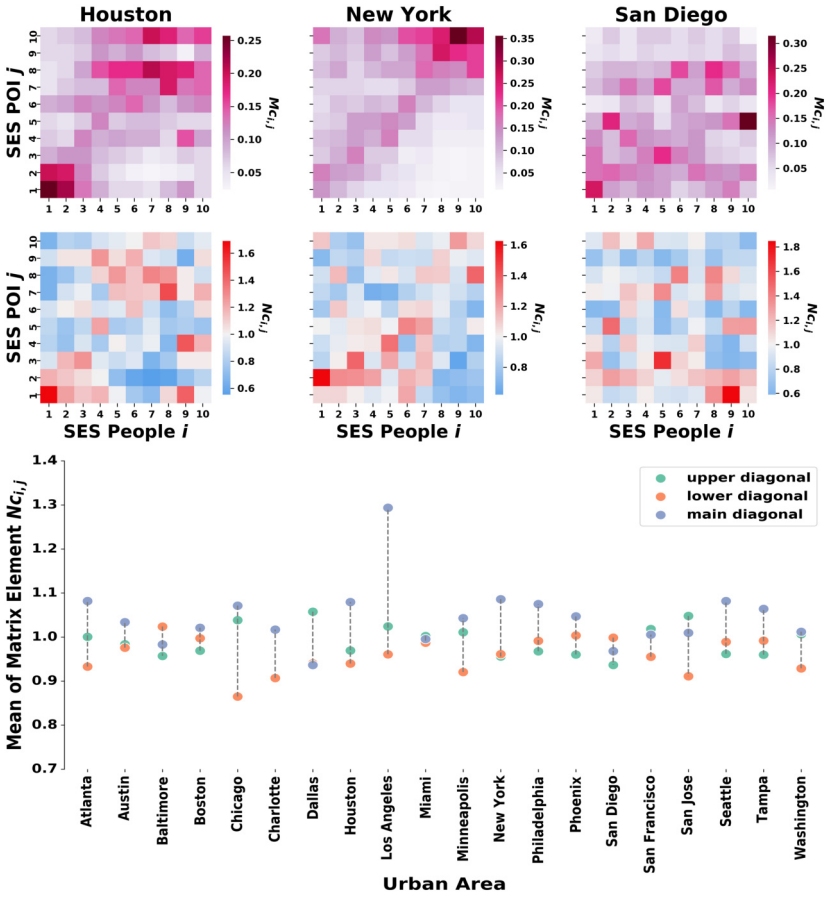

Figure 7.13: Out-of-class socioeconomic stratification matrices. Out-of-class socioeconomic stratification matrices are constructed based on individual trajectories or a set of locations visited by each individual after the removal of own census tract. (a) The out-of-class empirical stratification matrices $Mc_{i,j}$, showing the probabilities that individuals from a given class visit to places of different classes. The darker colour shades of bins represent larger visiting probability. (b) The out-of-class normalised stratification matrices $Nc_{i,j}$, defined as the fraction of the empirical and randomised stratification matrices without own census tract. We observe less diagonality in Houston (Fig. 7.13a and Fig. 7.13d), New York (Fig. 7.13b and Fig. 7.13e) and San Diego (Fig. 7.13c and Fig. 7.13f). Interestingly, upward bias tendency is still considerably present as seen in (Fig. 7.13g) where mean of upper diagonal matrix elements exceeds mean of lower diagonal matrix elements in 12 out of 20 urban areas, including Houston.

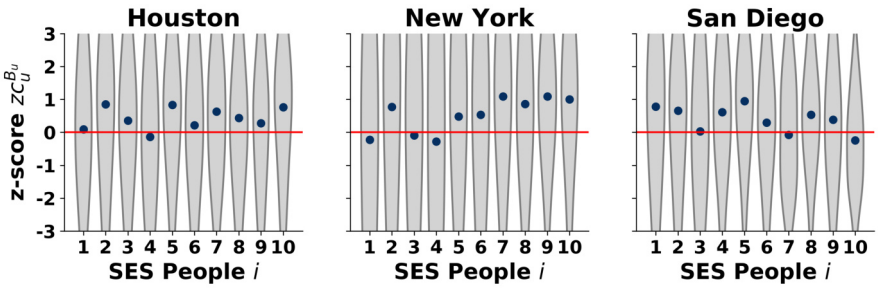

Figure 7.14: Out-of-class Individual Bias z-score $zc_u^{B_u}$. After removing own census tract, class level distributions and their median values are shown for each socioeconomic class. Horizontal red line at z-score=0 represents the condition of unbiasedness. Median values that appear on the top of red line suggests upward biases in terms of visiting patterns, otherwise downward biases persist. The blue dots show that while in general people from lower classes in New York (Fig. 7.14b) are less bias than expected, they could be as much biased regardless their socioeconomic classes in Houston (Fig. 7.14a) and San Diego (Fig. 7.14c).

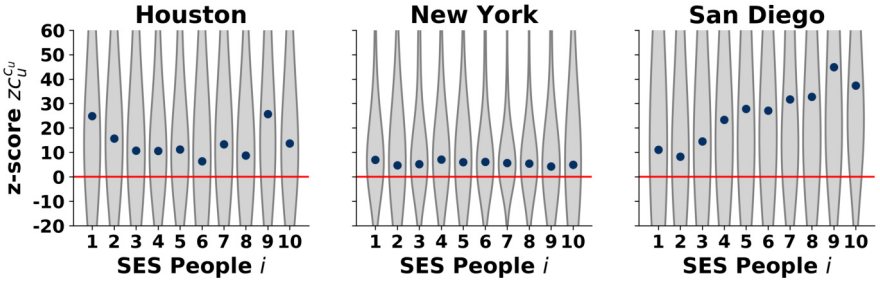

Figure 7.15: Out-of-class Class-level Bias z-score $zc_u^{c_u}$. Distribution of class-level biased z-scores as the function of socioeconomic classes. After removing own census tract, distributions are shown for each socioeconomic class with their median values as blue points for Houston (Fig. 7.15a), New York (Fig. 7.15b), and San Diego (Fig. 7.15c). All z-score values remain above the red unbiased median line, therefore signalling an upward visiting bias.

7.1.13 Homophily mixing measures

We formulate the dispersion index as a supplementary measure to the diagonality index which is discussed in Section 3.4. The dispersion index aims to measure the segregation tendency based on the normalised stratification matrix $N_{i,j}$ constructed as a 10-by-10 matrix with non-negative entries $a_{ij} \geq 0$, where the summation over all entries can be written as $S = \sum_{i=1}^{i} \sum_{j=1}^{j} a_{ij}$. Let $b = (1, 2, \ldots, 10)$ be a threshold and for each threshold b we sum up the diagonal and the off diagonals up to $S_b = \sum_{i_b=1}^{i} \sum_{j_b=1}^{j} a_{i_b j_b}$ where i_b, j_b belong to the first b diagonals (in both directions). It returns S_1 as the trace, and $S_c = S$ as the sum of all elements. The absolute dispersion measure is proposed as

$$D_b = \frac{S_b}{S} \tag{7.1}$$

and defined as the function of b. Consequently, the presence of large values concentrated around the diagonal contributes to a sharp increase along b. In contrast, we expect a marginal increment as we increase b in the case of a highly homogeneous matrix $N_{i,j}$.

To capture the extent to which such values differs from by chance, we compute a reference D_{rb}. In this regard, a reference homogeneous matrix $Mr_{i,j}$ is constructed where all entries are constant. The absolute dispersion measure of the reference matrix Dr_b is computed likewise. The dispersion in the matrix $N_{i,j}$ relative to a homogeneous reference point $Mr_{i,j}$ is measured by

$$\Delta D = Db - Drb \tag{7.2}$$

retaining the area between the curves of D_b and D_{rb} as a function of b.

The index value ranges from 0 to 1. The lower boundary indicates complete heterogeneous mixing in which users visit places across socioeconomic status and the upper boundary pinpoints complete homogeneous mixing in which users' mobilities are concentrated within their own socioeconomic status. Dispersion index provides consistent result with diagonality index as revealed in Fig. 7.16.

7.1.14 Kruskal-Wallis H Test

In order to check the statistical significance of the difference between the empirical stratification matrix $M_{i,j}$ and the randomised stratification matrix $R_{i,j}$, we employ a

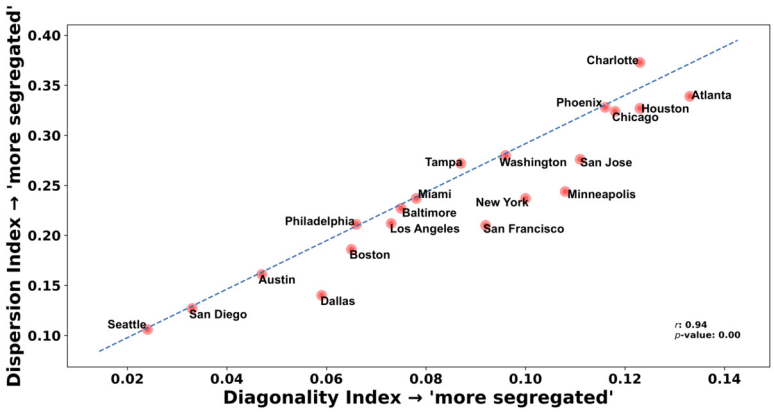

Figure 7.16: Homophily Mixing. Homophily mixing tendency presented as 2 dimensional measure: Diagonality and Dispersion Index. Charlotte and Seattle seem to be outliers because the two are located away from the rest. Houston has the highest main-homophily mixing tendency among other cities, as well as the strongest neighbouring-homophily mixing. New York is less segregated, while Dallas is even far from being homophily mixing.

Kruskal-Wallis H Test (non-parametric one-way ANOVA). Considering that the definition of the normalised stratification matrix $N_{i,j}$ does not take into account any statistical test, just a simple quotient between the real data $M_{i,j}$ and some random null model number $R_{i,j}$, this procedure aims to justify the significance of homophily mixing in the diagonal elements. The null hypothesis $(H0)$ is formulated as an equal median between $M_{i,j}$ and $R_{i,j}$. We cannot accept $(H0)$ if the p-value is smaller than the confidence level $\alpha = 0.05$, otherwise the alternative hypothesis (H_a) holds. Table 7.2 justifies the specification of proposed null model in testing the existence of homophily mixing. We obtain statistically significant difference between $M_{i,j}$ and $R_{i,j}$ for all cities.

