Measuring Spatial Inequalities in Public Transportation Systems: Three Key Perspectives

Submitted by Nandini Sudha Iyer, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Computer Science, March 2024.

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Abstract

For several decades, inequalities in public transportation systems have been of significant consideration to transport researchers and urban planners. Increasing concerns regarding the climate crisis, coupled with improved standardisation of public transit data, has led to a myriad of studies highlighting apparent disparities in transit service and infrastructure. However, research on transit inequalities typically focus on a singular urban mechanism. In this thesis, we present three perspectives, and their respective frameworks, that are crucial for assessing how disparities in public transit can be exhibited. We argue that, in order to develop a well-rounded understanding of inequalities in public transit infrastructure, one must consider how transit intersects with (a) disparities in residential-workplace dependencies, (b) experienced segregation, and (c) features of the physical environment. Focusing on cities in the United States of America (US), we analyse transit with respect to residential and employment landscapes, highlighting how housing and occupational disparities are exacerbated by transit service, further disadvantaging vulnerable communities. Then, we estimate experienced segregation in daily mobility patterns, underscoring the segregated nature of transit service in terms of mobility opportunity and empirical mobility patterns. Finally, we identify transit inequalities for neighbourhoods with similar physical environments, highlight how spatial characteristics elucidate the types of environment in which socioeconomic transit disparities arise. The presented frameworks emphasise how public transit heightens the inequalities that are present in housing landscapes, mobility patterns,

and the built environment, ultimately providing a quantitative perspective on how US transit systems provide opportunities for fulfilling different types of mobility desires.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and below. The following papers, the findings of which are used in this dissertation, have been published, presented at conferences, or are currently under review for publication:

Publications included in the Dissertation:

- Iyer, N., Menezes, R., & Barbosa, H. (2023). Network Entropy as a Measure of Socioeconomic Segregation in Residential and Employment Landscapes. Complex Networks XIV. CompleNet 2023. Springer Proceedings in Complexity. Springer, Cham. – **Chapter 4**

- Iyer, N., Menezes, R. & Barbosa, H. The role of transport systems in housing insecurity: a mobility-based analysis. *Under Review at EPJ Data Science.* – **Chapter 4**

- Iyer, N., Menezes, R., & Barbosa, H. (2023). Mobility and transit segregation in urban spaces. Environment and Planning B: Urban Analytics and City Science. https://doi.org/10.1177/23998083231219294 – **Chapter 5**

- Iyer, N., Menezes, R. & Barbosa, H. Public Transit Inequality in the Context of the Built Environment. Complex Networks XV. CompleNet 2024. Springer Proceedings in Complexity. Springer, Cham. *To appear April 2024.* – **Chapter 6**

Nandini Sudha Iyer

March 2024

Acknowledgements

I would like to start by thanking my first supervisor, Dr. Hugo Barbosa, and my second supervisor, Prof. Ronaldo Menezes for all their insight and guidance. You have both provided me with thoughtful feedback and support throughout the years and without your consistent encouragement and expertise, this thesis would not have been possible. I would also like to thank my examiners, Dr. Marcos Oliveira and Prof. Laura Alessandretti, for taking the time to provide insightful comments and ideas, which bolster the quality of this work.

My deepest gratitude goes to my parents and brother for being such grounding forces despite the physical distance and time difference between us. Your constant support has instilled me with the confidence to follow my research interests. Our calls have proven to be the highlights of my days here and I could always count on them to be refreshing and filled with laughs.

To Lee, Morgan, Jane, and Heather– your humour and creativity have been constant sources of inspiration to me. It warms my heart thinking about the efforts we have made to stay close and the many great experiences we have shared throughout the past few years. I find so much comfort in our friendships and I am so grateful to have you all in my life.

Finally, I would like to thank my friends I've made here in Exeter. I feel so appreciative to have made such cherished memories together. To Ana, Fredie, Trish, Clodomir, Arpi, Nathanael, Barbara, Emma, Liv, Andrea, Chico, Sima, and Ian, you have made Exeter feel like home and I feel so lucky to have met people as loving as you. I am so excited to keep hiking, crafting, baking, and getting tangled up in new shenanigans with you all.

Table of contents

List of figures

List of tables

List of abbreviations

US United States of America

MAUP Modifiable Areal Unit Problem

CSA Connection Scan Algorithm

RAPTOR Round-bAsed Public Transit Optimised Router

GTFS General Transit Feed Specification

R^5 Rapid Realistic Routing on Real-world and Reimagined networks

OSM OpenStreeMap

ORS Openrouteservice

OSRM Open Source Routing Machine

CBG Census Block Group

CT Census Tract

ACS American Community Survey

LODES LEHD Origin-Destination Employment Statistics

LEHD Longitudinal Employer-Household Dynamics

PPR Persons-Per-Room

PPB Persons-Per-Bedroom

CDR Call Detail Records

BE Built Environment

ICE Index of Concentration at the Extremes

LISA Local Indicators of Spatial Association

HS-Lo Highly-segregated, low-income

MS-Lo Mildly-segregated, low-income

LS Less-segregated

MS-Hi Mildly-segregated, high-income

HS-Hi Highly-segregated, high-income

Chapter 1

Introduction

Inequality has existed in urban areas for centuries, albeit at varying scales and forms. Inequality refers to an uneven distribution of resources, goods, or conditions [62]. Regardless of the various scales and types of inequalities that individuals face, the need to move is universal across all people. That is, mobility plays a crucial role not only in terms of commuting, but also for satisfying lifestyle necessities, leisure desires, and social connectivity [41]. However, policies that regulate land-use, such as zoning laws and investments in employment and retail areas, can lead to residential inequalities, such as segregation [264, 102, 71]. Furthermore, poor transit links between residential, retail and employment areas may limit individuals, who are already facing residential inequalities, from satisfying their mobility desires [281]. Accordingly, it is crucial to understand mobility in various urban contexts, as when mobility is a forced process, it can reflect deprivation, whereas when mobility is an option, it can be indicative of privilege [104, 203]. Moreover, given that lower income individuals tend to have higher rates of transit dependency [139], it is essential to clearly articulate the role that public transit plays in providing different demographics access to varying types of destinations. In doing so, policy-makers and transportation engineers can understand how changing public transit systems (i.e., increasing frequency of transit, adding routes in particular neighbourhoods) can enable mobility opportunities.

Mobility equity aims to explore this issue, with early studies analysing the spatial connectivity of residential and workplace locations [294, 143]. Since this initial research, increased access to high-resolution mobility data allowed mobility research to shift from a population level to an individual level [247]. Mobile network operators and smartphone applications have since been providing mobility datasets of varying scale and detail, allowing researchers to identify universal mobility patterns or disparities in mobility behaviour across particular demographics. For example, a recent study leveraged high resolution mobility data to unveil the distinct urban routines that neighbourhoods of various deprivation levels have [46].

Studies such as this allude to the concept of *mobility justice*, which encapsulates mobility disparities arising from the quality of experience, access to infrastructure, and temporalities of mobility [149, 262]. The latter refers to the impact that waiting for and overtaking other vehicles has on transit journeys. Thus, mobility justice refers to understanding why space connects certain individuals, while fragmenting others, in order to develop solutions that ensure that vulnerable demographics (low-income individuals, older adults, racial and ethnic minorities, etc.) are not socially excluded from the larger population due to a lack of mobility options.

Additionally, as issues concerning the climate crisis become more prominent, it is critical to consider how to *fairly* transition to more sustainable modes of transit, such that mobility remains affordable and accessible for all fragments of the population. However, distinguishing transport poverty as a distinct, systemic problem, rather than a consequence of individual economic poverty, remains an open challenge. [177]. Furthermore, research that does focus on transport poverty spans numerous topics, including mobility poverty, accessibility poverty, transport affordability, and exposure to transport externalities [176]. Research in these fields has shown that low-income residents face the highest public transport fares, but spend a larger portion of their earnings on owning cars because of inadequate public

transit options [69, 126]. The lack of clarity in defining transport poverty as a significant, stand-alone problem, makes it difficult to clearly articulate when, where, and for whom transport systems can be improved. This often leads to ineffective policies, such as providing transit fare subsidies or loaning private vehicles to vulnerable demographics [183, 177]. These approaches to reduce transport poverty typically fall short due to data availability, targeting the wrong demographics, and being unfeasible to adopt on a large scale [178]. Thus, in order to creates policies that are effective in targeting transit poverty, it is crucial to understand how transit systems serve different areas and demographics in a region, with respect to various mobility purposes.

1.1 Research Questions

The previous section demonstrates how transport poverty is still a nascent and developing concept. Moreover, understanding transport inequality as a subset of mobility justice motivates defining disparities in public transportation with respect to various types of mobility purposes. However, using mobility data to analyse transit poverty can lead to creating strong transit links between areas with high mobility flows, while providing those areas with few transit options to other parts of the city. Thus, as urban landscapes shift, support networks, employment hubs, and amenities may become fragmented from the communities that depend on them. Accordingly, in this thesis, we argue that transport inequality should be understood with respect to its ability to fulfil both potential and empirical travel demands, while also accounting for spatial disparities that arise due to various urban mechanisms, such as housing markets and amenity accessibility. Thus, we contend that it is critical to understand public transit in terms of how it (1) connects individuals to employment opportunities, (2) serves as a tool to overcome residential segregation by reducing levels of experienced segregation in other domains of urban life and (3) contributes to socio-spatial inequalities, when considering features of a neighbourhood's physical environment.

In doing so, we capture how disparities in transit service intersect with the following urban mechanisms: residential-workplace dependencies, human mobility, and the built environment. The frameworks we outline not only consider structural characteristics of commuting networks, segregated transit networks, and the built environment, but also provide insight regarding the experience of using transit in the context of these different urban functions. That is, we consider how housing insecurity, which measures the several dimensions in which housing can be inadequate, impacts the employment areas one can commute to using transit, within a reasonable time frame. Moreover, we assess how mobility can serve as a function to overcome residential segregation, allowing for points of interaction at amenities and while using transit systems. Finally, we analyse how socioeconomic inequalities arise in access to workplaces and essential amenities, when considering neighbourhoods with similar built environment features. We ask the following questions, with respect to various major cities in the United States of America (US). This region of analysis was chosen for three main reasons. First, the US has a wealth of socioeconomic, geographic, and mobility data available [43, 252]. Second, the US is comprised of many cities, exhibiting a large heterogeneity in urban landscapes and quality of transit infrastructure [293]. Finally, the US is unique in that it has openly accessible eviction data for numerous cities, which is useful for quantifying measures of housing insecurity [72].

Residential-Workplace Dependencies:

RQ1 How can commuting networks capture socioeconomic disparities in residential-workplace dependencies that are overlooked when using measures of place-based inequality? How do structural network properties correspond with conventional segregation metrics?

RQ2 How can we develop a comprehensive measure for housing insecurity that extends beyond measuring one dimension of it (i.e. affordability, stability)? How can we leverage available data to identify neighbourhoods that are particularly vulnerable to housing insecurity? How do estimations of housing insecurity correspond with various sociodemographic characteristics?

RQ3 In what ways does transit service inhibit employment accessibility? Are there similarities in commuting characteristics across housing demographics? How does public transportation facilitate job accessibility for different demographics?

Human Mobility:

RQ4 To what extent does residential segregation persist in travel behaviour? How do changes in segregation levels, from the residential to the mobility dimension, impact socioeconomic groups differently?

RQ5 How can we leverage transit modelling to estimate how transit systems contribute to existing levels of residential segregation? In what ways do different socioeconomic groups experience segregated transit service? How can we estimate the extent to which inequalities in mobility patterns create disparities in the transit routes that socioeconomic groups would use?

Built Environment:

RQ6 How does accounting for physical features of a neighbourhood reveal spatial disparities in transport service? How can we define neighbourhoods based on how its urban planning bolsters flexibility in mobility choices? Are there inequalities in how transit

systems serve neighbourhoods with different spatial characteristics?

RQ7 To what extent does incorporating both socioeconomic and spatial features reveal additional insights regarding how transit accessibility may disadvantage vulnerable demographics? In what types of neighbourhoods are socioeconomic transit inequalities prevalent? How are socio-spatial transit inequalities exhibited when considering access to workplaces, nearby amenities, and overall mobility opportunity within a city?

1.2 Document Structure

In this document, we explore inequalities in human mobility with respect to three urban perspectives: residential-workplace dependencies, experienced segregation, and the built environment.

The current chapter, **Chapter 1**, presents a high-level overview of inequality in mobility and public transit. It lists the research questions that will be addressed in the thesis and outlines the overall structure of the document.

Chapter 2 provides a literature review on transport poverty on a theoretical and quantitative level. It begins by exploring various political philosophies on justice. Then, we address static approaches to analysing urban areas, based on inequalities in residential and occupational landscapes. This is followed by a more dynamic conceptualisation of urban analytics, introducing the field of human mobility, along with accompanying models, methods, and applications. Finally, we narrow in on the concept of transport justice, which lies in the intersection of social justice, urban planning, and human mobility. The following three chapters explore how transit disparities intersect with residential-occupational mismatches, experienced segregation, and the built environment, respectively.

Chapter 3 provides an overview of data sources that are used in the thesis, along with details regarding fundamental methods used to define core metrics, outlining the segregation

metrics we use and the mobility data characteristics. We introduce measures that characterise urban areas based on their physical features. Finally, we use transport data to highlight the differences in open-source transport modelling tools, as well as the reasons for which they arise.

In **Chapter 4**, we explore the spatial connection between residential and employment locations, particularly highlighting inequalities and inefficiencies in public transport commutes. We begin by highlighting how structural inequalities in commuting networks capture disparities that conventional segregation metrics overlook; then, we introduce a framework that estimates housing insecurity to highlight how transit infrastructure restricts employment accessibility. This chapter underscores the significance of residential-workplace dependencies in the context of transit service.

Chapter 5 incorporates mobility data to elucidate how segregation spills over from the residential dimension, impacting experienced segregation on public transit lines and at various amenities. Furthermore, it demonstrates how transit service itself facilitates segregation by providing inadequate access to a diverse range of neighbourhoods. In doing so, we highlight the importance of measuring transit inequality with respect to how individuals experience segregation throughout multiple urban dimensions.

In **Chapter 6**, we identify disparities in transit service, comparing socioeconomic inequalities in areas that have similar urban features. We define each city based on how well the urban planning supports flexibility in mobility options. Then, we evaluate transit service with respect to how it provides access to all other neighbourhoods in a city, nearby amenities, and employment locations. Thus, we show how incorporating built environment features into transit analysis helps to pinpoint the types of neighbourhoods in which socioeconomic transport disadvantage is prevalent.

Finally, **Chapter 7** reconciles the three urban mechanisms, providing concluding remarks and discussing limitations in data and methodology. Additionally, we address how future work could build upon the methodologies introduced in this thesis.

Chapter 2

Literature Review

Shifts towards more sustainable modes of transportation have placed a large emphasis on improving public transit options for residents of urban and rural areas. During this transition, it is critical to ensure that "improvements" to transit infrastructure do not solely benefit those in privileged positions. This thesis outlines three spatial perspectives to consider when analysing inequalities in transit systems: residential-workplace dependencies, urban segregation, and the physical features of an environment. In this chapter, we highlight relevant research that emphasises the importance of each of the aforementioned perspectives. We begin by outlining theoretical arguments for social equity and justice, addressing the importance of mitigating inequality. Furthermore, we review how different schools of thought define equality in a society. Then, we discuss approaches to estimating inequality in urban areas, in terms of how urban spaces are used and designed. Finally, we centre our focus on inequalities in transit infrastructure, providing an overview of transport research.

2.1 Sociodemographic Inequality

Although inequality is typically considered from a normative perspective, not all individual characteristics should be analysed from this point of view. For instance, traits such as hair

colour may have uneven distributions, but are not necessarily reflective of an unfair society. Furthermore, it is important to distinguish between poverty and inequality, as the former focuses on the lower end of the entire distribution, while inequality consider the entire distribution altogether.

Some economists and researchers argue that inequality is an inevitable artefact that creates a healthily competitive market and leads to positive economic outcomes [114]. A bulk of these arguments rely on Simon Kuznet's hypothesis, which suggested that increasing inequality is a sign of development in a country [157]. Specifically, he suggests that inequality and development have an inverted U-shaped relationship, such that inequality will increase as a region develops, until it reaches a critical point, after which inequality will decrease. However, the return to high levels of inequality, particularly in developed countries such as the United States of America and the United Kingdom, has cast doubt on this hypothesis [234]. Moreover, others have demonstrated the importance of developing policies to mitigate excessive inequalities that extend beyond moral obligation [132, 281]. These arguments focus on negative consequences of rising inequality, such as health outcome disparities and imbalanced political power.

Disparities in urban areas have long been studied across multiple disciplines, with initial research focusing on place-based inequalities such as residential and occupational segregation, with segregation referring spatial inequalities in housing or employment opportunities are distributed. [150, 95, 143]. Previous research has identified that social inclusion fuels the productivity of cities [75, 292]. Accordingly, investing in improving the diversity of cities can lead to positive socioeconomic outcomes, fostering innovation and entrepreneurship [239]. These types of research highlight the instrumental value of equality by revealing the association between inequality and positive economic outcomes. However, numerous philosophers and economists support the notion that reducing inequality has intrinsic value as

well, arguing that equality is a basic human right and that ensuring equality of opportunities and outcomes is inherently good [184, 244, 140].

In this manner, inequality can be considered at either the opportunity level or the outcome level [12]. Equality of opportunity strives to provide individuals with the same set of opportunities, which can help them reach their preferred outcomes. Meanwhile, equality of outcomes refers to individuals having comparable levels of material wealth and economic capital. While a majority of inequality research focuses on reducing inequality of opportunity, to "level the playing field", economists, such as Atkinson, point out that inequality of outcomes deserves a similar level of scrutiny. This if for three reasons: (1) it is difficult to separate unequal outcomes due to chance from unequal outcomes due to hardship, (2) the perceived importance of equality of opportunities stems from the unequal distribution of rewards, and (3) inequality of outcomes directly impacts inequality of opportunity for future iterations of society. Thus, identifying inequality from both the opportunity and outcome perspective is crucial for developing a complete understanding of inequality in a region.

Theories of justice can be used to address inequality, as they advocate for diffrent way in which benefits should be distributed and the moral principles under which they should be distributed [218, 60]. Here, we cover three different theories of justice: utilitarianism, egalitarianism, and the capabilities approach. Utilitarianism focuses on maximising a particular dimension of human experience, such as happiness or human welfare [38]. Additionally, it has an outcome-oriented approach, in which the welfare of individuals is weighted equally. Meanwhile, Rawls' egalitarianism builds upon two principles, the first of which takes precedent over the second: (1) every individual is entitled to basic liberties and (2) resources and goods should be distributed to maximise benefit to most disadvantaged and should be a consequence of equal opportunity [148, 244]. Rawls' formulation of egalitarianism suggests that under the premise of equality of opportunity, inequalities in outcomes are justifiable as they are reflective of individual choice. While utilitarianism prioritises helping individuals who

benefit the most, egalitarianism focuses on helping individuals who are in a disadvantaged position [280]. At times, these two theories lead to the same outcome, particularly when the individuals who gain the most are also the most disadvantaged.

These two theories overlook "doings" (the activities in which individuals can engage) and "beings" (who individuals are able to be). Amartya Sen argues that distributive theories that focus solely on resources dismiss how different groups and social contexts have varying abilities to transform these goods into a beneficial outcome [259]. Moreover, he introduces the capabilities approach, which refers to the freedom individuals have to achieve their "beings" and "doings". Translating the capabilities approach into a quantitative analysis requires extending accessibility measures beyond travel time, and incorporating features such as residential inequality, affordability, safety, and reliability. Furthermore, it entails understanding disparities in how transit serves demographics, with respect to varying trip purposes. Michael Walzer, a political theorist, advocates for a more nuanced understanding of justice, which he terms "complex equality" [298]. Here, he argues in favour of categorising goods into various spheres (i.e., community membership, kinship and love, political power), such that each sphere has a separate distribution principle and dominance in one sphere does not diffuse into dominance in another sphere. Ultimately this overview underscores how despite being aligned under the pursuit of autonomy and moral equality, differences in conceptualising justice can still arise.

2.2 Place-based Inequalities

Understanding the effects of social integration has been at the forefront of sociological research for centuries, highlighting the many benefits of well-integrated communities, such as lower crime rates and better health outcomes [145, 258, 11]. In this section, we discuss the various way in which disparities, based on the characteristics of where individuals live or work, can be measured.

2.2.1 Residential Segregation

A majority of research on integration, and its counterpart, segregation, tend to focus on residential characteristics, assuming the individuals that one interacts with are likely to be from one's neighbourhood. Segregation is typically derived by comparing the empirical distribution of demographics in different neighbourhoods to an equally distributed version of the population across the entire region. This is often followed by disclaimers that an equal distribution may not be ideal or just from a social perspective [40], offering a subtle nod to sociological studies that elucidate how the tendency for individuals to associate with others that are similar to them (known as homopholy [196]), is not strictly a positive or negative force. Therefore, it is important to distinguish whether homophilic processes are occurring as a result of individual preferences or systemic constraints. Often, cities have highly concentrated pockets of individuals from similar backgrounds, which, depending on the environmental context, has the potential to provide a sense of community to residents while also exposing the broader population to different cultures and practices [228, 200, 152]. The disparity arises when urban infrastructure does not provide sufficient or adequate resources to facilitate journeys to other neighbourhoods, for individuals within these communities. Instances such as these allow systemic inequalities to shape segregation in mobility behaviour.

Residential segregation has been in the spotlight of sociological and urban studies for centuries, with research consistently churning out new methods for conceptualising and measuring how sociodemographic groups share spaces. Segregation estimates the extent of spatial separation for two or more demographics in a given region. It is often distilled into five dimensions: evenness, exposure, concentration, centralisation, and clustering [191].

Evenness

Given a region of interest, the evenness dimension captures how uniformly demographic groups are distributed across the sub-regions, often measured using the dissimilarity in-

dex, Gini coefficient, or mutual information index. The dissimilarity index is usefule for understanding the spatial spread of a demographic in a given region. Specifically, it conveys the percentage of a minority population that would have to change in order for each neighbourhood to have the same demographic distribution as the entire region [191]:

$$\text{Dissimilarity} = \frac{\sum_{i=1}^{n} \Big(t_i |p_i - P| \Big)}{2TP(1-P)} \tag{2.1}$$

where p_i and P reflect the fraction of minority demographics in a neighbourhood i and the entire region, respectively. T is the total population in the entire region, while t_i refers to the population in a neighbourhood i. For regions in which the minority population is equally distributed across neighbourhoods, $|p_i - P|$ will be smaller for all neighbourhoods, resulting in a lower numerator and, in suit, a lower dissimilarity index. Meanwhile, when a region has some neighbourhoods that significant minority over-representation (with other neighbourhoods consequently having noteable minority under-representation), $|p_i - P|$ will be larger, leading to higher values of dissimilarity. In this manner, larger values of the dissimilarity index, which is bounded between 0 and 1, indicate higher levels of segregation for the minority population.

The mutual information index, proposed by Theil, is a multi-group measure of diversity in a given region, based on information theory, that estimates how knowing a group's representation in a given neighbourhood impacts the uncertainty level of knowing its representation in another area [285]:

$$\begin{aligned} \text{Mutual Information} &= \frac{\sum_{i=1}^{n} \Big[t_i (E - E_i) \Big]}{ET}, \\ \text{where } E_i &= p_i ln(\frac{1}{p_i}) + (1-p_i) ln(\frac{1}{1-p_i}) \\ \text{and } E &= P ln(\frac{1}{P}) + (1-P) ln(\frac{1}{1-P}) \end{aligned} \tag{2.2}$$

where p_i coveys the fraction of the population in a spatial unit (neighbourhood), i, and P denotes the minority population across the entire region of analysis. Furthermore, T denotes the total population of the entire region. Thus, E captures the diversity score given an entire region, while E_i does the same, but for a specific neighbourhood, i, within the region. Thus, the mutual information index uses a weighted average, accounting for the population distribution across neighbourhoods, to estimate how much each neighbourhood's diversity deviates from that of the entire region. It is bounded between 0 and 1, with higher values indicating a greater deviance from the region's overall demographic composition (i.e., more segregation).

Exposure

Exposure reflects the likelihood of interaction between demographics, under the notion that living in the same area increases the likelihood of contact. The motivation for estimating segregation in this dimension is similar to the methodologies used in this thesis, because it attempts to measure the experience of segregation in a residential setting. The two most common metrics are isolation and interaction. Isolation captures the probability that a vulnerable demographic shares a neighbourhood with their own sociodemographic group. Meanwhile, interaction shows the likelihood that a vulnerable demographic shares a space with the majority demographic:

$$\text{Isolation} = \sum_{i=1}^{n} \left[\frac{x_i}{X} \frac{x_i}{t_i} \right]; \quad \text{Interaction} = \sum_{i=1}^{n} \left[\frac{x_i}{X} \frac{y_i}{t_i} \right] \tag{2.3}$$

where x_i, y_i, and t_i reflect the number of individuals in the minority, majority, and total population, respectively, in a spatial unit i. Meanwhile X reflects the entire minority population across the whole region of analysis.

With this in mind, isolation considers the minority population in a neighbourhood, with respect to the minority population in the entire region ($\frac{x_i}{X}$) and the total population in the

neighbourhood ($\frac{x_i}{t_i}$). Isolation has an upper bound of 1, when the entire minority population, X, makes up the entire population in neighbourhood i. On the other hand, interaction accounts for the fraction of a minority population that lives in a neighbourhood ($\frac{x_i}{X}$) as well as the majority representation in the same neighbourhood ($\frac{y_i}{t_i}$). Thus, an interaction value closer to 1 reflects a higher likelihood that an individual from the minority group lives in the same neighbourhood as someone belonging to the majority group.

Concentration

Concentration incorporates spatial attributes, identifying areas where minority groups reside in a small proportion of a region. The most common metric for estimating concentration is Delta, which measures the distribution of a vulnerable demographic, with respect to the distribution of space in a neighbourhood [129]:

$$\text{Delta} = 0.5 \sum_{i=1}^{n} \left| \frac{x_i}{X} - \frac{a_i}{A} \right| \tag{2.4}$$

where a_i is defined by the land area in a neighbourhood i and A measures the entire land area of the city.

In this thesis, we use the Index of Concentration at the Extremes (ICE), proposed by Douglas Massey, to understand how inequalities are spatially distributed, in regards to both the most affluent and most poor households in the relevant US cities [190]. This metric was conceived in an attempt to describe a region using both spatial attributes and the characteristics of polarised demographics. While most segregation metrics account for how segregated the minority group is in relation to the entire population, ICE incorporates both ends of the demographic spectrum: :

$$\text{ICE}_i = \frac{A_i - P_i}{T_i}, \tag{2.5}$$

where the ICE for a neighbourhood, i, is defined as the difference between the number of affluent residents, A_i, and the residents below the poverty line, P_i, over the entire population, T_i. While the numerator captures the imbalance between the extremes, the denominator expresses the degree of imbalance in relation to the entire population of neighbourhood i. Thus, ICE aims to measure the imbalance between affluence and poverty by measuring the concentration of both the extremely disadvantaged and advantaged in a given population. In Chapter 5 we compare ICE to many of the segregation indices outlined in this chapter, including the social similarity index.

The social similarity was introduced in a recent study, which used a rank-based approach for measuring the distance between socioeconomic groups [315]. It is particularly useful for incorporating perceived levels of inequality, with respect to different demographics (i.e. low income individuals may perceive segregation to be worse than their high income counterparts). This measure requires high resolution socioeconomic data, in which each individual x is associated with an income rank r_x, such that higher ranks imply higher socioeconomic status. Furthermore, it defines A_{ij} as the set of individuals who are closer to a socioeconomic rank i than a socioeconomic rank j, defined as $A_{ij} = \{x \Big| |r_x - r_i| < |r_i - r_j|\}$, for all individuals x in a population of size N. The social distance between two ranks i and j can be defined as:

$$d_{i \to j} = \begin{cases} \frac{||A_{ij}||+0.5}{N-1}, & \text{if there exists a rank k } (k \neq j) \text{ such that} |r_k - r_i| = |r_i - r_j| \\ \frac{||A||}{N-1}, & \text{otherwise} \end{cases} \tag{2.6}$$

In defining social distance in this manner, $d_{i \to j}$ may not be equal to $d_{j \to i}$, incorporating how various demographics may perceive socioeconomic distance differently. The social similarity can be calculated by taking the complement of social distance: $s_{i \to j} = 1 - d_{i \to j}$. Thus, a larger value of $s_{i \to j}$, which is bound between 0 and 1, implies more social similarity and, consequently, more segregation. In this thesis, we use income brackets to define the

socioeconomic rank of individuals in a neighbourhood, given the neighbourhoods income distribution across these brackets. Then, the social similarity that an individual, x, experiences in her neighbourhood can be described by calculating the average of her social similarity to all other individuals, y, in their neighbourhood:

$$\text{Social Similarity}_x = \frac{\sum_{j=1}^{m} s_{x \to y_j}}{m} \tag{2.7}$$

where m is the number of other individuals living in the same neighbourhood as x and y_j reflects the socioeconomic rank of a given individual. This can be aggregated to a spatial unit by measuring the average social similarity that each individual in a spatial unit experiences. We use the social similarity index, as well as the index of dissimilarity, isolation, and mutual information, to validate our use of ICE in Chapter 6.

Centralisation and Clustering

Finally, centralisation reflects residential proximity to the urban core, while clustering defines how much spatial clustering minority groups express in their residential distribution. While some dimensions are related to one another, each is conceptually distinct and important for characterising the state of segregation in a region. It is important to note that spatial versions of these segregation metrics have been proposed to incorporate a more nuanced understanding of how demographics are distributed across space. In this thesis, we use residential segregation as a baseline for comparison between more dynamic forms of experienced inequalities derived from high-resolution mobility data. Consequently, spatial extensions of the above segregation measures are not strictly relevant in this context, as they make assumptions, typically through the use of kernel density estimations, that enforce rules about how proximity to other neighbourhoods influences residential segregation. In this manner, we can account for both residential and mobility inequalities, while also being able to compare across these two dimensions of urban life.

2.2.2 Residential Mobility

While residential segregation measures sociodemographic inequality of a neighbourhood at a given time, neighbourhoods can be conceptualised as a more dynamic entity, with an inflow and outflow of households that define the demographic composition of a neighbourhood over time. Residential mobility is often an outcome of life stages, housing markets, and urban structure [61]. Moreover the authors of Ref. [61] show how residential mobility can broken down into two forces: the motivation to leave one's current house and choosing the area to which one will relocate. Disparities exist in both of these components. That is, the motivation to move can arise from shifts in household needs due to a change in household structure or from urban mechanisms such as gentrification and evictions, which lower-income residents tend to experience more frequently [70]. Similarly, residential choice has been shown to be associated with mechanisms such as neighbourhood attachment, which refers to a pull force between individuals and their physio-social environment [58].

Other works have shown how low-income households that are given cash in place of subsidies, move to areas with a similar sociodemographic composition to their original neighbourhoods. Furthermore, the households that do move to more integrated neighbourhoods still maintain their contacts from their previous neighbourhood [249]. As employment opportunities moved to the suburbs, researchers began evaluating how urban poverty may, in turn, suburbanise [128, 243]. While this could be interpreted as a consequence of less discrimination in housing policies, it is essential to accommodate these residential shifts by creating effective anti-poverty programs in the suburbs [27, 154]. Identifying the increasing suburbanisation of poverty is particularly interesting, considering that poverty had previously been associated with urban and rural areas. More recent research in this domain demonstrates how forces such as gentrification push more vulnerable demographics to less accessible regions [127, 167]. Thus, we see how inequalities arise in the residential dimension, both

when housing is interpreted as a static feature (segregation), and when residential mobility is considered.

2.2.3 Housing Insecurity

The rapidly increasing density of urban areas poses a threat to exacerbate the ever present housing crises around the world [318, 105]. Housing crises are often linked to economic and political forces, as seen by the Global Financial Crisis of 2007–2008 [57, 142, 307]. Social scientists and urban planners have consistently and carefully studied the sources and consequences of unstable housing using different granularities. In particular, the COVID-19 pandemic enforced the importance of studying housing insecurity. As travel restrictions limited mobility, individuals were confined to their places of residence, emphasising the significance of housing quality [232, 26]. Over the years, understanding the motivation for residential movements of disadvantaged groups has posed a considerable challenge. Desmond et. al suggest eviction, coupled with neighbourhood dissatisfaction, gentrification, and slum clearance as potential explanations for the high levels of moves among the urban poor [70]. Furthermore, housing insecurity has been shown to have a negative effect on job accessibility, wellbeing, and the stability of support networks [71, 73, 151].

Although these works focus on housing insecurity, they tend to quantify this concept using a single measure, such as eviction rates or rent burden, as a proxy. Furthermore, developing a consistent metric for housing insecurity has been a long-standing obstacle [164]. In an effort to provide a universal metric for housing insecurity, Cox et al. survey existing research in housing issues to define seven dimensions of housing insecurity: housing stability, housing affordability, housing quality, housing safety, neighbourhood safety, neighbourhood quality, and the last, optional dimension, homelessness [63]. Housing stability focuses on concepts such as overcrowding in houses, evictions, and frequent moves. Housing affordability encompasses financial aspects of housing such as rent burden, incomplete or late payments,

mortgage, and taxes. Meanwhile, housing quality focuses on the robustness of the house as a physical structure. This includes characteristics such as the functionality of appliances and how rundown a house's interior and exterior are. Housing safety differs from housing quality, in that is measures the presence of vital housing facilities such as heating, water access, and (a lack of) pests. The neighbourhood level characteristics portray the notion that residential locations are not based solely on the house, but also account for the environment in which the house exists. Neighbourhood safety can be described by crime rates, abandoned buildings, proximity to environmental threats, noise levels, and traffic. Neighbourhood quality encompasses urban infrastructure quality, such as amenity accessibility.

2.2.4 Beyond the Residential Dimension

The urban inequalities presented thus far assume that spatial inequality can be explained by residential characteristics. The emergence of connected enclaves (areas that share cultural or sociodemographic similarities, but are spatially separated) and new mobility data sources have resulted in considering how other aspects of urban life contribute to sociodemographic inequality [310]. The environment in which individuals live has been shown to shape the mobility choices they make [198]. Therefore, understanding inequality based on the physical context in which one resides can provide a clearer understanding of how urban inequality is expressed in a manner that is not solely derived from the demographic composition of one's neighbourhood. The urban experience, in terms of mobility, is largely comprised of trips to work [138], underscoring the importance of considering social inclusion not only from the residential dimension, but the employment perspective as well. In this sense, mobility, specifically commuting patterns, can be viewed as a potential way to improve diversity and integration by creating points of social inclusion in employment areas [111].

In the social sciences, occupational inequality has been a main factor for understanding economic inequality [246, 211]. Occupational inequalities can arise from differences in

productivity valuation of employees, worker-job skill mismatches, and variations in perceived reward from jobs [120]. Furthermore, it is well established that the nature of residential and occupational choice is deeply intertwined [304, 269]. Initial urban spatial models were developed to understand residential demand and how employment changes can spur residential relocation [174, 6]. However, the majority of these models assume that workplace choice affects residential choice [295, 260]. From a spatial perspective, misaligned employment and residential landscapes can lead to spatial inequalities. Metropolitan areas tend to have a close association between levels of segregation based on residential distributions and workplace segregation [113, 209]. It becomes clear, therefore, that solely analysing inequality from a residential dimension does not accurately depict how residents may experience inequality in their day to day lives.

The environment in which individuals reside and spend time can have a range of impacts on their well being, from affecting one's health through pollution and noise to influencing one's behaviours based on mobility options and urban design [161]. The physical area that is human-made is referred to as the built environment. It is comprised of land use patterns, transportation systems, and physical road and sidewalk infrastructure [115]. The built environment is conventionally characterised with respect to density, diversity, and design [49]. Density is often estimated with respect to the spatial unit being considered. It can be defined to estimate density from a population perspective (i.e., persons per area) or from a structural perspective (i.e., dwelling units per area). Diversity is often captured by measuring the variety of land uses in a region, which is found by applying Shannon entropy to land use distributions. Meanwhile, design pertains to quantifiable features of urban planning styles. This can be defined in numerous ways, from the density of intersections and pedestrian crosswalks to the grid-like patterns of urban roads, in contrast to sparser suburban streets that generally exhibit a larger degree of curvature. More recent research on the built environment has extended the 3D's to include features such as distance to transit,

and destination accessibility [85]. Distance to transit is often interpreted as the proximity to the closest transit stop or density of the transit system.

In this section, we have considered fairly spatial aspects of inequality in urban areas. While complex, and influenced by political, cultural, and numerous other forces, these inequalities are fairly static. The next section introduces the field of human mobility, which can provide fine-grained estimations of how inequality exists and is experienced in urban settings.

2.3 Human Mobility

Human travelling patterns underpin many urban and social processes such as information spread, disease dynamics, disaster response, and traffic management. While these processes can impact how individuals interact with their environment, individuals and communities can influence urban mechanisms to change. Although it is clear that the dynamics between individuals and their environment is symbiotic in nature, the extent to which these two forces are intertwined remains fairly ambiguous. The field of human mobility studies patterns in the movement of humans with respect to a variety of dimensions [15]. These dimensions include, but are not limited to, space, time, and economic resources, and can be applied to fields such as sociology and urban transportation science. Some believe mobility patterns arise solely as an artefact of individuals achieving a goal [283]. For example, mobility patterns to grocery stores arise from humans' dependence on food. Contrarily, others argue that mobility in itself can be an end goal [186].

Moreover, individuals from different socioeconomic classes may influence and be influenced by their urban setting to varying degrees. Understanding how different demographics coordinate with urban infrastructure is crucial, particularly when environmental contexts change. In the case of a natural disaster, a demographic's dependence on public transportation may leave them vulnerable to harm. Similarly, shifts in cultural values and norms may

impact common working hours for a certain demographic, narrowing the amenities that are accessible to them. Thus, it is crucial to understand whether sociodemographic disparities are an artefact of structural disparities or if they reflect behavioural preferences of a demographic.

2.3.1 Mobility Models

Mobility offers opportunities for overcoming the segregation that residential mechanisms, such as the housing market, impose on individuals. The prevalence of GPS (Global Positioning System) data and Call Detail Records has allowed for high-resolution analysis of travel behaviours [15, 4]. Studying mobility trajectories has highlighted the predictability of travel patterns, and, in turn, resulted in mobility models that can reproduce such behaviour [322, 267, 272, 226]. Human mobility models provide a quantitative perspective on human travel behaviour that help in understanding patterns in migration, traffic congestion, and information spread [15, 273, 106]. On one hand, they can be used to emulate individual trajectories and shed light on the mechanistic processes that fuel travel patterns. On the other hand, mobility models can capture travel patterns aggregated across the population, by drawing upon physics and network science to model human mobility as particle movements in a larger system.

The Gravity model, for instance, defines mobility flows between two regions as increasing when the population of either region increases [322]. Additionally, mobility flows decrease as the distance between the two regions increases. The Gravity model provides an important contribution to modeling heterogeneous mobility flows. Inspired by the radiation and absorption process from physical models, the Radiation model incorporates employment opportunities (proportional to resident population size), employment benefits (income, working hours, etc.), and distance between the job opportunity and one's residence to create a parameter-free model that predicts mobility patterns [268]. In doing so, it addresses some of the significant limitations of the Gravity model, such as its inability to capture travel fluctuations and the

theoretical gap in defining how distance impacts that probability of mobility flow between regions. The Intervening Opportunity model incorporates the radiation model's opportunity prioritisation by considering opportunities in the destination region as well as in regions between the source and destination [169].

Mobility models can also be constructed with respect to individual behaviour, capturing the stochasticity that emerges from free will [15]. The most fundamental of these models is the Random Walk, which forms a path defined by random displacements draw from a probability distribution at discrete time steps. Extensions to such models have been developed, modelling random walks over a continuous time period, accounting for individuals' predilection to return to previously visited locations or explore new destinations, and incorporating the effect that the frequency and recency of visits has on daily mobility trajectories [204, 272, 16]. It becomes clear that there are various ways to approach modelling human mobility patterns, and choosing which model to implement is determined by both the research question and the resolution of available data.

2.3.2 Spatial Networks

Although mobility models are commonly used to capture fundamental laws in travel behaviour, spatial networks are also often leveraged to understand the context in which mobility exists [19]. In spatial networks, nodes are located in a conventionally two-dimensional space. Meanwhile, edges may or may not be associated with space, depending on the type of network. For example, in social networks, encoding edges with spatial features may not be informative or relevant. Spatial networks are useful in scenarios where space and proximity is non-negligible. Street networks, for example, can be used to understand different routes individuals may take to travel between two locations. Typically, node ($V = v_1, v_2, ..., v_n$) reflect an intersection, while the edges ($E = e_1, e_1, ...e_m$), which are typically directed and weighted, reflect streets [19]. This method for modelling street networks is referred to as

a primal graph. Conventionally, an edge, e_{v_i,v_j} is weighted with a cost of driving from intersection v_i to intersection v_j, which can be measured based on time, speed, emissions, or many other cost objectives. The lack of distinction between intersections, road-ends, and network boundaries in primal graphs motivates alternative approaches for representing road networks. To address these limitations, dual graphs can be used to model road networks, by using nodes to represent streets and edges to capture intersections. When roads, themselves, are the focus of a study, using a dual graph is generally a suitable choice [236]. Whereas, a primal graph may be more contextual if one is focusing on analysing the locations in a city [23].

2.3.3 Inequalities in Human Mobility

In 1968, John Kain proposed the "Spatial Mismatch Hypothesis", which revealed how less-privileged individuals face disadvantages in their labour market outcomes due to their residential location [143]. Kain suggested that residential-workplace mismatches were the result of two urban processes in the US: employment opportunities moving to the urban periphery and discrimination against Black residents, preventing them from moving out of the city centre. Businesses, particularly those hiring low-skilled minorities, were incentivised to move to the suburbs, where the price of land was cheaper. Kain's original hypothesis focused on spatial mismatch in the context of racial discrimination, however, discrimination is not fundamental to the spatial mismatch hypothesis. That is, mismatches can arise from demographic preferences for particular residential characteristics or for specific types of jobs. These mechanisms are mentioned, not to downplay the consequences of discrimination, but rather emphasise how disparities in commuting origins and destinations can be a result of both structural inequalities and demographic preferences.

Regardless, long commuting distances can deter individuals from accepting a job with many benefits, if the cost of commuting is too high [102]. Moreover, low-skill jobs tend

to rely more on local and informal methods of hiring (i.e., word of mouth). Furthermore, individuals in areas with low transit access and car ownership rates may be limited by their job search radius.