Moreover, the Kruskal-Wallis H Test is also used in testing the individual bias z-scores distributions across SES. Given the socioeconomic stratification in our analysis with 10 SES groups, we are interested to determine statistically significant differences between two or more SES groups. It is built based on the null hypothesis $(H0)$ of equal median values of individual bias z-scores across all groups, with the alternative hypothesis (H_a) as the opposite condition. The decision rule states to reject the null hypothesis if p-value is less than confidence level in which we set $\alpha = 0.05$. Results in Table 7.3 show statistically significant differences in median values of individual bias z-scores between SES groups in observed areas except New York. It underlines the role of socioeconomic stratification in shaping individual bias z-scores. Differences are not significant in New York as initially indicated by the relatively flat median values as a function of SES class in Fig. 7.8.

Urban Area	H statistic	p-value
Atlanta	7433.403	0.000
Austin	12283.197	0.000
Baltimore	4784.554	0.000
Boston	8631.003	0.000
Charlotte	11163.451	0.000
Chicago	5585.874	0.000
Dallas	8181.791	0.000
Houston	6237.627	0.000
Los Angeles	1555.634	0.000
Miami	3840.331	0.000
Minneapolis	9687.547	0.000
New York	13333.814	0.000
Philadelphia	3757.849	0.000
Phoenix	3834.426	0.000
San Diego	905.343	0.000
San Francisco	7663.206	0.000
San Jose	7063.406	0.000
Seattle	10666.214	0.000
Tampa	3699.826	0.000
Washington	7318.923	0.000

Table 7.2: Kruskal-Wallis H Test on empirical stratification matrix $M_{i,j}$ and randomised stratification matrix $R_{i,j}$. We show that statistically significant difference between $M_{i,j}$ and $R_{i,j}$ is valid in all urban areas at $\alpha = 0.05$, signifying the proper use of null model and the significant presence of homophily mixing in the diagonal elements.

Urban Area	H statistic	p-value
Atlanta	141.489	$5.082x10^{-26}$
Austin	33.065	$1.301x10^{-4}$
Baltimore	60.753	$9.594x10^{-10}$
Boston	111.234	$8.241x10^{-20}$
Charlotte	69.144	$2.240x10^{-11}$
Chicago	146.474	$4.736x10^{-27}$
Dallas	63.167	$3.274x10^{-10}$
Houston	69.038	$2.349x10^{-11}$
Los Angeles	57.382	$4.269x10^{-9}$
Miami	127.727	$3.477x10^{-23}$
Minneapolis	61.857	$5.872x10^{-10}$
New York	17.350	$4.3504x10^{-1}$
Philadelphia	117.192	$5.015x10^{-21}$
Phoenix	114.045	$2.202x10^{-20}$
San Diego	75.688	$1.156x10^{-12}$
San Francisco	185.417	$3.741x10^{-35}$
San Jose	46.844	$4.196x10^{-7}$
Seattle	56.402	$6.575x10^{-9}$
Tampa	81.395	$8.535x10^{-14}$
Washington	121.341	$7.101x10^{-22}$

Table 7.3: Kruskal-Wallis H Test on individual bias z-scores. In all cities, the p-value is away lower than the confidence level at $\alpha = 0.05$. These statistics imply that socioeconomic stratification leads to statistically significant differences in individual bias z-scores in those areas. New York reveals a contrasting situation where the distribution of individual bias z-scores does not seem to differ across SES.

7.1.15 Dunn's Test

In this section, we employ an additional statistical test called Dunn's Test. After obtaining results of the Kruskal-Wallis H test to check whether all distributions of visiting patterns are the same between people across SES, we motivate this post-hoc procedure to determine exactly which SES of people produces statistically significant different individual bias z-scores. The null hypothesis ($H0$) states equality regarding their individual bias z-scores for each pair of SES i of users, against the alternative hypothesis (H_a) where one or more pair are statistically significantly different. If the p-value is less than the confidence level ($\alpha = 0.05$), we should reject the null hypothesis, implying that those pairs exist.

We capture variability of pairs of SES i of users that contribute to the difference in individual bias z-scores in each city as seen in Fig. 7.17 (dark blue blocks). For example, in Austin, the significant difference in individual bias z-scores is found between user with SES 1 and 8 as well between user with SES 1 and 10. The pairs that contribute to differences in individual bias z-scores vary across cities. This indicates that the individual biases are not homogeneously induced across socioeconomic classes. Given that no statistically significant differences between individual bias z-scores in New York are found, all blocks are coloured orange.

7.2 Mobility segregation dynamics during pandemic interventions

7.2.1 Summary statistics

In the dataset, Bogota retains longest temporal observation until May 2021, followed by London (February 2021), Jakarta (December 2020), and New York (July 2020). Each individual in every city has a set of trajectories constituting timestamps (start and end) whenever he is detected at a certain location (latitude and longitude). We focus on mobility traces of people whose home locations are successfully identified. In Bogota, there are approximately 55,000 people containing 25 million trajectories. The number of people fluctuates among cities, so do total trajectories: Jakarta (around 65,000 people/26 million trajectories), London (almost 200,000 people/ 115 million trajectories), and New York (about 277,000 people/30 million trajectories).

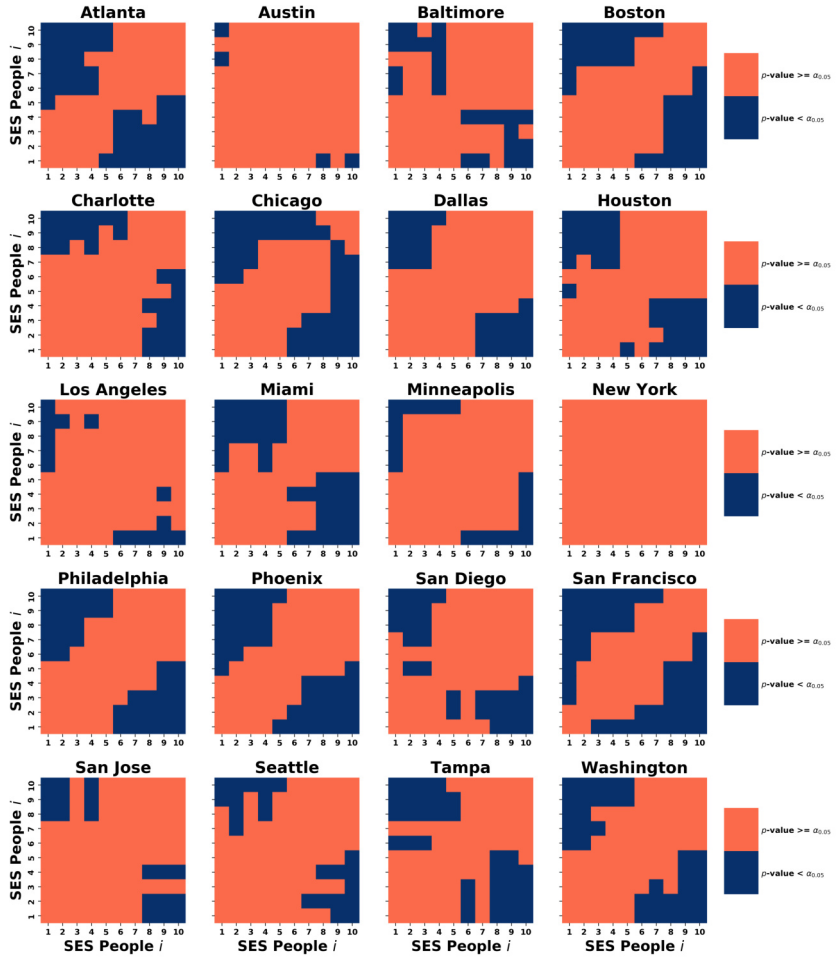

Figure 7.17: Dunn's Test. The dark blue blocks represent pairs of SES of people i that contribute to the difference in individual bias z-scores in each city; otherwise orange blocks indicate blocks of no contributions from those pairs. The value contained in each matrix element e_{ij} denotes an adjusted p-value for the difference between a pair of SES class i. A pair of classes is statistically significantly different if $e_{ij} < \alpha = 0.05$, as coloured in dark blue, otherwise orange.

Urban Area	Number of People	Number of Trajectory
Bogota	55,000	25 million
Jakarta	65,000	26 million
London	200,000	115 million
New York	277,000	30 million

Table 7.4: Sample size. We have different size sample across cities but preserves the temporal representation of pandemic cycle: before lockdown, lockdown, and reopening.

Human mobility captures multi-layer information with high spatiotemporal resolution. Not only physical movement from one point to million others, it resumes individual behavioural dynamics in exploring spatial boundaries. In order to make meaning-

ful observation related to individual mobility patterns within urban landscape, we map out socioeconomic condition of people and places they visit by inferring income-based metadata gathered from bureau of statistics of respected locations. This method allows us to comprehensively analyse two aspects of individual trajectory over places: spatial and socioeconomic status (SES) distribution. We construct a pipeline comprising data collection, data processing, and data analysis as depicted in Fig. 7.18.

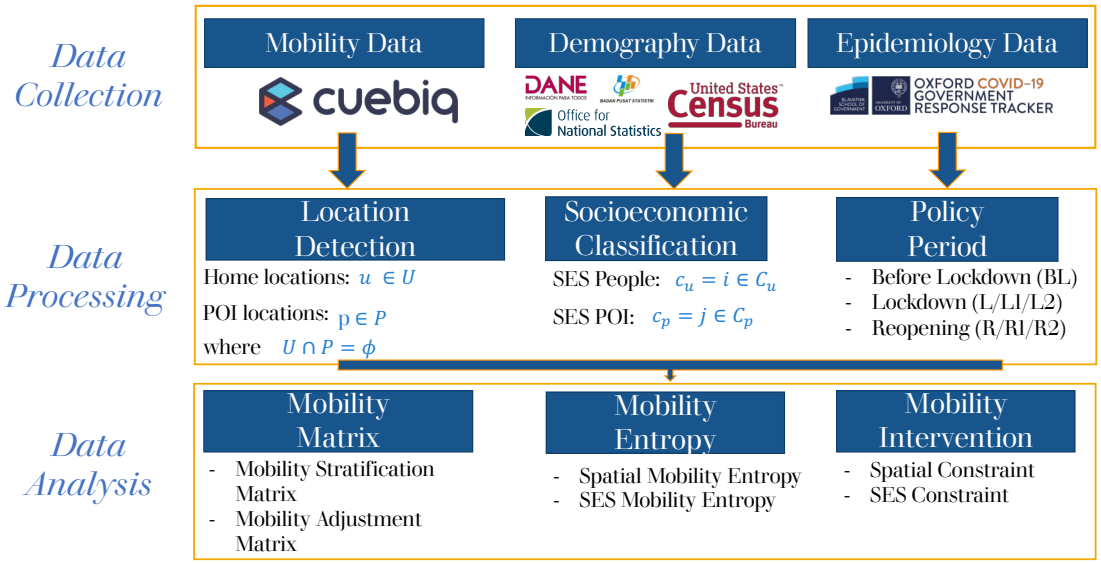

Figure 7.18: Data analytical pipeline. We observe mobility in Bogota (Colombia), Jakarta (Indonesia), London (United Kingdom), and New York (United States). Three types of data are used: mobility data (CUEBIQ), socioeconomic data (Bureau of statistics), and COVID-19 data (OxCGRT/national task force).

7.2.2 Mobility stratification matrix: All visits

Distribution of frequency visit with regards to socioeconomic stratification between SES People i and SES POI j is conceptually introduced in Section 5.2 as Mobility Stratification Matrix M_{ij}. Normalisation in performed by own SES (column-wise). Fig. 7.19 reveals the generic pattern in which assortative mixing increases during the lockdown as increasing r is found across cities. It reflects the extend individual responds to the pandemics by reorganising their typical mobility configuration. In the case of more than one period of lockdown appears (L1 and L2), the first seems to be stronger in inducing the isolation effect. As the reopening (R1) phase is started, the assortative visit remains higher than the level before lockdown (BL).

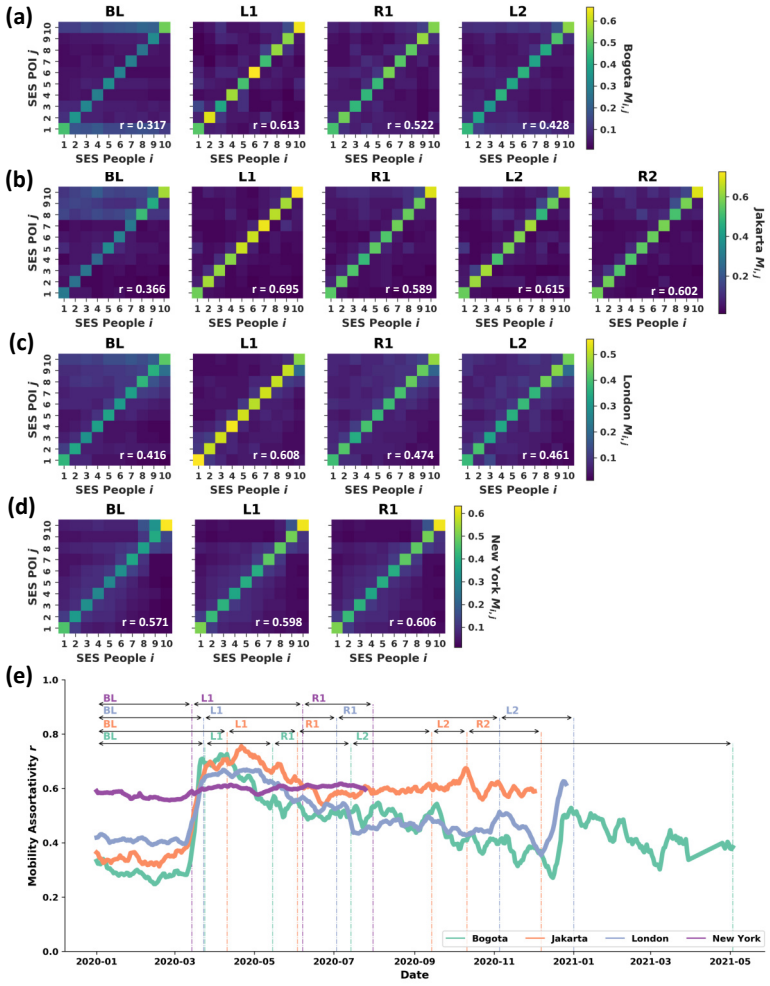

Figure 7.19: Mobility Stratification Matrix for all visits M_{ij}. Matrix elements in Fig. 7.19a-d represent the magnitude of frequency visits for each pair of SES People i and SES POI j where lighter colour shows larger visit proportion. All locations found in individual trajectories are taken into account.

7.2.3 Mobility stratification matrix: Without home area visit

We repeat the procedure used to generate Fig. 7.19 after excluding local visits to own neighbourhood to generate Mobility Stratification Matrix for visits outside home area Mc_{ij}. This step is considered as robustness control over the persistent assortative mixing. In Fig. 7.20 we see that the first lockdown is still the most stringent because it alters preference to visit more places within own socioeconomic class. Comparing to Fig. 7.20, assortativity coefficient r in general is away lower, indicating that short distance visit in the surrounding neighbourhood assumes considerable proportion on mobility pattern.

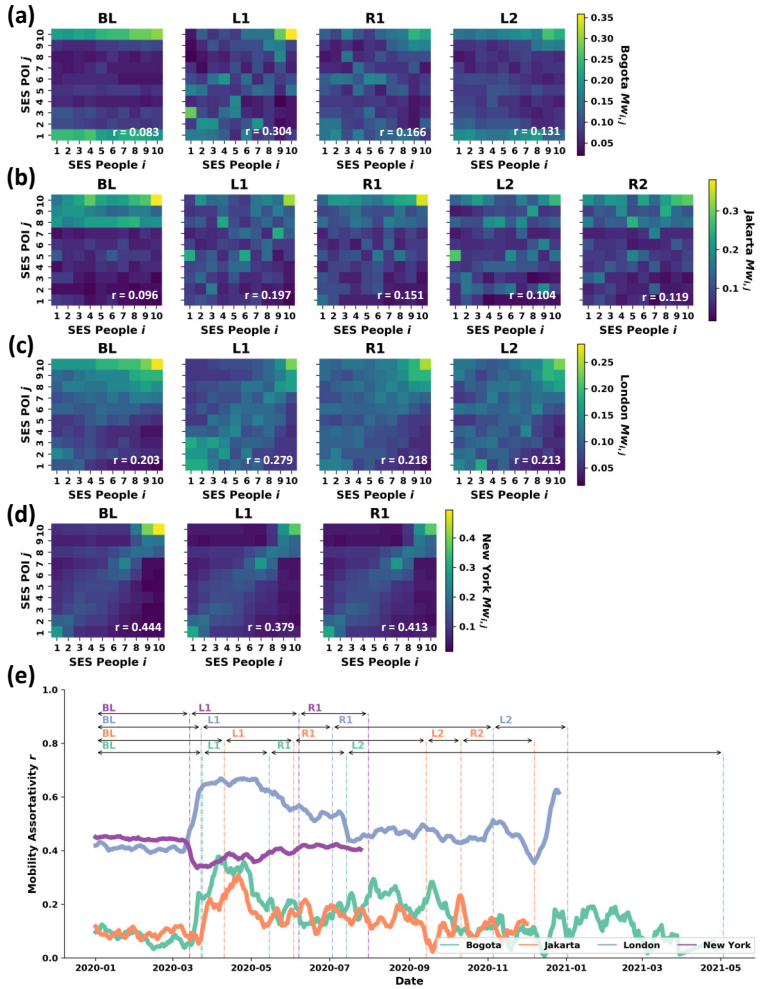

Figure 7.20: Mobility Stratification Matrix for visits outside home area Mc_{ij}. Proportion of frequency visit of people from SES i to places in SES j is computed after removing places located in own neighbourhood. The lighter bin colour, the higher visit probability is.

7.2.4 Mobility adjustment matrix: All visits

Mobility adjustment matrix S_{ij} is constructed to detect the indication of residual isolation effect. We operationalise the computation in Section 4.2.3 in which the difference in proportion of frequency visits between two consecutive periods is visible in Fig. 7.21. None of cities in this study exhibit full recovery after the occurrence of re-opening as the bin colour remains under brown shades, indicating larger visit ratio to places in own socioeconomic class as to compare with before lockdown period. Therefore, it leads to the notion of residual isolation induced by COVID-19 outbreak.

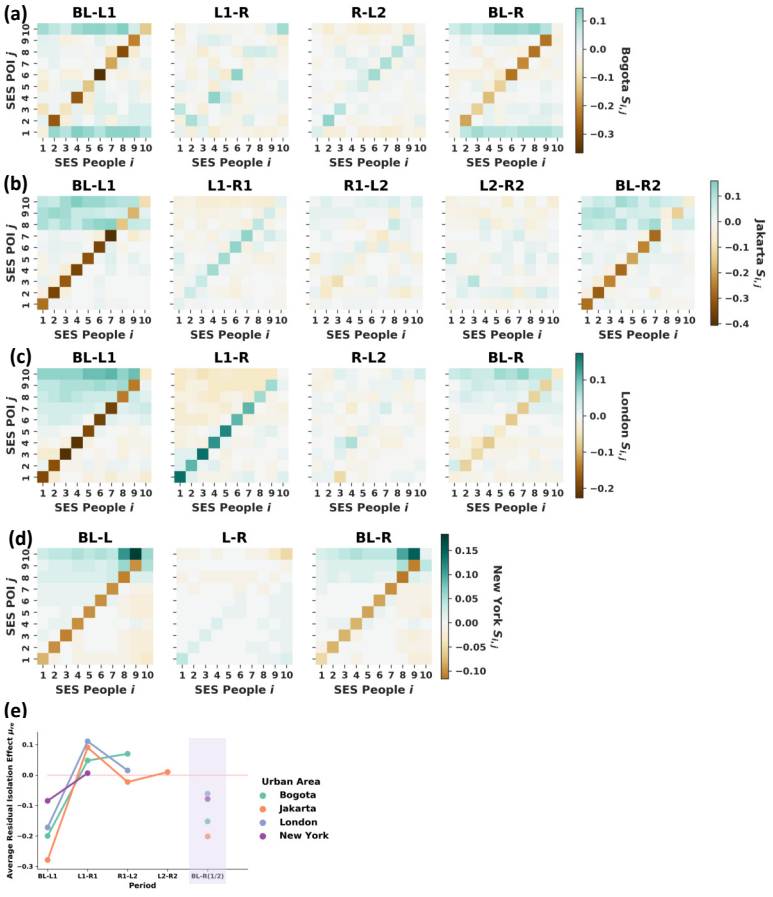

Figure 7.21: Mobility Adjustment Matrix for all visits S_{ij}. The difference in term of visit probability between a pair of two consecutive Mobility Stratification Matrix M_{ij} is measured. The presence of white bins indicates indifferent visiting pattern, while green shows more visits during the first period. Otherwise, brown shades appear. All locations found in individual trajectories are taken into account.