The three main approaches to addressing residential-workplace mismatches are (a) moving jobs to individuals, (b) moving individuals to jobs, or (c) connecting individuals and jobs [131]. The first two mechanisms directly address spatial mismatch. However, the last mechanism, despite being the least invasive intervention, serves as an indirect solution that minimises the consequences of spatial mismatch rather than the direct issue itself. Thus, it is necessary to understand the underlying forces that are driving spatial mismatch. If a high job search radius is the force disconnecting individuals and employment opportunities, then interventions could entail ensuring that ad agencies in city centres include job opportunities in the suburbs. Alternatively, improving efficiency and accessibility in public transit systems would benefit spatial mismatches caused by high commuting costs.

Extracting how individuals perceive travel poses a difficult challenge, but there have been efforts to use the necessity of trips, destination choice, and route choice as a proxy for this value [202]. These metrics provide valuable alternatives for incorporating individual-level motivation in human mobility models. This can be further extended to understand if and how trip purposes differ across sociodemographic groups [147]. Access to descriptive mobility data has provided insight into whether one's mobility patterns are influenced by their economic standing and gender [17, 3]. Through these findings, we emphasise that understanding human mobility requires a deep understanding of how individuals choose to fulfil different types of mobility desires.

The increased availability of mobility data has supported extending static measures of inequality to account for one's activity space. Activity spaces represent all the locations that an individual has direct contact with in her daily mobility trajectories. Segregation indices, such as isolation and exposure indices, have been modified to account for activity spaces [256,

312]. Furthermore, Farber et al. have presented a series of indices that incorporate trip lengths and space-time prisms to understand inequalities in mobility patterns [89, 90]. Moro et. al. use mobility data to build an extension of the Exploration and Preferential Return model, which identifies an association between experienced income segregation and individuals' level of place exploration [206]. In doing so, they demonstrate how experienced segregation is related to residential features and amenity visitation patterns. The methodology of these studies are in line with mobility justice, which introduces the new mobility paradigm, in which "activities are not separate from the places that happen contingently to be visited...[and] travel is not just a question of getting to the destination" [263]. Various research in mobility inequalities has identified exacerbated levels of income segregation following natural disasters and relationships between income inequality and segregation in various countries, hindering social mobility [316, 214].

While these studies highlight the benefits of considering segregation from both a residential and a mobility-based perspective, they are vulnerable to the Modifiable Areal Unit Problem (MAUP), which refers to variance of statistical results because of how spatial scales or boundaries are defined [220]. MAUP motivates the adoption of individual mobility data, with studies elucidating how smaller spatial scales typically lead to larger segregation values [311, 155]. Using individual-level mobility data, however, requires carefully considering how to define a neighbourhood to appropriately aggregate individual-level measures to a neighbourhood scale [317].

Furthermore, mobility data can consist of demographic, temporal, and spatial biases. Often, mobility data, such as GPS positions from third-party smartphone applications, can provide high-resolution trajectories of individuals. However, the demographics that are sampled in the data may not be representative of the region being analysed [241, 166]. This can lead to identifying mobility patterns that reflect only a portion of the entire population [306]. Furthermore, this data, as well as mobility data from Location Based Social Networks

that require actively checking in, are vulnerable to temporal biases. That is, certain applications may be used or avoided at certain times of day or during specific periods in a year (i.e., holidays, weekends) [3, 271]. If temporal fluctuations are overlooked, findings may reflect erratic behaviour, conveying mobility patterns during periods of atypically high or low application usage as general mobility behaviour [175]. Finally spatial biases can arise based on the extent of network coverage, leading to inaccuracies in reported GPS locations [320]. This bias is crucial to consider as it can lead to lower accuracy of findings in rural areas than compared to urban areas [166]. In essence, data is likely to consist of one or more of these biases, highlighting the need to engage with comprehensive preprocessing techniques (i.e. post-stratification weighting to correct demographic biases, temporal aggregation), before conducting any quantitative analysis.

A vast amount of research that focuses on human mobility and the built environment, quantifies mobility in terms of trip frequency, trip length, vehicle miles travelled, and vehicle hours travelled. These mobility characteristics are, then, compared to built environment features. Through an extensive literature review, Ewing and Cervero found that, generally, trip frequency is more reflective of socioeconomic factors than those of the built environment [85]. On the other hand, trip lengths have been found to mainly be a function of the built environment, with socioeconomic characteristics playing a less distinctive role. When considering inequalities in the built environment, issues pertaining to self-selection bias frequently arise [219]. That is, a high concentration of a demographic group in the urban core could be an expression of that group's preference to live in downtown areas. Alternatively, it could be reflective of urban constraints that limit where individuals in that group can live. Existing works on transport inequality show how more vulnerable groups tend to live in more accessible areas [5]. In line with the self-selection bias, discussed earlier, disentangling choice from constraint remains difficult.

Many urban analysts consider how housing supply and job distributions influence residential landscapes [120, 269, 260, 209]. Similarly, others analyse how the built environment impacts individuals' mobility choices [85, 5, 198]. Both residential distributions and mobility behaviour are vulnerable to self-selection bias, a prevalent issue in urban analytics [219]. For example, we might observe higher transit ridership in high-income neighbourhoods because that demographic chooses to live in neighbourhoods that prioritise transit, not because transit serves higher income neighbourhoods better. However, it is important to note that self-selection bias, in the context of urban dynamics, is not an equally distributed phenomena. That is, financial, social, and urban constraints can limit where less privileged groups can live or the areas to which they can reasonably travel, consequently restricting their ability to "self-select". Harvey reaches this conclusion in *Social Justice and the City*, stating "The rich, who have plenty of economic choice, are moveable to escape such consequences of monopoly, than are the poor whose choices are exceedingly limited. We therefore arrive at the fundamental conclusion that ... the rich can command space whereas the poor are trapped in it." [119]. Thus, even the concept of self-selection biases contains inequalities in terms of the extent to which the biases they can be expressed by different sociodemographic groups.

2.4 Inequalities in Public Transit

While inequalities in access to particular destinations and travel modes have been studied for some time, research on transport justice has only recently started to gain traction [294, 143]. In this section, we introduce transit justice from a theoretical perspective, addressing how various justice frameworks interpret inequalities in transit. We highlight features of public transit networks that distinguish them from other mobility networks. Then, we shift our focus from theory to applied studies that estimate transit inequality, paying particular regard to the three urban mechanisms underscored in this thesis.

2.4.1 Transport Justice

The focus on improving public transportation systems is motivated by environmental and social factors. Primarily, advocates of public transit highlight how it reduces car accidents, traffic, and pollution [189, 208]. However, transit also provides opportunities for interaction [178, 176]. Accordingly, many researchers have considered how various features of transit systems can give rise to sociodemographic inequalities, with respect to physical ability, gender, and numerous other demographics [178, 189]. Merriman et. al suggest that transport infrastructure has been developed for particular industries that do not necessarily reflect the values and ideals of everyone in a city [197]. Furthermore, transport researchers have revealed the contrast between stated political goals of shifting to sustainable urban mobility and urban planning decisions that prioritise cars [14]. Allegiance to car-oriented planning began as early as the 1930s in the United States [217]. The rise of 'jaywalking' as a concept withdrew the right to the street from pedestrians. Road construction and lane additions reflected how planning agencies prioritised automobiles as the ideal form of mobility. This shift in urban design has been identified to be spearheaded by oil lobbies and car manufacturers [197, 217]. Moreover, the prominence of cars facilitated suburbanisation processes [217].

Transport poverty, a field that has a large overlap with human mobility, has been gaining prominence in the urban planning realm. It is often conflated with terms such as transport affordability, mobility poverty, accessibility poverty, but these terms are merely a subset of what transport poverty represents [177].

Pereira et. al. highlight how transport research focus on three main components of transport disadvantage: uneven allocation of transport-related resources, daily mobility patterns, and transit accessibility [231]. Disparities in transit resources overlook the capabilities approach to justice, discussed in Section 2.1. For example, frequent transit service may not be useful if residents do not travel to the areas that transit serves. Similarly, individuals may not be able to use particular travel modes due to financial or physical ability-related

constraints. Empirical mobility patterns have been used to explain which transport modes and trajectories are fulfilled, rather than only considering mobility opportunities. However, similar to the issues discussed in Section 2.2.1, disentangling individual preference from urban constraints remains a challenge with this approach [119]. Moreover, empirical mobility patterns tend to overrepresent privileged individuals [166]. Thus, defining inequalities solely based on observed mobility behaviour may poorly reflect how mobility patterns align with mobility desires.

2.4.2 Modelling Transport Networks

Transportation networks model mobility routes for numerous travel modes, from pedestrian paths to railways. While graphs can also be used to model transportation networks, applying the same shortest-path finding algorithms used for road networks leads to worse computational performance [20]. Choosing the appropriate pathfinding algorithm is crucial as it is the primary bottleneck for developing efficient transit accessibility tools [59]). Shortest-path algorithms tend to appear in transit networks in two ways: frequency and schedule-based algorithms [170]. Frequency-based approaches condense transit timetables into travel times and headways (the duration between consecutive transit services). Meanwhile, schedule-based networks maintain departure and arrival information from transit data, using numerous nodes to reflect various departure and arrival times at transit stops. Thus schedule-based algorithms account for the variance in headways between services as well as the issue of competing transit lines serving the same locations [171].

Dijkstra's Algorithm is typically used to estimate the shortest path between two points in a road network [77]. The most efficient form of Dijkstra's Algorithm leverages a priority queue to identify the shortest path or distance between any two nodes in a graph. Given an origin node, O, and a destination node, D, the algorithm assigns infinity to all nodes except the origin node, which is assigned zero. This value represents the cost of travel, either

in length or time. Then, traversing all outgoing edges, each edge is relaxed, meaning the algorithm checks whether destination D is reachable by an outgoing edge, at a lower travel cost. After the current node is settled, the node with the smallest current travel cost is chosen, and the algorithm proceeds, relaxing all of the new node's outgoing edges. A node is settled when all of its edges have been relaxed. The algorithm is complete with the destination node, D has been settled. This method can be extended to a bidirectional search, in which Dijkstra's algorithm is simultaneously applied forwards to the source node and backwards to the target node until a node has been visited from the two directions [66]. Ultimately it is crucial to be intentional about how one chooses to represent road and mobility infrastructure. This idea is further explored in the Section 2.4, in which we discuss conceptual and quantitative differences between road and transport networks, when modelling and routing.

A prominent method for finding shortest paths in frequency-based algorithms is contraction hierarchies, a variant of Dijkstra's algorithm introduced in Section 2.3.2 [99]. Since frequency-based algorithms collapse travel times between nodes to a representative value, networks can be modelled similarly to road networks, such that nodes represent transit stops and pedestrian intersections. Then, edges between nodes can reflect travel distance or travel time. Depending on the type of source node (transit or pedestrian node), travel time is typically calculated as a function of distance, walk speed, and transport vehicle speed. Contraction hierarchies exploit the hierarchical nature of road networks. That is, motorways, country lanes, and national highways have different functions and features. In turn, roads that are more connected have a higher importance, allowing more nodes to be settled. Thus, contraction hierarchies depend on a two-phased process:

The first phase is preprocessing, which categorises nodes based on some formulation of importance and contracts nodes from least to most important. Although node ordering is known to be a difficult problem, simulating node contractions can give a reasonable result [24, 99]. This involves sorting nodes by their edge difference and storing the nodes in a

Algorithm 1 Shortest Path using Contraction Hierarchies

```
1: Input: Graph G = (V, E)
2: Output: Contraction Hierarchy CH
3: procedure GETCONTRACTIONHIERARCHY(G)
4:     CH ← G
5:     order ← compute node order                ▷ ordered using edge difference
6:     for all v ∈ order do
7:         shortcuts ← NODECONTRACTION(CH, v)
8:         CH ← CH ∪ shortcuts
9:     return CH
10: procedure NODECONTRACTION(CH, v)
11:     shortcuts ← ∅
12:     for all (u, v) ∈ CH ∧ (v, w) ∈ CH do
13:         if ISNECESSARYSHORTCUT(CH, u, v, w) then
14:             shortcuts ← shortcuts ∪ {(u, w)}
15:     Remove node v and its adjacent edges from CH
16:     return shortcuts
17: procedure ISNECESSARYSHORTCUT(CH, u, v, w)
18:     shortest_path ← SHORTESTPATHQUERIES(CH, u, w)
19:     return length of (u, v) + (v, w) < length of shortest_path
20: procedure SHORTESTPATHQUERIES(CH, u, w)
21:     Use bi-directional Dijkstra's to find the shortest path from u to w
22:     return length of the shortest path
```

priority queue. The edge difference for a node v can be found by simply subtracting the number of shortcut edges added when a node is removed, from the number of edges removed from v.

Then, the node contraction step follows, where U refers to the set of source nodes that have incoming edges for a node v and W represents the destination nodes to which v has outgoing edges. For each node v, from least to most important ordering, edge differences are recomputed for all nodes that have still not been settled. If the current node v no longer has the smallest edge difference, the priority queue is rebalanced and the first node in the updated priority queue is now considered the focal node. Then, a series of local shortest path queries are run from all incoming nodes in U on a subgraph that excludes node v. If the shortest path that is found is less than the `edgeweight(u,v)+edgeweight(v,w)`, then a shortcut edge is added between nodes u and w with the weight found by Dijkstra's algorithm. This process is repeated until the most important node is removed. Then, a new graph, G', is constructed such that it has the original nodes and edges, as well as the shortcut edges found through the contraction process.

The second phase consists of querying shortest paths from a set of origin and destination points, in which a bidirectional Dijkstra search is applied to the contracted graph, G'. Details on the bi-directional can be found if the following references [21, 66]. The main point to emphasise is how the modelling of transport networks diverges from the reality of using them. While contraction hierarchies are computationally efficient for routing on road networks, and are frequently used in the context of transit networks, the next paragraph offer potential alternatives that steer away from the use of priority queues and hierarchical speedups.

Contraction hierarchies leverage the hierarchical nature of road networks. This may be suitable when modelling transit networks using representative timetable data, however, constructing networks to incorporate the specifics of transport timetables requires considering alternate techniques, which incorporate the time spent waiting for transit and transferring

between travel modes. Public transit networks are fairly similar to road networks as both networks model spatial information. However, they diverge from road networks, in that transit networks need to account for the timetables of public transportation.

When modelling the temporal components of transit, the most straightforward approach is through using a time-dependent model [238]. Time-dependent models configure nodes to capture the arrival or departure of a given transit vehicle at a given station. Then, edges can reflect either waiting at a station between an arrival and departure event or travelling from a departure event in one station to an arrival event at another. Another approach is the time-expanded model, that builds off the time-dependent model, to incorporate the time it takes to transfer between transport vehicles [238, 20]. Thus, each node in the time-dependent model is duplicated in the time-expanded model, such that one node will represent being on board a vehicle at the station, and the duplicated node reflects being at the station after alighting the vehicle.

When we consider modelling transport networks on such a detailed level, three major issues with using contraction hierarchies for pathfinding arise [20]. First, bidirectional searches are more complex when applied to time-expanded constructions of transit networks, as usually the target station is known, but the node at the target station is still ambiguous. Second, the hierarchical structure of road networks does not simply translate to that of the time-expanded transit network, because it requires applying local searches to a significantly larger set of nodes. Third, the high node degree in transit networks does not lead to a significant speedup when contracting nodes. Additionally, contraction hierarchies optimise for the criteria of finding the shortest path without considering other objectives that are relevant when using public transit, like financial costs and the number of transfers.

Thus, solutions such as the Connection Scan Algorithm (CSA) and the Round-bAsed Public Transit Optimised Router (RAPTOR) were proposed [76, 65]. Notably, both these algorithms do not depend on the use of graphs. CSA models transit timetables by defining

the set of transit stops, transit connections, and pedestrian footpaths, in a given region, where connections refer to a transit vehicle departing a stop p_i at time t_x and arriving at stop p_j at time t_y:

Algorithm 2 Shortest Path using Connection Scan Algorithm [76]

Input: timetable, source stop s, target stop t, minimum departure time τ_s, maximum arrival time τ_t
Output: The set of all $(j_{dep\_time}, j_{arr\_time})$ over journeys j s.t.
- j departs at or after τ_s at s
- j arrives before or at τ_t at t
- the pair $(-j_{dep\_time}, j_{arr\_time})$ is Pareto-optimal among all journeys
- j contains at least one leg

for all stops x **do**
    $S[x] \leftarrow \infty$
for all trips x **do**
    reset $T[x]$
for all footpaths f from s **do**
    $S[f_{arr\_stop}] \leftarrow \tau + f_{dur}$
Find first connection c^0 departing not before τ using binary search
for all connections c increasing by c_{dep_time} starting at c^0 **do**
    if S[t] $\leq c_{dep\_time}$ **then**
        Algorithm is finished.
    if $T[c_{trip}]$ is set or $S[c_{dep\_stop}] \leq c_{dep\_time}$ **then**
        raise $T[c_{trip}]$
        if $c_{arr\_time} < S[c_{arr\_stop}]$ **then**
            for all footpaths f from c_{arr_stop} **do**
                $S[f_{arr\_stop}] \leftarrow \min(S[f_{arr\_stop}], c_{arr\_time} + f_{dur})$

Trips can, then, be defined by subsequent connections between the same vehicle. A single array represents every connection, sorted in order of departure time, with a value representing the earliest time a stop can be reached. When initialised, the only non-infinite representative value is the departure time of the source stop. Then, CSA iterates though each connection and, in the case that the arrival time at the current connection improves the arrival time at the destination, the representative value at the destination is relaxed. In this manner, the array is only scanned once. Although it does not require a computationally expensive preprocessing phase, its efficiency is dependent on the size of the timetable. Furthermore, CSA can be

extended to include multi-objective searches [76]. A particularly relevant limitation of CSA, that is not present in RAPTOR, is the fact that restrictions in features of the footpath graph (i.e., the need for it to be transitively closed) do not allow for instances that model unrestricted walking [296].

RAPTOR offers another efficient, non-graph based solution to routing within transit networks [65]. Using simple data structures and dynamic programming, RAPTOR works in rounds, where each round represents a transit transfer, ensuring that every route is only traversed at most once for each round. In this manner, RAPTOR can optimise not just shortest journeys in terms of travel time, but also in terms of transit-related objectives, such as number of transfers or transit fares (when using McRAPTOR, an extension of RAPTOR). To initialise the RAPTOR algorithm on a given transit timetable, each transit stop is defined with a label, $\tau^*(p_i)$. Here, p_i reflects a given stop and τ^* represents the earliest time one can arrive at stop p_i with at most $i-1$ transfers:

All lists are initialised to infinity, except for the source transit stop, p_s, which is initialised to τ, the departure time. Furthermore, the associated footpath, $\mathscr{F}$, models footpath transfers between transit stops, such that $\mathscr{F}$ is transitive and $l(p_i, p_j)$ reflects the walking time between stops i and j. As a reminder, each round, k, determines how to reach every stop with at most $k-1$ transfers. The RAPTOR algorithm can be succinctly summarised into three steps, that occur for a given round k, for which the goal is to compute $\tau_k(p)$, for all p. First, the algorithm applies the upper bound for $\tau_k(p)$ to be $\tau_{k-1}(p)$. In simpler terms, for a given number of transfers, this limits the earliest arrival time, at a given stop, to be no greater than the earliest arrival time at that stop, when we consider any smaller amount of transfers. Second, the algorithm finds all stops that are accessible and traverse the associated RAPTOR of those stops, updating any stops with improved arrival times, if fitting. Finally, RAPTOR accounts for any transfers using footpaths. That is, for any footpaths that connect the current stop p_i to another stop p_j, if $\tau_k(p_i) + l(p_i, p_j)$ is less than $\tau_k(p_j)$, $\tau_k(p_j)$ is updated to this

Algorithm 3 RAPTOR Algorithm [65]

Input: Source and target stops p_s, p_t, and departure time τ.
`// Initialization of the algorithm`
for all i **do**
    $\tau_i(\cdot) \leftarrow \infty$
$\tau^*(\cdot) \leftarrow \infty$
$\tau_0(p_s) \leftarrow \tau$
mark p_s
for $k \leftarrow 1, 2, \ldots$ **do**
    `// Accumulate routes serving marked stops from previous round`
    Clear Q
    for all marked stop p **do**
        for all routes r serving p **do**
            if $(r, p') \in Q$ for some stop p' **then**
                Substitute (r, p') by (r, p) in Q if p comes before p' in r
            else
                Add (r, p) to Q
            unmark p
    `// Traverse each route`
    for all routes $(r, p) \in Q$ **do**
        $t \leftarrow \perp$ `// the current trip`
        for all stops p_i of r beginning with p **do**
            `// Can the label be improved in this round?`
            `Includes local and target pruning`
            if $t \neq \perp$ and $\text{arr}(t, p_i) < \min\{\tau^*(p_i), \tau^*(pt)\}$ **then**
                $\tau_k(p_i) \leftarrow \tau_{arr}(t, p_i)$
                $\tau^*(p_i) \leftarrow \tau_{arr}(t, p_i)$
                mark p_i
            `// Can we catch an earlier trip at` p_i`?`
            if $\tau_{k-1}(p_i) \leq \tau_{dep}(t, p_i)$ **then**
                $t \leftarrow \text{et}(r, p_i)$
    `// Look at foot-paths`
    for all marked stops p **do**
        for all foot-path $(p, p') \in F$ **do**
            $\tau_k(p') \leftarrow \min\{\tau_k(p'), \tau_k(p) + l(p, p')\}$
            mark p'
    // Stopping criterion
    if no stops are marked **then**
        stop

value. RAPTOR provides an efficient solution that can be parallelised and can provide dynamic results due to its lack of preprocessing.

Having described the differences between various transit routing algorithms, we now introduce two, state-of-the-art, open source tools for modelling transit accessibility: `UrbanAccess` and `r5py` [31, 230]. We use both these tools throughout this thesis and despite them having similar functionality, they have distinct differences in design and methodology that are worth addressing. In terms of similarities, however, `UrbanAccess` and `r5py` depend on General Transit Feed Specification (GTFS) data and road network extracts, which we gather from OpenStreetMap [221]. GTFS refers to a popular data format that describes transit service by defining transit stops, routes, and schedules. OpenStreetMap is an open-source database of geospatial information, such as roads and amenity coordinates. Further details on OpenStreetMap and GTFS data structures are outlined in Chapters 3.3 and 3.4.1, respectively.

`UrbanAccess` configures nodes represent transit stops and street nodes. Thus, there are four types of edges: (1) edges connecting two transit stops (2) edges connecting two street nodes (3) edges connecting a street node to a transit node and (4) edges connecting at transit node to a street node. The directed, transit-transit edges, and corresponding edge weights, are created from the GTFS operational schedule, with respect to the user-specified day and time window. The directed, transit-pedestrian edges connect every transit stop to its closest pedestrian node, deriving travel time using distance and an assumed walking speed of three miles per hour. Meanwhile, the undirected, pedestrian-pedestrian edges use the distance between each node pair in the road network to weight the edges with travel time, again assuming a walking speed of three miles per hour. Finally, the directed, pedestrian-transit edges, similar to the transit-pedestrian edges, maps every transit node to its nearest pedestrian node. However the pedestrian-transit edges account for wait time at the transit stops based on average transit headway – the amount of time between transport vehicles at a stop. Thus, the pedestrian to transit edges differ from the transit to pedestrian edges, in that they account for

both the walking time from pedestrian to transit nodes and the wait time at the transit stop. Further details regarding the construction of the transit-pedestrian network can be found in [31]. To calculate transit accessibility, `UrbanAccess` uses `Pandana`, which leverages contraction hierarchies to efficiently calculate travel access [94]. With this package, it is possible to calculate travel metrics such as the travel times between points or the cumulative opportunities accessibility index for a given zone.

While `UrbanAccess` is built using C++, `r5py` is a Python library that seamlessly integrates with the Rapid Realistic Routing on Real-world and Reimagined networks (R^5) routing engine, which was created by Conveyal [230, 59]. `r5py` leverages RAPTOR, a dynamic pathfinding approach that is easily parallelisable, as opposed to Dijkstra variants, and can solve multi-objective tasks, which can include minimising travel times, wait times or the number of transfers [65].

2.4.3 Accessibility

Accessibility can take on multiple definitions, due to the fact that it spans the fields of socioeconomic inequality, urban planning, and transportation [78]. We point out that accessibility can sometimes be used to refer to how easily people with disabilities can enter or access different facilities. However, within the context of this thesis, accessibility describes how individuals can reach different areas, through various mobility services such as transport systems or street networks. Early formulations of accessibility aimed to encapsulate opportunity attractiveness by describing how land use intersects with transportation systems. Other early definitions include measuring how easy it is to reach any land-use activity from a given point [116, 56]. Over time, socioeconomic and temporal dimensions were incorporated, to develop a better understanding of whether more privileged groups command space in terms of the areas to which transit provide access, while less privileged demographics are constrained by it [119, 45, 28].

Most methodologies for defining transit accessibility can be classified as either **system accessibility**, **system-facilitated accessibility**, or **integral accessibility**. These categories are similar to the concepts of relative and integral accessibility, introduced by Ingram, where relative accessibility estimates how connected two locations are and integral accessibility reflects how connected a location is to all destinations in a region [133]. System accessibility analyses proximity to transport links, essentially examining topics surrounding first mile problems that evaluate the ease at which individuals can reach a transit stop to begin their desired journey via transit. System-facilitated accessibility assesses how individuals can arrive at their desired destinations using the transit system, contrasting system accessibility, which determines how easy it is to reach an entry point in the transit network that will presumably connect them to their destinations. Integral accessibility captures overall access to numerous endpoints, incorporating multiple forms of transit modes and types of destinations [193]. Each of these categories reflect different objectives individuals may have when considering to use public transportation. Of particular importance to this thesis is system-facilitated and integral accessibility. We leverage both concepts to provide a thorough overview of how transit accessibility changes depending on the theoretical perspective being considered. That is, system-facilitated accessibility can capture how transit serves particular types of journeys, be it commutes or amenity visitations, whereas integral accessibility combines the privilege of choice with transit service to understand general accessibility considering all potential destinations. Thus, we focus on describing modelling approaches to measure these two categories of accessibility. For details on models of system accessibility, which is not covered in this thesis, we refer readers to the following references [92, 55].

Models that analyse system-facilitated accessibility account for both proximity to entry points in the transit network and distance to particular destinations through the transit network. There are several methods for estimating shortest paths in a transport network, which is outlined in Section 2.4.2. As a result, depending on the chosen algorithm variations in

accuracy can arise. While some methods incorporate wait times, transfer times, and transit frequency, others model public transit as a static network and apply classical shortest-path algorithms, such as Dijkstra's, which is described in Sections 2.3.2 and 2.4.2 [289, 210]. Typically, system-facilitated methods measure travel cost in terms of travel time, which is function of travel mode and distance [185]. This avoids the conflation of differences in speed between pedestrians and transit vehicles that would occur if travel costs were measured in distance. Moreover, Aman and Smith-Colin introduce the comprehensive public transit accessibility (CPTA) score to estimate access at different spatial resolutions. This metric includes features of system and system-facilitated accessibility as well as characteristics of transit frequency, flexibility, and efficiency [7].

Integral accessibility models have proposed metrics for representing general accessibility to opportunities, but most tend to fall short at incorporating destination quality and importance into their measures. While numerous such models exist, a significant contribution is the approach introduced by Farber et. al., which incorporates temporal features of the transit system by measuring travel times to the ten closest supermarkets from 06:00 to 22:00 [88]. This method, similar to the approaches we use, leverages GTFS feeds and Dijkstra's algorithm. Integral accessibility models have also been used to highlight how transit inequalities exist when considering affordability, specifically showing the significant impact that financial costs have on accessibility [176, 80]. The following sections highlight relevant research in transport poverty regarding residential-workplace landscapes, segregation, and the built environment, to better frame how transit accessibility fits into each spatial perspective.

2.4.4 Transit-Oriented Commuting Behaviours

Having addressed different ways of measuring transit accessibility, we now shift to discussing how such metrics can be incorporated into understanding commuting behaviours. Many works have highlighted the various reasons how proximity plays a role in connecting employ-

ment and housing landscapes. Some explanations include the cost of commuting [108] and residential markets supplying housing to particular demographics, resulting in dense commuting flows between specific neighbourhoods [122]. Moreover, the interconnected nature of residential and workplace segregation emphasises how housing markets and employment opportunities further contribute to experienced inequalities [82, 282]. Suburbanisation of particularly low-earning jobs, concentration of low-income households in urban cores, and outdated transit systems that support suburb to city centre flows all contribute to the growing inaccessibility of employment opportunities [5, 243, 197, 167, 56]. Commuting burdens, in terms of time and distance, have been shown to be connected with numerous health problems, including depression [117].

The rising availability of transit data has allowed for temporal, high-resolution of commuting behaviour. This is useful when exploring socioeconomic disparities in how transit connects residential and employment landscapes. Temporal data is relevant as low-wage workers tend to have work schedules that diverge from the typical 9-5 work day, around which transit systems plan their service [159, 91]. Furthermore, low-wage workers tend to have the highest rates of transit dependence, underscoring the importance of reducing commuting transit burdens they face [139]. While decades of literature exploring the validity of spatial mismatch exist, few quantitative works have interpreted spatial mismatch from both a residential-employment landscape perspective and a public transportation one [84]. Some works that have studied the intersection of public transportation and spatial mismatch have used a gravity model, weighted by job demand, transit travel time, and transit costs, to measure job accessibility at a neighbourhood level [264, 168]:

$$A_i = \sum_{j=1}^{n} \frac{E_j f(O_{ij})}{S_j}, \tag{2.8}$$

$$f(O_{ij}) = exp(-\beta(wT_{ij} + C_{ij})), \text{ and} \tag{2.9}$$

$$S_j = \sum_{k=1}^{n} P_k f(O_{kj}). \tag{2.10}$$

In these equations, A_i measures the transit access for low-wage job seekers in a neighbourhood i. It is calculated as a function of the number of jobs (E_j) in a neighbourhood j, the travel friction between two neighbourhoods ($f(O_{ij})$), and the demand for jobs in j (S_j), for all neighbourhoods, J, in a region. The travel friction ($f(O_{ij})$) is simply a function of the minimum wage (w), travel time (T_{ij}), and travel cost (C_{ij}). Meanwhile, the demand for jobs in a neighbourhood j (S_j) can be derived by summing the product of the number of poor job seekers in a neighbourhood k (P_k) and the travel friction between the two neighbourhoods ($f(O_{kj})$, for all neighbourhoods, K, in a region. Consequently, the authors of Ref. [168] find that, for all 50 US metropolitan regions analysed, transit provides higher access to high-earning employment opportunities than to low-earning ones. Furthermore, they confirm hypotheses of job suburbanisation by identifying how vulnerable demographics that reside in the core of the city suffer from poor service to their jobs, which are located in the suburbs. These findings highlight how transit planning is still shaped around outdated urban landscapes, in which transit provided job access when jobs were concentrated in the core.

Apart from the aforementioned research, summarising the extent to which vulnerable demographics suffer due to spatial mismatch remains fairly unexplored from the transit perspective. Even more unexplored is the role that housing mechanisms, such as housing stability and affordability, play in spatial mismatch and transit service. An existing study analyses the relationship between availability of affordable housing and commute distances, hypothesising that more availability is associated with smaller commute distances [314]. To calculate the jobs-housing fit, they employ a linear distance decay function to estimate the ratio of jobs to rentals in proximate neighbourhoods. In doing so, they find that a lower stock of affordable housing relates to longer commute distances. This study explores residential landscapes in a more nuanced manner, by defining residential locations based on housing

availability rather than socioeconomic composition. However, it is limited in that it only interprets commutes in terms of distance, overlooking the significant role that transit service plays for providing mobility options to more vulnerable demographics.

While these papers use the same employment datasets as we do in this thesis (from the US Census Bureau), they are either lacking in nuanced modelling of transit service or detailed estimation of housing characteristics. Although the framework we present in Chapter 3 of this thesis does not account for transit fare, we analyse spatial mismatch in residential-employment landscapes in the context of transit, specifically contributing novel definitions of residential vulnerabilities, by proposing a metric to quantify the numerous dimensions of housing insecurity. Furthermore we estimate commuting times by using transit networks to model transit journeys from residential to employment locations. Accordingly, we provide a nuanced look as to how inadequate and unstable housing conditions intersect with employment opportunities and different travel modes for commuting.

2.4.5 Segregation in Public Transit

Public transit is a crucial component of urban environments, providing access to employment opportunities and amenities within a region. However, characteristics of transit systems, such as its urban layout and service frequency, can create pockets of transport deprivation, isolating particular neighbourhoods from conveniently accessing transit service [213]. This is dire for demographics that rely more on public transportation as their mode of transport [130, 139]. Lack of access to transport can impact how individuals perceive their activity space, by restricting or providing access to particular destinations [287]. Inequalities in transport systems, and the types of amenities and neighbourhoods they provide access to, is important to consider as it can impact the level of choice that disadvantaged groups have when using transit [287].

There has been a considerable body of literature that focuses on segregation, public transit and mobility [216, 1, 319]. The authors of Ref. [216] utilise difference-in-differences estimations to analyse the causality between rail transit investments and income segregation in the US, measuring segregation at both the neighbourhood and metropolitan scale. To offset bias, considering that the process of adding new rail stations to particular neighbourhoods is not random, the authors further incorporate propensity score matching to compare similar neighbourhoods. They measure segregation at the neighbourhood level using the index of ordinal variation, whereas they use the information theory index for calculating metropolitan-level segregation [158, 286]. They used a fixed effects model to ultimately determine that the addition of rail stops over a 20 year period, does not significantly impact the residential segregation of a metropolitan area. While this work analyses how changes in transit system accessibility are associated with changes in residential inequalities, other works have incorporated high-resolution mobility data to estimate relationships between transit infrastructure and more dynamic forms of segregation, based on activity spaces.

For example, Abbasi et. al. use smart-card data in Seoul, South Korea, to understand temporal patterns in destination-based socioeconomic segregation [1]. By applying segregation indices of dissimilarity, exposure, and diversity to the socioeconomic composition of destinations, they uncover temporal patterns of segregation in *Dongs*, the administrative spatial subunit in South Korea. Namely, they highlight how throughout the afternoon and evening more central areas tend to have lower levels of segregation across all social groups. Furthermore, weekends tend to have more homogeneity, in terms of social interaction. Finally, passengers with disabilities tend to be more segregated during the weekends, compared to other sociodemographic groups. Another study uses Safegraph, which is the same mobility data source used in this thesis, to understand experienced partisan segregation in the US [319]. After calculating the partisan composition of different destinations they calculate segregation at the place-level:

$$PPS_i = \frac{3}{4} \sum_{p=1}^{3} \left| \tau_p - \frac{1}{3} \right|, \tag{2.11}$$

where τ_p reflects the proportion of each group, p, in a place, i, considering Republicans, Democrats, and others as the three partisan groups. $\frac{3}{4}$ acts as a normalisation factor, such that *PPS* ranges from 0 to 1. This equation is inspired by previous mobility research that looks at experienced income segregation [206]. They further aggregate this to a community-level:

$$EPS_j = \sum_{i}^{M} PPS_i \cdot \eta_i, \tag{2.12}$$

where M is the number of destinations, and η_i is correlated with how many residents in j are visiting destination i. We point out these metrics as there are similarities across this methodology and the one we present in Chapter 5, particularly when aggregating from place-based segregation to a neighbourhood-level measure. In defining these metrics, the authors show that destination-based partisan segregation exhibits larger stratification that that of income or racial segregation. They highlight cities in Southern and Northeastern regions as having distinctively higher levels of partisan segregation. One limitation, that our framework also faces due to using the same data source, is inferring mobility patterns from neighbourhood-level features. Furthermore, the authors implement ordinary least squares regression models to predict place-based partisan segregation using metropolitan-level characteristics. In doing so, they find that geographically central, liberal, lower socioeconomic status, and majority black cities tend to have significantly higher levels of partisan segregation. While both these works apply segregation indices to mobility data, they consider segregation only at the destination-level. This leaves room to explore how residential and amenity layouts can lead to segregated mobility flows. In addition, one can analyse how transit facilitates these flows and how experienced segregation changes from these different urban contexts (i.e., residential, transit use, amenity).

A recent work, which studies transit inequality in Shenzen, China, uses two mobility datasets to compare disparities in activity spaces based on travel mode [98]. The authors leverage transit smart card data and license plate recognition data to analyse differences in accessibility based on travel mode. They reveal that individuals using private cars have higher access to amenity opportunities. They also showed how disparities in activity space were less present in the urban core, compared to the urban periphery. Finally, they find that areas that are vulnerable to transit-related social exclusion exist in the urban periphery. Both the methodology in Ref. [98] and the framework we introduce in Chapter 5 analyse social exclusion from a transit-oriented perspective. However, while the methods in Ref. [98] compare differences in mobility and accessibility between public transit and private vehicle users, our approach leverages amenity visitations data to understand how segregation can be experienced not only in terms of destinations, but also while using transit, due to residential landscapes, amenity locations, and transit dependence. Thus, our research focuses on how segregation is exhibited both in terms of transit service and transit use.

2.4.6 Transit and the Built Environment

Existing works on transport inequality show how more vulnerable groups, such as low-income households, tend to live in more accessible areas [5]. In line with the self-selection bias, discussed earlier, disentangling choice from constraint remains difficult. Another prominent concept within transportation research is transit-oriented development (TOD), introduced by Peter Calthorpe, which combines land-use, transportation, and urban planning to improve accessibility and use of transit systems [47]. TODs encourage making neighbourhoods mixed-use spaces that have access to transit within a 600 meter radius. The goal of TODs was to shift away from car-dependence and balance travel modes. Understanding how neighbourhood characteristics impact transit use has revealed how TOD residents are more likely to walk, cycle, and use transit compared to their non-TOD counterparts [144].

While these studies underscore relationships between transit-related features in neighbourhood and transit use, concerns regarding self-selection bias, discussed in Section 2.3.3, have lingered. In other words, individuals who prefer to use public transit may choose to live in neighbourhoods with better transit service. Consequently, neighbourhoods with high transit access may exhibit higher rates of transit use, not because better access motivates more travel by transit, but because these neighbourhoods attract residents who prefer using transit. In response to these concerns, Millard-Ball et. al. leverage the competitiveness of residential lotteries in San Francisco to control for self-selection bias in residential choice [198]. Thus, they show how having transit-rich neighbourhoods reduces car usage and that higher walk and cycle access similarly increases the chances of these travel modes being used. Interestingly, they also highlight how unavailability of parking spaces more significantly impacts car ownership, compared to transit access.

Although TODs offer potential solutions to encouraging transit use, other researchers have focused on understanding transit in the context of the urban built environment [51, 97]. Considering urban form is distributed differently across space and that transit largely relies on urban forms, the built environment is typically thought to have some degree of impact on transit service and use [290, 51]. However, it is not solely transit access that is spatially varied across different neighbourhood characteristics, but also residential locations (due to zoning restrictions and housing markets) and employment locations [314, 265]. This is a crucial point to consider, because just as similar individuals can cluster due to demographic-level preferences, discriminatory housing and employment policies can perpetuate residential distributions [150]. Keeping in mind that transit has been shown to better serve areas with more urban features and that numerous mechanisms can contribute to residential segregation, it is crucial to understand how transit inequalities exist in different types of urban settings [52]. Overlooking the overlap between the built environment and residential distributions can lead to identifying better transit access for low-income individuals [314, 5, 223]. These

works briefly acknowledge the importance of considering how low-income individuals living in the urban periphery may encumber multiple burdens as a result of their financial and mobility disadvantages.

Furthermore, these studies, that span the regions of Bogotá (Colombia), 8 metropolitan regions in Canada, and Seattle (USA), typically use variations of accessibility metrics discussed in Section 2.4.3 [314, 5, 223]. The study in Seattle defines transit access with respect to the density of transit stops, routes, and transit frequency as a proxy for good service in a housing cluster's buffer [314]. While these proxies are informative to an extent, frequent service can be rendered useless if it does not provide access to desirable locations. The studies analysing Colombia and Canada provide a more nuanced look at transit accessibility [5, 223]. The authors in Ref. [223] use a function of travel time, travel cost, and the number of non-work opportunities, such as amenities associated with healthcare and leisure. Additionally, this function is fit with a calibration parameter, β, which is based on least-square linear regression models that use empirical transit trips to estimate how travel costs relate to reduced accessibility. The limitations with this work, include travel data being derived from outdated survey data and dependence on linear relationships to describe accessibility. Finally, the authors of Ref. [5] focus on accessibility with respect to commuting behaviour. Thus, their measure of transit accessibility is a function of job opportunities, commute times, and workers in the employment location. Again, their main limitation, aside from not controlling for the intersection between residential and built environment landscapes, is how they compute travel times, as they depend on a inverse-power decay function that returns a value of 0.5 for commutes that are a half hour, and returns 1 for commutes that are zero minutes. Thus, shorter commutes have a larger weight. To address this gap, we introduce a framework that analyses socioeconomic inequalities in transit service, with respect to different types of urban neighbourhoods, leveraging built form characteristics to

define these categories. Moreover, we use more nuanced estimations of travel time via transit by integrating open-source transport modelling tools.

As a final point, in Section 2.2.4, we introduced the 3D's (design, density, diversity), which have been conventionally viewed as the three main components of the built environment. The majority of research in this field analyses how the built environment impacts transit ridership. Thus, their measure of the built environment includes expressions of transit service, such as distance to transit and destination accessibility. In this manner, these works can incorporate features of transit service into their built environment measure, to compare it against transit ridership characteristics.

2.5 Literature Gap

The previous sections provided a general background on inequality, human mobility, and transportation research, with a particular focus on urban areas. In this section, we leverage these concepts to draw attention to the two main research gaps our work aims to fill. At large, research on transit poverty tends to treat measuring complex inequalities and detailed transit modelling as mutually exclusive. We distil this larger issue into two distinct gaps in the literature. First, research on transit poverty that leverages detailed transit models tend to overlook the capabilities approach to justice. Second, few works in transit research that incorporate complex inequalities, such as housing instability or the effects of the built environment, tend to quantify transit systems using coarse proxies (i.e., transit stop density, number of routes). The lack of modelled complexity from either the sociodemographic inequality or transport system perspective makes research findings difficult to translate into effective policies that can reduce disparities in public transit.

The first research gap highlights how conventional approaches to transport research overlook how neighbourhoods with high transit access may still have embedded inequalities due to financial constraints, physical disabilities, and fear for safety, all of which are associ-

ated with vulnerable demographics [37]. As discussed in Section 2.4.3, much of transport research focuses on access to public transportation, be it from a system, system-facilitated, or integral perspective. Although these approaches reveal blatant spatial disparities in transit systems, they do not account for the capabilities approach to justice, discussed in Section 2.1. Specifically, they do not address how demographic groups, that already face other urban inequalities, may experience compounded burdens due to how transit serves them [185, 7]. Policies shaped from these findings, such as transit subsidies or the addition of new transit lines, likely reduce spatial disparities in transit accessibility, but disregard how constraints due to spatial dependencies or financial status may still prevent individuals from using transit.