7.2.5 Mobility adjustment matrix: Without home area visit

Mobility adjustment matrix S_{ij} is transformed to Sc_{ij} by eliminating visits to own neighbourhood. It runs on similar motivation in Section 7.2.5, namely as robustness check given the large local visits in the individual trajectory. Fig. 7.22 tries to uncover the main attribution of residual isolation effect by eliminating visits to own neighbourhood/home area. This procedure dilutes the magnitude of assortativity force, therefore we address the residual isolation effect as a longer term consequence of localised mobility due to COVID-19 restrictions. Interestingly, BL-R shows segregated pattern of visit where before lockdown people tend to explore more places in higher socioeconomic ranks (top rows/green shades) while during the reopening places in lower classes contribute more to visit proportion (brown shades) in every cities. Beyond that, Bogota exhibit bimodal segregation where dominant visit before lockdown does not only happen in upper class, but also lower class.

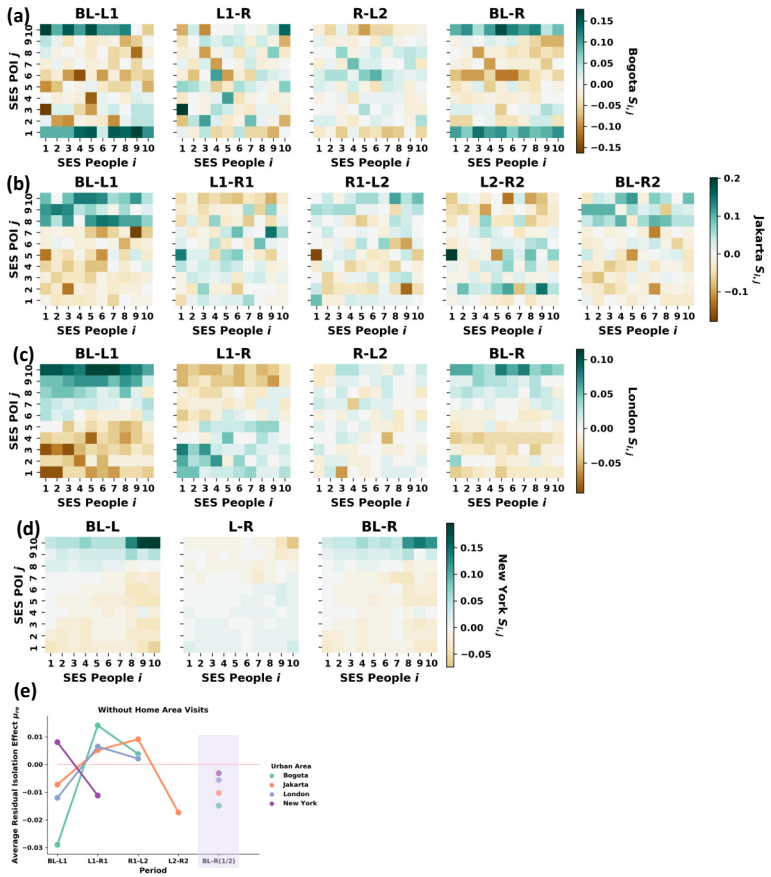

Figure 7.22: Mobility Adjustment Matrix for visits outside home area Sc_{ij}. Every Mobility Stratification Matrix for visits outside home area Mc_{ij} is paired with the one in the following period. There are three patterns to detect: no difference between those two periods (white), dominant visit in the first period (green), and dominant visit in the second period (brown).

7.2.6 Spatial mobility entropy

Heterogeneity of places visited by individual is quantified by computation of Spatial Mobility Entropy $Hm(X)$ proposed in Section 5.3. Dispersion of value may take either to the direction of 0, signifying strict preference on particular locations over the rest and making the trajectory more homogeneous spatial wise. In contrast, as the value takes closer to 1, no strict preference presumed and visits are widely distributed across locational space. We find that people become more restricted in deciding which locations to visit as the average value $Hm(X)$ hits the lowest point than ever in all cities. The introduction of reopening phase does not directly bounce the value back to the normal level before lockdown, in line with condition suggested in Fig. 7.19 and Fig. 7.21.

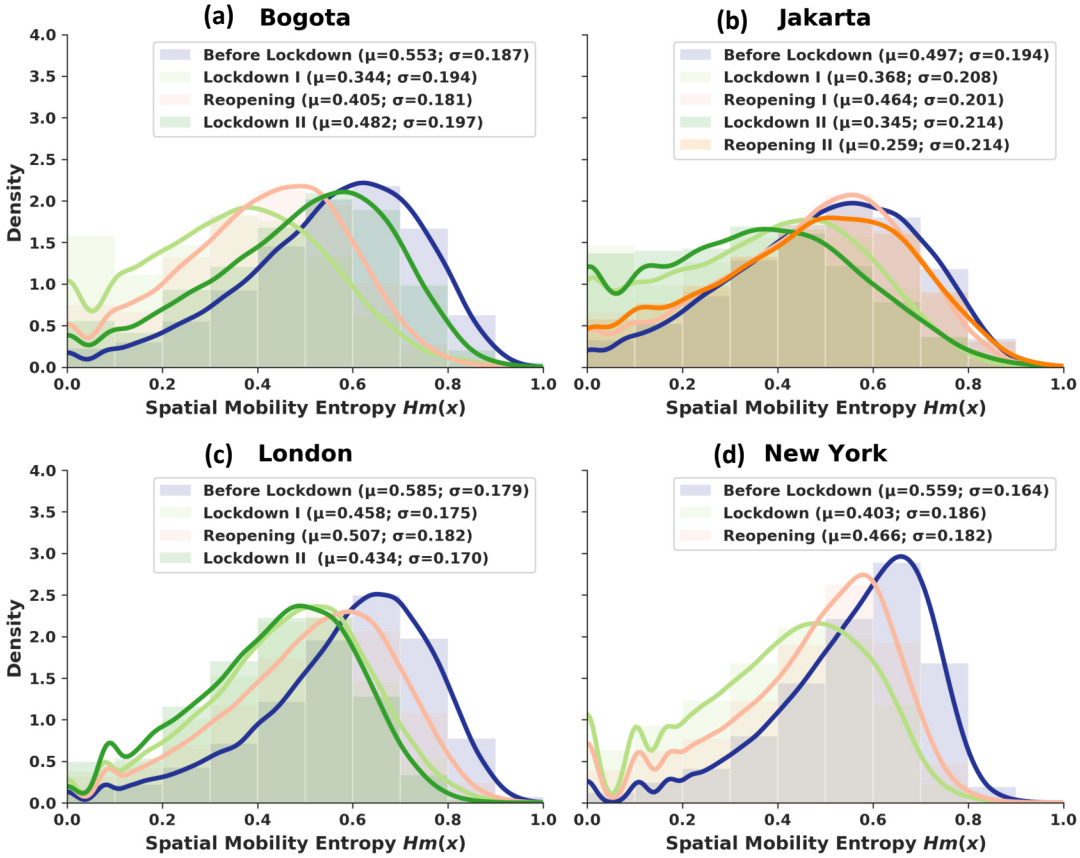

Figure 7.23: Spatial Mobility Entropy $Hm(X)$. We measure heterogeneity of individual preference regarding location of places visited. The presence of commonly repeated places pushes the value closer to zero, denoting lower degree of heterogeneity. On the other hand, higher variability of locations is represented by value near 1.

7.2.7 Socioeconomic mobility entropy

In this section, we redo the computation for trajectory heterogeneity in terms of socioeconomic factor based on entropy formulation in Section 5.3. To measure Socioeconomic Mobility Entropy $Hs(X)$, we substitute geolocation feature with SES of places. The result in Fig. 7.24 confirms previous finding where people have stricter preference over places during lockdown. It is beyond spatial boundary since socioeconomic profile of those places is now also heavily skewed, making average value $Hs(X)$ touches lowest record in comparison to other periods. Therefore, it reaffirms condition stipulated in Fig. 7.19, Fig. 7.21 and Fig. 7.23.

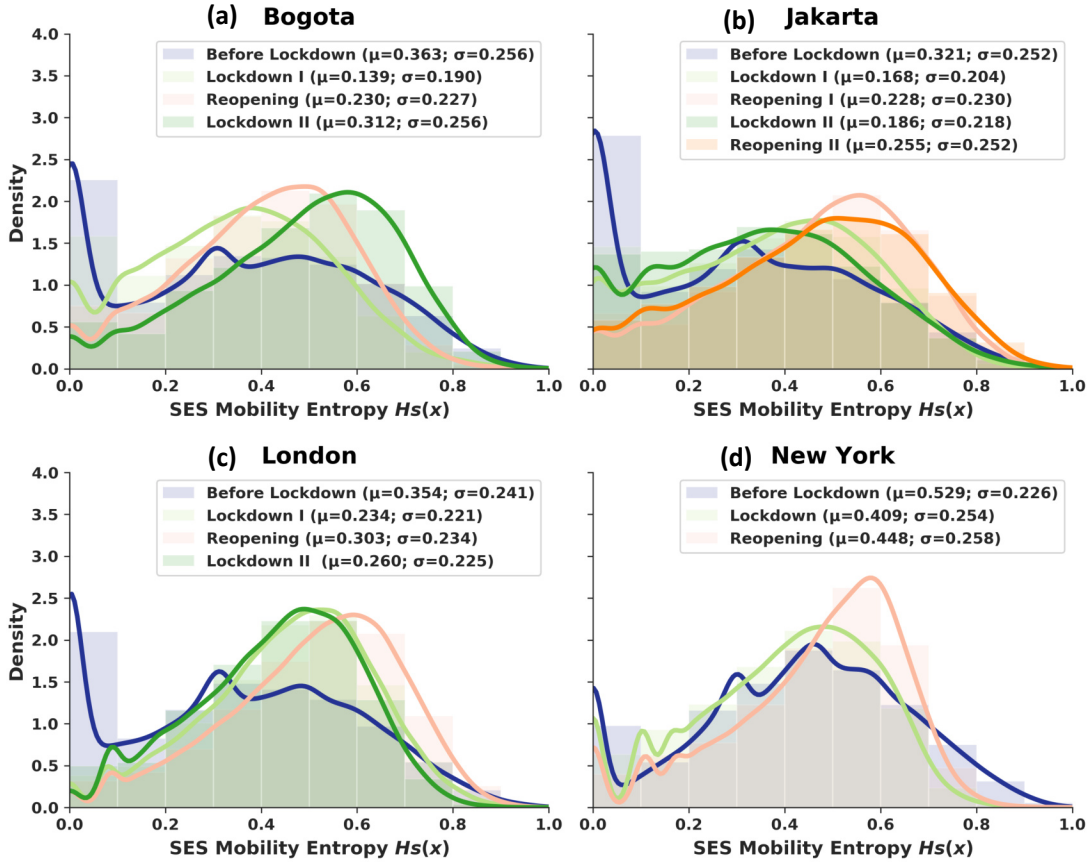

Figure 7.24: Socioeconomic Mobility Entropy $Hs(X)$. After replacing geolocation of places in individual trajectory by SES information, we recompute entropy. As the value skews to 0, visiting pattern tends to be concentrated on particular SES, otherwise it is somewhere close to 1.

7.2.8 Socioeconomic constraint

We provide technical note explaining the method used to determine the effectiveness of NPI in Section 5.4. There are 9 measures constituted in mobility interventions namely closings of schools and universities (C1), closings of workplaces (C2), cancelling public events (C3), limits on gatherings (C4), closing of public transport (C5), orders to stay-at-home (C6), restrictions on movement between cities/regions (C7), restrictions on international travel(C8) and presence of public information campaigns (H1). As effectiveness is defined as its significance in restricting mobility in term of socioeconomic exploration, we descendingly sort the rank $rk_{Hs(X)}$ based on $R-squared$ generated from each time series regression between Socioeconomic Mobility Entropy $Hs(X)$ (y-variable) and the stringency of respected intervention S_k (x-variable). As suggested previously in Section 4.3.5, public information campaigns (H1) is the most essential instrument before lockdown but tends to diminish over time. Other than that, we observe heterogeneity at city level across the globe.

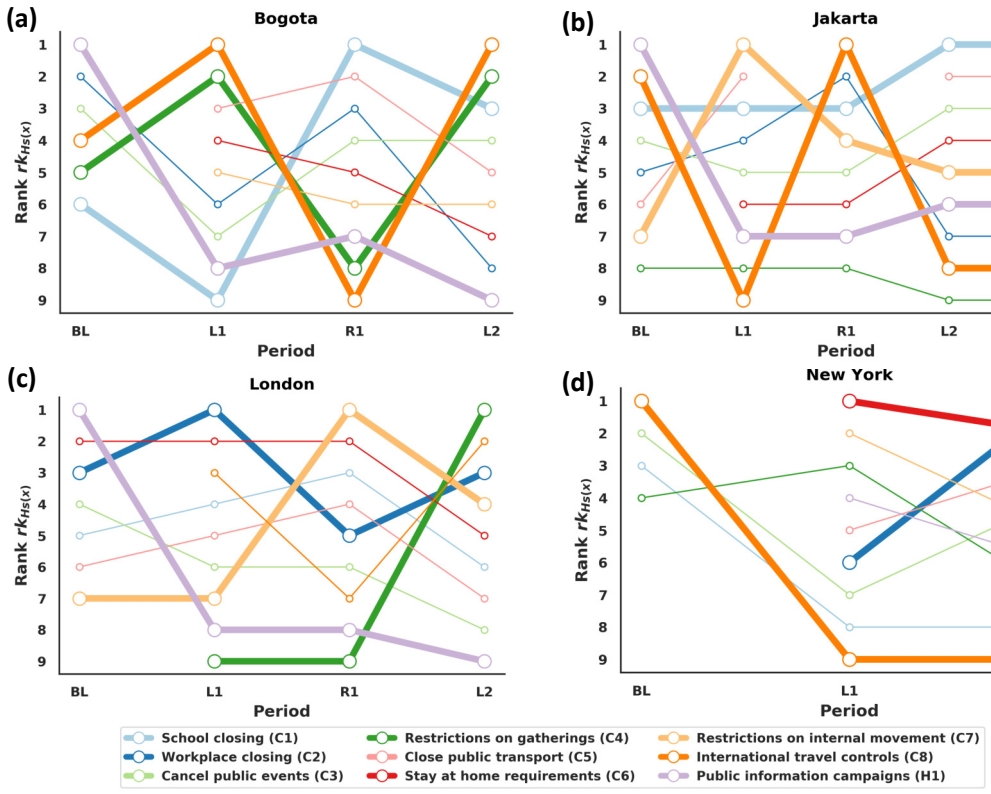

Figure 7.25: SES Constraint. The implementation of NPI also restricts exploration of places situated in diverse SES. The effectiveness of each NPI measure in limiting socioeconomic diversity in individual trajectory is presented as rank $rk_{Hs(X)}$.

Individual exploration occurs not only over physical space, but also beyond socioe-

conomic dimension. Therefore, enforcement of NPI (widely known as mobility restrictions) also reduces socioeconomic diversity of visiting places. In Bogota, locational (Fig. 4.7a) and socioeconomic diversity (Fig. 7.25a) of visiting places are identically shaped by consistent composition of restriction on the top rank. Before lockdown (BL), public information campaign (H1/red) is on the lead but toppled down by workplace closing (C2/dark blue) during the first lockdown (L1). The emergence of reopening (R1) marks the increasing importance of international travel control (C8/orange), while the second chapter of lockdown (L2) gives rise to the policy of public transport closing (C5/bright green). Looking at the ranking dynamics, information outreach gets lesser as more specific and targeted policy is preferred in the subsequent periods. There is counter cyclicality between workplace closing and international travel control. Closing public transport in longer period persistently and increasingly limits exposures to wider locations and socioeconomic mixture.

7.2.9 Robustness of mobility adjustment: All visits

We take into account the robustness check of isolation effect by applying Kruskal-Wallis H Test (non-parametric one-way ANOVA) on Mobility Stratification Matrix for both before ($M_{i,j}$) and after removing visits to own home area ($Mw_{i,j}$). The formulation of the null hypothesis ($H0$) could be defined as an equal median between before lockdown and another period that comes after. If the p-value appears to be smaller than the confidence level $\alpha = 0.05$, $H0$ is rejected. Otherwise, he alternative hypothesis (H_a) remains. Table 7.5 and 7.6 provide justification for the presence of different degrees of isolation effect due to the variability of mobility in response to the dynamics of mobility restrictions. New York stands on strikingly opposite pattern as statistically significant difference is seen after removing local visits to the area where home is located while other cities exhibit such pattern for broad visits to any locations.

Urban Area	Matrix Element	BL & L1	L1 & R1	R1 & L2	L2 & R2	BL & R1	BL & R2
Bogota	all	7.556*	3.567*	1.664	—	0.435	—
Bogota	diagonal	11.063*	2.063	7.406*	—	9.606*	—
Jakarta	all	9.135*	5.108*	0.748	0.043	1.504	2.720
Jakarta	diagonal	12.091*	10.079*	1.651	0.571	9.143*	9.606*
London	all	10.832*	12.362*	0.215	—	0.299	—
London	diagonal	14.286*	13.719*	1.286	—	9.143*	—
New York	all	1.404	5.970	—	—	1.381	—
New York	diagonal	7.406*	0.143	—	—	6.606*	—

*$p < 0.05$

Table 7.5: Kruskal-Wallis H Test on Mobility Stratification Matrix before removing visits to home area across pairs of policy period ($M_{i,j}$). Statistical significance could be implied in which the induced isolation effect largely takes place between before lockdown and the first lockdown (BL & L1). It happens in all urban areas (for diagonal elements) but New York (for all elements) as the p-value is away lower than the confidence level at $\alpha = 0.05$. Even after the introduction of the first reopening, the distribution of mobility pattern still does not revert to the pre-pandemic level (BL & R1)

7.2.10 Robustness of mobility adjustment: Without home area visit

Urban Area	Matrix Element	BL & L1	L1 & R1	R1 & L2	L2 & R2	BL & R1	BL & R2
Bogota	all	1.728	0.795	0.202	—	6.595^{*}	—
Bogota	diagonal	1.851	0.006	0.001	—	1.463	—
Jakarta	all	6.090^{*}	0.006	0.013	0.160	4.550^{*}	3.252
Jakarta	diagonal	1.286	0.051	0.001	0.281	0.691	0.966
London	all	0.294	0.199	0.001	—	0.638	—
London	diagonal	1.286	0.463	0.023	—	1.286	—
New York	all	0.119	0.084	—	—	0.001	—
New York	diagonal	0.206	0.051	—	—	0.206	—

$^{*}p < 0.05$

Table 7.6: Kruskal-Wallis H Test on Mobility Stratification Matrix after removing visits to home area across pairs of policy period ($Mw_{i,j}$). Mobility pattern differs significantly between before and during the first lockdown (BL & L1) in Jakarta (for all elements) but not apparent in other urban areas given the p-value is away lower than the confidence level at $\alpha = 0.05$. Similar direction also becomes visible between before and during the first reopening (BL & R1). Strict isolation along diagonal elements is not found anywhere. Therefore, levelling up the contribution of local visits in the surrounding of home locations to isolation.

7.2.11 Inter-mobility

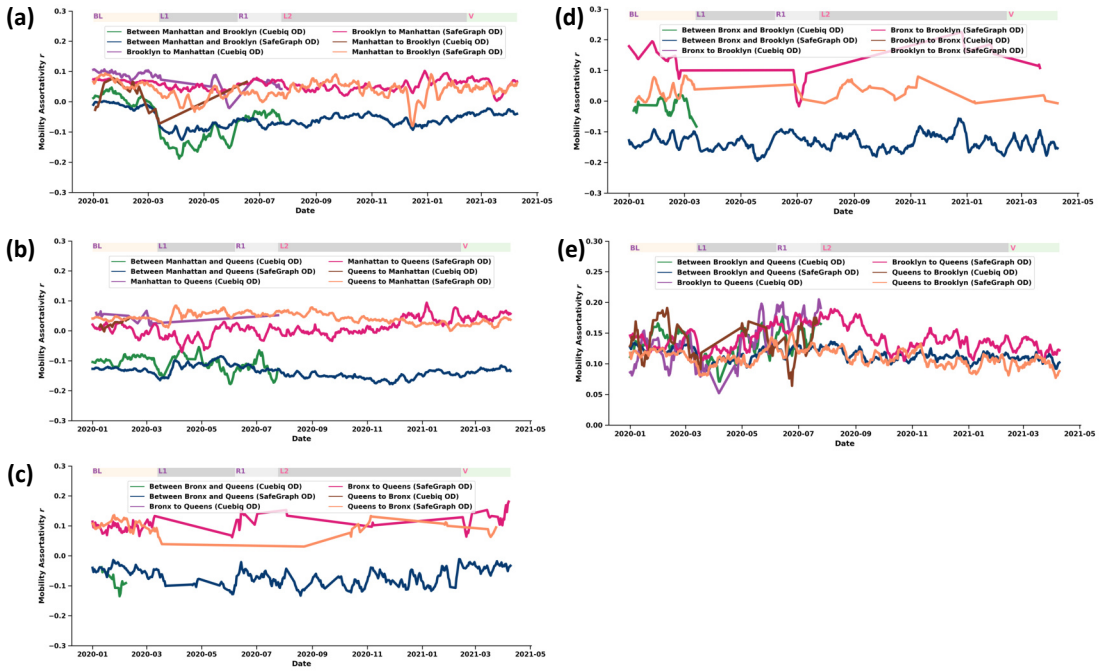

Figure 7.26: Inter-mobility assortativity r. The computation is performed for trips across a pair of boroughs at census-tract level OD matrix: Manhattan-Brooklyn (Fig. 7.26a), Manhattan-Queens (Fig. 7.26b), Bronx-Queens (Fig. 7.26c), Bronx-Brooklyn (Fig. 7.26c), and Brooklyn-Queens (Fig. 7.26d). In this category, mobility is mostly disassortative, leading to lower segregation level except flows of people between Brooklyn and Queens.