On the other hand, the second research gap emphasises the importance of detailed transit modelling. That is, research that accounts for compounded inequalities in various aspects of urban life (i.e., housing, employment, support network connectivity) tend to model transit service using methods that simplify public transportation infrastructure [1, 98]. In doing so, the nuances that come with using transit, such as the number of transfers or time spent waiting for transit, are overlooked, making it difficult to identify which transit journeys are reasonable for demographics that are trying to avoid long walks to transit links, high travel fares, or long wait times due to safety or financial burdens. Furthermore, while research in human mobility addresses inequalities in activity spaces of various sociodemographic groups, few works in transport research have interpreted transit inequality from a similarly dynamic perspective. Thus, despite measuring features of inequality, such as segregation, they fail to capture the complexity of transit systems and, importantly, how transit service fits into the larger context of various urban mechanisms (employment landscapes, urban sprawl). Consequently, it remains unclear how inequalities in transit service arise for different trip purposes.

Ultimately, it is crucial to incorporate intricate features from the perspective of sociodemographic inequalities and transit modelling. Forsaking one of these perspectives can lead to

misinterpreting where and why transit inequalities arise and can potentially lead to unsuccessful interventions. To contribute to filling these gaps in the literature, we incorporate theories and methods from the research fields of demographics and human mobility to consider structural and experiential transit inequalities within three urban mechanisms: housing insecurity, urban segregation, and the built environment. Accordingly, in the following chapters, we present three frameworks for understanding how spatial inequalities in urban areas intersect with transport justice.

Chapter 3

Materials and Methods

In this chapter, we introduce the main data sources that are used throughout this thesis. The data sources are delineated by four different categories: demographic, geospatial, mobility, and public transit. For each of the three perspectives explored in this thesis, transit and socioeconomic data sources are necessary for assessing disparities in public transport service. Meanwhile, housing data is particularly useful for characterising neighbourhood-level vulnerability to housing insecurity, which we leverage to highlight inequalities in residential-workplace dependencies. Furthermore, mobility data serves to parameterise agent-based models, which are used to estimate more dynamic forms of experienced segregation. Finally, spatial data informs features of physical spaces, serving to contextualise transit inequality in the context of the built environment.

3.1 Population Data

Residential neighbourhoods have long been studied to better understand the context in which individuals live [137, 233, 54]. Geodemographics, which entails deriving quantitative measures of individuals based on where they live, is often made available by national-level census surveys [225]. Smaller spatial subdivisions (for example, states or counties in the US

context) have also provided extensions to such data sources. In this section, we outline the relevant population-level surveys that can be used to inform the sociodemographic makeup of different regions. Furthermore, we discuss estimation errors that exist within such datasets as a result of anonymisation and protecting civilians' privacy. Finally, we focus specifically on characteristics related to housing conditions, which we use in 3 to define residential locations that are vulnerable to housing insecurity.

3.1.1 Socioeconomic Status

This thesis focuses on transit inequality in the context of the United States of America (US). While national census surveys exist for other countries, at varying scales, this section focuses on socioeconomic data provided by the US Census Bureau. Additionally, our work analyses disparities at the census tract and block group level. Census tracts are subdivisions of a county and aim to have a population of 4,000, although the population can range from 1,200 and 8,000 people. Meanwhile, census block groups CBGs, which are the smallest spatial unit that the Census Bureau publishes data for, typically consist of 600 to 3,000 individuals. When data is available, we conduct our analyses at the CBG-level, as it is a higher spatial resolution. Regardless, the following chapters will specify the spatial resolution that is chosen.

The American Community Survey (ACS) is a recent addition to census data, providing a more detailed overview of US residents, compared to the decennial census, which would provide national survey data of large sample sizes, but at a low frequency. [275]. While the decennial census is conducted less frequently, its large sample size tends to result in smaller margin of errors. However, because the ACS depends on monthly surveys of much smaller sample sizes, it suffers from less certainty. Thus, while it is helpful in terms of informing population changes at a higher temporal resolution, it should not be considered as the ground truth. Furthermore, two points should be noted. First, regardless of transparency, any survey that depends on sampling methods will be accompanied by some degree of uncertainty.

Second, both the decennial and ACS censuses provide complementary datasets informing the extent of error in sampling. Sampling errors tend to be exacerbated in more heterogeneous neighbourhoods due to the small sample size of the ACS [242, 275]. While uncertainty in the ACS is a result of sampling techniques and trade-offs made by the Census Bureau, it remains the best option for high-resolution geospatial data analysis.

In this thesis, we measure socioeconomic demographics using median household income [17, 10, 303] and household income distributions [309, 254, 299]. Median household income indicates, for a given spatial unit, the income level of which half the households earn less and the other have earn more (central tendency). Some limitations of the median household income include that it does not incorporate either features of household composition or regional characteristics reflecting the local cost of living. However, this omission allows cross-regional comparison across the US, as adjusting income levels using local features obscures differences in purchasing power across larger spatial regions. Furthermore, the median household income is useful, in that it is robust to outliers in income distribution. Income data from the US Census Bureau is typically reported at the household level, and can be found in Table B19013 of the ACS [42]. It includes the incomes of household members older than 14 years old and is derived from the distribution of income for all households in a region, even those that have no reported income.

We define segregation by leveraging information regarding the income distribution in an area. While the median household income is useful for conveying the socioeconomic level of the typical household in a region, household income distributions are helpful for identifying the extent of poverty or inequality within a neighbourhood. That is, while the median household income is convenient for inter-neighbourhood comparisons that are robust to outliers, the household distribution is sensitive to change, making it a convenient tool for identifying from which end of the distribution inter-neighbourhood disparities are arising. Thus, we define residential segregation by drawing upon Table B19001 in the 2020 American

Community Survey 5-Year Estimates (ACS), provided by the US Census Bureau. Household income distributions across 16 income brackets, from Table B19001 of the ACS, inform the economic composition of a neighbourhood, at the Census Block Group (CBG) level [43].

We define segregation with respect to income brackets provided by national economic indicators, as this allows for cross-regional comparisons and prevents any biases that may arise for local economic features. One such bias includes cities using different methodologies to measure and report socioeconomic indicators, leading to variances in accuracy across regions. Incorporating local income characteristics can provide a clearer picture of segregation within a particular city. For example, segregation measures for a wealthy city with high segregation may be diminished due to national economic indicators not accounting for local income distributions. However, using local indicators makes it harder to perform a nation-wide comparison.

Our analysis in Chapter 5 discusses how changing the income bracket thresholds, which refer to affluence and poverty, impacts measures of segregation. In short, stricter constraints can precisely identify segregation dynamics, considering how the most extreme ends of the income distribution share space. However, looser threshold constraints lead to larger variance, or spread, of segregation values and capture how a greater proportion of the population shares space. This could lead to dampened segregation values in non-residential aspects of urban life (i.e., destination segregation, segregation on transit routes). Thus, using a larger portion of the population to define the presence of affluent and poor household can introduce challenges when identifying segregation in dynamic aspects of urban life, which have been shown to exhibit more social mixing than residential segregation [216, 206, 90, 312, 256]

3.1.2 Housing Data

In short, housing insecurity can be distilled into seven categories: Housing Stability, Housing Affordability, Housing Quality, Housing Safety, Neighbourhood Safety, Neighbourhood

Quality, and Homelessness. In Chapter 2.2.3, we describe characteristics of these seven housing dimensions as defined by Cox [63]. Cox points out how many quantitative studies tend to use one dimension as a proxy for housing insecurity, and therefore only capture particular disadvantages. Accordingly, in this work, we attempt to define housing insecurity using as many dimensions as possible. We do not include the Neighbourhood Safety, Neighbourhood Quality, and Homelessness dimensions as the available data sources provide information at larger geographical units. Thus, incorporating these dimensions would require sacrificing the census tract granularity at which we measure housing insecurity. Cox states that the Homelessness dimension is optional in defining housing insecurity, bolstering the decision to not include it in our definition. Finally, we combine the Housing Quality and Safety dimensions because their data sources largely overlapped.

The majority of our data is sourced from the 2019 American Community Survey (ACS). We use 2019 data to define housing insecurity because eviction data in 2020 was distorted due to the temporary eviction moratoria that were enforced as interventions during COVID-19 [123]. The national scale of ACS data enables us to apply our analysis to various cities within the US. We define housing characteristics at the census tract level. Although the ACS provides housing data at a census block group level, which are statistical divisions of census tracts, the data availability of eviction rates is limited to the census tract spatial scale. An overview of the data sources outlined in the following sections is presented in Figure 3.1

Housing Affordability

To define the dimension of housing affordability, we use census tract-level data from the following ACS tables: (a) B25070: Gross Rent as a Percentage of Household Income (b) B25097: Mortgage Status by Median Value and (c) B25001: Housing Units. Rent as a Percentage of Household Income helps to define rent burdened households, which the U.S. Department of Housing and Urban Development (US HUD) define as households that spend

Dimension of Housing Insecurity	**Concept**	**Data Attributes**	**Data Source (*Temporal Resolution*)**	**Spatial Granularity**
Housing Affordability	# of severely rent-burdened households per surveyed households	Relative to survey sample	ACS Table 25070 (*2019*)	Census tract
	Median mortgage	Raw value	ACS Table 25097 (*2019*)	Census tract
	# of housing units per census tract population	Relative to population	2019 ACS Table 25001	Census tract
Housing Safety/ Quality	# of households lacking complete kitchen facilities per surveyed households	Relative to survey sample	2019 ACS Table 25051	Census tract
	# of households lacking complete plumbing facilities per surveyed households	Relative to survey sample	2019 ACS Table 25047	Census tract
	# of households lacking telephone service per surveyed households	Relative to survey sample	2019 ACS Table 25043	Census tract
Housing Stability	# of eviction court filings per tract population	Relative to population	January - March 2020 Eviction Lab	Census tract
	Median household size of all housing units	Raw Value	2019 ACS Table 25010	Census tract
	Median # of rooms	Raw Value	2019 ACS Table 25018	Census tract
	Median # of bedrooms	Raw Value	2019 ACS Table 25041	Census tract

Table 3.1 Summary of data sources for the three housing dimensions used in the clustering framework. Most of the data is provided by the US Census American Community Survey, with the Eviction Lab providing data on formal eviction rates to characterise the housing stability dimension.

30% or more of their income on rent [279]. They extend this definition, labelling households at least 50% of their income on rent as severely rent-burdened. We normalise the number of severely rent burdened households in each tract with respect to the total number of surveyed households in that tract. Moreover, we measure a census tract's housing stock by dividing the number of housing units by the population of each tract.

Housing Quality and Safety

Due to the range of data available from the ACS, we combine housing quality and housing safety into one dimension. Other potential data sources include the American Housing Survey and data hubs managed by individual counties or cities. The American Housing Survey provides data on neighbourhood safety and quality, but the finest resolution it provides is metropolitan areas. On the other hand, city-managed data sources provide highly detailed accounts of housing complaints, but each city has methodologies for cleaning the data and reporting it, introducing difficulties when expanding the analysis to different regions. Thus, we use the following tables from the ACS to define the quality and safety of housing: (a) B25051: Kitchen Facilities for All Housing Units, (b) B25047: Plumbing Facilities for All Housing Units, and (c) B25043: Tenure by Telephone Service Available by Age of Householder. Households must contain a sink, a stove, and a refrigerator to be considered as having complete kitchen facilities. Similarly, they must have hot and cold running water, a flush toilet, and a bathtub or shower to qualify as a household with complete plumbing facilities. Telephone service provides one way of measuring household isolation. This is pertinent for providing adequate medical and crime-related services to households. For a household to be considered as having available telephone service, they must have access to telephone service and have a functional phone. All of the above data is measured, for each census tract, as a fraction of households without these services over the total number of households surveyed.

Housing Stability

Finally, to quantify housing Sstability, we draw on eviction and overcrowding data. The American Housing Survey does not provide eviction data, leaving us to depend on county and city-level governments to make the data accessible. Cities like New York City, Dallas, Detroit, and San Francisco have processed eviction court filings to publish data referring to evictions. This introduces a significant array of issues. First and foremost, eviction data assumes that each eviction is processed through a judicial means. However, landlords often enforce informal evictions as they provide a more affordable means for the same outcome. This includes changing the locks or paying a family to move [118]. Other informal evictions range from cities declaring the housing inhabitable or the threat of foreclosure [25]. Documented evictions also impact tenants' abilities to rent in higher income neighbourhoods with lower crime rates, as landlords often check prospective tenants' housing history, giving landlords an often underestimated influence in neighbourhood composition and gentrification [74]. Yet another issue is data processing from the courts' perspective. Errors in data entry and nuanced rulings lead to inaccurate measures of evictions when compounded into accessible tables [237]. Despite all of these issues, leveraging eviction court filings is the most promising option for estimating forced moves. Projects like Anti-Eviction Mapping Project and the Eviction Lab are vocal about increasing eviction data availability. While both projects work with local organisations, the former provides a broader range of analyses and case studies spanning many North American cities. Meanwhile, the latter focuses on how evictions were impacted by COVID-19 and publish weekly and monthly eviction counts for various cities on a census tract and ZIP Code Tabulation Area (ZCTA) level. For this study, we focus on cities for which the Eviction Lab has processed eviction rates on a census tract level.

We return to the ACS to estimate overcrowding in housing. Drawing on studies by the US HUD, which measure overcrowding using Persons-Per-Room (PPR), Persons-Per-Bedroom (PPB), and Unit-Square-Footage-Per-Person, we use the following ACS tables:

(a) B25010: Average Household Size of Occupied Housing Units by Tenure, (b) B25018: Median Number of Rooms and (c) B25041: Bedrooms [30]. These tables capture PPR and PPB. However, the ACS does not provide data on the physical size of housing units. The American Housing Survey, which we previously alluded to in Section 3.1.2, does provide physical housing characteristics, but only on a metropolitan level, which is not fitting for the granularity of our studies. Thus, we capture housing stability by combining eviction filing data from the Eviction Lab and overcrowding metrics, such as PPR and PPB, derived from the ACS.

3.2 Mobility Data

Gathering mobility data is crucial for capturing human trajectories. It can be collected through various means at different magnitudes of resolution. Historically, aggregated travel flows were collected through national census data, local travel surveys, and tax revenue data [15]. The lack of standardisation in these approaches posed issues for studies focusing on different regions as well as research conducted over a long time range. As technological advances progressed, it became possible to extract mobility data by tracking bills and, more successfully, credit card transactions [39, 270, 162]. Compared to the low time resolution on surveys, financial transactions were accompanied by a geographic location and a specific time, increasing the level of detail in the data. However, the increasing prevalence of smartphones in daily life provided a new level of data resolution through anonymised Call Detail Records (CDR), which estimate caller locations based on geographic coordinates of cell towers. Similarly, mobile phone data can be procured through third-party applications and social media platforms [288]. Moreover, mobile phone indicators are a consistent data source across regions whereas the accuracy of travel survey data often depends on how structured the local government is. While CDRs provide high resolution mobility data on a large scale, issues such as inconsistent temporal frequency and cell tower density limit the

precision of identifying callers' locations [308]. Throughout this thesis, we use two main forms of mobility data: commuting data and amenity visitations. While the former informs mobility patterns with respect to residential-workplace locations, the latter provides insight into mobility trends for other domains of urban life that are not strictly associated with workplace. While these datasets do not have the level of detail that mobile phone records do, they have the advantage of being openly accessible, allowing for reproducibility of results.

3.2.1 Commuting

We use the LEHD Origin-Destination Employment Statistics (LODES) dataset from the United States Census's 2019 Longitudinal Employer-Household Dynamics (LEHD) program [44]. The LODES dataset captures the residential patterns of the surveyed workforce by measuring the number of individuals commuting from one census block group to another. However, for consistency across results in Chapter 4, we evaluate residential-employment trends on a census tract level. LODES data provides characteristics of survey participants with respect to the census tract that they live in and the census tract in which they work. This information includes income groups, industrial sectors, and age. By combining a census tract's housing and public transit characteristics with its employment attributes, we can explore how individuals from various housing demographics may have access to different types and magnitudes of employment opportunities.

Figure 3.1 captures the socioeconomic makeup of each city, according to the LODES data. Orange and purple reflect low-income (earning below $1,250 per month) and high-income (earning above $3,333 per month) workers, respectively. The dashed lines indicate what an even distribution across demographics would look like. Thus, we observe that, within the context of the LODES dataset, cities such as Boston and San Francisco have skewed representation of socioeconomic groups. To measure levels of residential segregation, we

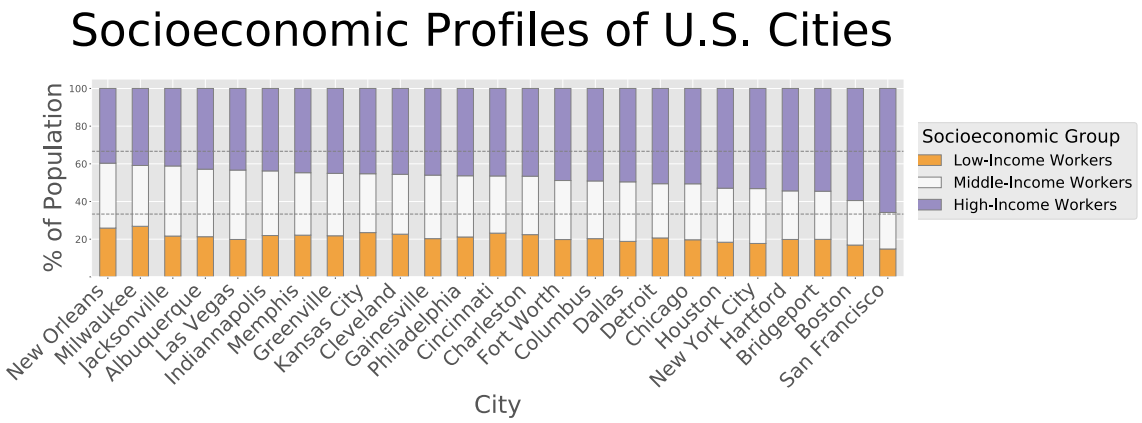

Fig. 3.1 **Representation of socioeconomic groups in the LODES commuting data.** Workforce distribution by income level split into three socioeconomic groups (low, middle and high-income) across 25 US cities.

use Table B19001 from the 2019 ACS 5-Year Estimates, which measures the population distribution across income brackets for each census tract [43].

3.2.2 Amenity Visitations

Our mobility data is sourced from SafeGraph, a data company that aggregates anonymised location data from numerous applications in order to provide insights about physical places, via the SafeGraph Community [252]. To enhance privacy, SafeGraph excludes census block group information if fewer than two devices visited an establishment in a month from a given census block group. The SafeGraph Weekly Patterns data provide visitation counts, on a weekly level, to amenities across the US, along with the distribution of CBGs from which the visitors came. Safegraph defines amenities as physical places that provide services or interests to individuals. This can range from restaurants and cafes to banks and religious facilities. Furthermore, the CBGs in which individuals reside are defined by SafeGraph, using users' locations from 18:00 to 07:00 over a six-week time frame. The SafeGraph data provides business visitation counts on a weekly level. Moreover, it includes the home Census

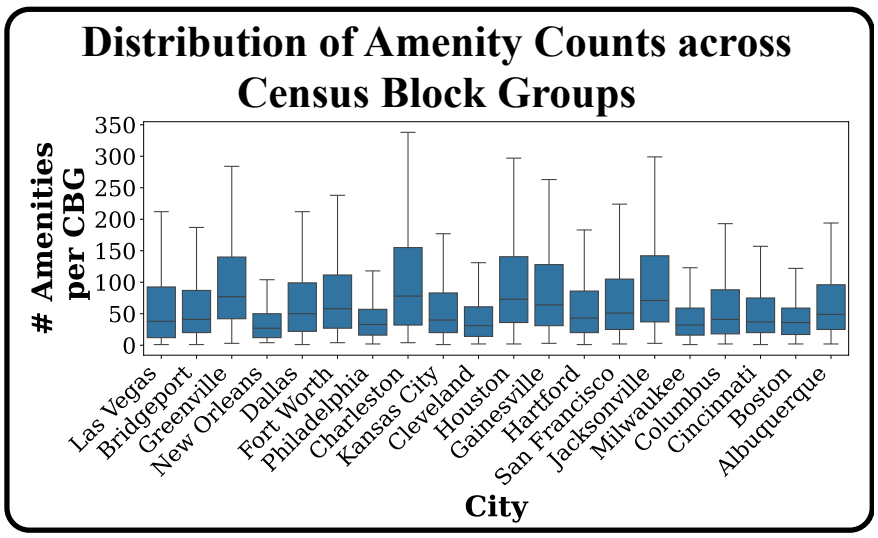

Fig. 3.2 **Prevalence of amenities in a neighbourhood, for a given city.** Each box plot reflects the distribution of the number of amenities in a census block group, with respect to each US city used throughout this work.

Block Group (CBG) of the business' visitors. Figure 3.2 shows the distribution of amenities in each CBG, for each city that amenity visitation data is used.

Furthermore, Table 3.2 shows the relationship between outgoing trips from a census block group to the population of that CBG. We also consider the relationship between SafeGraph mobility and population for the CBGs in the bottom and top Median Household Income quintiles. Thus, Table 3.2 supports research that shows differences in travel behaviour for socioeconomic groups [181, 17, 206]. The differences in how the population of a socioeconomic group corresponds to the mobility behaviour of a neighbourhood could potentially be attributed to under-representation of minorities in mobile location data [166].

3.3 Spatial Data

Two main approaches arise in the context of geospatial data analysis [153]. The first considers spatial data as an expression of social processes, while the other approach views geography as embodying social processes. That is, geospatial data science often focuses

Table 3.2 Correlation between number of outgoing visits and CBG population from the January 2021 Weekly Patterns from Safegraph and 2021 American Census Survey, respectively. The second column considers all CBGs in a city, while the the third column compares mobility rates and population from the bottom income quintile of CBGs. Similarly, the last column does the same, but for the top income quintile. Asterisks indicate the significance of correlation coefficients.

	Pearson Correlation Coefficient		
City	All CBGs	20th pctile	80th pctile
Dallas	**0.183*****	0.121	**0.254*****
Fort Worth	**0.196*****	-0.0	**0.387*****
Bridgeport	**0.316*****	**0.275****	**0.333*****
Gainesville	**0.331*****	0.167	**0.619****
Kansas City	**0.359*****	0.151	**0.367*****
Boston	**0.362*****	**0.383*****	**0.248***
Philadelphia	**0.395*****	**0.288*****	**0.442*****
Cleveland	**0.418*****	**0.197****	**0.557*****
Hartford	**0.418*****	**0.317*****	**0.615*****
San Francisco	**0.443*****	**0.469*****	**0.496*****
Columbus	**0.468*****	**0.507*****	**0.517*****
New Orleans	**0.469*****	**0.31****	**0.427*****
Jacksonville	**0.471*****	**0.344****	**0.619*****
Houston	**0.479*****	**0.331*****	**0.542*****
Albuquerque	**0.493*****	0.246	**0.684*****
Charleston	**0.505*****	**0.572*****	**0.762*****
Cincinnati	**0.506*****	**0.433*****	**0.569*****
Milwaukee	**0.532*****	**0.459*****	**0.651*****
Greenville	**0.654*****	**0.427****	**0.653*****
Las Vegas	**0.794*****	**0.55*****	**0.858*****

*p$<$ 0.05;**p $<$ 0.01; ***p$<$ 0.001

on quantifying demographic features of predefined statistical boundaries. However, an equally viable approach consists of defining spatial structure based on patterns that emerge from sociodemographic characteristics. The main distinction lies in how geography is used. Expressive methods use geography as a means for explaining social processes, while embodying methods leverage social structures to delineate spatial divisions [276, 96]. While spatial analyses conventionally concentrated on demographic distributions in residential contexts, more recent research has expanded to include activity spaces or egohoods to capture spatial patterns of how demographics interact with their environment [277, 172, 278]. It is critical to note, however, that the above approaches are all susceptible to the MAUP discussed in Chapter 2.3.3. Moreover, assigning quantitative features to spatial units allows for testing spatial autocorrelation on a global and local level [9, 248, 100, 222], a geostatistical tool we employ in Chapter 4.4. Quantitative human geography has shifted between prioritising spaces and places, where space refers to the physical features of an area and place captures how those areas are used [255, 141]. While space characterises an area as a setting in which activities take place, place considers how spaces are a product of society. Moreover, places frame space with respect to individual-level behaviours and decisions [141]. Our work aims to capture both spatial features of a neighbourhood, as well as how it used, through defining the built environment. In order to define regions in a city by their built environment characteristics, we retrieve street networks, building footprints, and points of interest from OpenStreeMap (OSM) to define street design, building density and amenity diversity, respectively [221].

OSM is an open-source database of geospatial information. Road layouts, intersection locations, enforced speed limits, buildings footprints, land-use, and amenities, are just a handful of the geographic information it provides [112]. In the context of the US, information pertaining to roads uses TIGER (Topologically Integrated Geographic Encoding and Referencing) shapefiles, provided by the US Census Bureau, as a foundation off of which to

build. Street network data from OSM is comparable to proprietary providers such as Tele Atlas and NAVTEQ [321]. The following metrics make use of the vast amount of geospatial data that OSM offers.

3.3.1 Street Network Design

The layout of streets in US cities, specifically their level of griddedness, has historical roots established by the US Homestead Act [302] and can be indicative of urban planning goals in a region [274]. As car-centric mobility preferences have grown, modern suburban roads layouts have evolved to match these mobility needs [34]. Thus, given that street design has changed over time, it is critical to interpret street layouts with respect to the time and geographical context in which they exist. Moreover, it is essential to note that grid-like streets are not directly reflective of the extent of urban planning, which is showcased by how various designs emerge as a result of differences in cultural mobility preferences [188, 250]. To calculate how grid-like a region's road layout is, we follow Boeing's methodology, denoting griddedness to be a function of the straightness of streets, street orientation order, and the proportion of four-way intersections [34]. For a given census tract, ct, straightness can be measured as follows:

$$\varsigma(ct) = \frac{D(ct)}{L(ct)} \tag{3.1}$$

where $D(ct)$ reflects the average great-circle distance between the endpoints of each street in ct. $L(ct)$ is the average length of the street segments in a neighbourhood ct. By defining straightness with the ratio in 3.1, ς captures the extent to which streets in a neighbourhood resemble straight lines, with values closer to one indicating straighter streets.

Street network order captures another aspect of how grid-like a city's street layout is [32]. Using Boeing's street network orientation definition, we apply entropy to the orientation of a

city's street network, and take the complement of this value:

$$\varphi(ct) = 1 - \left(\frac{H_{o,ct} - H_g}{H_{max} - H_g}\right)^2 = 1 - \left(\frac{-\sum_{i=1}^{36} P(o_i, ct)\ln(P(o_i), ct) - \ln(4)}{\ln(36) - \ln(4)}\right)^2 \tag{3.2}$$

where the order (φ) of a given census tract, ct, is described with respect to the entropy of the tract's street orientation distribution ($H_{o,ct}$), the entropy of the idealised city grid (H_g), and the maximal entropy (H_{max}). The empirical entropy for a tract is measured by binning street orientations into 36 groups, where each group reflects a range of 10°. Thus, the entropy can be measured with respect to the probability of a street, in a given census tract, ct, belonging to one of these groups ($P_{o_i,ct}$). The ideal city grid entropy, H_g,applies entropy to the scenario in which all streets are equally distributed across 4 groups, assuming each group represents an orthogonal orientation. In proposing this metric, Boeing noted the importance of shifting each bin by -5° to ensure that bins for common orientations, such as 90°, would include streets with orientations slightly lower or higher (i.e. 89.9° and 90.1°) in the same bin group. Furthermore, the maximal entropy, H_{max}, is calculated using an equal probability distribution across all 36 groups. Incorporating the ideal grid entropy and maximal entropy serve to normalise the empirical entropy value using min-max scaling. Finally, by taking the complement of the normalised value, we end up with a metric in which values approaching one denote streets that point in similar directions to one another.

The last component of the grid index is the proportion of nodes in the street network that are four-way intersections, γ. To derive a single measure that represents these three metrics, Boeing proposes taking the cube root of the product of each component:

$$G(ct) = \sqrt[3]{\varsigma(ct) * \varphi(ct) * \gamma(ct)} \tag{3.3}$$

where $\varsigma(ct)$, $\varphi(ct)$, and $\gamma(ct)$ denote the straightness, orientation order, and proportion of four-way intersections of streets in a census tract, ct, respectively. Since each component

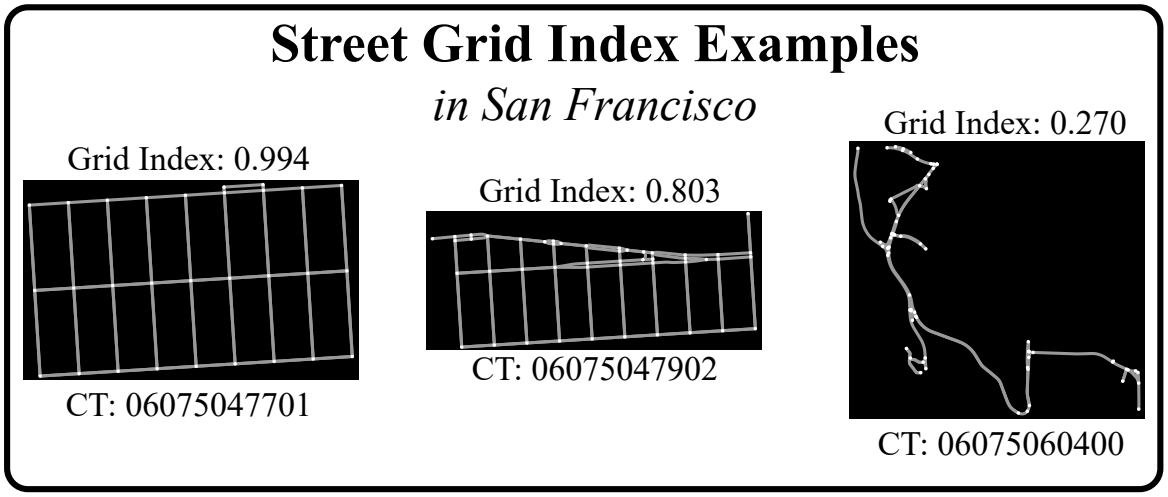

Fig. 3.3 **Examples of street networks and their corresponding grid indices, for different census tracts (CT) in San Francisco.** The FIPS code, which identifies each CT, is displayed at the bottom of each plot. The grid index is shown at the top. Nodes reflect intersection and edges denote streets in each CT's street network.

has a lower bound of 0 and upper bound of 1, $G(ct)$ is, consequently, bound between 0 and 1, with larger values being indicative of neighbourhood with more grid-like streets. Section C.1 in the Appendix discusses other possible ways of computing $G(ct)$, validates the approach in Equation 3.3, and explores how these alternate approaches would impact the results found in Chapter 6.

Figure 3.3 displays three Census Tract (CT)s in San Francisco with different grid index values. The left most street network in Figure 3.3 depicts a CT with a highly gridded street design, while the centre plot exhibits a gridded, but not fully orthogonal street network. The right most panel has a much lower grid index. It highlights how natural features, such as bodies of water or mountains, can impact street design, as this census tract is located along the coast of the Pacific Ocean and contains Lake Merced on the right of the displayed street network.

3.3.2 Building Density

Urban theory and research highlight how measures of density can capture urban vibrancy by measuring the concentration of an urban resource [107, 207]. We calculate building density using building footprints from OpenStreetMap. Specifically, we divide the cumulative area of all buildings in a census tract by the total land area, provided by US Census TIGER shapefiles, for that tract. Thus, building density reflects the proportion of total land area that consists of buildings. Since build footprint area cannot exceed the amount of land in a neighbourhood, this ratio is also bound between 0 and 1, with larger values reflecting areas with higher building density.

3.3.3 Amenity Diversity

We have established how density can capture the abundance of a particular resource and design can reflect the general mobility preferences that have been integrated through urban planning. However, we have yet to discuss how an environment can be defined by the variety of resources it has. This is a distinguishing feature as a dense, grid-like area may only consist of a limited set of amenities, and consequently, attract a specific demographic [87]. To estimate amenity diversity, we define amenities with respect to 10 categories: Food, Education, Healthcare Facilities, Finance, Religious Venues, Government Facilities, Recreational Areas, Entertainment, Retail, and Professional Services [109]. OpenStreetMap provides amenity tags to delineate the categories to which each amenity belongs. We, once again, apply normalised Shannon Entropy to the distribution of amenities in a census tract, across all categories:

$$H_{amenity}(ct) = \frac{-\sum_{i=1}^{10} P(ct,c)\ln P(ct,c)}{\ln(10)} \tag{3.4}$$

where $H_{amenity}(ct)$ captures the the amenity diversity in a census tract, t, and $P(ct,c)$ reflects the empirical probability of an amenity in tract ct belonging to category c. The denominator serves to normalise the entropy using the maximal entropy scenario in which there is an even distribution of all amenities in a tract. Thus, amenity diversity also ranges from 0 to 1, with higher values suggesting a more uniform distribution of types of amenities in a neighbourhood.

3.4 Public Transportation Data

3.4.1 GTFS Feeds

Akin to the increasing prevalence of mobile phone data, public transit data has quickly improved in terms of both standardisation and availability. Although the availability of this data depends on resources of local transit agencies, the usefulness of openly available transit data for transport planning is widely recognised [313, 48]. Only a couple decades ago, Google and TriMet (a Portland transit agency) worked together to introduce General Transit Feed Specification (GTFS) as a standard for transit data [195]. GTFS data has two forms: static and real-time. This work leverages static data, which provides information regarding transit schedules and stops. Meanwhile, real-time GTFS data describes traffic conditions and delays. The GTFS standard requires at least six text files: `stops`, `stop_times`, `trips`, `routes`, `agency`, and `calendar`. `stops` provides geographical references for where transit stops are, in a given region. `stop_times` describes the flow of transit through a given stop, while `trips` provide the frequency of particularly routes, which are defined in `routes`. `calendar` defines a high-level overview as to how trips vary across the week. Finally, `agency` delineates which transport agencies are in charge of particular routes in a city, which is specifically useful when considering cities that have multiple agencies and public transport options available. We leverage General Transit Feed Specification (GTFS) data, from the Mobility Database,

to build public transportation networks for 16 US cities [201]. The Mobility Database is an updated version of TransitFeeds, a commonly used source for acquiring GTFS data. However, the Mobility Database includes over 100 corrected sources and over 150 new sources, in comparison to TransitFeeds.

3.4.2 Transport Network Data and Features

Throughout this thesis, we measure features of public transit systems using `UrbanAccess` and `r5py`. While both tools are open source projects, the former leverages Djikstra-variants to estimate transit travel times, while the latter uses RAPTOR. The differences between these methods are discussed in 2.4.2. Furthermore, the first two spatial inequalities (residential-workplace dependencies and experienced segregation) use UrbanAccess to measure features of transit systems, while the last perspective, which incorporates built environment characteristics, depends on r5py estimates. Although the data sources for the two tools are the same, this section outlines how the frequency-based modelling approach of `UrbanAccess` and the schedule-based approach of `r5py` lead to divergent estimates in transit time. Below, we compare estimates of transit times between the two tools, identifying positive correlations between median travel times. However, we also discuss differences between the two that explain why each tool may give different travel time estimates between two tracts.

In both tools, we set the walking speed to 3 miles per hour, or 4.82803 kilometres per hour. Although r5py has a default walking speed of 3.6 kilometres per hour, for consistency, we ensured that both transport routing tools had the same baseline assumption for how fast pedestrians can travel. We use San Francisco as a case study to emphasise the similarities between the tools, despite their varying efficiency and pathfinding approaches. The three cities that we applied both r5py and UrbanAccess routing to were Boston, Philadelphia and San Francisco. That is, for these three cities we measure transit times between every pair of

Table 3.3 Comparing estimated transit times between transport routing tools, for Boston, San Francisco, and Philadelphia. The second column lists the Pearson correlation coefficient between transit time estimates for census tract pairs in each city. Asterisks indicate the degree of significance for the coefficients. The last two columns denote median transit travel times, considering all pairs of census tracts, as measured by UrbanAccess and r5py, respectively.

		Median Transit Times	
City	Correlation	`UrbanAccess`	`r5py`
Boston	**0.9*****	42.900	50.826
San Francisco	**0.944*****	38.394	39.790
Philadelphia	**0.918*****	40.329	53.400
*p< 0.05;**p < 0.01; ***p< 0.001			

census tracts in the respective city. The second column of Table 3.3 highlights the strong and significant correlations between transit time estimates.

The last two columns show the median transit times between all pairs of census tracts, as defined by UrbanAccess and r5py, respectively. We observe longer estimated transit times when using r5py. These differences are mostly due to variations in our implementation of the tools and inherent discrepancies in the pathfinding algorithms used by each tool. In terms of our implementation, when using UrbanAccess in Chapter 4, we use a deterministic approach, leveraging the internal point of each census tract, as defined by the US Census Bureau. The internal point, in most cases, reflects the centroid of a census tract or geographical area. However, in cases of irregularly shaped tracts, the internal point is the closest point to the centroid, favouring a point that is not on water[1] Accordingly, each census tract is associated with one geographical point. Thus, when we route between a pair of census tracts, we are only routing the path between the tracts' centroids. This contrasts our approach to estimating times using UrbanAccess in Chapter 5 and r5py in Chapter 6, in which we introduce a stochastic approach. Chapter 5 discusses, in detail, the approach to calculating transit times. Thus, in this section, we focus on the stochastic approach that we implement while using

[1]Details about the Census Bureau's definition of an internal point are available at <https://www.census.gov/programs-surveys/geography/about/glossary.html#par_textimage_3>.

r5py. For each census tract, ct, we sample ten random points, $\{ct^1, ct^2, ...ct^{10}\}$, within the tract that do not intersect with any bodies of water. Then, for a journey from census tract ct_o to ct_d, we estimate the average travel time:

$$J_{transit}(ct_o, ct_d) = \frac{\sum_{x=1}^{10} \sum_{y=1}^{10} J'_{transit}(ct_o^x, ct_d^y)}{100} \tag{3.5}$$

Thus, for each census tract pair we estimate a hundred potential transit paths that originate from ten points in the origin tract and ten points in the destination tract. Figure 3.4 uses San Francisco as a case study for showing differences in the implementation and methodology when using UrbanAccess and r5py. Panels A and C in Figure 3.4 highlight the difference between the deterministic approach used with UrbanAccess and the stochastic sampling used with r5py. In Panel A, there is only one source and one destination node, each representing the centroids of their census tracts. Thus, the transit-pedestrian network is used to calculate the shortest path between the pair of nodes. Panel C, on the other hand, has 10 origin and 10 destination nodes. Thus, RAPTOR is used to calculate 100 shortest path estimated for each potential pair of origin and destination nodes. In effect, we observe that different origin and destination points in a census tract can lead to various transit routes being quicker, as opposed to using the centroid as a representative point.

The second way in which the discrepancies in travel time estimation emerge is through the methodologies incorporated in the tools themselves. UrbanAccess generates a transit-pedestrian network, given geospatial and GTFS data. Then, it applies contraction hierarchies to the transit-pedestrian network to improve the efficiency of pathfinding. Given any pair of origin-destination points, UrbanAcccess finds the closest node in the network, calculates the shortest path between these two nodes, and then calculates the walking time between the nodes and the queried points. Alternatively, r5py uses Dijkstra's algorithm to route queried points to closest transit nodes. Then, it applies RAPTOR to find transit paths that solve multiple objectives such as minimising walk time, wait time, and number of transfers.

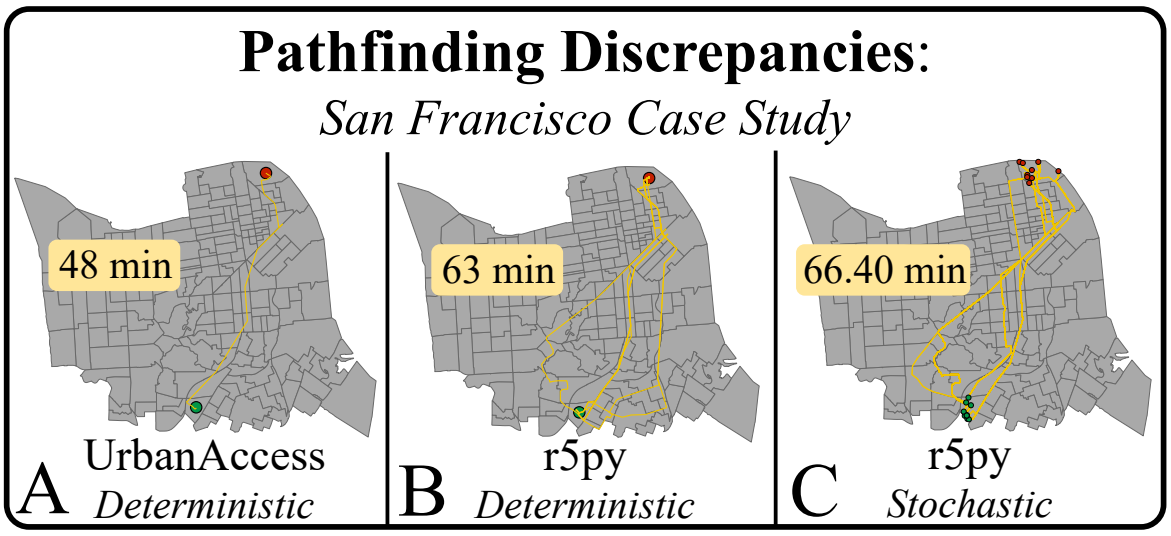

Fig. 3.4 **Differences in travel time estimates between different transport routing tools, using San Francisco as a case study.** The green and red points reflect the queried origin and destination points, respectively. Panels A and B both query one path between the internal points of the two census tracts. Panel A, uses UrbanAccess to estimate travel times, while Panel B used r5py. Panel C uses a stochastic approach to calculate 100 shortest paths between 10 randomly sampled points in the origin and destination tracts, each.