Bibliography

[1] Mimi Abramovitz and Richard J Smith. The persistence of residential segregation by race, 1940 to 2010: The role of federal housing policy. *Families in Society*, 102(1):5–32, 2021.

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

[3] Laura Alessandretti, Piotr Sapiezynski, Vedran Sekara, Sune Lehmann, and Andrea Baronchelli. Evidence for a conserved quantity in human mobility. *Nature Human Behaviour*, 2(7):485–491, 2018.

[4] Laura Alessandretti, Ulf Aslak, and Sune Lehmann. The scales of human mobility. *Nature*, 587(7834):402–407, 2020.

[5] Sabina Alkire, Usha Kanagaratnam, and Nicolai Suppa. The global multidimensional poverty index (mpi): 2018 revision. *OPHI MPI methodological notes*, 46, 2018.

[6] Linda Altieri, Daniela Cocchi, and Giulia Roli. A new approach to spatial entropy measures. *Environmental and ecological statistics*, 25:95–110, 2018.

[7] Carole A Ambler and James E Kristoff. Introducing the north american industry classification system. *Government Information Quarterly*, 15(3):263–273, 1998.

[8] Philip W Anderson. More is different: broken symmetry and the nature of the hierarchical structure of science. *Science*, 177(4047):393–396, 1972.

[9] Luc Anselin. Local indicators of spatial association—lisa. *Geographical analysis*, 27(2):93–115, 1995.

[10] Elsa Arcaute and José J Ramasco. Recent advances in urban system science: Models and data. *Plos one*, 17(8):e0272863, 2022.

[11] Dimitris Ballas. What makes a 'happy city'? *Cities*, 32:S39–S50, 2013.

[12] Andrea Baronchelli and Filippo Radicchi. Levy flights in human behavior and cognition. *Chaos, Solitons & Fractals*, 56:101–105, 2013.

[13] Marc Barthelemy. The statistical physics of cities. *Nature Reviews Physics*, 1(6): 406–415, 2019.

[14] M Batty. The emergence of cities: Complexity and urban dynamics. centre for advanced spatial analysis, university collage london. Technical report, working paper 64, URL: http://www. casa. ucl. ac. uk/publications . . . , 2003.

[15] Michael Batty. *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals*. The MIT press, 2007.

[16] Michael Batty. The size, scale, and shape of cities. *science*, 319(5864):769–771, 2008.

[17] Michael Batty. *Cities as Complex Systems: Scaling, Interaction, Networks, Dynamics and Urban Morphologies.* Springer, 2009.

[18] Michael Batty. *The new science of cities*. MIT press, 2013.

[19] Michael Batty and Paul A Longley. *Fractal cities: a geometry of form and function*. Academic press, 1994.

[20] Mariano G Beiró, André Panisson, Michele Tizzoni, and Ciro Cattuto. Predicting human mobility through the assimilation of social media traces into mobility models. *EPJ Data Science*, 5:1–15, 2016.

[21] Edward M Bergman and Edward J Feser. *Industrial and regional clusters: concepts and comparative applications*. Regional Research Institute, West Virginia University, 2020.

[22] Elena Besussi, Nancy Chin, Michael Batty, and Paul Longley. The structure and form of urban settlements. *Remote sensing of urban and suburban areas*, pages 13–31, 2010.

[23] Luis MA Bettencourt. *Introduction to Urban Science: Evidence and Theory of Cities as Complex Systems*. MIT Press, 2021.

[24] Paul Bishop and Peter Gripaios. Explaining spatial patterns of industrial diversity: an analysis of sub-regions in great britain. *Urban Studies*, 44(9):1739–1757, 2007.

[25] Eszter Bokányi, Sándor Juhász, Márton Karsai, and Balázs Lengyel. Universal role of commuting in the reduction of social assortativity in cities. *arXiv preprint arXiv:2105.01464*, 2021.

[26] Nibir Bora, Yu-Han Chang, and Rajiv Maheswaran. Mobility patterns and user dynamics in racially segregated geographies of us cities. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, pages 11–18. Springer, 2014.

[27] Willem R Boterman and Sako Musterd. Cocooning urban life: Exposure to diversity in neighbourhoods, workplaces and transport. *Cities*, 59:139–147, 2016.

[28] BPS. Garis kemiskinan, jumlah, dan persentase penduduk miskin di daerah menurut kabupaten/kota di provinsi dki jakarta 2018-2020, 2020.

[29] Stefano Breschi, Francesco Lissoni, and Franco Malerba. Knowledge-relatedness in firm technological diversification. *Research policy*, 32(1):69–87, 2003.

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

[31] Howard Brody, Michael Russell Rip, Peter Vinten-Johansen, Nigel Paneth, and Stephen Rachman. Map-making and myth-making in broad street: the london cholera epidemic, 1854. *The Lancet*, 356(9223):64–68, 2000.

[32] 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.

[33] U.S. Census Bureau. American community survey, 2012.

[34] U.S. Census Bureau. American community survey, 2018.

[35] Tim Butler, Chris Hamnett, and Mark Ramsden. Inward and upward: Marking out social class change in london, 1981—2001. *Urban Studies*, 45(1):67–88, 2008.

[36] Francesco Calabrese, Mi Diao, Giusy Di Lorenzo, Joseph Ferreira Jr, and Carlo Ratti. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation research part C: emerging technologies*, 26:301–313, 2013.

[37] Antoni Calvo-Armengol and Matthew O Jackson. The effects of social networks on employment and inequality. *American economic review*, 94(3):426–454, 2004.

[38] Manuel Castells. The space of flows. *The rise of the network society*, 1:376–482, 1996.

[39] Hsien-Yen Chang, Wenze Tang, Elham Hatef, Christopher Kitchen, Jonathan P Weiner, and Hadi Kharrazi. Differential impact of mitigation policies and socioeconomic status on covid-19 prevalence and social distancing in the united states. *BMC public health*, 21(1):1–10, 2021.

[40] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. Mobility network models of covid-19 explain inequities and inform reopening. *Nature*, 589(7840):82–87, 2021.

[41] Caroline Chapain, Phil Cooke, Lisa De Propris, Stewart MacNeill, and Juan Mateos-Garcia. Creative clusters and innovation. *Putting creativity on the map. London: NESTA*, 2010.

[42] Debashish Chowdhury and Andreas Schadschneider. Self-organization of traffic jams in cities: Effects of stochastic dynamics and signal periods. *Physical Review E*, 59(2):R1311, 1999.

[43] Merlin Chowkwanyun and Adolph L Reed Jr. Racial health disparities and covid-19—caution and context. *New England Journal of Medicine*, 383(3):201–203, 2020.

[44] Andrew David Cliff and J Keith Ord. *Spatial processes: models & applications*. Taylor & Francis, 1981.

[45] Carmela Comito. Human mobility prediction through twitter. *Procedia Computer Science*, 134:129–136, 2018.

[46] Comptroller. The office sector in new york city. `https://www.osc.state.ny.us/files/reports/osdc/pdf/report-11-2022.pdf`, 2022.

[47] Robert Cowan. *The dictionary of urbanism*, volume 67. Streetwise press Tisbury, 2005.

[48] Andrew Crooks, Dieter Pfoser, Andrew Jenkins, Arie Croitoru, Anthony Stefanidis, Duncan Smith, Sophia Karagiorgou, Alexandros Efentakis, and George

Lamprianidis. Crowdsourcing urban form and function. *International Journal of Geographical Information Science*, 29(5):720–741, 2015.

[49] Balázs Cs Csáji, Arnaud Browet, Vincent A Traag, Jean-Charles Delvenne, Etienne Huens, Paul Van Dooren, Zbigniew Smoreda, and Vincent D Blondel. Exploring the mobility of mobile phone users. *Physica A: statistical mechanics and its applications*, 392(6):1459–1473, 2013.

[50] CUEBIQ. Venue categories, 2020. URL `https://www.cuebiq.com/about/data-for-good/`.

[51] DANE. Medida de pobreza multidimensional municipal de fuente censal 2018, 2020.

[52] Shawna J Dark and Danielle Bram. The modifiable areal unit problem (maup) in physical geography. *Progress in Physical Geography*, 31(5):471–479, 2007.

[53] Juan de Dios Ortúzar and Luis G Willumsen. *Modelling transport*. John wiley & sons, 2011.

[54] Gert De Roo and Jean Hillier. *Complexity and planning: Systems, assemblages and simulations*. Routledge, 2016.

[55] Angus Deaton and Darren Lubotsky. Mortality, inequality and race in american cities and states. *Social science & medicine*, 56(6):1139–1153, 2003.

[56] Suma Desu. *Untangling the effects of residential segregation on individual mobility*. PhD thesis, Massachusetts Institute of Technology, 2015.

[57] Lewis Dijkstra, Aneta J Florczyk, Sergio Freire, Thomas Kemper, Michele Melchiorri, Martino Pesaresi, and Marcello Schiavina. Applying the degree of urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics*, 125:103312, 2021.

[58] Won Do Lee, Matthias Qian, and Tim Schwanen. The association between socioeconomic status and mobility reductions in the early stage of england's covid-19 epidemic. *Health & Place*, 69:102563, 2021.

[59] 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.

[60] 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), 2021.

[61] Gilles Duranton and Diego Puga. Diversity and specialisation in cities: why, where and when does it matter? *Urban studies*, 37(3):533–555, 2000.

[62] Gilles Duranton and Diego Puga. Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics*, volume 4, pages 2063–2117. Elsevier, 2004.

[63] Steven N Durlauf. Complexity and empirical economics. *The Economic Journal*, 115(504):F225–F243, 2005.

[64] European Commission. What criteria are used to define a city? https://urban.jrc.ec.europa.eu/thefutureofcities/what-is-a-city, 2019. URL `https://urban.jrc.ec.europa.eu/thefutureofcities/what-is-a-city`.

[65] Steven Farber, Tijs Neutens, Harvey J Miller, and Xiao Li. The social interaction potential of metropolitan regions: A time-geographic measurement approach using joint accessibility. *Annals of the Association of American Geographers*, 103 (3):483–504, 2013.

[66] Steven Farber, Morton O'Kelly, Harvey J Miller, and Tijs Neutens. Measuring segregation using patterns of daily travel behavior: A social interaction based model of exposure. *Journal of Transport Geography*, 49:26–38, 2015.

[67] Moshe Farjoun. Beyond industry boundaries: Human expertise, diversification and resource-related industry groups. *Organization science*, 5(2):185–199, 1994.

[68] Jo Foord. The new boomtown? creative city to tech city in east london. *Cities*, 33:51–60, 2013.

[69] K Frenken, FG Van Oort, T Verburg, and RA Boschma. Variety and regional economic growth in the netherlands: final report to the ministery of economic affairs. Technical report, Ministry of Economic Affairs, 2004.

[70] Koen Frenken, Frank Van Oort, and Thijs Verburg. Related variety, unrelated variety and regional economic growth. *Regional studies*, 41(5):685–697, 2007.

[71] John Friedmann. The world city hypothesis. *Development and change*, 17(1): 69–83, 1986.

[72] Riccardo Gallotti, Armando Bazzani, Sandro Rambaldi, and Marc Barthelemy. A stochastic model of randomly accelerated walkers for human mobility. *Nature Communications*, 7(1):1–7, 2016.

[73] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[74] Daniel Goya and Helena Cárdenas. Marshallian and jacobian externalities in creative industries. *Caribbean Quarterly*, 10(3), 2022.

[75] Mark Granovetter. Economic action and social structure: The problem of embeddedness. *American journal of sociology*, 91(3):481–510, 1985.

[76] Urban Morphology Research Group. Glossary, 2023. URL `http://www.urbanform.org/glossary.html`. Accessed: 2023-03-30.

[77] 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.

[78] Aino Halinen and Jan-Åke Törnroos. The role of embeddedness in the evolution of business networks. *Scandinavian journal of management*, 14(3):187–205, 1998.

[79] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*, volume 2. Springer series in statistics New York, New York, 2009.

[80] 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.

[81] J Vernon Henderson. Urban development: Theory, fact, and illusion. Technical report, Oxford University Press, 1991.

[82] Vernon Henderson. Medium size cities. *Regional science and urban economics*, 27(6):583–612, 1997.

[83] Martin Hess. ‘spatial’relationships? towards a reconceptualization of embedded ness. *Progress in human geography*, 28(2):165–186, 2004.

[84] César A Hidalgo, Bailey Klinger, A-L Barabási, and Ricardo Hausmann. The product space conditions the development of nations. *Science*, 317(5837):482–487, 2007.

[85] 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.

[86] John Iceland, Daniel H Weinberg, and Erika Steinmetz. Racial and ethnic residential segregation in the united states 1980-2000. *Census 2000 Special Reports*, 8(3), 2002.

[87] Jane Jacobs. *The economy of cities*. Random House, 1960.

[88] Jane Jacobs. *The Death and Life of Great American Cities*. Vintage Books, New York (A Division of Random House, Inc.), 1961.

[89] Jane Jacobs. *The Death and Life of Great American Cities*. Random House, 1961.

[90] Jeffrey Johnson. Cities: Systems of systems of systems. *Complexity theories of cities have come of age: An overview with implications to urban planning and design*, pages 153–172, 2012.

[91] Rosie Jones, Ravi Kumar, Bo Pang, and Andrew Tomkins. " i know what you did last summer" query logs and user privacy. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 909–914, 2007.

[92] Raja Jurdak, Kun Zhao, Jiajun Liu, Maurice AbouJaoude, Mark Cameron, and David Newth. Understanding human mobility from twitter. *PloS One*, 10(7): e0131469, 2015.

[93] Yuhao Kang, Song Gao, Yunlei Liang, Mingxiao Li, Jinmeng Rao, and Jake Kruse. Multiscale dynamic human mobility flow dataset in the us during the covid-19 epidemic. *Scientific data*, 7(1):390, 2020.

[94] Vincent Kaufmann, Manfred Max Bergman, and Dominique Joye. Motility: mobility as capital. *International journal of urban and regional research*, 28(4): 745–756, 2004.

[95] Jerome B Kernan and Grady D Bruce. The socioeconomic structure of an urban area. *Journal of Marketing Research*, 9(1):15–18, 1972.

[96] John Frank Charles Kingman. *Poisson processes*, volume 3. Clarendon Press, 1992.

[97] Kolmogorov-Smirnov Test. Two-sample kolmogorov-smirnov test. `https://www.mathworks.com/help/stats/kstest2.html`, 2022.

[98] Moritz UG Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Open COVID-19 Data Working Group†, Louis Du Plessis, Nuno R Faria, Ruoran Li, et al. The effect of human mobility and control measures on the covid-19 epidemic in china. *Science*, 368(6490):493–497, 2020.

[99] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, 2004.

[100] Lauren J Krivo, Heather M Washington, Ruth D Peterson, Christopher R Browning, Catherine A Calder, and Mei-Po Kwan. Social isolation of disadvantage and advantage: The reproduction of inequality in urban space. *Social Forces*, 92(1): 141–164, 2013.

[101] Karl Kropf. Urban tissue and the character of towns. *Urban Design International*, 1:247–263, 1996.

[102] Paul Krugman. Increasing returns and economic geography. *Journal of political economy*, 99(3):483–499, 1991.

[103] Paul Krugman. *Geography and trade*. MIT press, 1992.

[104] Paul Krugman. First nature, second nature, and metropolitan location. *Journal of regional science*, 33(2):129–144, 1993.

[105] Shengjie Lai, Nick W Ruktanonchai, Liangcai Zhou, Olivia Prosper, Wei Luo, Jessica R Floyd, Amy Wesolowski, Mauricio Santillana, Chi Zhang, Xiangjun Du, et al. Effect of non-pharmaceutical interventions to contain covid-19 in china. *nature*, 585(7825):410–413, 2020.

[106] Neil Lee and Andrés Rodríguez-Pose. Innovation and spatial inequality in europe and usa. *Journal of economic geography*, 13(1):1–22, 2013.

[107] Didier G Leibovici. Defining spatial entropy from multivariate distributions of co-occurrences. In *Spatial Information Theory: 9th International Conference, COSIT 2009 Aber Wrac'h, France, September 21-25, 2009 Proceedings 9*, pages 392–404. Springer, 2009.

[108] Keti Lelo, Salvatore Monni, and Federico Tomassi. Socio-spatial inequalities and urban transformation. the case of rome districts. *Socio-Economic Planning Sciences*, 68:100696, 2019.

[109] 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.

[110] Albert Levy. Urban morphology and the problem of the modern urban fabric: some questions for research. *Urban morphology*, 3(2):79–85, 1999.

[111] John R Logan and Brian J Stults. The persistence of segregation in the metropolis: New findings from the 2010 census. *Census Brief Prepared for Project US2010*, 24, 2011.

[112] Jed A Long and Chang Ren. Associations between mobility and socio-economic indicators vary across the timeline of the covid-19 pandemic. *Computers, Environment and Urban Systems*, page 101710, 2021.

[113] Malte Luebker. Income inequality, redistribution, and poverty: Contrasting rational choice and behavioral perspectives. *Review of Income and Wealth*, 60(1): 133–154, 2014.

[114] Feixiong Luo, Guofeng Cao, Kevin Mulligan, and Xiang Li. Explore spatiotemporal and demographic characteristics of human mobility via twitter: A case study of chicago. *Applied Geography*, 70:11–25, 2016.

[115] Gordon MacLeod. 'institutional thickness' and industrial governance in lowland scotland. *Area*, 29(4):299–311, 1997.

[116] Hanja Maksim and Manfred Max Bergman. *Mobilities and Inequality*. Routledge, 2016.

[117] Anders Malmberg and Peter Maskell. The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering. *Environment and Planning A: Economy and Space*, 34(3):429–449, 2002.

[118] Alfred Marshall. *Principles of economics: unabridged eighth edition*. MacMillan, 1920.

[119] Sallie A Marston. The social construction of scale. *Progress in Human Geography*, 24(2):219–242, 2000.

[120] Ron Martin. Economic theory and human geography. *Human geography: society, space and social science*, pages 21–53, 1994.

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

[122] Wes McKinney. Quantile-based discretization function, 2023. URL `https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.qcut.html`. Accessed: 2023-03-15.

[123] Graham McNeill, Jonathan Bright, and Scott A Hale. Estimating local commuting patterns from geolocated twitter data. *EPJ Data Science*, 6:1–16, 2017.

[124] 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.

[125] 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.

[126] Sahar Mirzaee and Qi Wang. Urban mobility and resilience: exploring boston's urban mobility network through twitter data. *Applied Network Science*, 5(1): 1–20, 2020.

[127] Seyed M Moghadas, Thomas N Vilches, Kevin Zhang, Chad R Wells, Affan Shoukat, Burton H Singer, Lauren Ancel Meyers, Kathleen M Neuzil, Joanne M Langley, Meagan C Fitzpatrick, et al. The impact of vaccination on coronavirus disease 2019 (covid-19) outbreaks in the united states. *Clinical Infectious Diseases*, 73(12):2257–2264, 2021.

[128] Alfredo J Morales, Xiaowen Dong, Yaneer Bar-Yam, and Alex 'Sandy'Pentland. Segregation and polarization in urban areas. *Royal Society Open Science*, 6(10): 190573, 2019.

[129] Barrie S Morgan. A distance-decay based interaction index to measure residential segregation. *Area*, pages 211–217, 1983.

[130] Esteban Moro, Alex Pentland, Dan Calacci, and Xiaowen Dong. Atlas of inequality, 2019. URL `https://inequality.media.mit.edu`. Accessed: 2021-06-28.

[131] 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.

[132] Lewis Mumford. What is a city?(1938). *Historic Cities: Issues in Urban Conservation*, 8:49, 2019.

[133] 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.

[134] NAICS. Economic census: Naics codes & understanding industry classification systems. `https://www.census.gov/programs-surveys/economic-census/year/2022/guidance/understanding-naics.html`, 2022.

[135] Frank Neffke and Martin Henning. Skill relatedness and firm diversification. *Strategic Management Journal*, 34(3):297–316, 2013.

[136] Frank Neffke, Martin Svensson Henning, Ron Boschma, et al. Surviving in agglomerations: Plant evolution and the changing benefits of the local environment. Technical report, Utrecht University, Department of Human Geography and Spatial Planning . . . , 2008.

[137] Vinicius M Netto, Maíra Pinheiro Soares, and Roberto Paschoalino. Segregated networks in the city. *International Journal of Urban and Regional Research*, 39 (6):1084–1102, 2015.

[138] Mark Newman. *Networks*. Oxford university press, 2018.

[139] Mark EJ Newman. Mixing patterns in networks. *Physical Review E*, 67(2): 026126, 2003.