The major difference, however, is that r5py incorporates the concept of a departure time window to address the modifiable temporal unit problem (MTUP). MTUP suggests that accessibility can significantly change based on the time frame being considered [229, 165]. This is relevant to transit routing as leaving a few minutes after the queried departure time may have a large impact on the available transit routes. r5py accounts for this by applying RAPTOR at multiple departure times. Specifically, r5py computes a travel time estimate for each minute, stopping when it reaches 10 minutes past the queried departure time. Panel B in Figure 3.4 exemplifies this concept, applying the deterministic approach in UrbanAccess to r5py. That is, only one pair of origin-destination points are queried, both of which are the internal points in their respective census tracts. Panel B showcases the different paths found within the departure time window beginning at 08:00AM on a Monday in January 2020. The value returned from querying the internal points of the origin and destination tract is the median of all the travel times calculated for every minute between 08:00AM and 08:10AM.

We can compare Panels A and B in Figure 3.4 to understand how the tools' methodologies, themselves, lead to variances in travel time estimates, since we use the same deterministic approach for querying paths. Thus, it becomes clear how r5py's departure timeframe solution, to address the MTUP, leads to different potential transit routes. Contrarily, UrbanAccess simply queries the shortest path from the specified departure time. However, both tools have their own strengths and are helpful in different contexts. Although UrbanAccess was released in 2017, it remains highly useful for analysing which specific transit lines are being used. This is particularly relevant for our analysis in Chapter 5, in which we estimate experienced segregation while using public transportation. On the other hand, r5py is particularly useful for multi-objective transit routing, since its use of RAPTOR can account for numerous mobility goals, such a minimising time spent waiting for transit or the number of transfers.

Finally, we provide network characteristics for the transit-pedestrian networks used in UrbanAccess. Table 3.4 includes details regarding the transit layer of the transit-pedestrian networks, as well as GTFS data specifications, such as the number of unique routes and unique trips, considering the GTFS data from all transport agencies in a city. These networks are used for estimating travel times with UrbanAccess. The second and third columns reflect the number of nodes and edges in the transit layer of the transit-pedestrian networks that we construct. As discussed in Chapter 2.4.2, contraction hierarchies will be applied to these multi-edge directed graphs to condense them into networks that are more efficient for solving routing queries.

We note that the networks vary in terms of the types of transit included. For example, San Francisco and New York City include transit by ferries, as it is a common mode of commuting for some residents. Moreover, some the suburbs of some cities tend to be better served than other. Chicago, for instance, has 231 unique routes in its transit system, 101 of which are run by Pace, a suburban bus service which provides connections to more central transit authorities [199]. The remaining 130 routes are run by the Chicago Transit Authority,

Table 3.4 Characteristics of the transit layer in transit-pedestrian networks for various US cities. These networks reflect the transit service schedule for Mondays, from 06:30AM to 10:30AM, in January 2020 This table only includes the 20 cities, for which graph-based routing algorithms were used.

City	# Nodes	# Edges	# Trips	# Routes
Houston	11845	439474	7616	112
Kansas City	4478	46294	1380	57
Milwaukee	4768	226099	4688	52
Boston	6957	76688	4278	204
Jacksonville	2963	28492	1036	39
Columbus	3866	25991	501	37
New Orleans	2931	18605	527	37
Dallas	7813	65250	1859	105
Fort Worth	2005	14228	545	38
Albuquerque	2517	14196	361	36
Greenville	462	7716	220	12
Philadelphia	12152	174578	3897	126
Cleveland	6684	95421	1739	45
Gainesville	1927	24038	1136	42
Hartford	6943	39562	1232	98
Bridgeport	1473	8299	205	16
Cincinnati	5284	286662	1379	50
New York City	10489	458024	15704	197
Detroit	11858	58195	1124	96
San Francisco	4488	65510	2310	99
Memphis	4274	17323	336	32
Charleston	1071	10288	482	23
Indiannapolis	3533	50029	1072	32
Las Vegas	8001	62668	1631	76
Chicago	24997	529034	11117	231

8 of which are train routes and 122 of which are bus routes. Thus, it is apparent that transit across US cities is fairly heterogeneous not only in terms of the transit service across different routes, but also in the travel modes that transit infrastructure accommodates. Although, we present networks for 25 US cities, we calculate shortest paths, using UrbanAccess, for 20 of these cities, barring Chicago, Detroit, Indianapolis, Memphis, and New York City. We omit these cities as their eviction data was not available from the Eviction Lab. However, as we move away from focusing on residential disparities, we reincorporate some of these cities, particularly those with prominent transit systems, for our analysis in Chapter 6.

3.4.3 Road Networks

We estimate travel times, when using a car, between two neighbourhoods to provide a baseline to which transit service can be compared. We accomplish this using three different tools: `Openrouteservice (ORS)`, `Open Source Routing Machine (OSRM)`, and `Pandana` [93, 179, 94]. All three of these tools apply routing algorithms to road networks, which are retrieved from `OpenStreetMap` (`OSM`) [221].`ORS` uses the A* pathfinding algorithm to query travel time matrices. A* follows the same general approach to routing as Dijkstra's Algorithm, which was introduced in Chapter 2.3.2. It differs from Dijkstra's, in that it includes a heuristic function that roughly estimates the distance remaining to the destination node. It should be noted that the heuristic should be developed in a manner, such that it never overestimates the cost to the destination node. Thus, the performance of A* depends greatly on an appropriate heuristic function. Meanwhile, `OSRM` and `Pandana` use Contraction Hierarchies, which was introduced in Chapter 2.4.2, to determine shortest paths. We use `ORS` in Chapter 4, `OSRM` in Chapter 5, and `Pandana` in Chapter 6.

We incorporate `OSRM`, despite its similarities with `Pandana`, because the stochastic approach used in Chapter 5 required significantly more computational power than the environment, in which Pandana was installed, supported. Although Chapter 5 will outline the

Table 3.5 Comparing driving time estimates for different routing tools in five US cities. The second to fourth columns reflect median driving times for all pairs of census tracts in city, using Openrouteservice, Open Source Routing Machine, and Pandana, respectively. The last three columns show the similarities in tract-level driving time estimates between the different routing tools. These values are reflective of the Pearson correlation coefficient, with asterisks denoting the significance of the correlation.

	Median Driving Times			Pearson Correlation between Routing Tools		
City	*ORS*	*OSRM*	*Pandana*	*ORS-OSRM*	*ORS-Pandana*	*Pandana-OSRM*
New Orleans	11.1	11.48	11.2	**0.977*****	**1.0*****	**0.974*****
Dallas	23.13	21.06	23.17	**0.846*****	**1.0*****	**0.846*****
Philadelphia	17.6	15.93	17.67	**0.964*****	**1.0*****	**0.964*****
Cincinnati	21.53	19.18	21.55	**0.869*****	**1.0*****	**0.871*****
San Francisco	11.68	10.65	11.75	**0.902*****	**1.0*****	**0.903*****

[1] *p<0.05; **p<0.01; ***p<0.001

methodology in further detail, we provide a brief overview in this section, in order to explain why differences in shortest path estimations may arise. We use `OSRM` to estimate average transit times between all CBGs pairs in the five aforementioned cities. In order to derive these estimates, we randomly sample 100 locations in each origin and destination CBG, and use `OSRM` to calculate the shortest paths between origin node n_o^i and destination node n_d^i, for $i \in 1...100$. Here, n_o^i belongs to the origin CBG and n_d^i is a point in the destination CBG. Then, we define the average time between the origin and destination CBG by simply determining the mean, given all 100 shortest paths.

To justify the use of all three tools throughout this thesis, we explore characteristics of how each tool estimates driving times. We compare driving times for the 5 cities that were studied in all the analyses: New Orleans, Dallas, Philadelphia, Cincinnati, and San Francisco. The second to fourth columns in Table 3.5 show the median driving times for all origin-destination pairs, in a city. For `ORS` and `Pandana`, this compares travel times via cars between every possible census tract combination. Meanwhile, for `OSRM`, this describes driving times for all possible census block group pairs, due the higher spatial resolution in Chapter 4's analysis. While `ORS` and `Pandana` vary in terms of routing algorithms, Table 3.5

highlights the strong correlation between the median driving times between census tracts. Since A* is merely a variation of Dijkstra, we expect to see strong correlations between the two measures. However, we explain the less strong correlations between `OSRM` and other tools, due to the differences in spatial resolution. That is, we use the same deterministic approach as we outlined for `UrbanAccess`, discussed in the previous section, to measure driving times between census tracts using `ORS`. Whereas, for `OSRM` and `Pandana`, we use a stochastic approach, in which origins and destinations are not determined by the centroids of their respective tracts, but through a random sampling of points over multiple iterations. Furthermore, when we measure the median driving time for a given tract using `ORS` or `Pandana`, we simply find the median value from a particular census tract, to all other tracts in the city. However, we use `OSRM` in Chapter 5 to understand transit at a census block group level. Thus, when we estimate the median driving time of an origin census tract to a destination tract, we consider the shortest paths from all CBGs in the origin tract to all the CBGs in the destination tract. Then, by finding the mean of all these values, we have the average travel time from one tract to another. Accordingly, we can apply the same methodology as we did when using `ORS` or `Pandana`, to determine the median driving time for an origin tract to all other tracts. Thus, we can see how despite different routing algorithms and computational efficiencies, all three tools (`ORS`, `OSRM`, and `Pandana`) provide similar estimates of travel times, when using a car.

3.4.4 Routing Notation

Throughout this thesis, we often use metrics derived from the transit networks. This section aims to establish a consistent notation, which we will refer to throughout the following chapters. First, we refer to census tracts and census block groups as *ct* and *bg*, respectively. We define the time, in minutes, that a transit journey between two points, $(i,\ j)$, takes as $J'_{transit}(i,j)$. Similarly, $J'_{driving}(i,j)$ reflects the number of minutes it takes to drive from point

i to point j. We can aggregate these values to reflect travel times between administrative boundaries using $J_{transit/driving}(ct_1, ct_2)$ to measure the transit or driving times from an origin census tract, ct_1, to a destination tract, ct_2. Furthermore, J can be derived through a stochastic approach, in which numerous journeys from origin points in ct_1 to destination points ct_2 are calculated using J', then aggregated to the tract level. Alternatively, $J_{transit/driving}(ct_1, ct_2)$ can be derived by simply calculating the J' value using the internal points of ct_1 and ct_2, respectively. The process of calculating J will be explained in each chapter.

We also aggregate $J_{transit/driving}(ct_1, ct_2)$ to the origin census tract level to convey mobility opportunity, average commute times, and amenity accessibility. The notation for these, in the context of transit times from a census tract ct, are $T^{opp}_{transit}(ct)$, $T^{comm}_{transit}(ct)$, and $T^{am}_{transit}(ct)$, respectively. The driving times are represented using $T^{opp/comm/am}_{driving}(ct)$. As a quick note, mobility opportunity captures the notion of integral access by averaging the travel times from a census tract to all other census tracts in a region. We elaborate on how these values are calculated, and the concepts they convey, as we encounter them in the chapters. Although we use census tracts as an example in this section, they can be replaced with other administrative boundaries, such as census block groups, depending on the spatial resolution of the analysis.

3.5 Discussion

In this chapter, we described the different data sources we will utilise throughout the thesis in order to highlight how transport inequality is exhibited within various urban mechanisms. In particular, we introduce demographic, mobility, spatial, and transportation data. All the datasets, barring SafeGraph, are openly accessible. SafeGraph is a safegaurded dataset, however it is accessible to academics for research purposes. As a brief summary, we analyse transit inequalities in residential-workplace dependencies (Chapter 4) at a census tract level for 2019. We choose 2019 because the temporary eviction moratoria led to low rates of evictions throughout 2020, which reflected COVID-19 interventions, rather than eviction

dynamics [123]. Furthermore, the spatial scale of the eviction and commuting data limits our analysis to the census tract level. Chapter 5 is conducted at the census block group level, as both the SafeGraph mobility data and the American Community Survey (ACS) data are available at this high spatial resolution. We use the 5-Year estimates from 2020 for this analysis, as our transport data is from the first month of 2021 and, at the time of this analysis, the 2021 ACS survey had not been released. Finally, we continue using 2020 ACS 5-Year estimates to explore transit disparities with respect to the built environment, in Chapter 6. Similar to the analysis in Chapter 4, our use of commuting data limits the spatial resolution in Chapter 6 to the census tract level. In addition to showing data characteristics of each dataset in this chapter, we discuss their limitations, from data collection to analysis. The following chapter leverages the sociodemographic, commuting, and transit data to explore how transit facilitates commutes between residential and employment locations.

Chapter 4

Inequalities in Residential-Workplace Dependencies

4.1 Introduction

Throughout this thesis, we analyse transport justice in the context of different urban mechanisms, which refers to processes that drive functionality and change in an urban area (i.e., housing markets, demographic trends, development of infrastructure). In doing so, we ensure to consider complexities from both a sociodemographic inequality perspective and a transit modelling one. This chapter investigates expressions of transit inequality when considering residential-workplace dependencies. In this manner, we study how transit inequality aligns with commuting processes, while also emphasising how measures of residential-workplace dependencies and housing insecurity can reveal disparities in transit service, that are novel to the field of transit poverty.

This chapter addresses long standing nuances regarding how public transportation intersects with the spatial mismatch hypothesis. Research has shown how long commute times and low workplace accessibility can lead to negative labour market outcomes for low-income individuals [143, 102]. Thus, we apply measures of network entropy to commuting networks,

in order to understand disparities in residential and employment locations. Specifically, we consider how differences in the heterogeneity of employment locations, between low-income and high-income workers, can be more informative in characterising different types of residential-workplace dependencies, compared to conventional measures of segregation. Furthermore, we incorporate empirical transit times to see how the spatial inequalities in residential-workplace dependencies translate to time spent commuting via public transit.

Finally, we analyse transit inequalities in the context of residential disparities, proposing a clustering approach for quantifying the complexities of housing insecurity. By applying geospatial analysis to vulnerable housing demographics, and their respective commuting patterns, we assess transit times to different types of employment areas. Thus, analysing system-facilitated transit access provides insight as to how transit service can contribute to the burden of housing insecurity, with respect to empirical employment locations and to areas associated with better job opportunities. We note that this section leverages work that has been previously published, as well as work that is currently under review for publication [134, 136].

4.2 Network Entropy in Commuting Networks

Networks are particularly useful for analysing commuting behaviours, as they capture structural patterns that other approaches may overlook [173]. Specifically, network entropy can capture the concentration of labour supply and demand as well as the level of diversity of where workers are commuting to or from [187]. Entropy has been used in commuting networks to explain economic growth [103], identify spatial inequalities [163], and measure social assortativity [35]. However, the majority of previous research on entropy in commuting networks analyses the networks of an entire population. Thus, we not only consider the commuting networks of an entire population, but also commuting networks comprised of individuals from particular socioeconomic groups. Disaggregating commuting networks by

demographics allows us to study whether disparities in structural diversity could serve as an indicator of social exclusion.

We use the LODES data, introduced in Chapter 3.2.1, to construct commuting networks for 25 cities in the United States, where every node reflects a census tract and directed, weighted edges depict the number of individuals commuting from one tract to another. These 25 cities cover a wide range of population sizes and socioeconomic characteristics. The LODES dataset also provides information about how the total commutes from a pair of census tracts are distributed across lower, middle, and higher-income demographics. The low-income group consists of individuals earning less than $1,250 per month, while the minimum monthly income for the high-income group is $3,333. In addition to building the entire commuting network of a city, we also build a low, middle, and high-income network, which have the same nodes as the network for the entire city. However, the networks for each socioeconomic group, which we refer to as *disaggregated networks*, have different edge weights depending on the number of people in a socioeconomic group that commute between a pair of tracts.

Throughout our analyses, we use the term *commuting destinations* to refer to the work-places that a residential population commutes to, while *commuting origins* describe the residential areas from which an employment area's workforce commutes. Furthermore, *labour supply* refers to employment areas that supply jobs, while *labour demand* reflects the employable population of a neighbourhood. We use Shannon's entropy, which captures the level of information that can be extracted given a probability distribution, as a measure of network entropy to estimate diversity in commuting destinations and origins [261]. Moreover, network entropy can be applied at different network resolutions (i.e. local or global) and can focus on the commuting in-flow to a work area or the commuting out-flow from a residential region.

Global entropy of in-flow and out-flow captures the urban concentration of labour supply and demand, respectively, by characterising the degree of monocentricity. That is, when a city has one area that supplies most of the labour opportunities, there is more certainty in predicting commuting destinations, corresponding with a lower global in-flow entropy value for the entire city. Thus, global in-flow entropy (H_{GN}^{in}) leverages the node strength, $\sum_i p_{ij}$, of all incoming commutes to each tract, j in a city:

$$H_{GN}^{in} = \frac{-\sum_{\forall j}\left(\sum_{\forall i} p_{ij}\right) log\left(\sum_{\forall i} p_{ij}\right)}{log(|\mathbf{CT}|)}, \tag{4.1}$$

where **CT** reflects the set of census tracts in a city, with each tract corresponding to a node in the commuting network. Moreover, p_{ij} reflects the edge weight (or number of workers) commuting from tract i to j. Thus $|\mathbf{CT}|$ is defined by the number of census tracts in a given city. Global out-flow entropy captures the distribution of labour demand, such that low entropy values depict a scenario in which most workers live in a few areas and high entropy values indicate where residential locations of workers are more evenly distributed across nodes. In contrast to H_{GN}^{in}, global out-flow entropy (H_{GN}^{out}) calculates the out-degree node strength, $\sum_j p_{ij}$, for all census tracts, i in a city.

Global entropy defines a city with respect to how labour supply or demand is distributed across all census tracts. Meanwhile, local entropy defines each tract based on how evenly distributed all its incoming or outgoing commutes are. A high local in-flow entropy (shown in the right-most network of Figure 4.1A) for a census tract implies that the commuting origins for individuals who work in that tract are evenly distributed across all the potential origins. Local in-flow entropy ($H_L^{in}(j)$) for a census tract j accounts for the probability, $p_{(i|j)}$, of tract j receiving commutes from a census tract i, for all possible commuting origins, i:

$$H_L^{in}(j) = \frac{-\sum_{i\in\mathbf{CT}} \frac{p_{ij}}{p_j} log \frac{p_{ij}}{p_j}}{log(|\mathbf{CT}|-1)}. \tag{4.2}$$

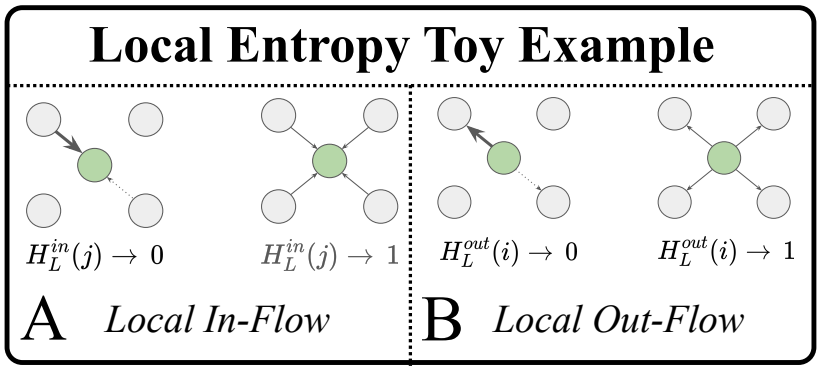

Fig. 4.1 **Higher values of local in/out-flow entropy reflect a less concentrated dependence of commuting origins/destinations on employment/residential areas.** Panel A shows how employment areas with a concentrated commuting flow from specific residential neighbourhood leads to lower values of local in-flow entropy. Similarly, Panel B conveys how lower local out-flow entropy values arise when individuals in a residential neighbourhood tend to concentrate their commutes to a particular employment area.

where p_j reflects the weighted in-degree of node j, normalised with respect to the sum of edge weights in the entire network. Local out-flow entropy ($H_L^{out}(i)$) can be defined similarly, except rather than consider the probability $p_{(i|j)}$ for incoming commutes, it calculates the probability, $p_{(j|i)}$ of workers living in tract i to commute to a census tract j, for all possible workplaces, **CT**:

$$H_L^{out}(i) = \frac{-\sum_{j \in \mathbf{CT}} \frac{p_{ij}}{p_i} log \frac{p_{ij}}{p_i}}{log(|\mathbf{CT}| - 1)}. \tag{4.3}$$

where p_i conveys the weighted out-degree of node i, normalised with respect to the sum of edge weights in the entire network. The right-most network in Figure 4.1B shows how a concentrated distribution of commuting origins leads to lower values of local out-flow entropy. In this manner, we leverage network entropy measures on a global and local scale to understand how commuting characteristics of low-income versus high-income workers indicate levels of segregation not only in terms of where each demographic group lives, but also in the context of the areas where they work. The denominators in Equations 4.1, 4.2 and

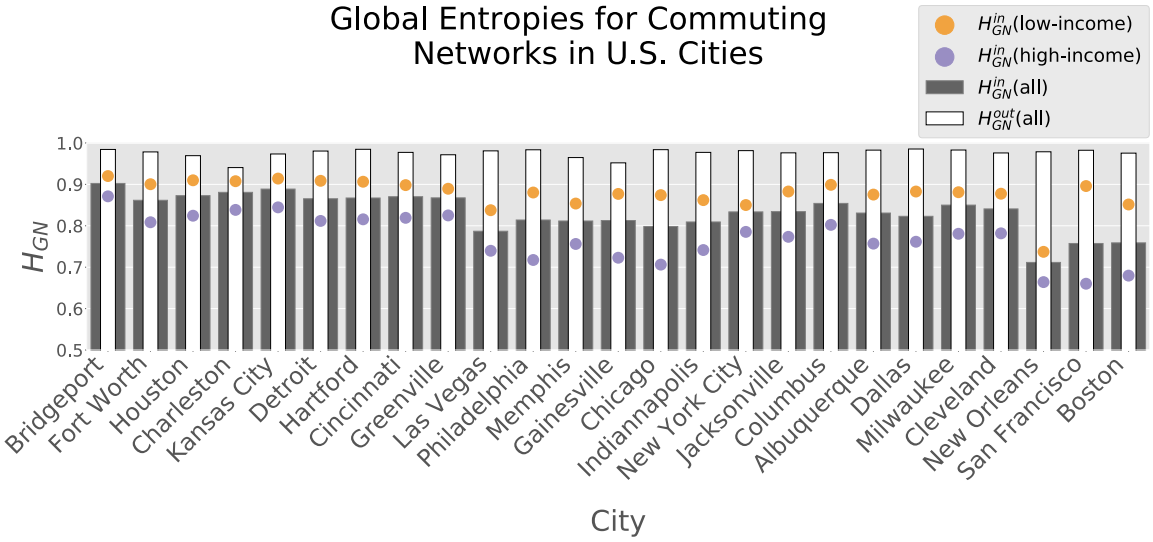

Fig. 4.2 **Employment areas (global in-flow entropy) tend to be more concentrated than residential locations (global out-flow entropy), as seen by higher H^{out}_{GN} values for all 25 US cites**. The grey and white bar capture global entropy of the entire commuting network for in-flow and out-flow commutes, respectively. The orange point measures H^{in}_{GN} for the low-income network, while the purple point does the same for the high-income network.

4.3 serve to normalise the entropy values for cities of varying sizes, based on the number of tracts in a city (Eq. 4.1), and the focal node's maximal possible degree (Eq. 4.2 and 4.3).

4.2.1 Global Entropy to Compare Disparities across Cities

We begin by applying global entropy measures to the commuting networks of the 25 different cities. We reiterate that because entropy values are normalised with respect to network size, we can make comparisons not only between cities, but also between commuting networks of different socioeconomic groups. Figure 4.2 elucidates how, for every city, the global entropy of commuting out-flows is consistently larger than the global entropy for commuting in-flows. This pattern is to be expected, as employment hubs tend to emerge in particular areas of a city, whereas residential locations tend to appear through most areas of a city [187]. Notably, the lower values of global in-flow entropy in cities such as New Orleans, San Francisco, and Boston indicate the presence of larger employment hubs.

By disaggregating the commuting networks based on workers' economic profiles, we can analyse the networks of low-income and high-income workers separately. Interestingly, within the cities we analyse, the global in-flow entropy of high-income commuters is persistently lower than that of the low-income group. These lower values imply less structural diversity in high-income commuting origins, which translates to a higher level of monocentricity for high-income jobs. What is clear is that there is a distinct difference in global entropy values for commuting in-flows when considering networks of different socioeconomic groups. Whether these differences are indicative of socioeconomic inequality is explored in the following sections.

When comparing normalised global entropy across cities' commuting networks, it is crucial to consider how common the occurrence of zeros are in the distribution of in-degrees and out-degrees of nodes [156]. If a given distribution has a prevalence of zeros, it essential to modify the normalisation process such that the adjusted entropy value is divided by the number of census tracts that have a non-zero node strength (Eq. 4.1). Not accounting for this can lead to artificially low entropy values. Tables A.1 and A.2 in Appendix A reflect the characteristics of the commuting networks for each city, ensuring that the presence of zeros does not drastically change the results of the analysis. This can be seen, as the largest change between the reported and adjusted global in-flow is 0.110%, for New Orleans. Furthermore, Boston is the only city that has the presence of any zeros in outgoing node strength, which leads to a 0.092% change from the reported global out-flow entropy in Figure 4.2. Thus, we show that our normalised measures of global in-flow and out-flow entropy are not skewed by zeros in the distribution and can, consequently, be compared across cities.

4.3 Conventional Socioeconomic Segregation in Urban Landscapes

We analyse socioeconomic features of neighbourhoods to better understand the mechanisms that drive the identified differences in commuting networks across demographics. Thus, we characterise urban areas in terms of the concentration of socioeconomic groups, using the ICE, introduced in Chapter 2.2.1. As a reminder, ICE can be calculated using the following formula:

$$\text{ICE}_{ct} = \frac{A_{ct} - P_{ct}}{T_{ct}}, \tag{4.4}$$

where the ICE for a census tract, ct, is defined as the difference between the number of affluent residents, A_{ct}, and the residents below the poverty line, P_{ct}, over the entire population, T_{ct}. We measure residential segregation using the household income distribution data from the ACS (Table B19001), introduced in Chapter 3.2.1. We define A_{ct} and P_{ct} as the number of workers with yearly incomes above \$125,000 and below \$20,000, respectively, for a given tract ct.

In a similar vein, we can use Equation 4.4 to measure the level of segregation for a census tract, ct, from an employment perspective, defining A_{ct} and P_{ct} as the number of workers commuting to tract ct from the high-income and low-income group, respectively. As a reminder, the high-income group consists of workers earning above \$3,333 per month, while low-income workers have monthly earnings of less than \$1,250. Thus, T_{ct} captures the total number of individuals commuting to ct, regardless of demographics. Accordingly, we can use ICE to capture segregation levels at both a residential (ICE_{res}) and employment (ICE_{emp}) scale. We use the LODES dataset to calculate ICE_{emp} as it provides unique information regarding demographic characteristics of a workforce. Meanwhile, we use the ACS household

income distributions to calculate residential segregation as it provides a high granularity of income levels to characterise tracts by its residents.

4.3.1 Local Entropy as a Measure of Segregation

In this section, we aim to disentangle whether overall structural diversity entails other forms of diversity, such as lower levels of segregation. In doing so, our goal is to clarify if the monocentricity of job opportunities, which is more present in high-income networks, is a privilege or a burden.

Residential segregation

To better understand the differences in how housing and employment landscapes intersect for socioeconomic groups, we evaluate whether a relationship exists between residential segregation and the diversity of employment destinations for residents of a particular tract. The fourth column of Table 4.1 lists the Pearson correlation coefficients when comparing the local out-flow entropy (H_L^{out}) with the residential ICE value (Eq. 4.4) of census tracts in a city. We recall that the higher the ICE_{res} value, the larger the proportion of high-income residents in that area. All but three of the 25 cities have a significant, negative correlation. This reveals that census tracts characterised by more affluent residents tend to have lower local diversity values (i.e., concentrate commutes to fewer tracts), which indicate their dependence on particular tracts for supplying labour to their residents. The red scatter plots in the top row of Figure 4.3 illustrate the negative trends between residential segregation and the diversity of commuting destinations, using Milwaukee, San Francisco, New York City, and Detroit as examples.

On one hand, one can argue that these higher diversity values for less affluent tracts makes them less vulnerable to any shortages in labour supply, such that if an employment location stops providing opportunities, they have other options of commuting destinations.

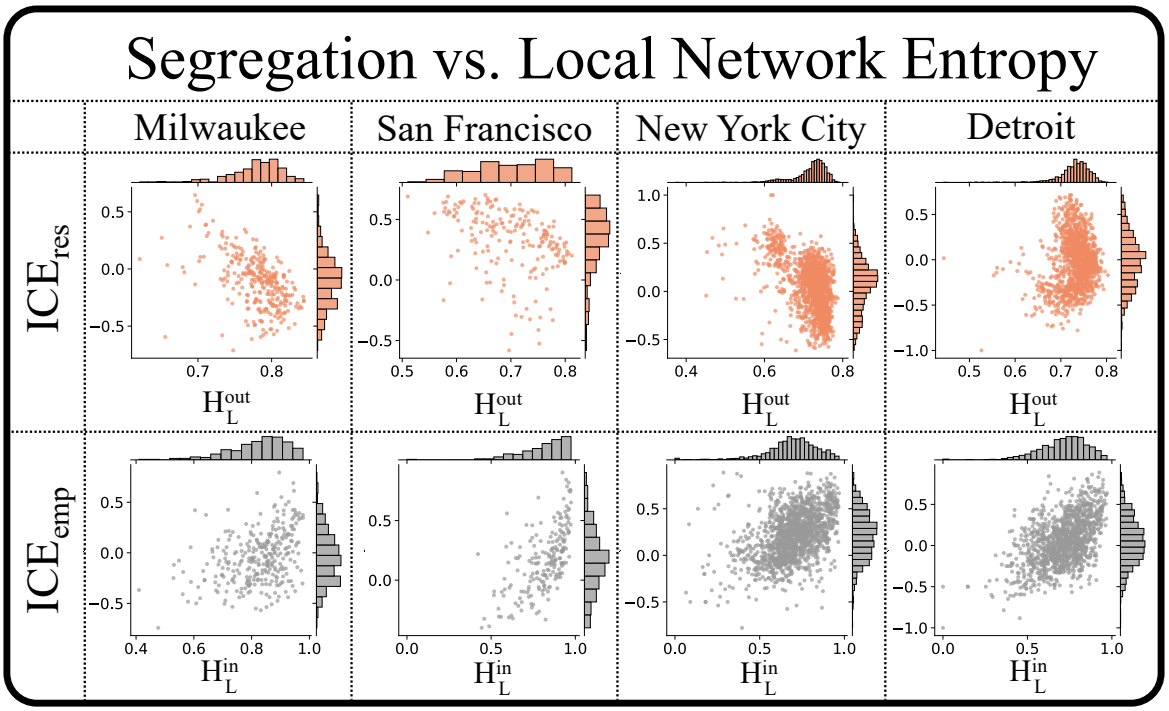

Fig. 4.3 **More diversity of commuting destinations tends to be associated with residential segregation for low-income individuals, while higher diversity of commuting origins is associated with employment segregation of high-income workers.** Each column reflects the relationship between segregation and local network entropy for Milwaukee, San Francisco, New York City, and Detroit. The top row of red scatter plots conveys the relationship between residential segregation (y-axis) and commuting destination diversity (x-axis), measured using local out-flow entropy. The bottom row of grey scatter plots shows how employment segregation (y-axis) related to commuting origin diversity (x-axis), measured using local in-flow entropy.

However, this negative correlation could also indicate the presence of inequalities in the housing landscape (i.e. through the diversity of commuting origins). Thus, we proceed to analyse the opposite dynamic, comparing the diversity of commuting origins (H_L^{in}) to the degree of employment segregation (ICE_{emp}), to explore whether the identified negative correlations imply socioeconomic disparity on a residential or employment level.

Employment segregation

For completeness, we have included the results for the correlation patterns between segregation of a tract's workforce (ICE_{emp}, Equation 4.4) and the diversity of commuting origins for that workforce (H_L^{in}) in the fifth column of Table 4.1, which shows significant positive correlations for 20 of the cities. When considering the high-income group, we see that, from an employment perspective, tracts in which more affluent employees work are more diverse, expressing heterogeneity in commuting origins. However, from a residential perspective, tracts with more affluent residents are less diverse, expressing homogeneity in commuting destinations. Thus, we show that higher values of structural diversity do not necessarily imply an socioeconomic advantage. The relationship between employment segregation and commuting origin diversity can be seen in the grey scatter plots (bottom row of Figure 4.3)

This section elucidated how the structural diversity of census tracts often corresponds with their urban characteristics. We measure the local in and out-flow entropy of entire commuting networks to highlight how diversity of the employment landscape can be reflective of employment and residential segregation. The next section wraps up this analysis by disaggregating the entire commuting network of each city into low-income and high-income networks. This allows us to examine how unequal labour distribution may be exacerbating existing inequalities.

Table 4.1 General network properties and Pearson correlation coefficients for different forms of local entropy and socioeconomic segregation. The last column refers to local entropy differences between the disaggregated networks, discussed in Section 4.3.2. Boldface is used to indicate significant correlations, with asterisks reflecting p-value.

City	Network Properties		H_L^{out} vs. ICE_{res}	H_L^{in} vs. ICE_{emp}	ΔH_L^{in} vs. ICE_{emp}
	Nodes	*Edges*	r[1]	r[1]	r[1]
Charleston	85	6,162	**-0.313****	0.116	**0.453*****
San Francisco	196	25,886	**-0.373*****	**0.614*****	**0.341*****
Gainesville	56	2,803	-0.243	**0.313***	**0.496*****
Greenville	111	10,431	**-0.386*****	0.033	**0.596*****
Albuquerque	153	18,871	**-0.385*****	0.116	**0.632*****
New Orleans	176	13,680	**-0.309*****	0.140	**0.843*****
Houston	921	398,876	**-0.287*****	-0.018	**0.716*****
Boston	204	20,608	**-0.401*****	**0.652*****	**0.646*****
Indianapolis	224	33,147	**-0.494*****	**0.231*****	**0.754*****
Las Vegas	487	124,083	0.027	**0.093***	**0.818*****
Philadelphia	384	69,364	**-0.547*****	**0.289*****	**0.754*****
Columbus	347	76,995	**-0.486*****	**0.272*****	**0.735*****
Hartford	224	33,107	**-0.507*****	**0.430*****	**0.710*****
Jacksonville	173	23,610	**-0.441*****	**0.473*****	**0.722*****
Cincinnati	222	32,801	**-0.185****	**0.306*****	**0.751*****
Milwaukee	297	50,210	**-0.517*****	**0.377*****	**0.829*****
Cleveland	446	85,609	**-0.182*****	**0.330*****	**0.834*****
Bridgeport	210	28,259	**-0.425*****	**0.278*****	**0.665*****
Fort Worth	357	76,883	**-0.483*****	**0.280*****	**0.798*****
Memphis	221	31,507	**-0.135***	**0.212****	**0.816*****
Chicago	1,318	441,406	**-0.287*****	**0.358*****	**0.855*****
New York City	2,164	990,302	**-0.478*****	**0.455*****	**0.837*****
Detroit	1,163	385,185	**0.139*****	**0.527*****	**0.865*****
Dallas	529	129,486	**-0.569*****	**0.350*****	**0.830*****
Kansas City	283	49,377	**-0.165****	**0.328*****	**0.781*****

[1]*p<0.05; **p<0.01; ***p<0.001

4.3.2 Socioeconomic Disparities in Diversity of Commuting Origins

In this section, we illustrate how local in-flow entropy measures of disaggregated networks can capture experienced segregation. For improved readability, we show the results for four representative cities of different sizes and socioeconomic profiles (Milwaukee, San Francisco,

New York City and Detroit). Nevertheless, our findings are based on the analyses of the 25 cities.

We proceed, analysing how differences in heterogeneity of commuting origins coincide with levels of employment segregation. We begin by splitting the commuting network of the city into separate networks that measure the residential-work patterns of socioeconomic groups separately. Specifically, we define the entire commuting network for a given city, which is provided by the LODES data, as $G^{all} = (V, E)$. Here, V reflects all census tracts in a city, while an edge $e_{ij} \in E$ conveys the number of individuals that living in census tract v_i and work in tract v_j. With this in mind, we can focus on the commuting flows for specific income demographics. We denote the low-income commuting networks as $G^{lo} = (V, E^{lo})$, in which any given edge $e^{lo}_{ij} \in E^{lo}$ reflects the number of low-income workers commuting from tract i to j, as defined by the LODES data. Similarly, $G^{hi} = (V, E^{hi})$ represents the high-income commuting network, $e^{hi}_{ij} \in E^{hi}$ representing the number of high-income workers commuting from tract i to j. For a given edge in low-income and high-income commuting network, the sum of their edge weights should never exceed the value of the corresponding edge in the entire commuting network ($e^{lo}_{ij} + e^{hi}_{ij} \leq e_{ij}$).

By considering commuting networks for different socioeconomic demographics, we can understand the extent to which the diversity of commuting origins differs between the low-income and high-income workers within a given census tract. Panels A-D in Figure 4.4 plot the local in-flow entropies of every tract in the low-income network ($H^{in}_{L,lo}$, x-axis) against their respective entropy values in the high-income network ($H^{in}_{L,hi}$, y-axis), for four of the analysed cities. The black line expresses the case in which a census tract has equal diversity of commuting origins in both socioeconomic networks, which we see a few cases of in New York City and Detroit, indicated by the white points on the diagonal. Orange points reflect census tracts in which the low-income individuals have a more even distribution of commuting origins than the high-income workforce. The purple points capture census tracts

with the opposite characteristics: greater diversity in residential locations for the affluent workforce.

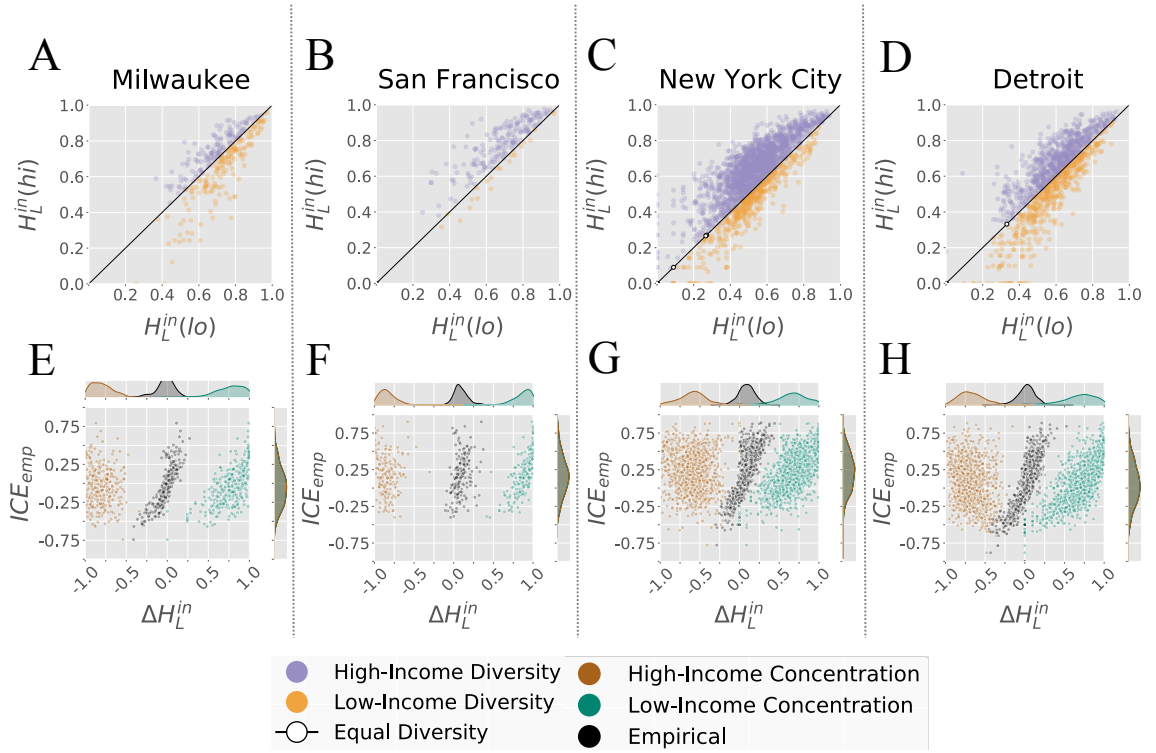

Fig. 4.4 **Comparing the local in-flow entropy between high-income and low-income commuting networks helps to identify cities in which high-income workers tend to have more diversity in commuting origins and reveals how network entropy can used to reveal disparities in residential-workplace dependencies.** Panels A-D compare entropy values for tracts in the low-income and high-income network, with points below the diagonal reflecting tracts in which low-income workers have more diverse commuting origins. Panels E-H show how the differences in these values (black points) compare to hypothetical scenarios derived from Figure 4.5

We observe that most census tracts in San Francisco and New York City tend to have higher homogeneity in commuting origins for low-income workers than compared to high-income workers. It is worth noting that because the LODES commuting dataset uses national income levels to define high and low-income brackets, city-level economics (i.e., differences in cost of living between cities) is not considered. Incorporating local income distributions in San Francisco, for example, which has more representation of higher-income workers

in its population (Figure 3.1, would create more strict constraints on what it means to a high-income worker. Consequently, the total edge weight in the high-income commuting network, G^{hi}, would likely decrease, potentially impacting the distribution of commuting origins and lowering the diversity of commuting origins for affluent workers.

We note, however, that incorporating city-level economics introduces separate challenges, with each city having different methodologies for standardising and identifying income distributions. Furthermore, tract-level commuting data that can be decomposed with respect to regional income levels is scarce and would limit the reproducibility of this chapter. Although defining socioeconomic demographics using national income distributions only portrays one perspective of commuting dynamics, it allows for straightforward inter-city comparisons. Future work could build off of these results by incorporating commuting data accounts for local economic features, to construct a more comprehensive understanding of socioeconomic inequality in residential-workplace dependencies.

We now focus on extending our understanding of segregation from a residential dimension to an employment one as well. In order to accomplish this, we evaluate how socioeconomic disparities in the homogeneity of residential locations for a region relate to the employment segregation within that region. For each tract in a city, we compare a census tract's in-flow entropy value in the high-income and low-income networks:

$$\Delta H_L^{in}(ct) = H_{L,hi}^{in}(ct) - H_{L,lo}^{in}(ct), \tag{4.5}$$

where $H_{L,hi}^{in}(ct)$ captures the local in-flow entropy of census tract ct in the high-income commuting network, whereas $H_{L,lo}^{in}(ct)$ describes the in-flow entropy for i in the low-income commuting network. Thus, $\Delta H_L^{in}(ct)$ can range from -1 to 1, where negative values represent more heterogeneity of commuting origins for the low-income population. Positive values capture scenarios in which higher-income workers have more heterogeneous commuting origins than lower-income workers.

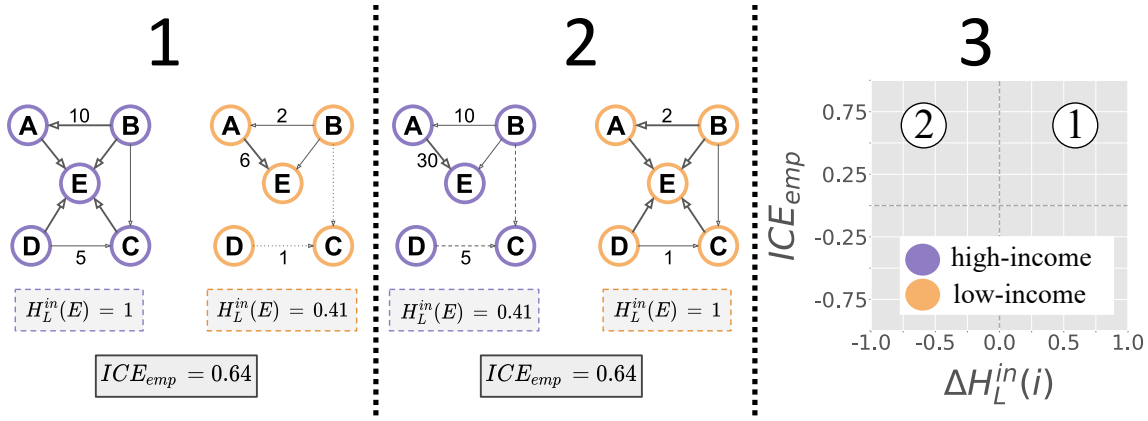

Fig. 4.5 **Toy example highlighting the distinction between socioeconomic disparities in commuting origin diversity ($\Delta \mathbf{H_L^{in}(ct)}$) and employment segregation ($\mathbf{ICE_{emp}}$.)** Panel A shows concentration of low-income commuting origins, while B captures homogeneous origins for the high-income group. Panel C conveys how the two scenarios of commuting diversity relate to employment segregation.

Comparing Local Entropies in Disaggregated Networks

For each of the 25 cities, we find significant positive correlations between the difference in local entropy values ($\Delta H_L^{in}(ct)$, Eq. 4.5) to the level of segregation in employment areas (ICE_{emp}, Eq. 4.4). These correlations are outlined in the last column of Table 4.1. We use Figure 4.5 to explain how entropy values are not strictly correlated with in-degree weights. The purple network captures the high-income commuting network and the orange network reflects that of the low-income workers. We set the total number of individuals working in node E to be 50. For scenarios 1 and 2, we can observe that the socioeconomic composition of individuals working in node E remains the same (40 high-income workers, 8 low-income workers). Thus, values of employment segregation (ICE_{emp}) are consistent throughout the examples. What changes across the scenarios are the values of H_L^{in}, which describe how evenly commuting origins are distributed across other nodes in the network. Scenario 1 (4.5, left panel) depicts a case where low-income commutes have more concentrated origins and high-income commuting origins are more diverse. Meanwhile, scenario 2 (4.5, centre panel) captures homogeneity in high-income commuting origins and heterogeneous commuting

origins for the low-income workers in the node E. We can, then, understand that $\Delta H_L^{in}(ct)$ serves to measure differences in heterogeneity of commuting origins between the high-income and low-income networks. The right panel in Figure 4.5 uses scenarios 1 and 2 to elucidate how positive correlations between ICE_{emp} and $\Delta H_L^{in}(ct)$ are non-trivial. Furthermore, while the employment segregation (ICE_{emp}) for both scenarios is the same, scenario 1 captures higher commuting origin heterogeneity for high-income workers (i.e. positive values of $\Delta H_L^{in}(ct)$), while scenario 2 does the same, but for low-income workers (i.e. negative values of $\Delta H_L^{in}(ct)$) In this manner, $\Delta H_L^{in}(ct)$ can reveal the disparity in commuting origin diversity between low-income and high-income individuals that work in a given census tract.

Modelling Hypothetical Scenarios of Disparities in Residential-Workplace Dependencies

The black scatter plots in Figure 4.4E-H illustrate the aforementioned positive correlations between socioeconomic differences in the heterogeneity of residences ($\Delta H_L^{in}(ct)$) and employment segregation ICE_{emp}, for four cities. However, what remains unclear is whether the positive correlations are an artefact of the larger representation of high-income workers in the commuting data, as seen by Figure 3.1. More explicitly, we pose the question: is the difference in commuting origins of low and high-income commuting networks ($\Delta H_L^{in}(ct)$) simply conveying the concentration of socioeconomic demographics in their workplaces? We can reflect on this question by maintaining the socioeconomic composition of employment areas and manipulating the distribution of commuting origins for either the high-income or low-income commuting networks. In this manner, we can assess whether the over-representation of high-income workers leads to inequalities in commuting origin diversity ($\Delta H_L^{in}(ct)$) consistently expressing positive correlations to employment segregation, regardless of extreme disparities in residential-workplace dependencies.

To address this issue we model two hypothetical scenarios, inspired from the examples in scenarios 1 and 2 of Figure 4.5, to emphasise how the observed positive correlations are not

a result of over-representation of high-income workers. Both scenarios elucidate how the correlations between employment segregation and local entropy differences are not always significant or positive. Each modelled scenario retains the in-degree of the disaggregated, empirical commuting networks (number of workers in a census tract), only changing how evenly a node's incoming edges are distributed across other nodes in the network.

The Low-Income Concentration scenario explores the relationship between employment segregation and inequalities in commuting origin diversity, when we enforce more diversity of commuting origins for the affluent population than the low-income group, captured by positive values of $\Delta H_L^{in}(ct)$. We construct this model, such that the residential locations for a tract's high-income workforce is uniformly sampled across all other tracts. The origins of the low-income workforce are defined by one, randomly sampled tract, producing smaller values of $H_{L,lo}^{in}(ct)$ All the while, we maintain the empirical workplace composition for each census tract, thus retaining empirical values of ICE_{emp} and aligning the Low-Income Concentration model with scenario 1 in the toy example. Meanwhile, the High-Income Concentration scenario, inspired by scenario 2 in the toy example, captures negative values of $\Delta H_L^{in}(ct)$ by simulating high diversity of low-income commuting origins and homogeneity for high-income commuting origins.

We emphasise that, in both hypothetical scenarios, the only difference to the empirical commuting network comes from the distribution of incoming edges to a tract. The results of the Low-Income and High-Income Concentration scenarios can be seen in the blue and brown scatter plots, respectively, shown in Figure 4.4. The equivalent distributions for ICE_{emp}, shown along the y-axis, demonstrate how measures of employment segregation remain consistent across the empirical and hypothetical scenarios, despite having disparities in residential-work patterns. On the other hand, the distributions along the x-axis, illustrate how $\Delta H_L^{in}(ct)$ can capture these structural inequalities by identifying tracts in which socioeconomic groups express stark differences in their dependence on commuting origins.

While the empirical scatter plot does not show signs of extreme structural disparities, as expressed by the hypothetical scenarios' scatter plots, the positive correlations indicate that areas which have more heterogeneity of residential locations for a particular socioeconomic group tend to be areas in which that socioeconomic group is more concentrated. The sign of $\Delta H_L^{in}(ct)$ specifies which demographic is segregated, with values less than zero implying low-income employment segregation. Thus, this section highlights how measuring disparities in structural diversity of disaggregated networks can expose dimensions of segregation in residential-workplace dynamics, that conventional metrics of segregation may overlook.

Local Entropy as a Potential Indicator for Commuting Times

Having evaluated how local in-entropy differences compare to employment segregation, we shift our attention to attributes of the residential dimension. That is, we focus on measuring $\Delta H_L^{out}(ct)$, to understand how differences in commuting destination heterogeneity correspond with transit commuting times:

$$\Delta H_L^{out}(ct) = H_{L,hi}^{out}(ct) - H_{L,lo}^{out}(ct), \tag{4.6}$$

which is defined similarly to Equation 4.5. In this case, negative values derived from Equation 4.6 imply that low-income workers in tract, ct, have more heterogeneous commuting destinations that high-income works in ct. Positive values indicate that low-income workers in ct have more concentrated commuting destinations than their high-income counterparts.

We calculate empirical transit commuting times, $T_{transit}^{comm}(ct)$, by computing an average of transit times from ct to all the destinations to which workers in ct commute, weighted by the number of commuters travelling to each destination. Table 4.2 reveals that 18 of the 25 cities have significant positive correlations between commuting destination inequalities and average transit commuting times. Positive correlations indicate that neighbourhoods, in which low-income commuters have more concentrated destinations than high-income

Table 4.2 Comparing socioeconomic differences in heterogeneity of commuting destinations to commuting transit times. The first column reflects the city of analysis, while the second column defines the average commuting time across all census tracts in a given city. The last column reveals the relationship between commuting destination diversity and the average commuting transit time, with respect to a given census tract. Positive Pearson correlation coefficients imply that in neighbourhoods where high-income residents have larger diversity of workplaces, commuting times tend to be longer. Asterisks indicate the significance of the correlation coefficients in the last column, with non-bolded entries being non-significant.

City	$\overline{T^{comm}_{transit}(ct)}$	$\Delta H^{out}_{L}(ct)\ vs. T^{comm}_{transit}(ct)$
San Francisco	30.326	-0.051
New Orleans	54.757	-0.05
Dallas	92.822	-0.016
Albuquerque	83.614	0.034
Jacksonville	103.541	0.05
Boston	34.649	0.111
Fort Worth	142.325	**0.146****
Charleston	108.557	0.164
Greenville	140.364	**0.232***
New York City	80.712	**0.237*****
Memphis	111.954	**0.27*****
Indiannapolis	68.408	**0.275*****
Milwaukee	31.242	**0.321*****
Philadelphia	34.473	**0.34*****
Cincinnati	53.127	**0.383*****
Las Vegas	111.746	**0.441*****
Gainesville	115.301	**0.453*****
Houston	148.409	**0.454*****
Chicago	64.757	**0.483*****
Detroit	130.127	**0.519*****
Bridgeport	138.524	**0.549*****
Hartford	77.785	**0.586*****
Cleveland	61.901	**0.586*****
Kansas City	126.180	**0.656*****
Columbus	112.148	**0.673*****

*p<0.05; **p<0.01; ***p<0.001

commuters, tend to have longer mean commutes via transit, under the assumption that all commuters are using transit. We do not account for commutes by car as we want to focus on how transit infrastructure supports job accessibility, with respect to structural inequalities in residential-workplace dependencies. However, the following sections, which analyse more complex features of residential disparities, integrate driving times into the analyses to act as a baseline for comparison.

By incorporating transit times into our measure of commuting destination inequality, we can understand how disadvantage can be exhibited in different ways, depending on the socioeconomic group. That is, we observe smaller commuting times for neighbourhoods in which low-income workers have a larger diversity of commuting destinations. This can be indicative of transit systems that are facilitating commutes for low-income neighbourhoods. Alternatively, it can be a signal of residential constraints that force low-income individuals to live in areas that have diverse employment opportunities and transit service. Chapter 6 analyses transit inequality with respect to built form features, allowing us to disentangle issues surrounding the self-selection bias. Furthermore, the following sections apply a more nuanced analysis to residential-employment disparities by integrating sociological concepts into characterising residential features and geospatial perspectives for defining employment locations. Specifically, we analyse how transit serves particular employment hotspots for those vulnerable to housing insecurity.

4.4 Measuring Housing Insecurity

The previous sections highlighted how structural features of socioeconomic commuting landscapes can reveal disparities that are obscured in conventional segregation metrics. However, the goal of this thesis is to highlight the importance of analysing transit inequality in the context of various urban mechanisms, with respect to structural and experiential inequalities. Thus, having introduced a structural framework for analysing spatial mismatch,

we shift to a more humanistic approach. That is, we focus on understanding commuting patterns of neighbourhoods, based on their vulnerability to insecurity. Numerous studies have highlighted how housing insecurity can negatively impact how one experiences urban life [71, 73, 151]. Thus, we analyse how neighbourhoods that are especially vulnerable to housing insecurity may suffer from poor transit links to employment opportunities. In order to accomplish this we, first, have to measure housing insecurity. Typically, urban studies use a single feature, such as rent burden or eviction rates, to act as a proxy for housing insecurity [124, 33]. To develop a more comprehensive estimate of housing insecurity, we leverage sociological research that underscores the multifaceted nature of housing insecurity. Accordingly, in this section, we detail our multidimensional approach for defining census tracts based on their level of vulnerability to housing insecurity. We apply spectral clustering to the data introduced in Section 3.1.2 and then evaluate the validity of the identified clusters using a range of sociodemographic indicators. In doing so, we establish a comprehensive measure of vulnerability to housing insecurity. We incorporate this variable throughout the rest of the chapter to understand how transit systems facilitate employment access for neighbourhoods particularly vulnerable to housing insecurity. We note that we limit our analysis to 20 US cities, due to data availability of eviction rates. These 20 cities still represent different geographical areas across the US, while also capturing hetergeneous housing, transit, and employment features.

4.4.1 Defining Vulnerable Housing Regions: A Clustering Approach

In order to examine the state of housing in various US cities, we adopt an unsupervised learning approach to define three different housing categories: most vulnerable, mildly vulnerable, and less vulnerable. We apply spectral clustering on the housing features that we outline in Section 3.1.2. Details about these features are summarised in Table 3.1. We begin preprocessing the housing features by replacing all negative values with placeholder

values (NaNs). Typically, negative values appear in the American Community Survey to indicate insufficient data, due to sample sizes, for example. Then, we retain the data from census tracts that have a population greater than zero and that have data on at least one of the housing affordability features (rent burden, mortgage, or housing stock). In doing so, we only consider census tracts in which individuals reside. We standardise the data such that, for a given city, each housing feature has a mean value of zero and a standard deviation of one. This is done to ensure that features with larger raw values (i.e. mortgage) will not have a greater influence on the results than features with smaller values (i.e. evictions per capita).

We implement spectral clustering to measure levels of housing insecurity within different urban areas. Spectral clustering is a particularly useful approach for clustering high dimensional data, as it makes no assumptions about the shapes of clusters, whereas approaches such as K-Means assumes that data clusters are spherical. Alternative approaches for clustering such as Expectation-Maximisation and K-Means are extremely sensitive to initialisation, requiring numerous iterations to derive high quality clusters.

The process of spectral clustering can be broken into three steps. First, one must extract an affinity matrix, A, from a graph that is built using the data points. This entails constructing an $|N| \times |N|$ similarity matrix where $|N|$ is the number of records, or census tracts. This can be done using a radial basis function, which applies a Gaussian function to the Euclidean distances between each data point, in which distance reflects the similarity of housing features. Since each node in our graph represents the housing characteristics of a particular census tract, we tend to have high values of $|N|$. Thus, we use a K-Nearest Neighbours similarity graph to create a more sparse matrix, in which we consider 10 of the nearest neighbours to each tract.

The next step is spectral embedding, which leverages properties of the Graph Laplacian to represent data points in a low-dimensional space. The diagonal matrix of degrees, D, is necessary to calculate the Laplacian Matrix, L, for an affinity matrix, A, which was defined

in the first step. D is simply an identity matrix, where each diagonal value is the row-wise summation of edge weights for the respective node in the graph. The Laplacian Matrix can, then, be defined in numerous ways. For the approach of identifying different housing vulnerability clusters, we use the Graph Laplacian, which is calculated by subtracting the diagonal matrix, D, from the affinity matrix, A. In this manner, all values on the diagonal of L capture the weighted degree of each node, while each cell l_{ij} is the negative edge weight, if an edge exists, or zero otherwise. One of many fascinating properties of the Laplacian matrix is that its rows and columns sum to zero. Moreover, the eigenvalues of a graph's Laplacian matrix informs structural properties of the graph, namely the number of components it has. If the first x eigenvalues are zero, this indicates that the respective graph built from A has x connected components. Eigenvalues that are near-zero indicate a loosely-connected graph, which couldn't be split into separate components with minimal cuts. Conversely, a very dense graph would have eigenvalues near $|N|$. Figure A.1 in the Appendix shows the eigenvalues, and respective spectral gap, for the Graph Laplacian in each city.

The last step of the algorithm involves applying a classical clustering algorithm, typically K-means, to partition the embedded data into respective clusters. It becomes clear, then, that eigenvalues can be leveraged to identify the number of clusters into which the data, embedded into a graph, can be clustered.

Since we are considering a variety of urban areas, the number of K-Means clusters that are appropriate for each city's housing characteristics differs. Thus, the spectral gap for each city (Fig A.1) informs the number of clusters for which the K-Means algorithm is applied. To extract meaning from each cluster, we rank the housing clusters based on the mean values of their housing features, where larger values denote worse housing conditions. The group that has higher ranks across the housing dimensions is deemed most vulnerable to housing insecurity. Then, to address the varying number of clusters across cities, we partition the ranked clusters into three housing demographic groups, in which each group

contains a similar number of clusters. These final three groups reflect the tracts that are the most vulnerable, mildly vulnerable, and less vulnerable to housing insecurity, in the context of each city. In cases where the number of clusters for a city is not evenly divisible by three (to equally partition into the three housing demographics), we assign the remainder number of cluster to the mildly vulnerable demographic. Figure 4.6 illustrates the housing features for the resulting housing demographics in Milwaukee (Figure 4.6A) and Cleveland (Figure 4.6B). Each column represents the final housing demographic groups (Less, Mildly, and Most Vulnerable), while each row illustrates the housing characteristics that were used to define the housing demographic with respect to a city. In this manner, each cell can be defined by the following equation:

$$HV_{f,h} = \frac{1}{|\mathbf{CT}_h|} \sum_{t}^{T_h} HF(f,t) \tag{4.7}$$

where $HF(f,t)$ refers to the value of a housing feature, f, for a tract, t, which belongs to a housing demographic, h. $\mathbf{CT}_h$ indicates the set of census tracts in a housing demographic, h. Accordingly, for a given housing feature (row), f, and housing demographic (column), h, a cell's value ($HV_{f,h}$) is defined by averaging the housing feature for each census tract in $\mathbf{CT}_h$. Then, we apply row-wise normalisation to compare the differences in demographics within each city. The upper left cell of a heat map, for example, conveys the mean percentage of severely rent burdened households across all census tracts in the less vulnerable housing cluster.

Figure 4.6 portrays how the housing features for Milwaukee and Cleveland map to their final housing demographics. That is, the rows represent the features that were considered when applying the clustering framework to end up with the final, shown by the columns. Thus, we can see the average characteristics of the three housing demographics, with respect to each considered feature. Figure 4.6A shows how the housing demographics (columns) are clearly distinguishable, in terms of having consistent levels of housing insecurity across most features,

in the context of Milwaukee. In these types of cities, using a single housing feature as a proxy for housing insecurity could be an adequate estimation. However, the housing characteristics in Cleveland (Figure 4.6B) emphasise the need for a multidimensional approach to defining housing insecurity, illustrating how neighbourhoods may be vulnerable to various forms of housing insecurity, ultimately underscoring the complexities of housing conditions. For instance, the less vulnerable census tracts in Cleveland have higher insecurity than the most vulnerable tracts, in terms of housing stock within the city. Moreover, the mildly vulnerable tracts have the higher insecurity when considering housing stock, and overcrowding in the housing stability dimension, represented by the bottom-most group of heat maps. When considering each housing feature, however, the most vulnerable tracts have more occurrences of the highest levels of insecurity. The intricacies of housing conditions, then, becomes clear, with Figure 4.6 emphasising the importance of considering the multidimensional nature of housing.

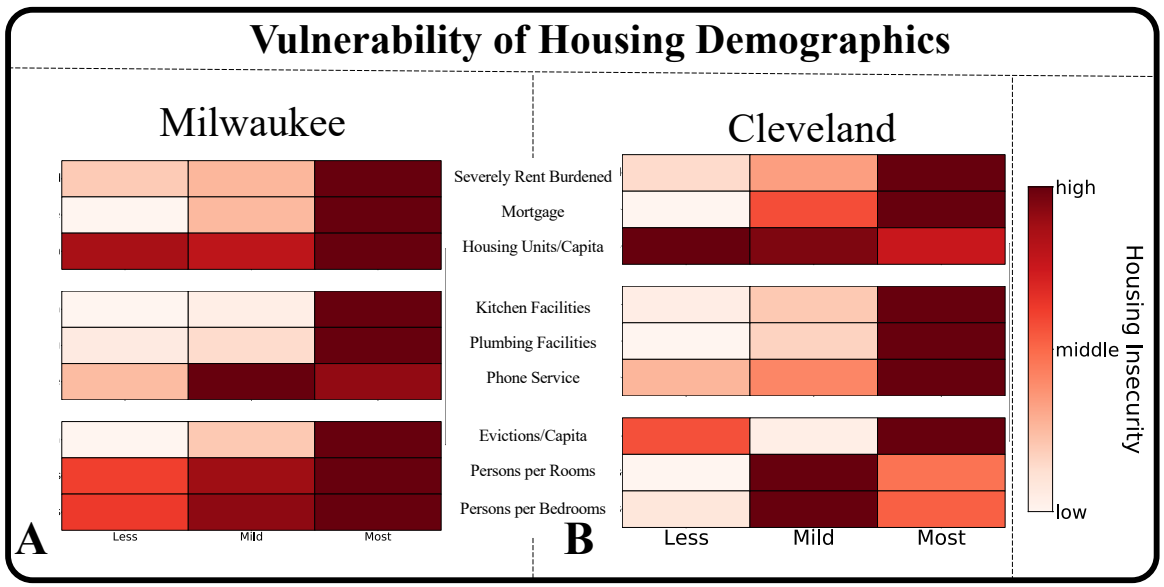

Fig. 4.6 **Incorporating multiple dimensions of housing insecurity reveal how some cities have clear delineations between housing demographics (Milwaukee), while others unveil less straightforward dynamics across dimensions in the housing landscape (Cleveland).** Each panel visualises the average housing characteristics for each housing demographic. From top to bottom, each heat map captures housing insecurity levels for the dimensions of affordability, safety/quality, and stability, respectively. Columns are the housing demographics, while rows are the housing features. Darker hues of red indicate higher levels of housing insecurity. On one hand, Milwaukee (Panel A) reflects a case where the most vulnerable census tracts consistently have the highest rates of housing insecurity. On the other hand, Cleveland (Panel B) presents a convincing case against using a single housing feature as a proxy for housing insecurity.

Ultimately, we partition the clusters within each city into three groups, based on their housing characteristics. These final three groups reflect the most vulnerable, the mildly vulnerable, and the less vulnerable census tracts for each urban area, with regards to housing insecurity. Figure A.2 in the Appendix shows the population distribution across each of the three housing demographics. While each of the housing demographics consists of a similar number of clusters, it be comes clear that this is not necessarily indicative of an even population distribution. Although cities such as Bridgeport and Columbus have less than ten percent of the population living in neighbourhoods that are the most vulnerable to housing

insecurity, it is still important to consider how urban infrastructure serves these demographics to ensure that they do note face exacerbated levels of social exclusion [176].

4.4.2 Socio-demographic Characteristics of Housing Groups

We validate our clustering approach by exploring how the housing demographics we defined in the previous section relate to socioeconomic variables associated with employment, wealth, and commuting. Specifically, with respect to each tract, we gather census data capturing (a) the percentage holding a professional degree, (b) the median household income, (c) unemployment rates, (d) poverty rates, (e) the percentage commuting using public transportation, and (f) the percentage of the tract with a commute longer than an hour. Figure 4.7 compares these socioeconomic variables across the tracts in the less vulnerable and most vulnerable housing groups for Bridgeport, Albuquerque, Philadelphia, Milwaukee, Cleveland, and Dallas. Meanwhile, Tables A.3 to A.8 in the Appendix list the sociodemographic characteristics of the most and less vulnerable housing demographics for all 20 cities. We choose these 6 cities as they span a range of geographic and public transport characteristics.

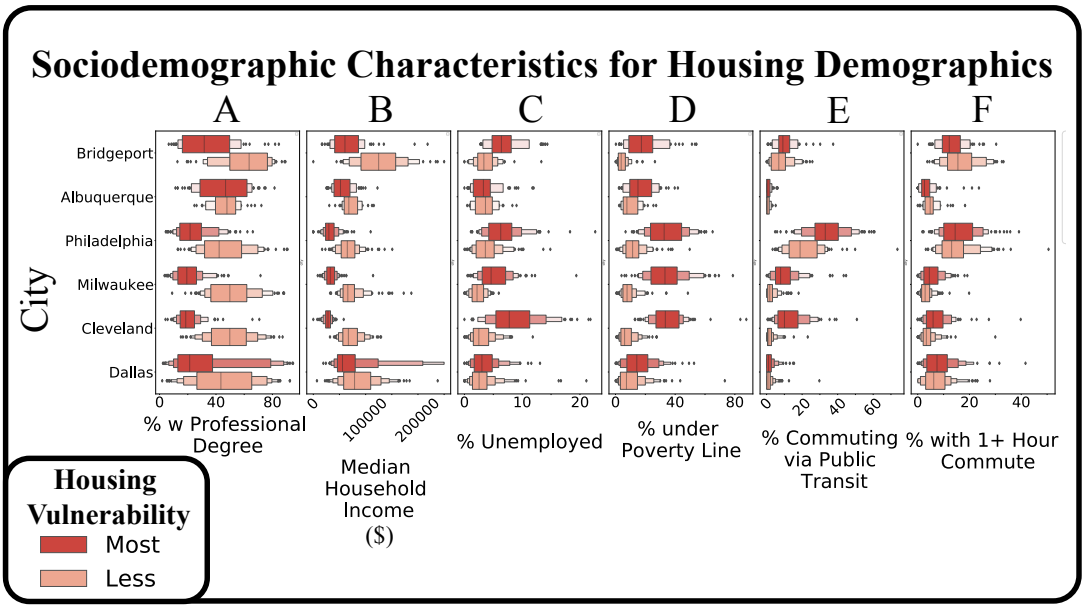

Fig. 4.7 **Neighbourhoods that are most vulnerable to housing insecurity tend to have the lower rates of educational attainment (Panel A), lower median households incomes (B), higher unemployment (Panel C), higher poverty (Panel D), more dependence on transit for commuting (Panel E) and higher percentages of commutes over an hour long (Panel F).** Dark and light red box plots reflect sociodemographic characteristics for the most and less vulnerable housing demographic, respectively.

Panel A in Figure 4.7 illustrates how census tracts that are the least vulnerable to housing insecurity tend to have a higher fraction of its residents holding a professional degree. This considers Associate's, Bachelor's, Master's, Professional schools, and Doctorate degrees as qualifications for estimating the level of educational attainment in a region. Similarly, Panel B highlights how the most vulnerable neighbourhoods have lower median household incomes than the census tracts that are less vulnerable. On the other hand, Panels C and D consider deprivation indicators such as unemployment and poverty. In both these cases we see associations between housing vulnerability and higher rates of unemployment and poverty, for all of the 6 cities considered. When we analyse commuting characteristics of housing demographics, we generally see more dependence on transit systems for the most vulnerable neighbourhoods. We note that Albuquerque and Dallas have much lower

percentages of transit commutes than the other 4 cities. Finally, when considering commute times, irrespective of commute mode, we observe less distinct disparities between the housing demographics. We hypothesise that this has to do with spatial organisation of residential and employment areas for each housing demographic. That is individuals organise their residential and employment locations to be close to one another, an idea that is consistent in urban commuting literature [125]. While, at first glance, it does not appear that significant disparities exist when comparing transit commuting features of various housing demographics, we leverage detailed transport modelling tools to better understand the spatial system-facilitated employment access for individuals living in neighbourhoods that are most vulnerable to housing insecurity. In doing so, we question whether the lack of transit commuting disparities is reflective of effective transit service, or is an artefact of segregated residential and employment landscapes. Ultimately, this section highlights the pertinence of understanding the nuances in housing conditions to determine residential characteristics. In the next section, we define the efficiency of cities' transport infrastructure to understand how urban services connect different parts of a city. Doing so builds the foundation for analysing the spatial relationship between housing and employment landscapes, as well as how public transit intersects with commuting behaviour.

4.5 Assessing Transit Service in the US

In the previous section, we introduced a clustering framework for identifying census tracts that are vulnerable to housing insecurity. To begin exploring whether indirect policies, such as transportation accessibility, pose further obstacles to individuals in precarious housing situations, we draw upon GTFS feeds to characterise public transportation systems. We begin by defining a metric for public transportation efficiency, by comparing transit and driving times of various journeys. With this metric, we proceed to define cities based on their transit characteristics, categorising their public infrastructure as highly efficient, adequately

efficient or inefficient public transportation infrastructure. Furthermore, we highlight how transit inequalities may be overlooked by neglecting to account for the spatial organisation of the housing landscape. This section aims to elucidate how transit systems serve housing demographics by measuring efficiency between and within cities and exploring residential attributes associated with proximity to transport infrastructure.

4.5.1 Comparing Transit Efficiency across the US

We begin our analysis by examining the current landscape of public transportation across all 20 cities, with a particular emphasis on how effectively transit serves each city. To investigate differences in efficiency of transportation systems, we build a transit-pedestrian network using `UrbanAccess` and the GTFS feeds outlined in Section 3.4. Each city's network consists of transit nodes and pedestrian nodes. Edges linking transit nodes reflect transit lines, while edges between pedestrian nodes represent paths in the road network, which is informed by `OpenStreetMap`. Building off these two networks, `UrbanAccess` connects the transit and pedestrian networks by mapping each transit node to the closest pedestrian node. Accordingly, the travel time via public transit from any two points in an urban area can be calculated as a series of transit and/or pedestrian paths. It should be noted that `UrbanAccess` assumes a walking speed of 3 miles per hour (4.83 kilometres per hour) to calculate pedestrian travel times. With a given city's transit network, we can calculate the time it would take to travel using from one census tract to another ($J_{transit}(ct_1, ct_2)$) in a given day, during a given time frame. We construct a network for each of the 20 North American cities from 06:30AM to 10:30AM on Mondays in January 2020. This window of time captures the bulk of commutes during rush hour [41]. Transit time, alone, is not particularly informative when comparing cities of different sizes, as travel time is a function of distance and road networks. Thus, we define the efficiency of a city's transportation system by measuring how much longer a trip takes using public transit, compared to driving, with $J_{driving}(ct_1, ct_2)$ reflecting the time it

takes to drive between two census tracts. We refer to this concept as travel impedance. The impedance, $\mathscr{Z}$, from a census tract ct_1 to a tract ct_2 can be formally defined as:

$$\mathscr{Z}(ct_1, ct_2) = \frac{J_{transit}(ct_1, ct_2)}{J_{driving}(ct_1, ct_2)} \tag{4.8}$$

`Openrouteservice` provides the data for estimating the time it takes to travel from one census tract to another [93] and is discussed at length in chapter 3.4.3. A travel impedance of one implies that driving between two points takes as long as using public transit during the specified day and time range. A travel impedance, t, greater than one suggests that transit trips take longer than driving trips by a factor of t. To compare the efficiency of public transportation systems across cities, we define the efficiency of a city's transport system as the mean efficiency for all potential commutes (all possible pairs of census tract origins and destinations in a city). Mathematically, this is calculated by averaging the travel impedance between each pair of census tracts, (ct_1, ct_2), where **CT** reflects the set of census tracts in a city, c:

$$\text{eff}_c = \frac{\sum_{ct_1, ct_2}^{ct_1, ct_2 \in \mathbf{CT}} \mathscr{Z}(ct_1, ct_2)}{|\mathbf{CT}|^2} \tag{4.9}$$

Figure 4.8 captures the efficiency of transport infrastructure for each of the cities we analyse, calculated using Equation 4.9. Darker hues of green reflect more efficient systems, while cities with whiter hues reflect regions where using transit takes significantly longer than driving. The efficiency values range from 1.896 (Milwaukee) to 7.261 (Fort Worth), with a median of 4.95 and a mean of 4.64 across all cities. For further details, Table 4.3 lists the corresponding transit efficiency for each city. Notably, Milwaukee is the only city with a transit system that, on average, serves its residents in less than double the time it takes to drive. On the other hand, the public transportation in Fort Worth, Bridgeport, and Greenville generally takes more than six times as long as driving.

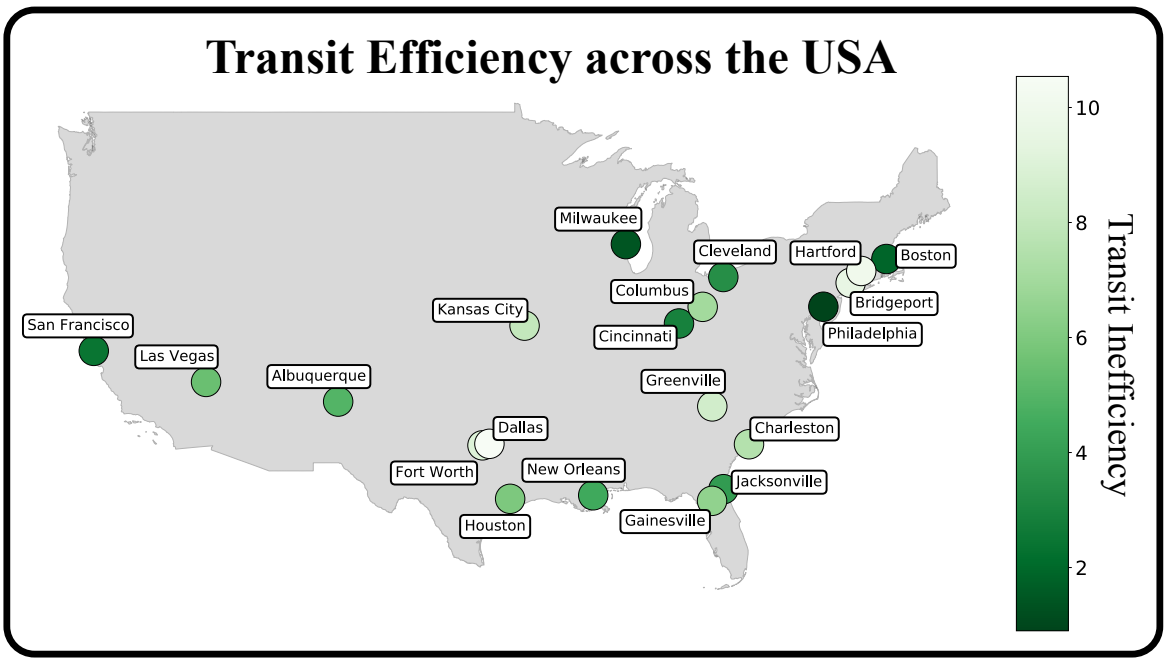

Fig. 4.8 **Map of the United States, highlight the 20 cities in our analysis and their corresponding levels of transit efficiency**. Darker hues of green indicate more efficient public transport systems.

The above results demonstrate that using public transit in US cities typically results in longer travel times than driving would. Moreover, Figure 4.9A illustrates how cities with less efficient transit systems tend to have higher rates of car ownership, with a Pearson correlation coefficient of 0.66 between the two variables. These results are consistent with research that reveals how the quality and reliability of transport infrastructure impacts the frequency with which residents use public transit [300, 36].

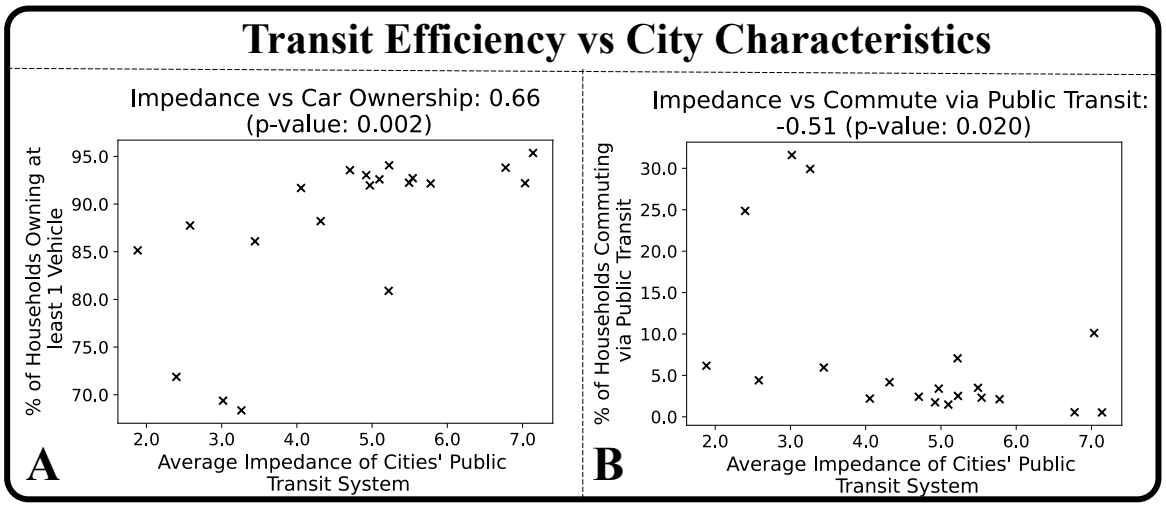

Fig. 4.9 **When consider 20 US cities, transit efficiency is associated with less car ownership and more transit dependence for commuting.** Panel A shows a positive relationship between inefficient transit systems and the percentage of households with one or more vehicles. Panel B highlights how more efficient transit systems tend to have a higher proportion of households that commute using public transport. The Pearson correlation coefficients for car dependence and transit commutes are 0.66 and -0.51, respectively.

Moreover, Figure 4.9B shows the negative correlation (Pearson correlation coefficient of -0.51) between transit inefficiency and the percentage of the population that uses transit for commuting. Thus, we observe that less efficient transit systems are associated with a higher dependence on cars and lower levels of transit commutes. In this manner, we can reflect on how less efficient transit systems contribute to the burden of insecure housing, as the financial cost of cars depletes resources that could be otherwise invested in savings or spent on higher quality housing, a phenomena referred to as forced car ownership [64, 192]. Moreover, choosing to commute using inefficient transit is costly from a time perspective [177].

4.5.2 Incorporating Distance into Measures of Transit Efficiency

Our analysis of the state of public transportation in the USA has been at the city level, allowing us to compare cities to one another. However, the cities in this analysis vary

Table 4.3 Statistical properties of transit efficiency and area (km^2) for 20 US cities, ranked by decreasing efficiency.

City	Transit Efficiency			Area (km^2)
	Mean	*Median*	*Std. Dev.*	
Milwaukee	1.896	1.826	0.318	625.435
Philadelphia	2.398	2.385	0.251	347.782
Cincinnati	2.580	2.404	0.590	1050.000
Boston	3.361	3.351	0.455	150.863
Cleveland	3.459	3.398	0.464	1184.094
San Francisco	3.462	3.428	0.291	121.478
Jacksonville	4.150	3.949	0.888	1975.201
Hartford	4.355	4.341	0.555	1903.544
Dallas	4.741	4.480	0.878	2259.440
Las Vegas	5.117	4.811	1.754	20439.277
Albuquerque	5.122	5.024	0.868	3007.645
Houston	5.238	4.793	1.189	9350.383
New Orleans	5.468	5.444	0.663	438.831
Gainesville	5.492	4.973	1.391	2267.635
Columbus	5.540	5.083	1.310	3833.131
Charleston	5.750	5.553	1.207	2377.463
Kansas City	5.820	5.173	1.700	5487.102
Greenville	6.776	6.557	0.969	2033.481
Bridgeport	7.239	6.998	1.357	1618.659
Fort Worth	7.261	7.165	1.158	2236.864

largely in size, ranging from 121 km^2 to 20,439 km^2. It should be acknowledged that there is a possibility that cities which err on the side of transit inefficiency may have effective transportation, but are larger, therefore obscuring the density and quality of the transit system. Table 4.3 shows how smaller cities are not necessarily the cities with the more efficient transit systems, with New Orleans having an impedance of 5.468. Similarly, Table 4.3 contains large cities that are both efficient and inefficient, indicating that region size may not be a confounding factor. To further address this potential issue, we analyse travel impedance as a function of distance. We accomplish this by creating 6 classes of transit journeys, with each category defined by how long a journey between two census tracts is. We refer to each class as a *distance group*.

We map each pair of census tracts to its respective distance group, based on how far the tracts are from one another. Then, for each distance group, we find the average impedance for all trips within that group. In doing so, we identify three signatures of transit efficiency with respect to trip distance, which correspond with the overall transit quality in cities. This is highlighted in Figure 4.10, which uses Cleveland, Albuquerque, and Bridgeport to exemplify each of the discovered trends for the most, moderately, and least efficient transport systems respectively.

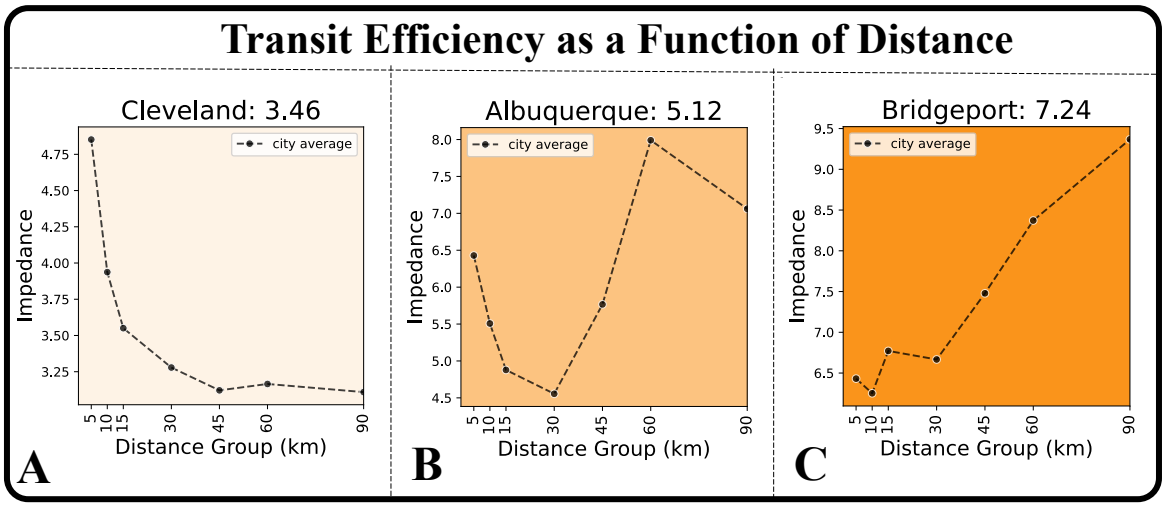

Fig. 4.10 **Transit efficiency can be understood with respect to different trip distances**. Each panel shows one of the three signatures we identify when analysing transit efficiency as a function of trip distance. Cleveland (Panel A) reflects efficient systems in which travel impedance decreases as trip distance increases. Albuquerque (Panel B) is a case of moderately efficient transit service, in which the relationship between travel impedance and trip distance switches from negative to positive at a given trip distance threshold. Panel C uses Bridgeport as an instance of inefficient transit systems, with increasing travel impedance as trip distance increases.