[140] Mark EJ Newman. Complex systems: A survey. *arXiv preprint arXiv:1112.1440*, 2011.

[141] Jan Nijman and Yehua Dennis Wei. Urban inequalities in the 21st century economy. *Applied geography*, 117:102188, 2020.

[142] Timo Ohnmacht, Hanja Maksim, and Manfred Max Bergman. Mobilities and inequality–making connections. In *Mobilities and inequality*, pages 21–40. Routledge, 2016.

[143] Lourdes Diaz Olvera, Dominique Mignot, and Christelle Paulo. Daily mobility and inequality: the situation of the poor. *Built Environment*, 30(2):153–160, 2004.

[144] ONS. Ons model-based income estimates, msoa, 2016.

[145] S Openshow. A million or so correlation coefficients, three experiments on the modifiable areal unit problem. *Statistical applications in the spatial science*, pages 127–144, 1979.

[146] Scott G Ortman, José Lobo, and Michael E Smith. Cities: Complexity, theory and history. *Plos one*, 15(12):e0243621, 2020.

[147] David O'Sullivan, Steven M Manson, Joseph P Messina, and Thomas W Crawford. Space, place, and complexity science, 2006.

[148] Anssi Paasi. Place and region: looking through the prism of scale. *Progress in Human Geography*, 28(4):536–546, 2004.

[149] 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.

[150] Manish Pareek, Mansoor N Bangash, Nilesh Pareek, Daniel Pan, Shirley Sze, Jatinder S Minhas, Wasim Hanif, and Kamlesh Khunti. Ethnicity and covid-19: an urgent public health research priority. *The Lancet*, 395(10234):1421–1422, 2020.

[151] Jaehyuk Park, Ian B Wood, Elise Jing, Azadeh Nematzadeh, Souvik Ghosh, Michael D Conover, and Yong-Yeol Ahn. Global labor flow network reveals the hierarchical organization and dynamics of geo-industrial clusters. *Nature communications*, 10(1):3449, 2019.

[152] John B Parr. Spatial definitions of the city: four perspectives. *Urban studies*, 44 (2):381–392, 2007.

[153] Géraldine Pflieger and Céline Rozenblat. Introduction. urban networks and network theory: the city as the connector of multiple networks, 2010.

[154] Santi Phithakkitnukoon, Zbigniew Smoreda, and Patrick Olivier. Socio-geography of human mobility: A study using longitudinal mobile phone data. *PloS one*, 7(6):e39253, 2012.

[155] Andy Pike, Arnoud Lagendijk, and Mário Vale. Critical reflections on "embeddedness" in economic geography: the case of labour market governance and training in the automotive industry in the north-east region of england. *Restructuring Industry and Territory. The Experience of Europe's Regions. London: The Stationery Office*, pages 59–82, 2000.

[156] Colin Pooley. Mobility, transport and social inclusion: Lessons from history. *Social Inclusion*, 4(3):100–109, 2016.

[157] ME Porter. The competitive advantage of nations. *Harvard Business Review*, 68 (2):73–93, 1990.

[158] Michael E Porter et al. Clusters and the new economics of competition. *Harvard Business Review Boston*, 76(6), 1998.

[159] Antony Potter and H Doug Watts. Evolutionary agglomeration theory: increasing returns, diminishing returns, and the industry life cycle. *Journal of economic geography*, 11(3):417–455, 2011.

[160] Laura Elena Raileanu and Kilian Stoffel. Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence*, 41(1):77–93, 2004.

[161] Akhil Anil Rajput, Qingchun Li, Xinyu Gao, and Ali Mostafavi. Revealing critical characteristics of mobility patterns in new york city during the onset of covid-19 pandemic. *arXiv preprint arXiv:2102.01918*, 2021.

[162] George Ritzer. *Encyclopedia of Social Theory*. Sage publications, 2013.

[163] Salvador Rueda Palenzuela, Albert Santasusagna Riu, Berta Cormenzana Izquierdo, Joan Tort Donada, and Xavier Úbeda. Understanding urban complexity via the spatial diversity of activities: An application to barcelona (spain). *Sustainability*, 14(3):1298, 2022.

[164] 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, 2005.

[165] P Rutten, GA Marlet, and FG van Oort. Creative industries as a flywheel. Technical report, Utrecht University, 2011.

[166] Safegraph. Core places. `https://docs.safegraph.com/docs/places-data-evaluation?utm_campaign=poi_questions&utm_medium=referral&utm_source=quora`, 2022.

[167] Robert J Sampson. Urban sustainability in an age of enduring inequalities: Advancing theory and ecometrics for the 21st-century city. *Proceedings of the National Academy of Sciences*, 114(34):8957–8962, 2017.

[168] Saskia Sassen. The global city: New york, tokyo and london. *NJ: Princeton University Pres*, 1991.

[169] Saskia Sassen. The global city: Introducing a concept. *Brown J. World Aff.*, 11: 27, 2004.

[170] Mike Savage, Fiona Devine, Niall Cunningham, Mark Taylor, Yaojun Li, Johs Hjellbrekke, Brigitte Le Roux, Sam Friedman, and Andrew Miles. A new model of social class? findings from the bbc's great british class survey experiment. *Sociology*, 47(2):219–250, 2013.

[171] 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.

[172] Stefan Schönfelder and Kay W Axhausen. Activity spaces: measures of social exclusion? *Transport Policy*, 10(4):273–286, 2003.

[173] Amartya Sen. *Inequality reexamined*. Harvard University Press, 1995.

[174] Nihal Senlier, Reyhan Yildiz, and E Diğdem Aktaş. A perception survey for the evaluation of urban quality of life in kocaeli and a comparison of the life satisfaction with the european cities. *Social indicators research*, 94:213–226, 2009.

[175] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1):3–55, 2001.

[176] Shih-Lung Shaw. Urban human dynamics. *Urban Informatics*, pages 41–57, 2021.

[177] Eunha Shim, Amna Tariq, Wongyeong Choi, Yiseul Lee, and Gerardo Chowell. Transmission potential and severity of covid-19 in south korea. *International Journal of Infectious Diseases*, 93:339–344, 2020.

[178] Daniel Silver, Ultan Byrne, and Patrick Adler. Venues and segregation: A revised schelling model. *PLoS one*, 16(1):e0242611, 2021.

[179] Wayne Simpson. *Urban structure and the labour market. Worker mobility, commuting and underemployment in cities*. Clarendon Press, 1992.

[180] Gideon Sjoberg. The origin and evolution of cities. *Scientific American*, 213(3): 54–62, 1965.

[181] Jeffrey A Smith, Miller McPherson, and Lynn Smith-Lovin. Social distance in the united states: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004. *American Sociological Review*, 79(3):432–456, 2014.

[182] Jungyul Sohn and Soo Kyoung Oh. Explaining spatial distribution of the middle class: a multiple indicator approach with multiple explanatory dimensions. *Applied Spatial Analysis and Policy*, 12(4):871–905, 2019.

[183] Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.

[184] Karl E Taeuber and Alma F Taeuber. *Residential segregation and neighborhood change*. Transaction Publishers, London, 2008.

[185] Tiit Tammaru, Szymon Marcin´ Czak, Raivo Aunap, Maarten van Ham, and Heleen Janssen. Relationship between income inequality and residential segregation of socioeconomic groups. *Regional Studies*, 54(4):450–461, 2020.

[186] Jinjun Tang, Fang Liu, Yinhai Wang, and Hua Wang. Uncovering urban human mobility from large scale taxi gps data. *Physica A: Statistical Mechanics and its Applications*, 438:140–153, 2015.

[187] Arnold Toynbee. Capital cities: Their distinctive features and capital cities: Melting-pots and powder-kegs. *Cities on the Move*, pages 67–78, 1970.

[188] John Urry. Mobilities polity press, 2007.

[189] Leo Van den Berg, Erik Braun, and Willem Van Winden. Growth clusters in european cities: An integral approach. *Urban studies*, 38(1):185–205, 2001.

[190] Maarten Van Ham, Masaya Uesugi, Tiit Tammaru, David Manley, and Heleen Janssen. Changing occupational structures and residential segregation in new york, london and tokyo. *Nature human behaviour*, 4(11):1124–1134, 2020.

[191] Maarten Van Ham, Tiit Tammaru, Rūta Ubarevičienė, and Heleen Janssen. Rising inequalities and a changing social geography of cities. an introduction to the global segregation book. *Urban socio-economic segregation and income inequality: A global perspective*, pages 3–26, 2021.

[192] 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.

[193] Xiang-Wen Wang, Xiao-Pu Han, and Bing-Hong Wang. Correlations and scaling laws in human mobility. *PloS One*, 9(1):e84954, 2014.

[194] Warren Weaver. Science and complexity. *American scientist*, 36(4):536–544, 1948.

[195] Joakim A Weill, Matthieu Stigler, Olivier Deschenes, and Michael R Springborn. Social distancing responses to covid-19 emergency declarations strongly differentiated by income. *Proceedings of the national academy of sciences*, 117(33): 19658–19660, 2020.

[196] James Q Wilson and George L Kelling. *Broken windows: The police and neighborhood safety. The Atlantic*. Washington, DC, 1982.

[197] David WS Wong and Shih-Lung Shaw. Measuring segregation: An activity space approach. *Journal of Geographical Systems*, 13(2):127–145, 2011.

[198] World Bank. Urban development. https://www.worldbank.org/en/topic/urbandevelopment/overview 2022. URL `https://www.worldbank.org/en/topic/urbandevelopment/overview`.

[199] Lun Wu, Ye Zhi, Zhengwei Sui, and Yu Liu. Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PloS One*, 9(5): e97010, 2014.

[200] Yang Xu, Alexander Belyi, Iva Bojic, and Carlo Ratti. Human mobility and socioeconomic status: Analysis of singapore and boston. *Computers, Environment and Urban Systems*, 72:51–67, 2018.

[201] Dingqi Yang, Daqing Zhang, and Bingqing Qu. Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(3):1–23, 2016.

[202] Ngai Ming Yip, Ray Forrest, and Shi Xian. Exploring segregation and mobilities: application of an activity tracking app on mobile phone. *Cities*, 59:156–163, 2016.

[203] Daniel Zachary and Stephen Dobson. Urban development and complexity: Shannon entropy as a measure of diversity. *Planning Practice & Research*, 36(2): 157–173, 2021.

[204] Chen Zhong, Xianfeng Huang, Stefan Müller Arisona, Gerhard Schmitt, and Michael Batty. Inferring building functions from a probabilistic model using public transportation data. *Computers, Environment and Urban Systems*, 48: 124–137, 2014.

[205] Jelena Živković. Urban form and function. *Climate action*, pages 862–871, 2020.