The first signature we observe is for the cities with efficient transit systems (Figure 4.10A): Philadelphia, Boston, San Francisco, Cleveland, Jacksonville, Hartford, and Dallas. In these cities, the travel impedance of longer journeys (30 km or more) is, on average, lower than shorter trips, indicating that the transit system is generally more efficient for trips of larger distances. These cities all tend to be more efficient, with the mean efficiency of all cities

following this signature being 3.704 and the median being 3.462. The transit efficiency of these cities never exceeds 5, conveying that transit times in these cities are typically upper-bounded at 5 times as long as driving times. Another signature we unveil is for cities with inefficient public transit (Figure 4.10C): Gainesville, Kansas City, Greenville, Fort Worth, and Bridgeport. These cities exhibit an increasing travel impedance as the distance of trips increases, implying public transit becomes less effective than driving when journey distances increase. The mean and median efficiencies for cities in this signature are 6.518 and 6.776, respectively.

The final signature we identify is a combination of the first two signatures and is found in cities that have moderately inefficient public transportation (Figure 4.10B): Milwaukee, Cincinnati, Albuquerque, Las Vegas, Charleston, Houston, Columbus, and New Orleans. These cities reveal characteristics of the first signature until a particular distance threshold. That is, travel impedance decreases as trip distances increase for shorter length trips in that region. Trips that are longer than the distance threshold follow the behaviour of the second signature, displaying increasing travel impedance with trip distance. Cities in the signature have a mean transit efficiency of 4.589 and a median of 5.180. For a more comprehensive look at the results for all 20 cities, we refer readers to A.3 in the Appendix.

4.5.3 Exploring the Intersection of the Housing and Transit Landscape

Prior to understanding the role transit plays in constraining or facilitating mobility for various housing groups, we analyse how housing demographics are spatially distributed around the public transportation infrastructure. Specifically, we consider how close each tract is to the transit system's centre of mass, which we refer to as the transit core. We use the centre of mass as a proxy for the location that has the most access to different transit stops in the region. We define the transit core by calculating the average longitude and latitude for all transit stops in the system, weighting for frequency of trips through each transit stop.

Figure 4.11 illustrates the disparity in residential locations for the most and less vulnerable housing groups. That is, in the vast majority of cities, the tracts that are most vulnerable to housing insecurity tend to be situated closer to the *transit core*, with the exceptions being San Francisco, Boston, Dallas, and Greenville. Cities are ordered in increasing transit efficiency, with Milwaukee having the most efficient transportation system and Fort Worth being characterised by the least efficient transit infrastructure. The mean distances to each city's transit core for all housing demographics are listed in Table A.9 in the Appendix.

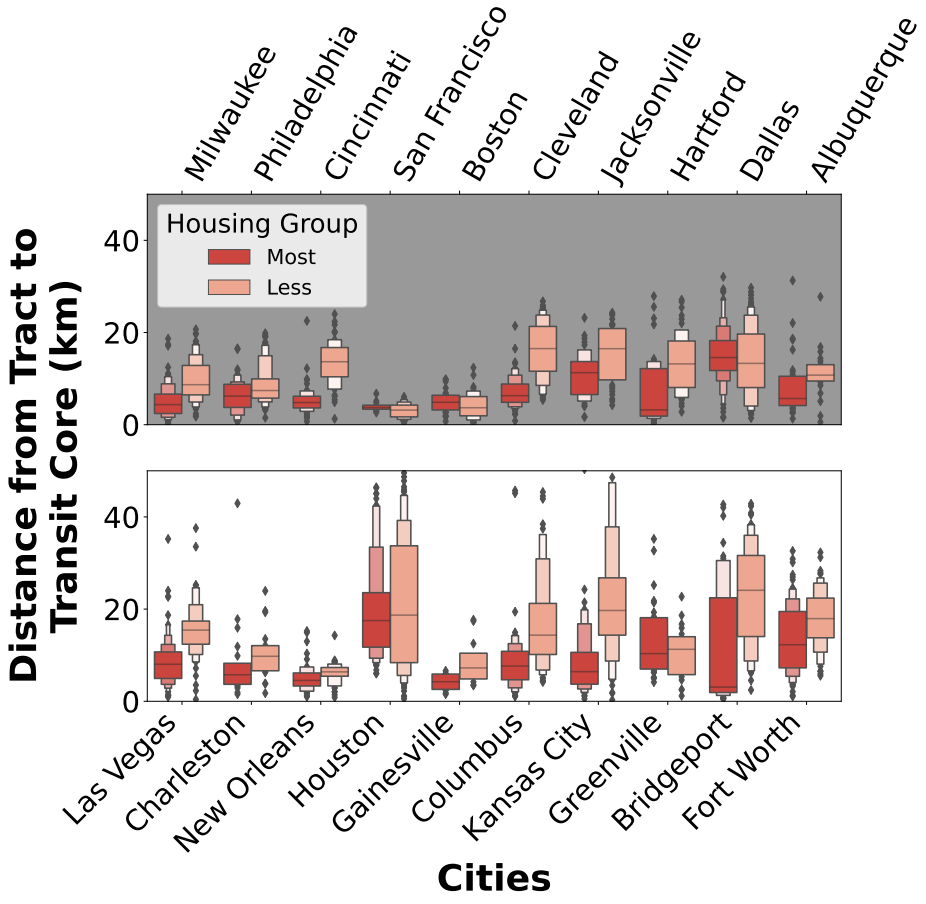

Fig. 4.11 **Neighbourhoods that are most vulnerable to housing insecurity tend to live closer to the urban core than their less vulnerable counterparts.** Each pair of box plots examine differences between housing demographics, based on their proximity to the transit system's centre of mass. The y-axis is the distance in kilometres, while the x-axis reflects the distribution for each city. Lighter red represent the less vulnerable housing demographic while dark red symbolises tracts that are most vulnerable to housing insecurity.

In this section, we have introduced the metrics of travel impedance and transit efficiency to compare cities to one another, based on the characteristics of their public transportation systems. This analysis paints a clearer picture of how cities with less efficient transport infrastructure also have higher rates of car ownership. Furthermore, cities with inefficient transit systems provide less efficient service to areas further away, which could contribute to additional obstacles for vulnerable demographics living in the urban periphery. However, we observe how the most vulnerable housing demographics tend to live closer to the transit core than their less vulnerable counterparts. This trend is consistent with transportation poverty studies that identify less privileged residents living closer to urban infrastructure and services [5]. These findings emphasise the importance of space in the housing landscape, as urban constraints may force vulnerable populations to live in the core of a city. Consequently, these demographic groups will appear to have better access to resources, yet the lack of choice is obfuscated. This concept is further explored in Chapter 6 of this thesis, in which we analyse transit inequalities with respect to different physical urban features. The next section focuses on incorporating more complex spatial relationships. In doing so, we evaluate system-facilitated transit accessibility for vulnerable housing demographics, with respect to employment locations that are associated with better economic outcomes.

4.6 Public Transit and Job Accessibility

This section aims to shine light on how urban infrastructure may impose additional constraints on individuals who are already encumbered by housing insecurity. In doing so, we combine the housing demographics defined in Section 4.4.1, the transit characteristics outlined in Section 4.5, and employment data defined in Section 3.2.1 to illustrate how the combination of employment landscape, housing market, and transit system limits employment accessibility using transit. To accomplish this, we investigate how public transit contributes to the

burden that individuals facing housing insecurity experience, presenting additional barriers to accessing better employment opportunities.

We begin analysing the intersection of the housing and employment landscape by assessing whether areas with similar workforce characteristics express a notion of spatial proximity. Then, we identify particular census tracts that employ an unusually high concentration of its workforce from a particular housing demographic. Specifically, we use spatial autocorrelation on a global and local level to define census tracts based on the labour force that works there. This builds off of the previous sections, as we define census tracts using the residential characteristics of their employment composition to understand how employment and housing landscapes intersect. To highlight the integral role transport infrastructure plays in job accessibility, we explore how commuting times change when individuals in vulnerable housing areas start working in employment areas that are associated with better housing conditions, which is associated with higher earnings [13].

4.6.1 Defining Housing and Employment Hotspots

Numerous works have underscored the important role transit plays in connecting residential and employment locations [108, 177, 282, 82]. In this section, we argue that, of the urban and economic forces that contribute to segregated experiences, transport infrastructure should, at the very least, not add to such constraints. It should, ideally, provide sustainable alternatives to accessing better opportunities. Thus, we identify residential and employment hotspots for each housing demographic to understand how transit connects these clusters.

We apply exploratory spatial data analysis techniques on a global and a local scale. The Moran's I statistic, a common method for assessing global spatial autocorrelation, tests whether spatial clustering of a specified metric exists in a geographic data set. The metric we consider is the the workforce composition of a census tract, defined by the percentage of a workforce that is made up by a particular housing demographic (i.e. the percentage of a

workers who commute from neighbourhoods that are most vulnerable to housing insecurity, with respect to the total number of workers in that employment tract). The extent of clustering is highly dependent on a spatial weights matrix, which characterises the spatial proximity between two areas in a city. Moreover, each census tract can be defined by how much the tract's employment of a housing group deviates from the mean value, across all tracts. Thus, by combining the tract-level data with the spatial weights matrix, one can derive the degree of spatial clustering in a city, for the workforce composition of a housing demographic. This provides some insight as to whether clustering is a spatial pattern for the entire city. However, it does not define where the high rates of employment occur in the city.

To identify these clusters, we use Local Indicators of Spatial Association (LISA) to analyse spatial autocorrelation on a local level. Thus, we can determine the census tracts that have high values of employment for a housing vulnerability group, that are also surrounded by tracts with similarly high employment rates for that demographic. In this manner, LISAs can pinpoint, what we refer to as, *employment hotspots*, which indicate regions that employ a high percentage of individuals that live in a particular housing demographic. Both the local and global analysis are inferential statistics, comparing the empirical data to their randomized counterparts, in which the empirical values are maintained, but are assigned to random locations to determine the significance of spatial clustering in the data.

Global spatial autocorrelation, in this context, assesses whether a housing demographic relies on particular areas of a city for job opportunities. To accomplish this, we define census tracts by the percentage of individuals working there that belong to a particular housing group. Then, we compare the employment rates of each census tract to its neighbours, characterising neighbours using Queen contiguity, in which neighbouring tracts are those that share a vertex with the focal tract. Table 4.4 lists the Moran's I statistic, with respect to employment rates for each housing demographic, in which bold cells reflect statistically significant values. For example, the second column conveys the extent of spatial concentration, in regard to how

many individuals from less vulnerable residential areas make up the workforce composition. Meanwhile, the last column captures the spatial autocorrelation of areas that employ similar rates of individuals from the most vulnerable census tracts. Table 4.4 is sorted by Moran's I value for the most vulnerable demographic, showing the notable role that space plays when considering the worker composition of individuals who live in neighbourhoods that are highly vulnerable to housing insecurity.

Table 4.4 Moran's I statistic, conveying the global spatial autocorrelation across census tracts based on the fraction of individuals from each housing demographic that work there.

City	Moran's *I* for Housing Demographics' Workplaces		
	Less	Mild	Most
Gainesville	**0.219****	0.094	**0.115***
New Orleans	**0.192*****	0.014	**0.132****
Greenville	**0.250*****	**0.516*****	**0.345*****
Fort Worth	**0.381*****	**0.388*****	**0.393*****
Dallas	**0.329*****	**0.303*****	**0.406*****
Boston	**0.348*****	**0.540*****	**0.435*****
Las Vegas	**0.429*****	**0.253*****	**0.437*****
Charleston	**0.368*****	0.052	**0.453*****
San Francisco	**0.543*****	**0.654*****	**0.499*****
Albuquerque	**0.203****	**0.273*****	**0.506*****
Cleveland	**0.587*****	**0.249*****	**0.552*****
Hartford	**0.287*****	**0.385*****	**0.558*****
Cincinnati	**0.424*****	**0.582*****	**0.560*****
Columbus	**0.467*****	**0.644*****	**0.615*****
New York City	**0.709*****	**0.645*****	**0.621*****
Jacksonville	**0.352*****	**0.512*****	**0.631*****
Milwaukee	**0.622*****	**0.309*****	**0.638*****
Houston	**0.430*****	**0.562*****	**0.640*****
Philadelphia	**0.620*****	**0.559*****	**0.656*****
St Louis	**0.650*****	**0.423*****	**0.678*****
Bridgeport	**0.483*****	**0.205*****	**0.720*****
Kansas City	**0.630*****	**0.523*****	**0.767*****

[1] *p<0.05; **p<0.01; ***p<0.001

Higher values of Moran's I in the last column of Table 4.4 indicate that neighbourhoods that are close to one another, tend to have similar employment rates of individuals that live in

the most vulnerable tracts. To distinguish between areas that have high and low employment rates of each housing group, we apply local spatial autocorrelation using LISAs [8]. In this context, LISAs use the variance of employment rates and the associated spatial weights of a region to identify clusters with a high concentration of employment of individuals from a specific housing group, which we refer to as *employment hotspots*. Figure 4.12A illustrates the housing landscape in Philadelphia and Bridgeport, with darker hues of red corresponding to census tracts that are more vulnerable to housing insecurity. Meanwhile, the purple geovisualisations convey employment hotspots for each of the housing groups, with darker hues of purple reflecting employment hotspots for individuals from the most vulnerable tracts. When we focus on the employment hotspots and residential tracts for the most vulnerable housing group, indicated by the dark purple and red, respectively, we can observe how home and workplace locations are often proximate to one another.

4.6.2 Transit Service to Better Job Opportunities

We leverage the housing demographics, transit networks, and employment hotspots, defined in earlier sections, to examine how transit infrastructure interfaces with upwards social mobility. We focus on social mobility because wealth is often the underlying constraint preventing individuals from improving their housing conditions [13, 301, 240]. We define upwards social mobility for individuals living in the most vulnerable neighbourhoods as having reasonable transit access to better job opportunities. Furthermore, we refer to better job opportunities, as a shorthand for employment hotspots for the mildly vulnerable housing demographic. This choice is based of the assumption that mildly vulnerable housing demographics have higher incomes and employment benefits, keeping in mind that income largely determines housing conditions. To implement this, we reassign the employment tracts of individuals from the most vulnerable demographic to randomly sampled employment hotspots for the mildly vulnerable demographic. In doing so, we assume that the hotspots for the mildly and less

vulnerable housing demographics provide better economic compensation compared to that of the most vulnerable housing group. This assumption stems from the positive relationship between median household income and lower levels of vulnerability in Section 4.4.1 and shown in Figure 4.7B. Figure 4.12B visualises how changing the workplaces of individuals commuting from the most vulnerable housing tracts impacts commuting characteristics.

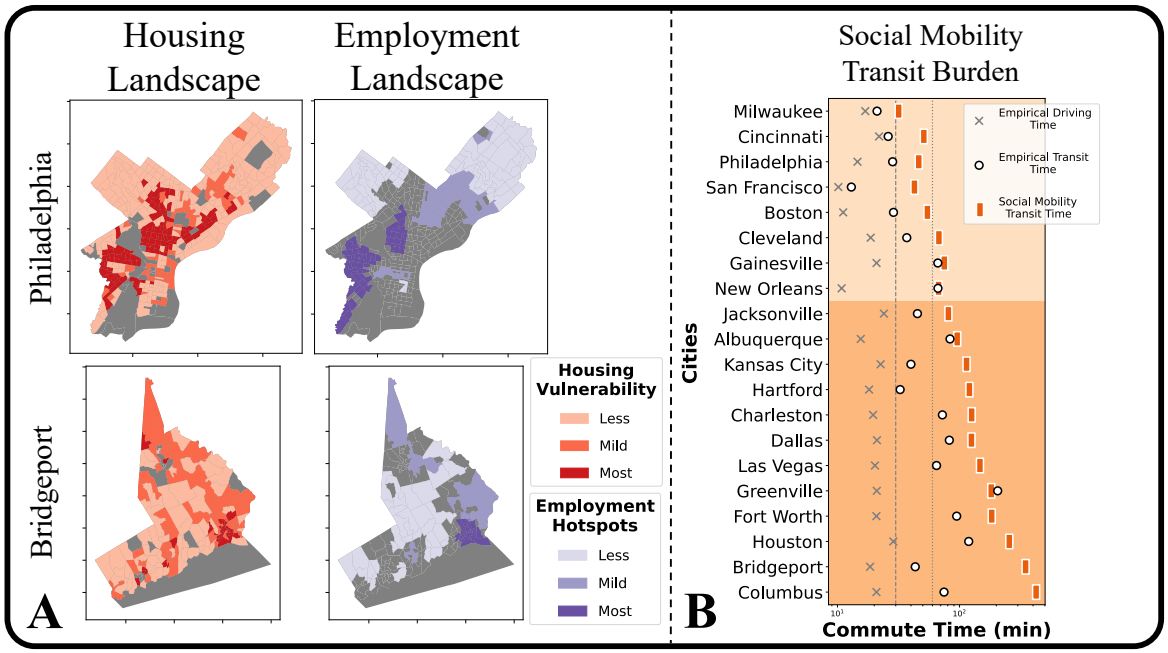

Fig. 4.12 **In 15 of the 20 analysed cities, neighbourhoods that are most vulnerable to housing insecurity have median transit commuting times of over an hour, when accessing workplaces associated with better opportunities.** Panel A reflects the spatial distribution of the housing and employment landscapes in Philadelphia and Bridgeport. Darker shades of red convey higher levels of housing insecurity, while darker shades of purple reflect employment hotspots for the more vulnerable housing demographics. Panel B depicts how transit commute times change for individuals living in the most vulnerable neighbourhoods if they were to start working in employment areas that would facilitate social mobility. The grey cross and white circle indicate empirical median commute times via driving and transit, respectively. Meanwhile, the orange rectangle refers to median transit commute times, averaged over 1000 of the social mobility simulations.

Figure 4.12B compares the commuting times for different scenarios, across each of the 20 cities in our analysis. The grey crosses reflect median empirical driving times of individuals that both live in census tracts that are the most vulnerable to housing insecurity

and commute to the most vulnerable employment hotspots. Similarly, the white circles represent the median empirical transit times for the same set of individuals. Meanwhile, the orange rectangles symbolise the median transit times for simulated commutes to better job opportunities. Over 1,000 iterations, we reassign the workplaces of the same set of individuals to randomly sampled mildly vulnerable employment hotspots. We note that the x-axis is in logarithmic scale, emphasising differences between shorter commutes. The dashed line indicates a 30-minute commute, whereas the dotted line marks an hour-long commute. Moreover, we reiterate that driving times do not account for traffic, but are a reflection of the cities' road networks. Figure 4.12B underscores the dependence between housing and employment locations, as median driving times is approximately a half hour or less, for all cities.

We observe how commuting times via transit increase for all cities, when compared to commute times using cars. However, for Milwaukee, Cincinnati, Philadelphia, San Francisco, and Boston, transit commute times remain under a half hour. The change from empirical driving to empirical transit times in San Francisco, Cincinnati, and Milwaukee is under five minutes. Moreover, shifting from the empirical data to the social mobility scenario reveals how commuting using public transit to areas with better opportunities leads to even longer commute times, barring Greenville and New Orleans. We note that transit commutes in the social mobility simulations for Gainesville, Milwaukee, and Albuquerque only increases travel time by less that 15 minutes, in comparison to its empirical counterparts. While empirical transit commuting times remain under an hour for the 10 of the 20 cities, only 5 cities maintain this characteristic in the social mobility context. Similarly, only Milwaukee transit infrastructure provides access to improved employment opportunities within approximately a half-hour transit commute (31.15 minutes). By using housing demographics and commuting behaviour to simulate potential for social mobility, we show how the majority of cities in our analysis do not have the adequate transport service for

supporting commutes, which fall under an hour-long journey, to workplaces that provide better employment opportunities. Furthermore, we reveal how in half of the 20 cities we analyse, individuals in the most vulnerable housing demographic (a demographic which tends to rely more on transit for commuting – 4.7E), have transit commute times of over an hour.

4.7 Discussion

This chapter provides a structural and experiential approach to understanding how transport inequalities can be expressed in the context of residential-employment landscapes. In particular, we analyse structural inequalities by highlighting how network entropy can capture sociodemographic inequalities in commuting patterns that conventional segregation metrics fail to detect. Furthermore, we provide a experiential perspective by proposing a framework to quantify the various dimensions of housing insecurity. Then, based on the same commuting data sources, we consider how transit facilitates access to jobs associated with better living conditions.

In analysing structural inequalities, we compare global in-flow entropy of 25 commuting networks across the U.S. to identify cities, such as San Francisco and Boston, that have higher degrees of monocentricity. Then, by incorporating local entropy measures for the entire commuting network, we uncover that census tracts with a higher concentration of affluence have residents that travel to more homogeneous workplaces. Meanwhile, tracts that attract a higher-income workforce express trends of heterogeneity in commuting origins. By splitting cities' commuting networks into high and low-income networks, we demonstrate the strength of network entropy in identifying disparities in residential-workplace trends across socioeconomic groups. We find that larger socioeconomic differences in commuting in-flow entropy values, which measures income disparities of a workforce in terms of their commuting origin distribution, correspond with higher levels of employment segregation.

Finally, we highlight how more diversity of commuting destinations for high-income workers is associated with longer average commute times from the origin census tract.

This brings into question what has long been deliberated in sociology: how the consequences of segregation may change depending on which demographic is segregated [119]. Future work can examine how these disparities in commuting flows reflect in other aspects of urban life, by considering other trip purposes that are not related to work. Specifically, other hypothetical scenarios can be explored to understand what mechanisms may be fuelling the strong correlation between entropy and segregation. Moreover, this framework can be applied to other mobility networks in the context of various demographic dimensions, such as gender or age.

Having introduced a network science approach to identifying spatial and inequalities in residential-employment locations, we shift our focus to understanding more nuanced forms of inequality in the housing dimension. We, first, introduce a classification framework that adopts a comprehensive approach to estimating levels of housing insecurity, accounting for the various dimensions of housing conditions. Existing approaches to defining housing insecurity, in the context of urban analytics, include using a specific housing feature as a rough proxy, such as rent burden or forced moves [124, 33]. We focus on Cox's definition as it captures financial, physical, and social forces that influence the state of housing. It is important to note that the seven dimensions proposed by Cox stem from a Global North perspective, with its definition based on housing policies in the USA. Attempts to develop a comprehensive measure for Global South incorporate features such as sanitation and water access [251]. The distinction between these two definitions is imperative, considering that different histories, cultures, and environments can re-frame the relevance of a housing dimension, and the features that can be used to estimate said dimension. Moreover, this work is limited in data availability of housing conditions, in that the neighbourhood and homelessness dimensions are yet to be incorporated. However, given the flexibility of the

proposed framework, introducing these dimensions is simply a matter of modifying the rank-based approach for extracting meaning from the generated clusters. Potential neighbourhood characteristics can be defined using crime data sources for safety or built form metrics for quality. Ultimately, our approach aims to capture various mechanisms that contribute to poor housing experiences, which we validate by comparing to a range of socioeconomic characteristics such as income, educational attainment, and mobility behaviour.

Then, we leverage open source tools to construct public transit and street network data. We use this data to characterise cities in the USA, based on their transport infrastructure. Moreover, we observe three types of transit systems based on how transit efficiency relates to trip distance. In line with research that demonstrates the decreasing significance of distance due to improved transit systems, we find that the cities with the most efficient transit service tend to have equally, if not more, efficient transit impedance for trips of longer distances than compared to shorter transit trips within the same city [235].

Finally, by incorporating the commuting data from the structural analysis, we unveil how transit infrastructure can impose additional hurdles to accessing workplaces that provide better financial opportunities. Studies have shown that targeted efforts in improving transit access to job opportunities has a positive effect on individual employment probability and individual income, particularly improving employment probabilities for lower-income individuals [253, 22]. However, we explore the geospatial layout of employment opportunities and residential landscapes to see how these efforts may also perpetuate inequalities in accessing jobs with different characteristics. Thus, this analysis contributes to research that motivates exploring inequality analyses from a spatial perspective, emphasising the importance of transit opportunity in social processes. Ultimately, housing conditions impact the level of comfort and belonging individuals experience within their environment [79, 224]. Thus, we aim to highlight how the disadvantage of housing insecurity is exacerbated by urban features that can hinder vulnerable populations from breaking out of the cycle of poverty. In

this manner, we show how transit infrastructure, coupled with residential and employment disparities, can impede individuals' abilities to live in improved housing conditions, for various cities.

By coupling network science, machine learning, geospatial analysis, and transport routing, we unveil inequalities in residential-employment disparities. In essence, we highlight how transit can act as a source of friction in accessing better job opportunities, for vulnerable demographics that face residential constraints.

Chapter 5

Mobility and transit segregation in urban areas

5.1 Introduction

The previous chapter analysed transit inequalities in residential-workplace dependencies. This chapter focuses on exploring socioeconomic transit disparities with respect to amenity visitations, rather than commuting patterns. That is, we analyse transport justice in the context of urban segregation. In order to do so, we incorporate mobility data to derive empirical mobility flows. Then, we can understand how transit passengers may experience segregation within transport infrastructure, while travelling to their destinations as well as how individuals experience segregation at the amenities they visit. In Chapter 2.4.5, we discussed relevant literature that combines mobility data to understand segregation in relation to public transportation. Analysing segregation at the destination-level has become increasingly prominent in human mobility research [1, 319, 206]. However, understanding how public transit provides service to destinations of varying segregation levels remains fairly unclear. It is crucial to consider transit within the reference frame of experienced segregation as public transport offers sustainable mobility options to individuals, particularly

for those who are unable to own a private vehicle. The few works that considered segregation, human mobility, and transit inequality focused on comparing access between transit riders and private vehicle users [98]. The approach we present in this chapter focuses on urban segregation based on the types of neighbourhoods that transit provides access to and the segregation experienced while using transit to fulfil empirical mobility demands.

In order to accomplish this, we combine census, mobility and transit data to analyse how transportation systems intersect with segregation in different aspects of urban life. We first define the state of residential segregation, using US Census data. Then, with anonymised mobility patterns from SafeGraph, we define segregation levels for amenities, based on the socioeconomic composition of its visitors. Finally, drawing upon open source resources, such as The Mobility Database, OpenStreetMap, and UrbanAccess, we construct transit networks to identify disadvantages within the system, analysing both the transport service and experience of using transit routes as potential sources of inequality. Moreover, we use Open Source Routing Machine (OSRM) to generate driving times and distances between any pair of coordinates in a city, given an OpenStreetMap (OSM) extract. OSRM is a high-performing routing engine that integrates well with OSM to find shortest paths on a road network. Driving times serve as a baseline for travel time, allowing us to compare how much longer trips take using transit than by driving. While cars and public transit vehicles both use road networks, transit vehicles must adhere to determined schedules and routes. From this perspective, cars have much fewer constraints as to how they can traverse the road network. Thus, driving times are useful for understanding the impedance, in terms of travel time, of using the transit system.

5.2 Defining Socioeconomic Segregation using the Index of Concentration at the Extremes

Socioeconomic segregation is generally defined as the extent to which various income-groups live apart from each other [191]. As discussed in Chapter 2.2.1, measures of residential segregation have been extended to account for spatial features. In this work, we quantify segregation using the Index of Concentration at the Extremes (ICE), which reconciles the diverging studies of concentrated affluence and concentrated poverty, to ultimately interpret them as one continuum. This is achieved by comparing how many households or individuals from the most deprived and privileged groups share the same residential area. As proposed by Douglas Massey [190], for a given region n, with T_n total households, A_n affluent, or privileged households, and P_n households in poverty, the ICE can be calculated as follows:

$$ICE_n = \frac{A_n - P_n}{T_n}, \tag{5.1}$$

Values can range from -1 to 1, reflecting extreme concentration of disadvantaged and privileged households, respectively. Thus, the ICE can capture levels of imbalance given the sociodemographic composition of a region. It is frequently applied to identify inequalities in public health, incarceration, and natural resource quality [50, 309, 284]. Figure 5.1 demonstrates how ICE values correspond to other common segregation measures. We compare how socioeconomic segregation, measured using the Index of Concentration at the Extremes, relates to other metrics of segregation. Specifically we consider ICE with respect to Dissimilarity, Isolation, Mutual Information Index, and Social Distance, all of which were introduced in Chapter 2.2.1. To avoid repetition, we group cities by their mean ICE value, where each group is a row of scatter plots. Thus, the first row depicts ICE correlations for cities with the lowest average segregation, while the last row does so for cities with the highest ICE segregation across all neighbourhoods. The blue scatter plots capture segregation

correlations at the census tract level. Census block groups (CBGs) are the highest resolution for which household income distributions can be openly accessed. Conventional segregation measures estimate inequality in an area using the distribution of individuals across its subareas. Thus, conventional measures such as dissimilarity (capturing the unevenness dimension) and isolation (depicting the exposure dimension) are constrained to the census tract level, which has a lower spatial resolution than CBGs, in terms of geographic boundaries.

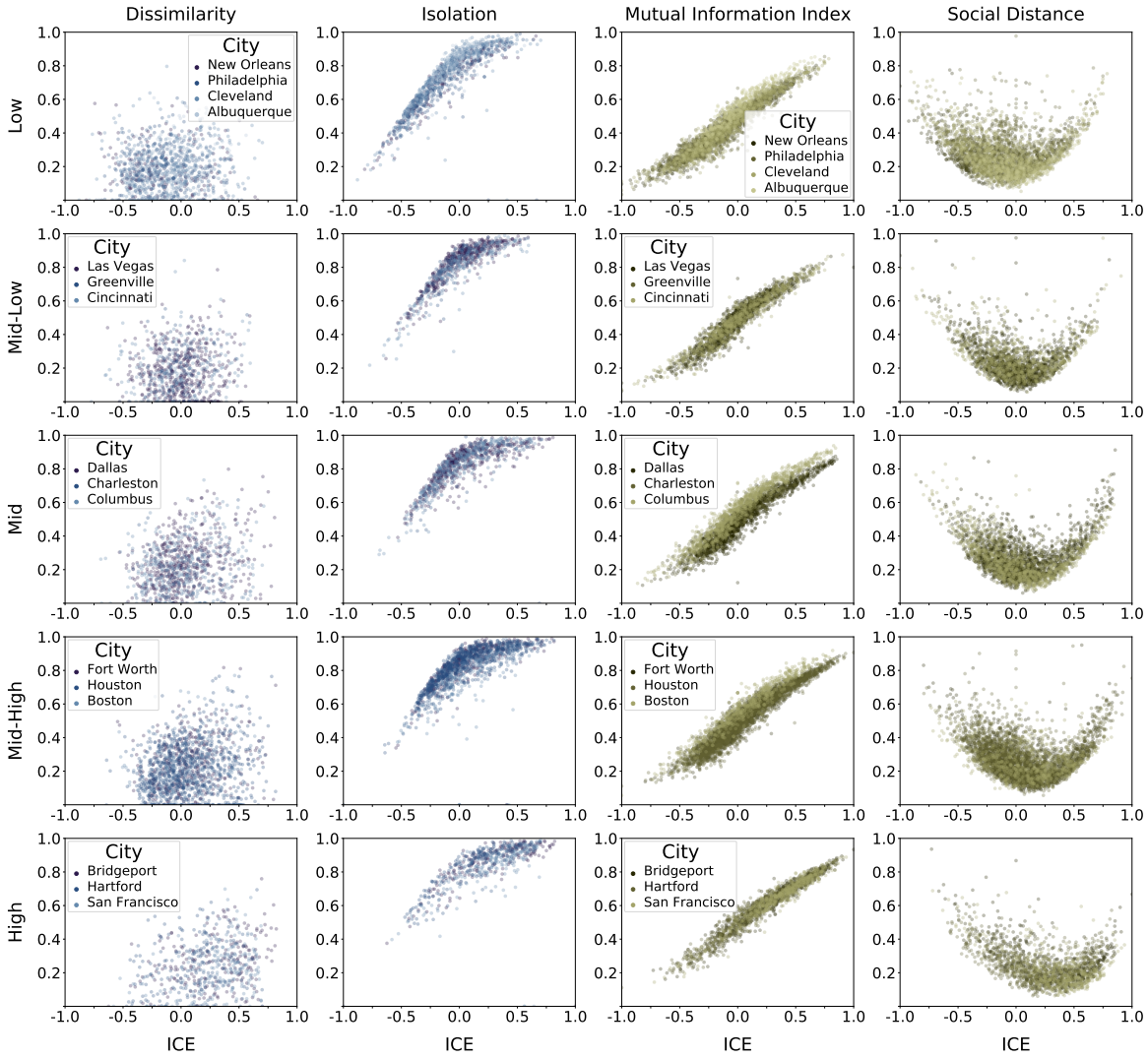

Fig. 5.1 **Various residential segregation metrics tend to capture different aspects of socioeconomic inequality.** Correlations between residential socioeconomic segregation (ICE) and other segregation metrics. ICE is compared to the index of Dissimilarity, Isolation, Mutual Information and Social Distance in the first, second, third, and fourth column, respectively. The x-axis of each scatter plot reflects ICE values, while the y-axis depicts values for the respective segregation metric to which ICE is being compared. Blue plots reflect comparisons at the census tract level, while green plots captures segregation for census block groups. Each row reflects the ICE correlations for a group of cities that are partitioned based on their mean ICE values.

We see that Dissimilarity and ICE do not exhibit any apparent relationship, while isolation expresses a positive association to ICE. Thus isolation is more useful at capturing disparities in highly segregated low-income tracts. However, isolation values have similar values

for positive ICE values, making it hard to identify segregation in terms of high-income concentration. The Mutual Information Index, derived from information theory, most closely captures the inequalities that ICE does, except it does so from a range of 0 to 1 [286, 245]. As a reminder, social distance uses a fractional rank-based approach to measure residential inequalities [315]. We see a consistently parabolic relationship between this measure and ICE, which highlights the utility of ICE in distinguishing between segregation of the most and least privileged demographics. Future work could focus on exploring how the use of different segregation metrics impacts identified segregation levels throughout urban dimensions. However, we continue with our analysis, using ICE, as it clearly distinguishes between segregation of the most and least privileged demographics. Moreover, ICE allows for spatial analysis at the CBG level.

Typical measures of segregation focus on inequalities experienced in residential areas. In this work, using the Index of Concentration at the Extremes (ICE) metric from Equation 5.1, we define segregation with respect to the socioeconomic concentration in three urban contexts: (a) residential (b) amenities and (c) public transit.

5.3 Sociodemographic Residential Segregation

We begin by exploring the state of residential segregation in 16 US cities. We take a closer look at the relationship between racial and socioeconomic composition to develop a better understanding of the residential landscape throughout the US. In doing so, residential segregation acts as a baseline, to which we can compare segregation levels in the other urban dimensions we consider in the coming sections.

We define the socioeconomic composition of a CBG using data from the American Community Survey, specifically Table B19001, which was introduced in Chapter 3.1.1. This data provides the number of households belonging to each of the 16 income brackets, as defined by the US Census Bureau, for every census block group. We denote the lower three

income brackets, that earn less than $20,000 per year, as households in poverty. Meanwhile, the upper three income brackets indicate affluent households, which earn more than $125,000 per year. We define these brackets as high income households. Middle class households are reflected by the middle 10 income brackets (earning between $20,000 and $125,000 per year). Thus, we can categorise each of the 16 income brackets into 3 income classes: low-income, middle-income, and high-income. Using these cutoffs to define income classes is common practice when measuring ICE at the CBG-level [160, 29, 297]. However, in Figure 5.2, we explore how ICE distributions would change within each city if we were to shift these income-bracket cutoffs.

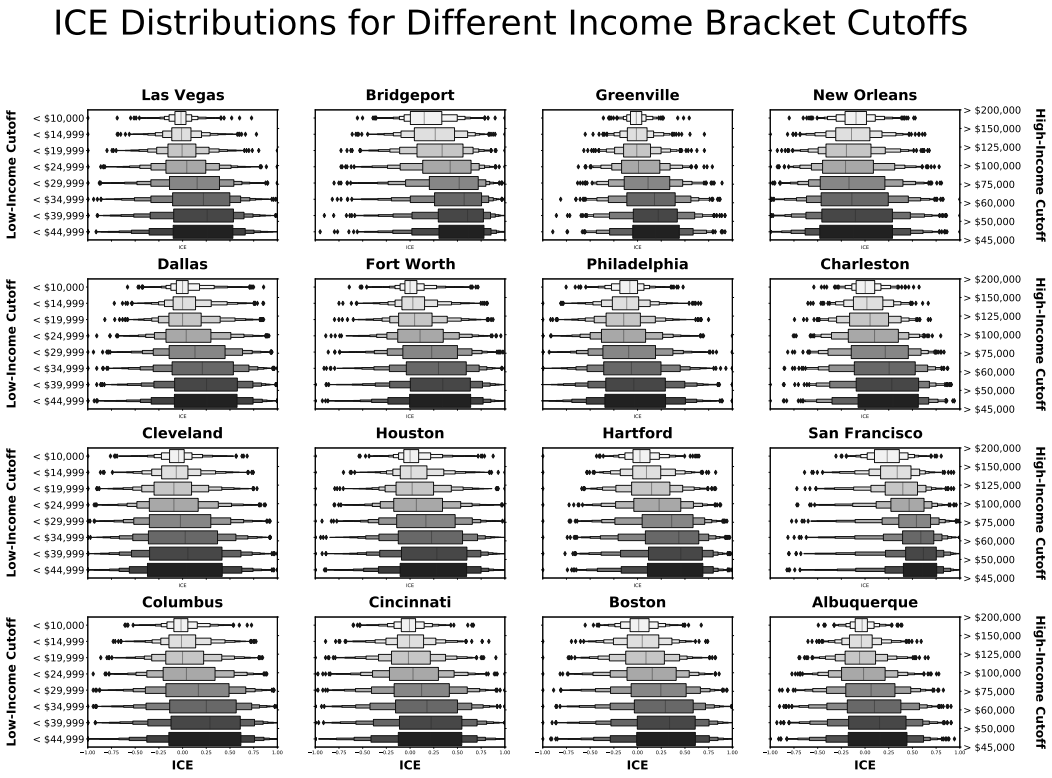

Fig. 5.2 **Less strict thresholds for defining extreme affluence and poverty leads to a larger range of ICE values, while also increasing the median ICE value in a city** Each scatter plot reflects distribution of ICE values in a city at the census block group level. Each box-plot refers to the ICE distribution (x-axis) for a particular definition of what constitutes a low and high income household. The definition of low-income cutoffs is on the left axis of the figure, while that of the high-income is on the right axis

Figure 5.2 explores the impact of using different income percentiles to define the number of low income and high income households in a neighbourhood. Each plot shows how the distribution of ICE values in a city changes when we shift the boundaries that define low and high income groups. We observe that by creating broader definitions of what it means to be an extremely low or high-income household, neighbourhoods tend to have a larger range of ICE values. Conventionally, when defining ICE, low and high income groups are defined by the 20th and 80th percentile of the income distribution, respectively. The 20th and 80th percentile correspond to the 3 lowest and 3 highest brackets. Thus, Figure 5.2 conveys how broader definitions of high and low-income lead to an increase in cities' median ICE values. While having looser constraints does lead to more variation across neighbourhoods, it shifts the focus away from the most and least privileged socioeconomic demographics in a city, which is the aim of this analysis. Furthermore, considering that non-residential aspects of urban life (i.e., amenity visitation, transit use) tend to have higher rates of social mixing [216, 206, 90, 312, 256], using less strict definitions of what it means to be affluent or poor would likely decrease measures of ICE as both demographics would represent less extreme fragments of a region.

With this in mind, we use the ACS data to define $ID_{bg,i}$ as the the number of households belonging to an income class, i, in CBG, bg. We combine this data, with our measure of segregation (Eq. 5.1) to define residential segregation for a census block group, bg:

$$ICE_{res}(bg) = \frac{ID_{bg,hi} - ID_{bg,lo}}{\sum_{i \in I} ID_{bg,i}} \tag{5.2}$$

where I reflects the set of three income classes (lo, mid, and hi). Conceptually, we are computing the extent to which households from affluent and poor households live in the same neighbourhood. The denominator reflects the total number of households in a neighbourhood, serving to normalise the segregation metric and to account for the middle-income population. We note that an ICE_{res} value of 0 can indicate either an equal, non-zero amount of low and

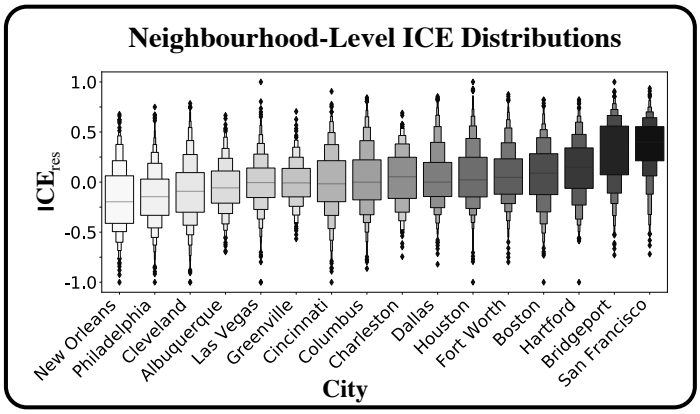

Fig. 5.3 **Socioeconomic residential segregation in 16 US cities, calculated using ICE**. Each box plot reflects the distribution of ICE values in census block groups, for a given city.

high-income households in a CBG, or a neighbourhood that is completely comprised of middle-income households. We emphasise that the overall socioeconomic composition varies across cities, with San Francisco housing a larger number of affluent households, as indicated through Figure 5.3. Each box plot in Figure 5.3 reflects the distributions of socioeconomic ICE values for all neighbourhoods in the corresponding city. Accordingly, when evaluating segregation within the context of one city, it is important to consider ICE values with respect to the overall composition of the city.

For example, the ICE for the entire population in San Francisco is 0.3, while for New Orleans it is -0.1. A San Francisco neighbourhood with a 0.1 ICE would reflect an area with a larger concentration of relatively lower income households, compared to San Francisco's baseline of 0.3. Meanwhile, a New Orleans neighbourhood with the same 0.1 ICE would exemplify the alternate case: an area with relatively high-income concentration. Despite both neighbourhoods having the same ICE value, the economic make-up of the city in which each area is situated, is what ascribes the severity of segregation that the neighbourhood experiences. Therefore, it is important to refer to the economic profiles as a baseline for each city, when interpreting the results presented.

Similarly, we compute the residential segregation of Black and White residents using Equation 5.1, such that A_{bg} and P_{bg} represent the number of White and Black residents,

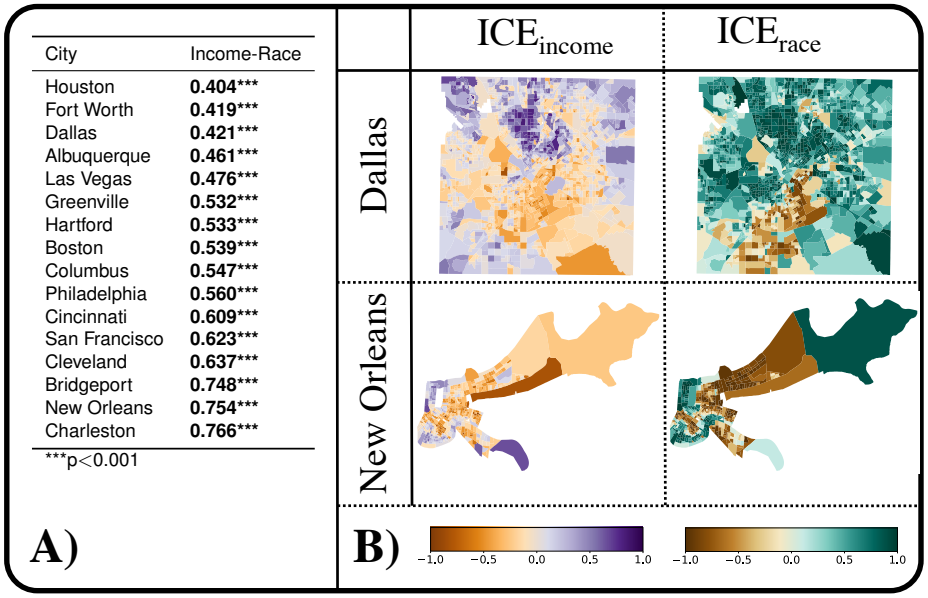

Fig. 5.4 **US cities exhibit positive correlations between income and racial segregation, albeit to different extents.** A) Pearson correlation coefficients, emphasising the relationship between economic and racial segregation in 16 US cities. B) Maps illustrating the entangled nature of economic and racial segregation for Dallas and New Orleans. Orange and purple reflect low and high-income concentrations, respectively. Brown and blue capture a high residential concentration of Black and White residents, respectively.

respectively, in a neighbourhood *bg*. We leverage Table B02001 from the American Community Survey, which captures the racial distribution of individuals in a CBG. Considering US history, and its persistent discrimination against the Black population, we select Black and White racial groups to represent the extremes in the context of racial segregation [95]. Accordingly, values of -1 indicate a high concentration of Black residents while +1 ICE levels reflect a large share of White residents. The second column in Figure 5.4A captures the relationship, using Pearson correlation coefficients, between a neighbourhood's economic and racial segregation level, for 16 US cities. A positive correlation indicates that neighbourhoods, or CBGs, with a large share of affluent households also have a high concentration of white residents. While the degree of correlation between the two demographic types varies across cities, we observe that all cities do have a positive, significant correlation between racial and economic segregation, indicated by the asterisks.

Figure 5.4B visualises the spatial landscape of economic and racial segregation in cities with lower and higher correlation coefficients (Dallas and New Orleans, respectively). Comparing ICE_{income} to ICE_{race} in Dallas reveals that areas with a large concentration of low-income residents are not directly translatable to highly segregated Black or highly segregated White neighbourhoods. On the other hand, New Orleans shows strong associations between neighbourhoods that are low-income and segregated (orange CBGs) and those that are largely composed of Black residents (brown CBGs). By comparing CBG-level segregation measures, with respect to socioeconomic and Black-White composition, we illustrate how residential patterns for income and racial groups can be intertwined. Measuring residential segregation levels characterises neighbourhoods based on the sociodemographic composition of its residents. Moving forward, we focus on analysing income segregation, however, the proposed methodology can be applied to explore whether racial segregation, or segregation for other sociodemographic groups, persists in different urban dimensions. Ultimately, Figure 5.4 conveys the state of residential segregation in 16 US cities, acting as a baseline against which we can compare other dynamics forms of segregation by leveraging mobility patterns and the characteristics of transit systems.

5.4 Mobility Segregation

In this section, we extend our analyses of residential segregation in US cities, to explore whether residents from segregated neighbourhoods tend to exhibit segregation in their amenity visitation patterns. To do so, we define the segregation at amenities based on the socioeconomic composition of its visitors during January 2021. Then, we explore whether residents from the most segregated neighbourhoods tend to travel to amenities that have similar levels of segregation compared to their residential neighbourhoods. This approach allows us to understand mobility differences between highly-segregated high-income neighbourhoods and highly-segregated low-income neighbourhoods.

5.4.1 Segregation at the Amenity Level

We begin analysing segregation from a human mobility perspective, by measuring segregation based on the socioeconomic makeup of an amenity's visitors. We define ID' as the normalised form of the income distribution, ID, defined by the ACS data:

$$ID'_{bg,i} = \frac{ID_{bg,i}}{\sum_{i' \in I} ID_{bg,i'}}, \text{ where } I = \{lo, mid, hi\}. \tag{5.3}$$

$ID'_{bg,i}$ defines the fraction of households in CBG, bg, that belong to income class, i. This is achieved by dividing the number of households in a CBG, bg, and income group, i, ($ID_{bg,i}$), by the total number of households in the CBG ($\sum_{i' \in I} ID_{bg,i'}$).The SafeGraph data provides information regarding the mobility flow from CBGs to amenities, while the ACS data denotes the socioeconomic composition of a CBG. Using the SafeGraph amenity visitations from the Weekly Pattern data set, we construct a $|\mathbf{BG}| \times |\mathbf{A}|$ sized mobility matrix, $\mathbf{M}$, for January 2021, where $\mathbf{BG}$ and $\mathbf{A}$ are the set of CBGs and amenities in a city, respectively. $\mathbf{M}_{bg,a}$ reflects the number of trips from CBG, bg, to amenity, a, during the month. SafeGraph reports any visitation counts that are between two and four as four, to support anonymisation. Accordingly, we replace any cell value of four in $\mathbf{M}_{bg,a}$ with a uniformly sampled value of two, three, or four. Combining the two sources, we can estimate the socioeconomic composition of mobility flows between CBG-amenity pairs, by performing a weighted sampling of ID', indicated by the $\sim$ symbol:

$$\mathbf{C}_{bg,a} = \{C_{bg,a,1} ... C_{bg,a,v}\}, \quad \text{where } C_j \sim ID'_{bg},\ v = \mathbf{M}_{bg,a}. \tag{5.4}$$

Here, C_j represents an individual from CBG, bg, who visits amenity, a. Each visitor, C_j, belongs to an income class $i \in I$, which is sampled from ID'_n. The size of $\mathbf{C}_{bg,a}$ is determined by the corresponding visitor count in $\mathbf{M}_{bg,a}$. Due to the level of anonymisation in the SafeGraph data, this method of sampling assumes that individuals in a neighbourhood

have an equal likelihood of travelling to each amenity. The segregation level of individuals visiting an amenity is determined by a neighbourhood's socioeconomic distribution and the volume of its residents that travel to the amenity. We note that even under this assumption, we identify segregation in visitation patterns in the following sections. To account for the stochastic nature of this approach, we perform the weighted sampling 100 times, where $C^x_{bg,a}$ reflects the economic composition of visitors from CBG bg visiting amenity a, during the x^{th} iteration.

We can modify $\mathbf{C}^x_{bg,a}$, to reflect the socioeconomic composition of each amenity, based on its visitors, such that $\mathbf{C}^x_a = \{\mathbf{C}^x_{bg,a} \mid bg \in \mathbf{BG}\}$ and $|\mathbf{C}^x_a| = \mathbf{M}_a$. Each element in $\mathbf{C}^x_a$ resembles the income group of a visitor from amenity a, during the x^{th} iteration. As a reminder, in defining Equation 5.4, each individual was assigned an income group that was sampled from the income distribution of her residential CBG. Thus, $\mathbf{C}^x_{a,i}$ captures the set of individuals visiting amenity a, in income group i, during iteration x. We can define the level of segregation, in terms of mobility patterns, at amenity a with the following equation:

$$ICE_{amenity}(a) = \frac{1}{100}\sum_{x=1}^{100}\frac{|\mathbf{C}^x_{a,hi}| - |\mathbf{C}^x_{a,lo}|}{|\mathbf{C}^x_a|}, \tag{5.5}$$

where $|\mathbf{C}^x_{a,hi}|$ and $|\mathbf{C}^x_{a,lo}|$ reflect the number of high and low income visitors, respectively, to amenity a in iteration x. Equation 5.5 computes the average segregation of an amenity's visitor composition using ICE, such that the economic makeup of visitors is determined through a stochastic, weighted sampling with respect to visitation frequency and the socioeconomic characteristics of visitors' origins. Having used the Index of Concentration at the Extremes and visitation patterns to measure amenity segregation, we define mobility segregation from the perspective of residents in a neighbourhood, based on the level of segregation they experience, on average, at the amenities they visit. We refer to this measure as traveller amenity segregation (TAS) as it captures segregation based on residents' mobility patterns.

Traveller amenity segregation, for a neighbourhood, only considers segregation at the amenities its residents visit. Then, the *TAS* of a neighbourhood, *bg*, is computed as the weighted average of these amenities' segregation levels, with respect to the frequency with which its residents visit the amenities:

$$TAS(bg) = \frac{\sum_{a \in \mathbf{A}} \mathbf{M}_{bg,a} * ICE_{amenity}(a)}{\sum_{a \in A} \mathbf{M}_{bg,a}} \tag{5.6}$$

Where $\mathbf{M}_{bg,a}$ is the number of visits from CBG *bg* to amenity *a* and $\mathbf{A}$ is the set of all amenities in a city. Thus, the amenities that are more frequently visited will play a bigger role in characterising the average segregation of amenities to which residents travel.

5.4.2 Comparing Levels of Residential and Mobility Segregation

From a broader perspective, *TAS* aims to depict how individuals experience segregation based on where they travel, by considering their amenity visitation patterns and the segregation levels at the amenities they visit. With this in mind, we examine the relationship between residential segregation and traveller amenity segregation (TAS). We can calculate how much a segregation value for a given neighbourhood, *bg*, changes between two contexts of urban life (i.e., changes between residential and employment segregation), given it's *ICE* value, in urban context *x*, and its *ICE* value in a different dimension, *y*:

$$\Delta ICE_{bg}(x,y) = \begin{cases} ICE_x(bg) - ICE_y(bg), & \text{if } ICE_x(bg) < 0 \\ -(ICE_x(bg) - ICE_y(bg)), & \text{if } ICE_x(bg) \geq 0 \end{cases} \tag{5.7}$$

ΔICE ranges from -2 to 2, where negative values signal a decrease in segregation levels. If a neighbourhood has a large concentration of high income residents, such that its ICE_{res} is positive, then having a lower *ICE* value in the mobility dimension is indicative of a decrease in segregation levels when shifting from the residential and mobility dimension. However,

having a lower *ICE* value in the mobility dimension, for neighbourhoods with negative ICE_{res} values, reveals that its residents are travelling to amenities that have a larger concentration of low-income individuals, than compared to their neighbourhood's residential composition. Such instances signify an increase in segregation from the residential to mobility dimension. Thus, the two cases account for the sign of the original *ICE* value. We denote the difference between $ICE_{res}(bg)$ and $TAS(bg)$ in a neighbourhood, bg, as $\Delta ICE_{TAS}(bg)$, which expresses how segregation levels change between the residential and mobility dimension.

By using the ICE as a segregation metric, we can distinguish neighbourhoods with high concentration of affluent households from those with a large share of households in poverty. For each city, we split neighbourhoods, using residential ICE values, into five, equally sized segregation groups: (1) Highly-segregated, low-income (**HS-Lo**) (2) Mildly-segregated, low-income (**MS-Lo**) (3) Less-segregated (**LS**) (4) Mildly-segregated, high-income (**MS-Hi**) and (5) Highly-segregated, high-income (**HS-Hi**). Figure 5.5 illustrates, for each city, the distribution of ICE values that belong to each group.

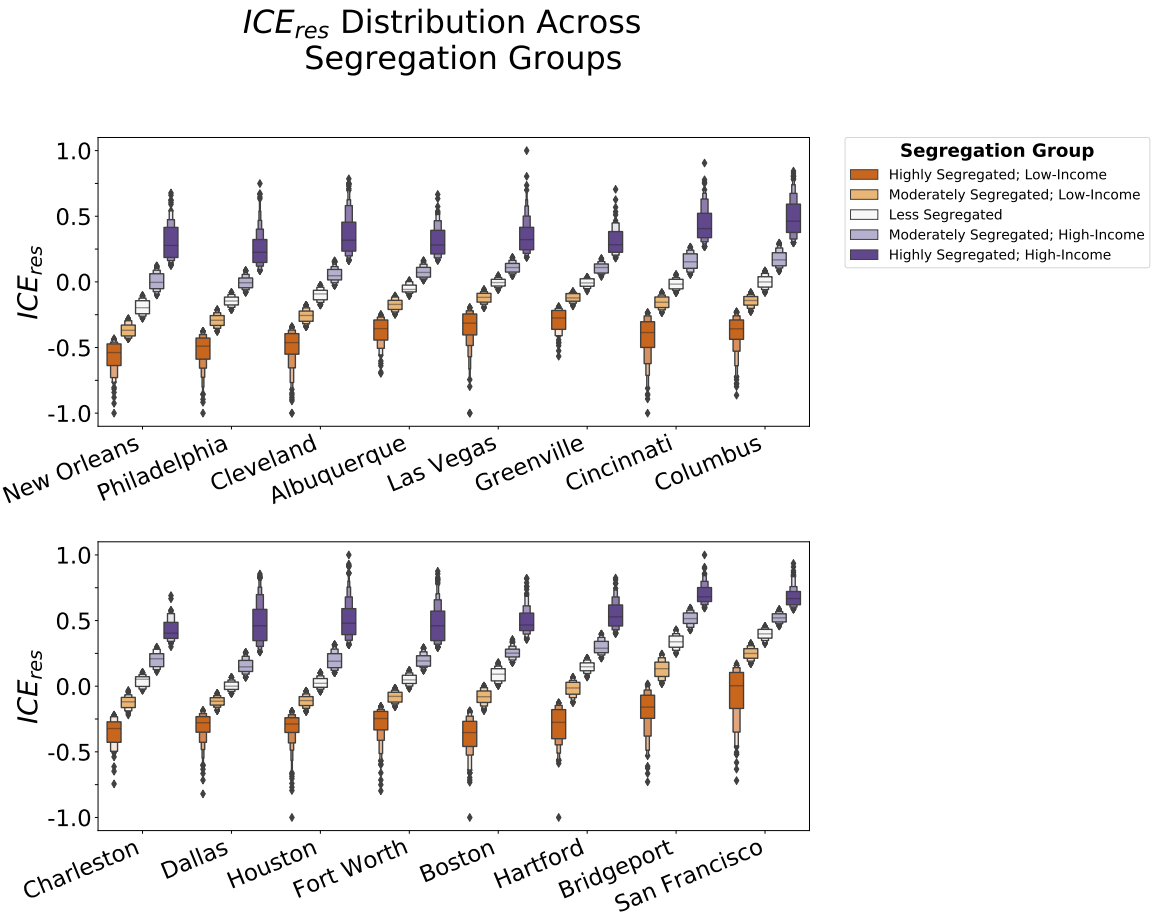

Fig. 5.5 **ICE distribution values within each segregation group, for a given city.** Orange box plots depict low-income concentration segregation groups, while purple reflects segregation groups of high-income concentration. Darker colours capture higher segregation. The y-axis shows the ICE values of census block groups belonging to a particular segregation group.

Since we create the segregation groups with respect to each city, neighbourhoods belonging to the highly-segregated, low-income group do not necessarily have high segregation magnitudes. Rather, they reflect neighbourhoods that have a high concentration of low-income households, relative to the segregation distribution of that city.

Through Figure 5.5, it becomes clear that a HS-Lo segregation group in one city does not carry the same meaning when applied to another city, due to differences in the economic composition of cities. For example, the median ICE value of the Highly-segregated, low-income group in New Orleans is approximately 0.25, while the median ICE value of the

same segregation group in San Francisco is approximately 0.6. Therefore, while the ICE *values* of two neighbourhoods from different cities are comparable, the segregation *groups* of neighbourhoods from different cities can not be compared as the groups are calculated with respect to each city's ICE distribution.

We focus on the highly segregated low-income and high-income neighbourhoods, represented in Figure 5.6 by the orange and purple points, respectively. The x-axis captures the magnitude of residential segregation ($|ICE_{res}(bg)|$), which allows us to compare the most segregated neighbourhoods in a city within the same frame of reference. For cities with particularly skewed economic compositions, such as San Francisco, we observe larger differences regarding where each group resides along the x-axis. As mentioned in section 5.3, and illustrated in Figure 5.3, we would expect to see this in a city such as San Francisco because its composition is largely made up of high-income households. Thus, the highly-segregated low-income neighbourhoods, within San Francisco, tend to have smaller magnitudes.

The consistently negative slopes in Figure 5.6 can be attributed to the positive correlations between residential segregation and *TAS*, shown in Table 5.1. Figure 5.6 reveals that the majority of CBGs tend to have negative ΔICE_{TAS} values. These smaller values suggest a decreased level of segregation when comparing the residential segregation individuals experience in their neighbourhood, to their *TAS* values that measure mobility segregation.

The key takeaway of Figure 5.6, however, is that, for most cities, the highly-segregated, low-income neighbourhoods exhibit larger decreases in segregation (shown by larger magnitudes of ΔICE_{TAS}) than their high-income counterparts, when considering changes from the residential to the mobility dimension. Thus, Figure 5.6 reveals that, generally, segregated low-income neighbourhoods tend to travel to amenities with a much different economic composition, than compared to segregated high-income neighbourhoods. We hypothesise that this finding could be a reflection of how individuals from lower income neighbourhoods correct for their level of segregation in the residential dimension by modifying their mobility

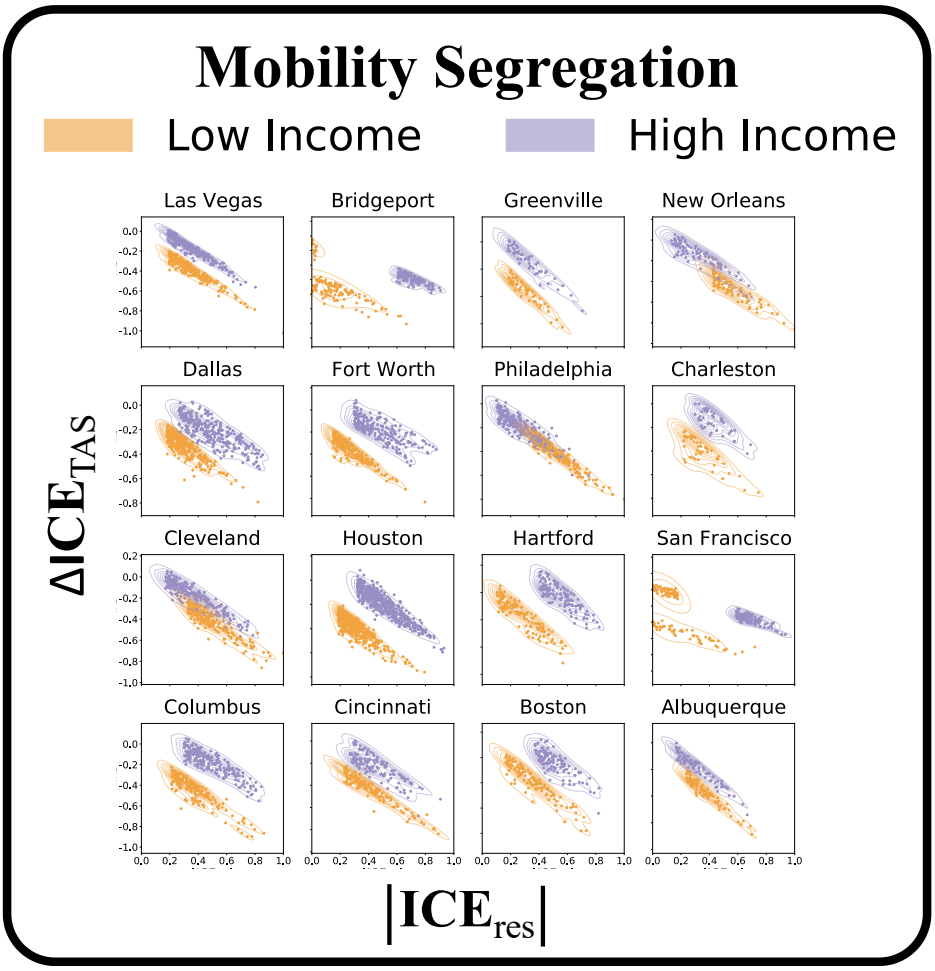

Fig. 5.6 **When considering how segregation changes from the residential to the mobility dimension, highly-segregated, low-income neighbourhoods tend to exhibit a larger decrease than their high-income counterparts.** Differences in segregation levels in the residential and mobility domain, where each scatter plot resembles one of the 16 US cities. The x-axis depicts the magnitude of residential segregation, for the HS-Hi neighbourhoods (purple) and the HS-Lo CBGs (orange). The y-axis shows the direction and magnitude of change in a CBG's mobility segregation level, compared to its residential segregation level.

behaviour [315]. However, this could also be an artefact of transit systems effectively facilitating social mixing for individuals in highly-segregated, low-income neighbourhoods. TO further explore this concept, the next section studies how transit systems provide service to neighbourhoods in different segregation groups.

Table 5.1 Pearson correlation coefficients between socioeconomic residential segregation of a neighbourhood and the experienced segregation of its residents based on the amenities they are visiting, for 16 US cities.

City	ICE_{res} α ΔICE_{TAS} r[1]
Las Vegas	**0.7526*****
Charleston	**0.7640*****
Greenville	**0.7788*****
Fort Worth	**0.7969*****
Cincinnati	**0.7976*****
Columbus	**0.8001*****
Houston	**0.8051*****
Albuquerque	**0.8116*****
Dallas	**0.8197*****
Philadelphia	**0.8249*****
Boston	**0.8262*****
Hartford	**0.8378*****
New Orleans	**0.8547*****
Cleveland	**0.8597*****
Bridgeport	**0.8661*****
San Francisco	**0.9107*****

[1] $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

This can be observed by considering how each group's points are distributed along the y-axis. Yet, two cities emerge as exceptions to this trend. A subset of highly-segregated, low-income neighbourhoods in San Francisco and Bridgeport exhibit distinctive patterns in which they experience an increase in ΔICE_{TAS}, pointing to an increase in mobility segregation levels, despite already having high levels of residential segregation. These findings emphasise the importance of considering segregation from various urban dimensions, as mobility can be used as a means to decrease the overall segregation that one experiences. While, we find

that mobility segregation has lower magnitudes than residential segregation, this section highlights how segregation continues to exist when considering amenity visitation patterns, finding strong associations between the two domains.

5.5 Transport Segregation

Having demonstrated the role that segregation plays in the residential and mobility facets of the urban experience, we begin to consider the intersection between segregation and public transportation systems. In this section we leverage public transit networks to analyse how structural properties of transportation systems coincide with the residential landscape. Employing the SafeGraph mobility data, we model potential transit use to estimate the level of segregation one would experience while using transit to satisfy her mobility demands. Due to the computational complexity of stochastically modelling transit use, we use five of the 16 cities as an example for how levels of structural and experiential segregation can be assessed in the transit system. We specifically choose New Orleans, Philadelphia, Cincinnati, Dallas, and San Francisco as the 5 focal cities, as each city spans different parts of the socioeconomic residential composition, as depicted in Figure 5.3.

5.5.1 Public Transport as a Tool for Overcoming Residential Segregation

We assess segregation in the context of transport by, first, examining how transit systems serve neighbourhoods with various segregation levels. To do so, we consider travel times between every possible pair of neighbourhoods in a city. For every CBG pair, (bg_o, bg_d) we sample 100 locations from the origin census block group, $\{bg_o^1 ... bg_o^{100}\}$, and 100 from the destination CBG, $\{bg_d^1 ... bg_d^{100}\}$, using UrbanAccess. Since edges in the transit-pedestrian networks are weighted by travel time, we can find the shortest path length on the transit-pedestrian

networks between points bg_o^x and bg_d^x, for $x \in \{1...100\}$. We calculate the corresponding driving times for the same set of coordinate pairs, using Open Source Routing Machine. As a result, we can measure the average time it takes to travel between two neighbourhoods in a city, using public transportation or a car. Accordingly, we denote $J_{transit}(bg_o, bg_d)$ and $J_{driving}(bg_o, bg_d)$ to reflect the average travel time for a journey between neighbourhoods bg_o and bg_d, when using public transport and cars, respectively.

Next, we explore whether transit systems facilitates overcoming residential segregation by providing service between neighbourhoods of different segregation levels. We use the previously defined travel times to evaluate which neighbourhoods can be accessed within a travel time threshold. Specifically, we use the previously defined segregation groups, which are derived by partitioning a city's neighbourhoods into 5 equally sized groups based on their residential segregation levels. As follows, $\mathbf{BG}_s$ refers to the set of neighbourhoods in segregation group s. Given a travel time threshold, t, and a CBG, $bg \in \mathbf{BG}_s$, we define $\mathbf{BG'}_{bg}$ as the neighbourhoods that can be reached from bg within t minutes. That is, we only consider neighbourhoods, bg', for which values of $J_{transit}(bg, bg')$ are less than the travel time threshold, t. Subsequently, we calculate the average segregation level of all neighbourhoods that are accessible from each CBG in $\mathbf{BG}_s$, within t minutes, for all neighbourhoods in segregation group s:

$$NA_{transit}(s,t) = \frac{\sum_{bg}^{\mathbf{BG}_s} \sum_{bg'}^{\mathbf{BG'}_{bg}} ICE_{res}(bg')}{\sum_{bg}^{\mathbf{BG}_s} |\mathbf{BG'}_{bg}|} \tag{5.8}$$

where $NA_{transit}$ conveys the average socioeconomic profiles of neighbourhoods to which transit systems provide access, within a time threshold, t, for CBGs belonging to a particular segregation group, s. This is achieved by determining the average segregation level of areas a segregation group can reach via transit, within a given time frame. We can calculate the same metric, but with respect to driving times, by using $J_{driving}$ to compute the set of

reachable neighbourhoods, for a given segregation group and time threshold. We refer to this measure of segregation in driving access as $NA_{driving}$. We visualise both metrics in Figure 5.7, for time thresholds from 5 to 60 minutes, at 5 minute intervals. For every matrix, the top row illustrates the changes in segregation characteristics of neighbourhoods that are accessible by the highly-segregated, low-income neighbourhoods (HS-Lo), for various time thresholds. Meanwhile, the bottom row captures average segregation levels, based on which neighbourhoods the highly-segregated, high-income neighbourhoods can reach. The top row of matrices illustrates how segregation of accessible neighbourhoods changes for transit time thresholds, measured using $NA_{transit}$. Meanwhile the bottom row of matrices captures driving accessibility ($NA_{driving}$), and serves as a baseline for comparison, as driving times calculated with OSRM are void of any transit schedule, route, or traffic constraints that are within the GTFS data used to build the transit-pedestrian networks. It is apparent that, when comparing transit to driving access for each city, segregation values for neighbourhoods accessible by car converge to reflect the city's overall socioeconomic composition much quicker than their public transit counterparts.

To some extent, we would expect transit segregation to have different values across segregation groups, especially for smaller time thresholds, as a reflection of spatial auto-correlation in residential segregation. However, the differences in transit segregation levels persists beyond 60 minute journeys for Dallas, Cincinnati and New Orleans, revealing apparent structural inequalities in the transit systems of those cities. The driving access matrices in Figure 5.7 emphasise the disparities in transit service, using driving times to convey the possibility for transit services to provide less segregated accessibility.

Here, we approach human mobility as an means for accessing neighbourhoods with various socioeconomic profiles. By investigating how transit systems connect neighbourhoods of different segregation levels to one another, we uncover structural inequalities in transit service, particularly in Dallas, Cincinnati and New Orleans. To conclude our analysis of

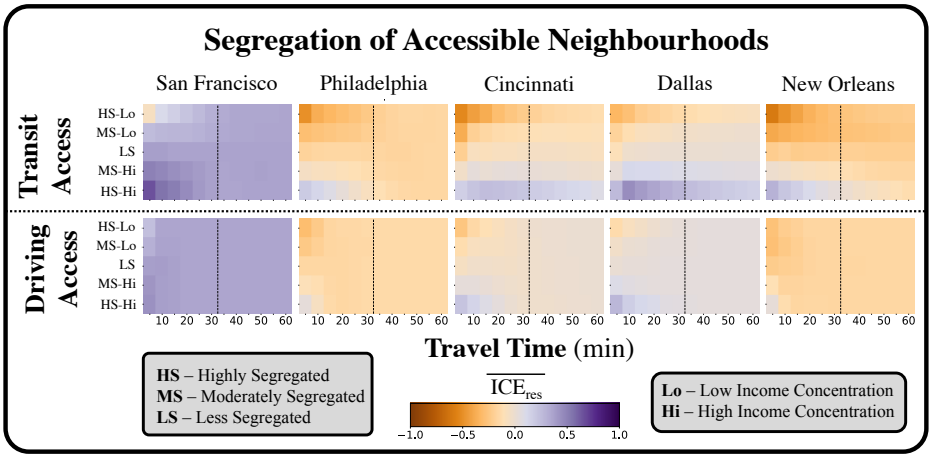

Fig. 5.7 **Public transit provides limited access to neighbourhoods of different socioeconomic backgrounds.** Average residential segregation level of neighbourhoods that are reachable within a given travel time, by public transit (top row) and car (bottom row), for 5 US cities. The y-axis shows neighbourhood accessibility for different segregation groups, while the x-axis defines different time thresholds. Orange cells reflect accessibility of neighbourhoods that are more segregated, with a concentration of low-income households. Purple cells capture transit service to neighbourhoods with a higher income concentration.

transit segregation, we approximate the level of segregation experienced while using public transit, estimating transit usage based on transport networks and travel behaviours derived from the SafeGraph mobility data.

5.5.2 Modelling Transit-Use Segregation

We continue our analysis of urban segregation by analysing how the areas to which individuals travel can impact the level of segregation experienced when using the transit system. We proceed, utilising the transit-pedestrian networks to calculate the shortest route, within the transit system, between a neighbourhood and an amenity. Specifically, we sample $\mathbf{M}_{n,a}$ points from CBG n's geographic boundaries, resembling the origin coordinates of visitors The destination points are defined by the longitude and latitude coordinates of amenity a. We note that by defining the economic composition of visitors by sampling neighbourhood income distributions, we assume uniform use of the transit system across socioeconomic groups. Thus, transit use segregation becomes an artefact of four main

mechanisms: (1) the neighbourhood's income distribution, (2) the amenities its residents visit, (3) the frequency with which residents visit said amenities, and (4) how the subset of the transit lines its residents use to visit their amenities intersects with the transit use of residents from other neighbourhoods. While this approach does not account for the fact that more vulnerable demographics may rely more on transit [101, 130], we hypothesise that incorporating disparities in transit reliance would exacerbate the levels of transit segregation we identify under this uniform-use assumption. Future work can use higher resolution mobility data to understand transit use, based on the speed of individual trajectories.

We define the set of edges in the transit layer of the transit-pedestrian network that visitor v, in $\mathbf{C}^x_{bg,a}$ traverses when moving from their sampled origin coordinate to amenity a as $\mathbf{P}^x_{bg,a,c}$. When an individual travels on a transit edge, her socioeconomic background contributes to the level of segregation experienced by all travellers using that transit segment. Thus, we can define the economic composition of a transit edge, e, for an iteration x as:

$$\mathbf{C}^x_e = \{v \mid bg \in \mathbf{BG}, a \in \mathbf{A}, v \in \mathbf{C}^x_{bg,a}, e \in \mathbf{P}^x_{bg,a,v}\} \tag{5.9}$$

To calculate segregation at the transit edge-level, we define $\mathbf{C}^x_{e,i}$ to reflect the set of individuals from the low, middle or high-income group that travel on an edge e, where $i \in \{lo, mid, hi\}$. In this manner we can define the level of segregation experienced on an edge, e, as an average of edge-level segregation across all stochastic iterations:

$$ICE_{edge}(e) = \frac{1}{100}\sum_{x=1}^{100}\frac{|\mathbf{C}^x_{e,hi}| - |\mathbf{C}^x_{e,lo}|}{|\mathbf{C}^x_e|} \tag{5.10}$$

To compare how segregation levels change across the residential, amenity, and public transport domains, we aggregate edge-level transit segregation to the census block group level. For a given neighbourhood, bg, we define the average segregation level experienced while using public transport, by residents in CBG bg, as **traveler transit segregation** (TTS):

$$TTS(bg) = \overline{\mathbf{T}_{bg}}, \textit{ where } \mathbf{T}_{bg} = \{ICE_{edge}(e) \mid a \in \mathbf{A}, v \in \mathbf{C}^{x}_{bg,a}, e \in \mathbf{P}^{x}_{bg,a,v}\} \tag{5.11}$$

where $\mathbf{T}_{bg}$ reflects the transit line segregation (ICE_{edge}) that each traveller from neighbourhood, bg, experiences on the edges she traverses to reach amenity, a. Accordingly, $TTS(bg)$ conveys the average experienced segregation when using transit for residents in CBG, bg, with respect to empirical amenity visitation patterns.

Thus, we have estimated how, residents in a given neighbourhood experience segregation from a residential (Eq. 5.2), amenity (Eq. 5.6), and public transport (Eq. 5.11) perspective. Figure 5.8 highlights how segregation levels persist across the three dimensions, focusing on the highly-segregated, low-income and highly-segregated, high-income segregation groups. Regardless of city-level economic composition, the parallel plots convey that segregation continues to exist in the transit and destination dimension, albeit to a lesser extent. The smaller range of segregation experienced in the mobility and transit dimensions reveals potential approaches to reducing the negative effects of segregation, that extend beyond changing housing policies. For example, one such solution could include extending transit service to connect areas of different segregation levels so that the segregated service identified in Figure 5.5.1 is reduced. Moreover, it could be worth exploring the role that the spatial distribution of amenities has on segregation measures in the mobility and transit dimension. We explore the latter concept in the following section

The Role of Amenity Distributions on Transit Segregation

To gain a deeper insight regarding how disparities in mobility destinations impact segregation while using the transit system, we develop a null model, which hypothesises that transit segregation is an artefact of disparities in the amenity landscape. To model this, we retain the same distribution of trip counts across neighbourhoods in a city. However, we modify the

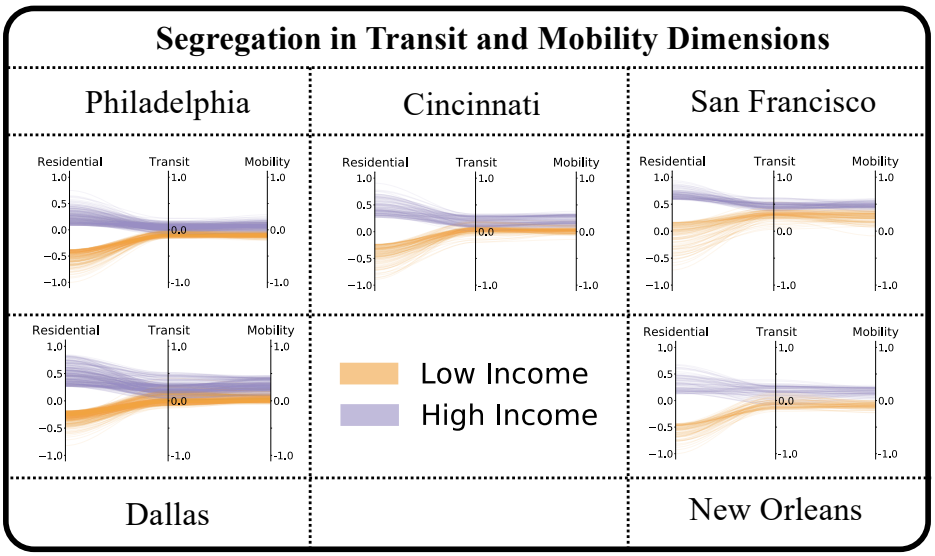

Fig. 5.8 **Levels of experienced segregation, while visiting amenities and while using transit to reach amenities, are lower than residential segregation measures.** Urban segregation levels in Philadelphia, Cincinnati, San Francisco, Dallas and New Orleans. From left to right, each axis in a parallel plot conveys experienced segregation in the residential, transit, and mobility domains, respectively. The purple and orange lines reflect segregation levels for the highly-segregated high and low-income groups, respectively.

destinations of every trip in the SafeGraph amenity visitations data, by randomly sampling a coordinate within a randomly sampled CBG. In other words, we maintain empirical mobility flows from a neighbourhood, but randomise the destinations to which individuals in a neighbourhoods travel and the locations of amenities. By doing so, we create an amenity landscape such that amenities are evenly distributed within a neighbourhood and across all neighbourhoods in a city. Moreover, this null model assumes an equal probability for a given individual visiting an amenity, removing mobility patterns such a preferential return and visitation recency [4, 16]. We construct a mobility matrix, $M^{null}_{bg,a}$, using the sampled amenity destinations, and apply the same workflow to determine amenity and transit edge segregation. Again, to account for the stochasticity of sampling destinations, we perform this process for 100 iterations.

We can, then, compare segregation estimated from the empirical data, to that of the null model, which eliminates apparent disparities in the amenity dimension, illustrated in Figure

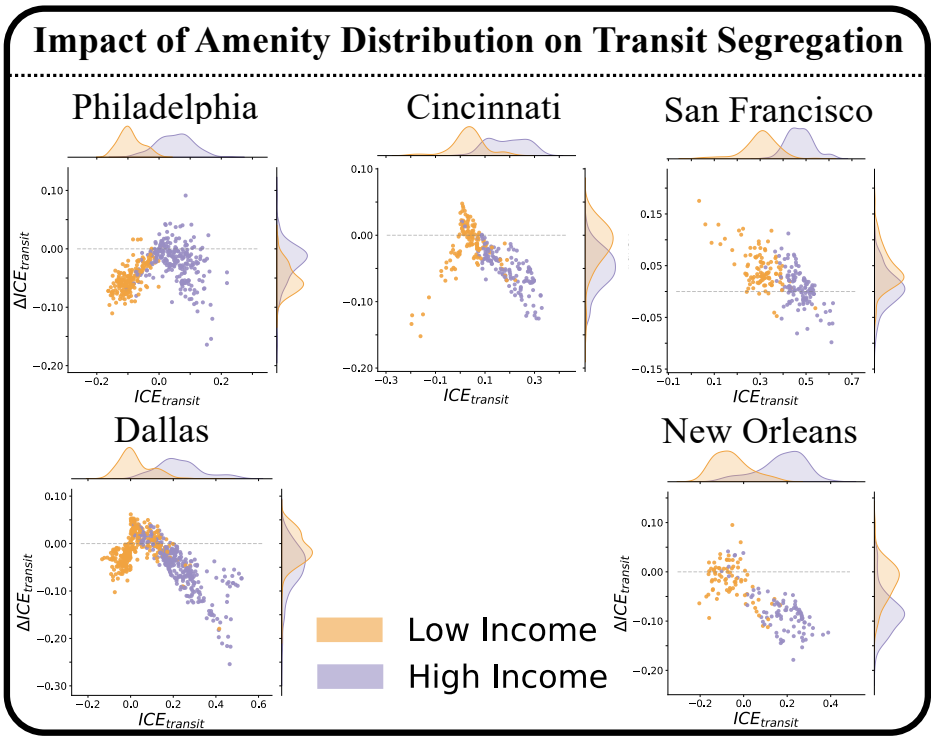

Fig. 5.9 **Inequalities in amenity distribution and visitation patterns contribute to experienced transit segregation, for Philadelphia, Cincinnati, Dallas, and New Orleans.** We compare empirical transit segregation to that of a null model, which is characterised by uniform amenity distributions and visitation patterns, shown on the y-axis. The y-axis compares changes between empirical transit segregation and transit segregation derived from the null model, where negative values reflect amenity landscapes and mobility behaviour contributing to measures of transit use segregation. The x-axis shows empirical levels of transit segregation as a baseline. The purple and orange points refer to segregation levels for the highly-segregated high and low-income groups, respectively.

5.9. The scatter plots shine light on the differences in CBG-level transit segregation between the empirical and null model.

The x-axis and its corresponding distribution above it, convey the empirical transit use segregation, for each CBG in a city. Thus, the distributions above the x-axis more clearly visualise the transit axis in the parallel line plots (middle axes in Figure 5.8). Meanwhile, the y-axis, and respective distribution on the right, emphasise how transit segregation measures change when removing inequalities in amenity visitations and the amenity landscape. This is achieved by calculating $\Delta ICE_{transit}$, as defined in Equation 5.7, comparing a neighbourhood's

empirical transit segregation to that measured in the null model. Negative values reveal that uniform spatial distribution and mobility behaviour (i.e., preferential return) contribute to the empirical transit use segregation.

San Francisco remains an exception, implying that the empirical amenity landscape and socioeconomic inequalities in amenity visitation contribute to lower levels of segregation that individuals experience while using its transport system. For most cities, we observe the majority of neighbourhoods having decreased levels of transit segregation when removing amenity visitation inequalities – indicated by points appearing below the dashed line. These decreases suggest that the spatial distribution of amenities, coupled with inequalities in which demographics are visiting certain amenities. To reduce the impact of transit segregation due to spatial proximity, approaches to mitigating experiential transit segregation can focus on improving service to amenities that exhibit high segregation in visitation patterns. The findings in Figure 5.9 also highlight how experienced transit segregation is not solely an artefact of segregated transit service, but can also be impacted by other urban features, which in this case refers to the spatial distribution of amenities.

Moreover, we observe that the high income groups experience larger decreases in transit segregation for most cities, as seen by the distribution on the y-axis (Philadelphia being an exception). This suggests that the transit segregation experienced by individuals in highly-segregated, low-income neighbourhoods remains consistent, despite the characteristics of their destinations. This finding highlights how transit segregation experienced by highly-segregated, high-income individuals is more shaped by empirical amenity visitation and distribution patterns, than for their low-income counterparts. Coupling this finding with the results in the previous chapter, transit service can specifically aim to better connect individuals from highly-segregated, high-income neighbourhoods to a variety of amenities.

The low magnitudes of $\Delta ICE_{transit}$ in the scatter plots elucidate how removing inequalities in amenity visitations and the amenity landscape does not significantly change segregation in

the transit realm. Thus, we observe that experienced transit segregation is likely to be more significantly shaped by other urban features (i.e. how the socioeconomic residential landscape intersects with the transport service and layout). Ultimately, we identify inequalities in how transit systems connect neighbourhoods from different socioeconomic backgrounds. We compare the average segregation of neighbourhoods that are reachable withing a given time, between trips taken using public transport versus cars. We note that San Francisco and Philadelphia allow residents from different segregation groups to reach a wider array of neighbourhoods within an hour long trip. Moreover, we stochastically model the transit lines individuals would use to satisfy their mobility demands. Finally, we test our empirical results against a null model to find that while disparities in amenity distribution and travel behaviour increase the level of economic concentration on transit lines, mobility and amenity inequalities do not fully account for the level of experienced transit segregation that we do identify.

5.6 Discussion

This chapter aims to shine light on how urban segregation intersects with how transport justice, when considering non-commuting mobility patterns (i.e., amenity visitations). That is, we build upon concepts used to measure residential segregation to derive more dynamic estimates of experienced segregation. Previous works have used mobility data to understand segregation at the amenity level [215, 67, 266, 206]. However, framing experienced segregation in the context of transport justice provides a means for understanding how transit infrastructure provides service to areas characterised by varying levels of segregation. Our work puts forward a framework for defining inequalities in transit systems in terms of *where* transit provides access to and *how* individuals experience segregation while using transport. In doing so, our results reveal that residential segregation levels persist through other aspects of the urban experience, namely amenity visitations and transport usage. These results are

consistent with research that shows residual effects of residential segregation in school, work, and mobility dimensions [215, 67, 266].

Furthermore, we show how highly segregated, low-income neighbourhoods tend to decrease their extreme levels of residential segregation through their mobility patterns. This is in line with findings that unveil demographic associations with social exploration of amenities [206]. Bridgeport and San Francisco serve as two exceptions to this trend, where a subset of the neighbourhoods with low-income segregation tends to visit amenities with high levels of segregation. In these cases, it is imperative to develop adequate urban infrastructure, that is designed to benefit the disadvantaged groups that have low amenity accessibility [5]. We also find that transit systems can hinder access to neighbourhoods, limiting the potential of exposure to individuals from different backgrounds. These results underscore findings that underprivileged demographics, be it immigrant or ethnic minorities, tend to have more constrained activity spaces than their privileged counterparts [121, 266]. It is unclear whether mobility patterns are dictated by limited transit access to other neighbourhoods. However, our findings reveal that by limiting exposure to different types of neighbourhoods, transit systems impose constraints on the activity space and urban experience of individuals, namely those without access to personal vehicles.

Limitations of this work include the assumption that the economic composition of a neighbourhood's travellers directly reflects the neighbourhood's income distribution. Although it is striking that we identify inequalities under this assumption, which removes demographic mobility preferences within a neighbourhood, higher resolution mobility data can provide closer approximations of urban segregation. Thus, this work can be further developed to analyse how segregation experienced within transit lines is impacted by empirically informed levels of socioeconomic transit usage. Moreover, using higher-resolution mobility data, such as those that tag mobility trajectories with the associated demographics of the traveller, could shine light on further disparities in how transport and amenity landscapes intersect.

Additionally, the proposed methodology can be applied to data spanning a larger time frame, to analyse temporal features of mobility and transit segregation. We emphasise that this framework can be applied to any region, given transit feeds for modelling transport networks and mobility data which includes or can be merged with demographic characteristics. This framework can provide a clearer insight as to how cultural differences in mobility patterns and the level of transit infrastructure can impact inequalities experienced in various facets of the urban environment.

In essence, we consider segregation from multiple urban dimensions to highlight the benefit of analysing segregation as an experience rather than a static variable. Moreover, identifying inequalities within transit systems is the first step in providing improved transit service, particularly to individuals from especially vulnerable demographics. By studying segregation from multiple perspectives, we can observe whether mobility is used as a tool to try and overcome residential segregation. Ultimately, this work motivates developing an improved understanding for how transport can facilitate access to different areas in a city, particularly for vulnerable demographics, while also providing sustainable transit modes that are integrated and inclusive.

Chapter 6

Urban Inequalities: A Built Environment Perspective

6.1 Introduction

The previous chapters have evaluated transit inequality with respect to residential-workplace inequalities and socioeconomic segregation. Chapter 4, for instance, revealed how vulnerable housing demographics tend to live closer to the transit system's centre of mass (Figure 4.11). However, a growing number of studies have identified recent trends in the suburbanisation of poverty [128, 243, 127, 167]. Proximity to areas that are rich with transit resources, intuitively, improves accessibility [227]. Yet, urban mechanisms, such as housing and employment landscapes, can impact the areas in which particular demographics can live. As a result, observed transit inequalities may be more reflective of inequalities in housing markets or employment opportunities. To address this issue, this chapter aims to disentangle transit inequalities with respect to the built environment (BE). In doing so, we explore how the extent of inequality changes when we consider disparities in different types of neighbourhoods. This is a critical perspective to consider as vulnerable demographics may face a larger degree of setbacks in less accessible neighbourhoods. Thus, we define a city's

residents with respect to their social and spatial characteristics, to investigate which types of environments tend to express larger transit inequalities. After analysing how transit service relates to socioeconomic features and BE characteristics independently, we combine income and BE features to define socio-spatial demographics. We contend that a just transit system not only provides adequate access to amenities and employment areas, but also provides neighbourhoods with comparable levels of transit service, facilitating opportunities to travel to any other neighbourhoods in the city. Thus, we assess disparities in integral access, system-facilitated access, and employment access, across socio-spatial characteristics. We note that an initial version of the findings presented in this chapter is set to be published in the proceedings of an upcoming conference [135].

6.2 A Built Environment Characterisation of Neighbourhoods

This section outlines our methodology to define the Built Environment (BE) of a neighbourhood by using building **density**, urban street **design** and amenity **diversity**, commonly referred to the 3D's of the built environment [49]. Chapter 3.3 describes our approach to measuring these built form components. As a brief overview, street network design computes how grid-like a neighbourhood road layout is by measuring the geometric mean of the straightness of streets, the street orientation order and the proportion of four-way intersections (Eq. 3.3). Building density measures the proportion of building area with respect to the total land area within a neighbourhood. Finally, amenity entropy captures how even the distribution of types of amenities are in a census tract, by using normalised entropy across ten amenity categories: Food, Education, Healthcare Facilities, Finance, Religious Venues, Government Facilities, Recreational Areas, Entertainment, Retail, and Professional Services (Eq. 3.4). These categories were defined by following the taxonomy used by Graells-Garrido

et. al. [109]. Information pertaining to the street attributes, building footprints, and amenity characteristics of a city are queried for October 11, 2021 from `OpenStreetMap` [221].

We estimate these measures to study transit service in different types of built environments for eight major US cities: New York City, Philadelphia, Chicago, San Francisco, Boston, Portland, Minneapolis and Seattle. We choose these cities as they are all in the top 30 of the most populated cities in the US [43]. Transit *commuters* in New York City, San Francisco, Boston, Chicago, and Philadelphia tend to be from higher income demographics [182, 227] Furthermore, New York City and Boston reflect cities with significant urban development before the popularisation of cars [110]. Meanwhile, Portland and Minneapolis reflect cities in which observed decline in ridership is not associated with the displacement of lower income individuals to the urban periphery whereas Seattle is one of the few cities in which transit ridership has remained consistent, even through the COVID-19 pandemic [83, 227]. Thus, these eight cities represent a range of transit and urban characteristics across populous areas in the US.

Similarly to how we derive the grid index, we define the built environment (BE) index for a census tract, ct, by taking the geometric mean of the tract's grid index, building density, and amenity diversity:

$$BE(ct) = \sqrt[3]{G(ct) * D(ct) * H_{amenity}(ct)} \tag{6.1}$$

where $G(ct)$, $D(ct)$, and $H_{amenity}(ct)$ reflect the grid index, building density, and amenity diversity, respectively. We choose the geometric mean because it allows for comparison across different units and is non-substitutable [291, 194]. Consequently, lower values in one BE feature (i.e. diversity) are not offset by higher values in another (i.e. density, design). Thus, smaller BE features retain their relevance, or weighting, when using the geometric mean, making it an appropriate choice for aggregating BE features into a single value. Section C.1 in the Appendix discusses the validation process, considering other approaches

Table 6.1 Built environment characteristics across eight US cities. The second to fourth columns reflect the average values for each component of the BE index. The fifth column defines the city-level average for the BE index. Finally, the last column indicates the presence of spatial autocorrelation in BE features, for each city, with asterisks indicating the level of significance.

	Mean Built Environment Characteristics				
City	*Grid Index*	*Building Density*	*Amenity Entropy*	*BE Index*	*GSA*
Minneapolis	0.578	0.103	0.647	0.307	**0.718*****
Philadelphia	0.687	0.187	0.598	0.384	**0.683*****
Boston	0.456	0.226	0.621	0.384	**0.330*****
Portland	0.701	0.202	0.700	0.456	**0.442*****
Seattle	0.715	0.228	0.733	0.485	**0.441*****
Chicago	0.829	0.265	0.580	0.491	**0.335*****
New York City	0.765	0.297	0.613	0.505	**0.386*****
San Francisco	0.758	0.354	0.655	0.548	**0.372*****

$^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

to computing the BE index. Given that all of the metrics used to calculate the BE index range from 0 to 1, the BE index is also bounded by this range. Neighbourhoods with larger BE indices suggest that their residents benefit from a greater flexibility in mobility options. That is, grid-like street networks provide a range of routes to travel between two points, they have multiple detour options in the face of traffic or construction, and they are associated with lower pedestrian fatalities and higher transit access [86, 18]. Furthermore, higher building density supplies numerous possibilities for origin and destination points, while larger values of amenity diversity suggest that more types of trip purposes can be fulfilled based on the amenity landscape of a neighbourhood. Thus, higher BE indices suggest more mobility options in terms of routing possibilities, origin-destination points, and trip purposes. We refer to this concept as **mobility flexibility**. Table 6.1 describes the average BE index, and its corresponding components, considering all census tracts in each city.

The fifth column of Table 6.1 reveals that Minneapolis and Philadelphia have lower average BE features, while New York City and San Francisco have larger mobility flexibility. While we can compare built form features between cities, we can also evaluate how neigh-

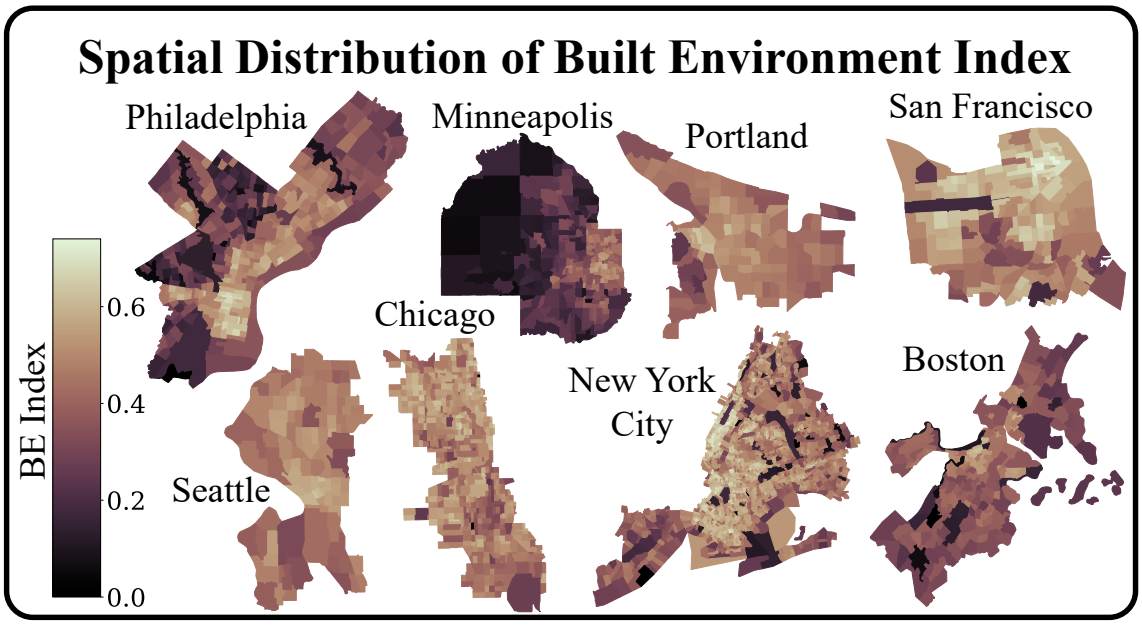

Fig. 6.1 **Spatial distribution of the BE index, across all census tracts in each of the analysed cities.** Lighter hues reflect larger indices, implying a higher flexibility in terms of mobility options.

bourhoods within a given city vary in terms of their physical environments. The last column of table 6.1 denotes the Moran's I value when applying global spatial autocorrelation (GSA) to census tracts in each city. The positive values highlight how neighbourhoods with similar BE indices tend to be located close to one another.

Additionally, the significance of the Moran's I values conveys that when the spatial features of the BE indices are randomly shuffled, spatial autocorrelation is not present. Thus, the asterisks highlight how the identified GSA is not purely an artefact of randomness. Furthermore, Figure 6.1 shows the spatial distribution of BE indices within each city, with darker hues conveying less mobility flexibility through the BE index. Through this figure, we can visualise how cities such as Minneapolis and Philadelphia have larger GSA magnitudes, while cities such as Boston and Seattle exhibit GSA to smaller extent.

Having defined neighbourhoods by their built form features, we can now explore whether socioeconomic transit disparities are present in areas with similar physical environments. We

Table 6.2 Description of how BE features relate to socioeconomic and transit characteristics. Columns two to four describe the average median household income for the Low, Moderate, and High BE terciles, for each city. The second to last column lists the Pearson correlation coefficient between a census tract's BE index and its transit mobility opportunity. The last column shows the correlation between tracts and their average commuting time, assuming all workers use transit. Significance of the correlation coefficients for a city are marked using asterisks, with significant values in bold.

	Average Median Household Income for BE Groups			**Transit-BE Index Correlation**	
City	*Low*	*Moderate*	*High*	*Opportunity*	*Commute*
Chicago	47,007	63,262	77,996	-0.063	**-0.232*****
New York City	69,228	67,960	81,695	**-0.261*****	**-0.064****
Philadelphia	45,991	52,444	58,044	**-0.316*****	**-0.488*****
San Francisco	132,552	124,389	124,613	**-0.467*****	**-0.503*****
Boston	73,774	76,468	80,928	**-0.485*****	**-0.65*****
Seattle	105,579	108,977	102,013	**-0.498*****	-0.16
Portland	77,324	83,793	91,802	**-0.607*****	**-0.208***
Minneapolis	102,004	83,587	67,353	**-0.783*****	**-0.19****

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

incorporate economic characteristics by using Table B19013 in the American Community Survey, which provides the median household income of a given census tract. We create built environment (BE) groups by splitting the distribution of BE indices, for a particular city, into three, equally-sized groups representing the neighbourhoods with low, moderate, and high mobility flexibility. The second to third columns of Table 6.2 convey the average median household income for all tracts belonging to a particular BE group. We observe that in Chicago, Philadelphia, Boston, and Portland, neighbourhoods with lower mobility flexibility (in the Low BE group) tend to have lower median household incomes than neighbourhoods with higher BE features. In contrast, San Francisco, Seattle, and Minneapolis tend to have lower median household incomes in areas with more mobility flexibility. These different trends highlight how BE features and socioeconomic characteristics uniquely intersect for various cities. This could potentially be explained by urban mechanisms such as housing markets or employment opportunities that can limit where individuals can afford to live

[127, 167]. As discussed in Chapter 2.4, transit-dependent individuals, who tend to be lower income, may be especially vulnerable to transport poverty if they live in neighbourhoods with lower mobility flexibility and transit service. In the next section, we incorporate transit data to explore how transit service varies across different BE groups.

6.3 Transit Service in the Context of the Built Environment

This section discusses the transit data sources we use and how we apply them to understand transit service in terms of accessibility and commuting patterns, for different BE Groups. We construct transit networks using GTFS feeds from The Mobility Database and street network data from OpenStreetMap. The time frame for which this data are gathered is October 11, 2021. We use r5py, discussed in Chapter 2.4.2 and 3.4, to query transit routes for Monday, October 11, 2021. To ensure accurate transit time estimates for commuting, we refine the network to model transit flow from 07:30 AM to 09:30 AM. Passing an origin and destination coordinate pair into this transit network will return the median travel time between the points, running five routing iterations for each minute of the defined time frame. Thus, given the transit networks we build for each city, the travel time between two coordinates reflects the median time over 600 Monte Carlo runs.

Then, we use journeys to define transit times for empirical commuting patterns, integral accessibility, and system-facilitated accessibility. As a reminder, integral accessibility, introduced in Chapter 2.4.3, denotes how connected an area is to all other areas in a given boundary. Its counterpart, relative accessibility, highlights how connected two specific locations are, by estimating transit travel times between the two locations. By measuring public transit service in terms of how it provides access to every neighbourhood in a region (integral accessibility), we aim to estimate how a city's transit system provides individuals with the opportunity to travel to a range of destinations. In order to arrive at these estimates, we randomly sample 100 origin and destination points, for every potential census tract pair

in a city. We pass these potential mobility trajectories into the transit networks to find the weighted shortest path between each origin-destination (OD) pair, (i, j). We use $J'_{transit}(i, j)$ to describe the time it takes, in minutes, to travel on a transit route from origin point i to destination point j. Then, we aggregate these values to the census tract level, by averaging the transit times for all origin-destination pairs beginning in census tract ct_o and ending in census tract ct_d:

$$J_{transit}(ct_o, ct_d) = \frac{\sum_{i \in ct_o, j \in ct_d} J'_{transit}(i, j)}{100} \tag{6.2}$$

where $J_{transit}(ct_o, ct_d)$ reflects the average transit time it takes to travel between two census tracts in a city. Thus, we can define the transit mobility opportunity ($T^{opp}_{transit}$) for a given census tract ct:

$$T^{opp}_{transit}(ct) = \frac{\sum_{ct' \in \mathbf{CT}} J_{transit}(ct, ct')}{|\mathbf{CT}|} \tag{6.3}$$

where **CT** reflects the set of census tracts in a city. The values measured by $T^{opp}_{transit}(ct)$ convey transit mobility opportunities, in that it estimates, for a given neighbourhood, the average time it takes to travel, considering all other neighbourhoods in the city. In essence, Equation 6.3 measures integral accessibility. The fifth column of Table 6.2 shows the Pearson correlation coefficients between the mobility opportunity of a census tract ($T^{opp}_{transit}(ct)$) and the tract's BE index, with asterisks indicating the significance of the coefficient. The negative coefficients suggest that neighbourhoods in a city that have higher mobility flexibility (or BE indices) tend to have smaller average transit times to all other neighbourhoods. Chicago is the only city for which a significant, negative correlation is not present.

Furthermore, we use 2021 commuting data from the LODES data set, discussed in Chapter 3.2.1, to measure the areas to which individuals in a census tract commute. Then,

we capture the average commuting time for a census tract:

$$T^{comm}_{transit}(ct) = \frac{\sum_{w \in \mathbf{W}_{ct}} J(ct, ct_w)}{|\mathbf{W}_{ct}|} \tag{6.4}$$

where $\mathbf{W}_{ct}$ is the set of all workers living in census tract ct and ct_w captures the census tract that a worker w in $\mathbf{W}_{ct}$ commutes to. Thus, $J(ct, ct_w)$ conveys the transit commute time for a given worker that lives in ct. Accordingly, $T^{comm}_{transit}(ct)$ denotes the average transit commute time for all workers that live in a census tract ct. Similar to the analysis in the previous chapters, this metric is calculated under the assumption that all commuters use transit to reach their workplace. However, we argue that this assumption is relevant to transit and climate goals. That is, $T^{comm}_{transit}(ct)$ does not try to estimate real commuting times with respect to empirical commuting mode choices. Instead, $T^{comm}_{transit}(ct)$ aims to measure employment accessibility using transit, in the context of shifting to more sustainable travel modes. Table 6.2 evaluates the correlation between a neighbourhood's average transit commute time and it's BE index, using the Pearson correlation coefficient.

In line with the identified association between integral access and the built environment, we find that all cities, barring Seattle, have negative, significant associations between commuting and the built environment. That is, neighbourhoods with higher BE indices tend to have shorter commute times. It is important to note, that while New York City has a significant and negative coefficient, its low magnitude suggests a weak correlation. Furthermore, Figure 6.2 shows the distribution of mobility opportunity ($T^{opp}_{transit}(ct)$) and transit commuting times ($T^{comm}_{transit}(ct)$) for census tracts in each BE group.

In Figure 6.2, we can observe that transit times, in terms of opportunity and commuting, tend to be longer for neighbourhoods with low BE features. Meanwhile, areas with high mobility flexibility benefit from shorter transit times for accessing all other neighbourhoods as well as empirical employment destinations. This is not necessarily the case when considering mobility opportunity in Chicago and New York City. Moreover, this trend is not apparent

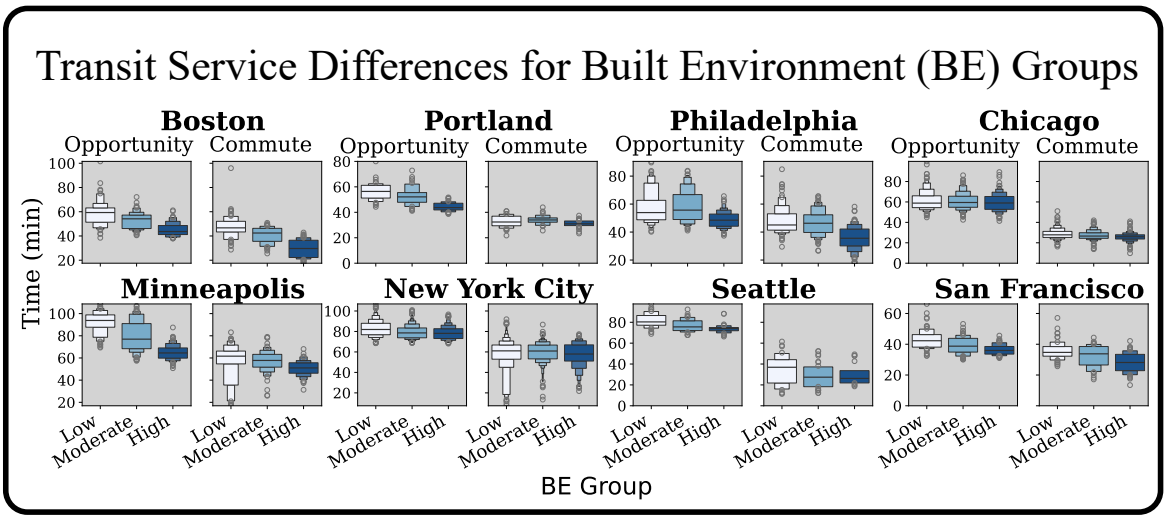

Fig. 6.2 **Generally, neighbourhoods in the High BE group tend to have shorter transit times compared to those in the Low BE group, in terms of commuting and general mobility opportunity.** The left most plot for each city reflects the distribution of integral transit access (mobility opportunity) across Low, Moderate, and High BE Groups. The right most plot shows average transit commuting times across the same BE groups. The BE groups are further distinguished by hues of blue, with the darkest hue depicting the High BE group and the lightest representing the Low BE group.

in the context of commuting times for Seattle, New York City, Chicago, and Portland. A potential explanation for this is that, in these four cities, housing, employment, and transit mechanisms may align to provide either short distances between residential-workplace areas for all neighbourhoods or strategic transit service that connects residential and employment areas, based on commuting workflows.

The differences in transit times distribution across BE groups can be validated from a statistical perspective in Tables 6.3 and C.2. These tables confirm the results discussed in the previous paragraph. To accomplish this, we use the Kolmogorov-Smirnov test to compare the similarity of distributions. By measuring the maximum absolute difference between the empirical cumulative distribution functions for each sample, we can determine whether to reject the null hypothesis that the two samples come from the same distribution. Thus, Table 6.3 reveals cities that have statistically significant distributions of transit times between Low

City	**Opportunity** Low-High	**Commute** Low-High
Boston	**0.618*****	**0.77*****
Portland	**0.76*****	0.309
Philadelphia	**0.344*****	**0.543*****
Chicago	0.114	**0.271*****
Minneapolis	**0.87*****	**0.537*****
New York City	**0.228*****	**0.121*****
Seattle	**0.667*****	0.3
San Francisco	**0.56*****	**0.502*****

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 6.3 Kolmogorov-Smirnov test statistic when comparing the distribution of transit times between the Low and High BE group, for mobility opportunity transit times (second column) and transit commuting times (last column). Bold values reflect instances in which the compared distributions are not the same.

and High BE neighbourhoods, which is reflected by significant values (bolded) that have high magnitudes.

6.4 Expressions of Transit Inequality in Different Built Form Types

So far, we have defined neighbourhoods based on their built environment features and the extent of transit service they receive. The previous section showed how integral accessibility and commuting transit times relate to BE characteristics in eight US cities. This section incorporates socioeconomic data from the US Census Bureau's American Community Survey to assess how socioeconomic transit inequality varies across BE groups. In doing so, we derive socio-spatial groups that are defined by a neighbourhoods BE group and income terciles. We begin by applying methods from network science to understand socioeconomic transit disparities, in the context of integral accessibility. That is, we leverage network assortativity to understand whether socioeconomic homophily is present in transit service

Table 6.4 Distribution of the number of census tracts in each socio-spatial group. Each column reflects a socio-spatial demographics, while rows denote a city.

Socio-Spatial Group / **City**	Low Income			Middle Income			High Income		
	Low BE	Moderate BE	High BE	Low BE	Moderate BE	High BE	Low BE	Moderate BE	High BE
Boston	18	17	14	20	20	20	18	18	21
Chicago	120	52	36	60	89	66	35	73	112
Minneapolis	9	30	47	26	34	28	54	25	14
New York City	201	184	204	237	229	185	193	217	242
Philadelphia	54	36	27	41	39	40	23	42	51
Portland	18	7	9	7	15	10	6	9	12
San Francisco	16	15	22	18	15	14	14	18	12
Seattle	13	9	10	7	8	11	5	8	4

networks that are created for different travel time thresholds. Then, we incorporate relative accessibility by comparing differences in access to essential amenities across socio-spatial groups. Finally, we compare socio-spatial inequalities, when considering the accessibility of employment areas.

6.4.1 Defining Socio-Spatial Demographics

In previous chapters, we split neighbourhoods into five, equally-sized quintiles, based on their median household income. However, in this chapter, we combine income and BE groups to determine socio-spatial demographics. Accordingly, we split the median household income distribution of each city into three, equally-sized terciles, which represent Low, Middle, and High income groups. We choose income terciles rather than income quintiles so that both BE and income groups are derived using terciles. Moreover, for cities with a particularly low number of census tracts, such as Minneapolis, Seattle, and Portland, creating 15 socio-spatial groups leads to an insignificant number of neighbourhood within each demographic. Table 6.4 conveys the number of neighbourhoods in each socio-spatial demographic.

The first column in Table 6.4 indicates which city is being analysed. The remain columns reflect the number of census tracts in a given socio-spatial demographic, with the socioeconomic and spatial (BE) group defined by the first and second row, respectively. Table 6.4 shows how even when we create 9 socio-spatial demographic groups, Seattle only has four high-income neighbourhoods that have high mobility flexibility. Thus, when interpreting the following results, it is crucial to keep in mind the neighbourhood representation across socio-spatial demographics in each city, particularly in the cases of Seattle and Portland.

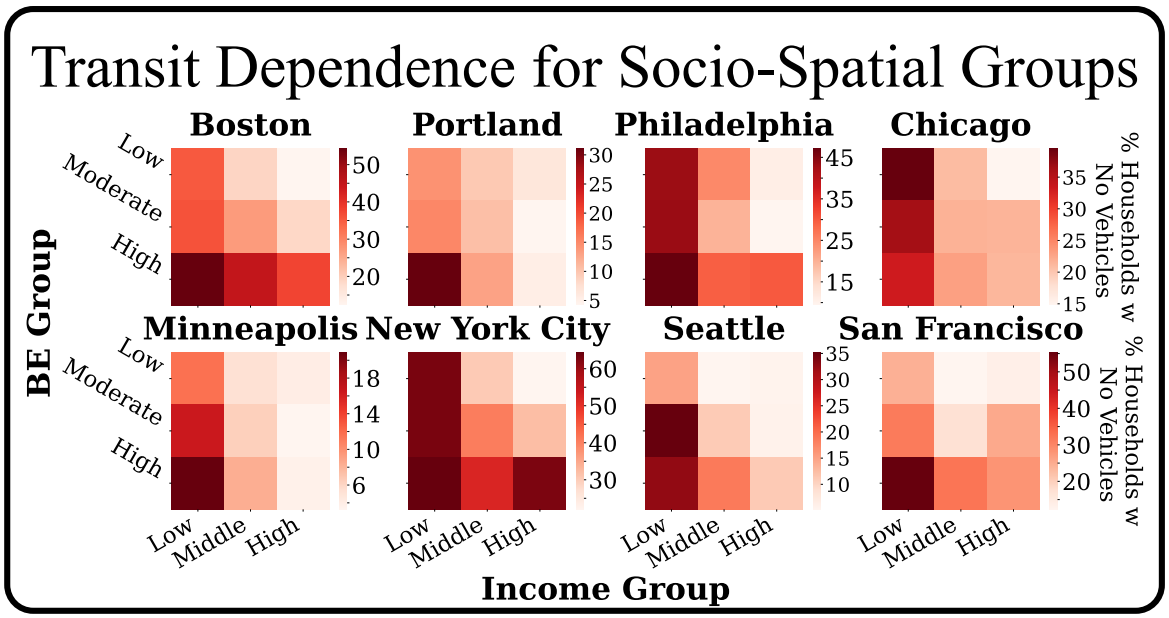

Fig. 6.3 **High BE neighbourhoods tend to express higher transit dependency, while high-income neighbourhoods tend to have lower rates of transit dependence.** The y-axis of each heat map reflects the spatial demographic (BE groups), while the x-axis represents the socioeconomic demographic (income group). Accordingly, each cell represents a socio-spatial demographic, with darker hues of red indicating a higher percentage of households that have no vehicles.

Figure 6.3 shows how transit dependence varies across the 9 socio-spatial demographics that we define, for all eight cities. The x-axis reflects the three income groups (Low, Middle High) and the y-axis represents the Low, Moderate, and High BE Groups. The colour bar indicates the percentage of households in a given socio-spatial demographic that do not own any vehicles. We observe that the left most column of each city's grid tends to be the darkest,

suggesting that within the same spatial demographic, the lower socioeconomic group tends to have more households that are dependent on transit, taxis, or potentially ride-hailing services. When we consider each row of a city's heat map, we see that car ownership is higher in Low BE groups, where transit service times are larger, as shown by Figure 6.2. Ultimately, Figure 6.3 aims to provide context to rates of potential transit dependence among socio-spatial demographics, highlighting how low-income neighbourhoods generally have lower rates car ownership across all spatial demographics. Throughout the following sections, we compare transit inequalities between groups at opposite ends of a particular demographics. That is, we analyse transit disparities across the highest and lowest socioeconomic groups, in areas with the most and least mobility flexibility, as measured by BE groups.

6.4.2 Integral Accessibility

The previous section shines light on potential transit dependence across socio-spatial demographics, underscoring a notable percentage of households, in areas with less mobility flexibility, that do not have access to cars, particularly in neighbourhoods that are low-income. Accordingly, this section focuses on exploring how transit systems connect neighbourhoods of different socioeconomic backgrounds, with respect to the built environment. In this manner, integral accessibility can convey overall mobility opportunity, which is measured using $T^{opp}_{transit}(ct)$. Integral accessibility, as a reminder, measures how transit serves a particular neighbourhood, with respect to all other destinations in a city. Previous chapters analysed mobility opportunity in order to define transit efficiency (Chapter 4.5) and the socioeconomic composition of neighbourhoods that are accessible by transit (Chapter 5.5.1). We return to this concept, with a different question at hand: Are there socioeconomic disparities in how transit systems facilitate mobility opportunity for various spatial demographics? To answer this question, we analyse how accessing neighbourhoods of different socioeconomic backgrounds changes across spatial demographics, when using transit.

We accomplish this by constructing transit service networks, which reflect neighbourhoods that are reachable within a given time threshold. We build a network, $G_{BE}(t)$, in which nodes represent census tracts in a given city. An edge from node ct_1 to node ct_2 indicates that census tract ct_2 is reachable from tract ct_1 within a travel time threshold of t minutes. Similar to the approach in Chapter 4.2, we create 'disaggregated' networks that reflect transit service from nodes belonging to a particular built environment group. That is, we create a network $G'_{BE}(t,b)$, such that nodes are the same set of census tract as in $G_{BE}(t)$. Furthermore, we only add edges between a pair of nodes, (ct_1, ct_2), if ct_1 is a tract in built environment group, b, and the transit travel time from ct_1 to ct_2 does not exceed the time threshold, t.

We leverage the income terciles to inform node attributes, in both $G_{BE}(t)$ and $G'_{BE}(t,b)$. Specifically, each node in a network represents a census tract and is, accordingly, assigned a node attribute reflecting the income group to which the respective tract belongs. Then, we leverage network properties to analyse whether transit systems connect neighbourhoods in the same income group. To do so, we apply network assortativity, which captures the tendency of nodes to connect with 'similar' nodes. By measuring similarity using income groups, we can assess whether transit service connects neighbourhoods that have analogous economic characteristics to a larger extent than neighbourhoods with different economic features. This is accomplished by creating a normalised mixing matrix, $M_{BE}[i][j]$, which captures the number of edges from nodes in income group i to nodes in income group j. This matrix, M_{BE} is normalised by dividing by the total number of edges in the graph, G_{BE}. Then, we use a_i and b_i to define the proportion of outgoing and incoming edges to nodes in income group i:

$$a_i = \sum_{j}^{\mathbf{I}} M[i][j], \quad b_i = \sum_{j}^{\mathbf{I}} M[j][i] \tag{6.5}$$

where **I** is set of three income groups (Low, Middle, and High). Finally, we can define the assortativity coefficient based on the income attributes of each node:

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} \tag{6.6}$$

where $\sum_i e_{ii}$ reflects the proportion of accessible edges that provide transit access between neighbourhoods in the same income group. In a perfectly assortative network, that is, a network in which transit only provides service between neighbourhoods in the same income group, r will be equal to one. Contrarily $r = 0$ when there is no assortative mixing. Perfectly dissasortative networks, in which transit does not connect any neighbourhoods in the same income group, will have negative values for r, since the term $\sum_i e_{ii}$ will be 0. While the range is bounded by -1, perfectly disassortative networks are similar to random mixing (where $r = 0$), thus the magnitude of a perfectly disassortative network should be closer to 0 than that of a perfectly assortative network [212].

By applying network assortativity to built environment networks, $G_{BE}(t)$, for a given time threshold, t, in a city, we can assess how transit systems connect neighbourhoods of similar economic backgrounds. We note that using transit travel times to connect neighbourhoods, rather than considering direct transit lines, allows us to capture the experience and realities of transit service, as opposed to their structural features. The grey lines in Figure 6.4 show how the network assortativity coefficient (r), shown along the y-axis, changes for different travel time thresholds. The time thresholds along the x-axis range from 5 minute to 90 minute transit journeys. For all eight cities, we observe a decrease in assortativity as the travel time threshold increases. Networks constructed to connect neighbourhoods that are reachable within 90 minutes tend to have an assortativity that approaches 0, with Seattle have the largest value (0.032) and Philadelphia having the lowest value (-0.003). This implies that neighbourhoods of different socioeconomic backgrounds are reachable, given that individuals have a travel budget of 90 minutes. A travel budget of 30 minutes, however,

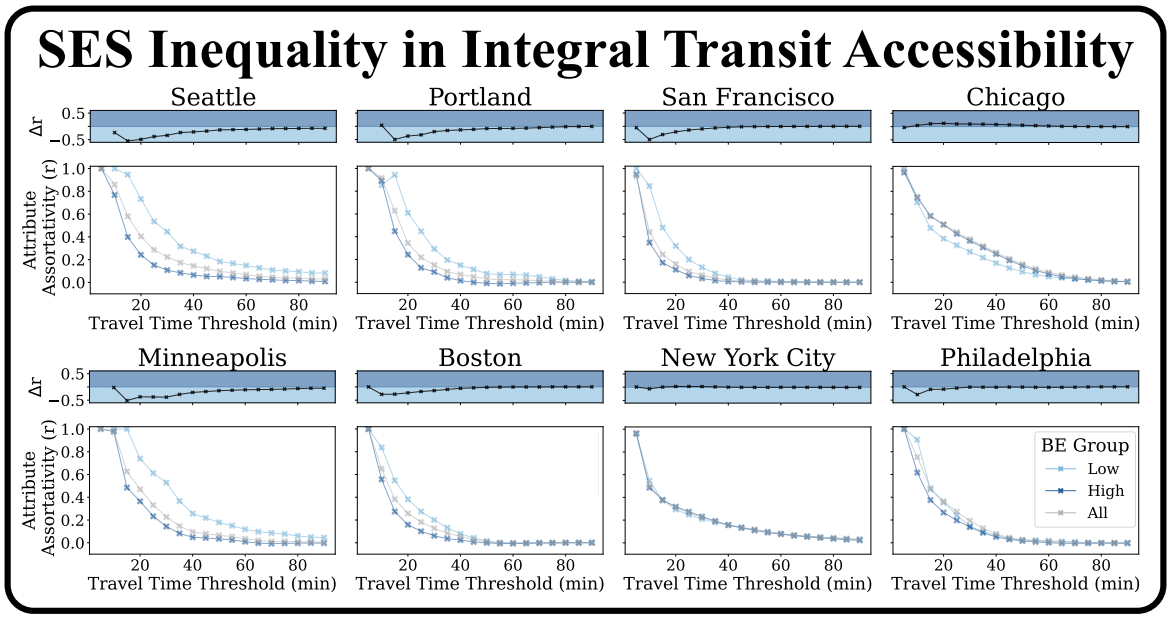

Fig. 6.4 **Socioeconomic homophily in transit service decreases with longer travel times, but tends to decrease at a faster rate for neighbourhoods in High BE areas.** The bottom panel for each city conveys how socioeconomic homophily changes given a travel time threshold, shown by the grey line. The light and dark blue lines depict whether transit serves neighbourhoods of similar economic compositions, for tracts in the Low and High BE group, respectively. The top panel illustrates the difference between the assortativity of the Low and High BE group for each time threshold, with negative values indicating more socioeconomic homophily in transit service for neighbourhoods in the Low BE group.

shows socioeconomic homophily in transit service between census tracts, with Chicago having the highest coefficient of 0.378 and San Francisco exhibiting the lowest coefficient of 0.062. Thus, Figure 6.4 conveys that,while all cities express a decreasing behaviour in assortativity as the time threshold increases (grey line), each city decreases at its own rate.

We can further extend this analysis by leveraging the networks, $G'_{BE}(t,b)$, introduced earlier. $G'_{BE}(t,b)$ is identical to $G_{BE}(t)$, except that edges in $G'_{BE}(t,b)$ only exist if the source node belongs to the BE group, b. By creating these networks, we can see how transit connects neighbourhoods of different socioeconomic backgrounds, with a specific focus on integral accessibility for neighbourhoods in the Low and High BE groups. Accordingly, we apply network assortativity to both $G'_{BE}(t,Low)$ and $G'_{BE}(t,High)$, for all eight cities. The assortativity coefficients for the Low and High BE groups are shown with the light blue and

dark blue lines in the bottom panels of Figure 6.4, respectively. The top panel compares the assortativity between the Low BE and High BE networks (Δr), with negative values denoting more socioeconomic homophily in transit service for neighbourhoods with lower built environment features. We observe a decreasing trend of network assortativity, similar to our analysis of the entire transit service network.

For each city, we maintain the same number of nodes (which represent census tracts), regardless of the travel time threshold we are considering. It should also be noted that each income group (i.e. low, middle, and high) has been created such that they consist of an equal number of census tracts (nodes). However, a city's transit access network is expected to have fewer edges for a smaller travel time threshold than for a larger one. As a result, networks associated with smaller travel time thresholds are expected to have more components, since less neighbourhoods are reachable from each node. This raises concerns when comparing assortativity for two reasons. First, the relationship between edge density and giant component formation depends on the degree of network assortativity [212]. Second, with this assumption of the giant component formation in mind, the largest component for networks of smaller travel time thresholds may not have equal group representation of the income group as it is representing a subset of neighbourhoods that may be skewed to a particular demographic. Both these effects can misconstrue measures of network assortativity, making comparisons across different time thresholds invalid due to variations in group representation, edge density, and the number of components [212, 146].

To address these concerns, we build a configuration model to test whether the observed assortativity coefficients in Figure 6.4 are merely reflecting the change in number of components or a change in group representation. For a given city, we build a configuration model $G_{BE}^{conf}(t)$ for every graph $G_{BE}(t)$, such that $G_{BE}^{conf}(t)$ has the same in-degree and out-degree distribution as $G_{BE}(t)$. For each graph, $G_{BE}(t)$, we construction 100 configuration models based on the empirical degree distribution. Then, we can compare, for each city, the dif-

Table 6.5 Understanding the relationship between spatial autocorrelation of income and differences in socioeconomic homophily between the Low and High BE group. The second column reflects the average difference in the assortativity coefficient between Low BE and High BE transit access networks. Meanwhile, the last columns conveys Moran's I as an indicator of the degree of socieoconomic spatial autocorrelation.

City	$\overline{\Delta r_{t\leq 10}}$	GSA
Chicago	0.001	**0.761*****
New York City	0.016	**0.626*****
Minneapolis	-0.05	**0.513*****
Philadelphia	-0.076	**0.54*****
Boston	-0.081	**0.431*****
Portland	0.157	**0.26*****
Seattle	-0.212	**0.311*****
San Francisco	-0.28	**0.422*****

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

ference in network assortativity between $G_{BE}(t)$ and its configuration model counterpart $G_{BE}^{conf}(t)$. The solid blue and yellow lines in Figure C.5 reflect network assortativity of the empirical graph ($G_{BE}(t)$) and the mean assortativity, consider all 100 of $G_{BE}(t)$'s respective configuration model ($G_{BE}^{conf}(t)$). The dotted blue lines highlight how both the configuration model and empirical graph have a similar decay of weakly connected components as the . Regardless, the decay of assortativity is unique to empirical graph, with the configuration models express assortativity coefficients near zero. This result can also be observed when comparing the networks that measure socio-spatial homophily in transit access networks ($G'_{BE}(t,Low)$ and $G'_{BE}(t,High)$) to their respective configuration models. (Figures C.6 and C.7)

However, by incorporating socio-spatial demographics into the transit network, we reveal that Seattle, Minneapolis, Portland, Boston, and San Francisco express a higher degree of socioeconomic homophily for neighbourhoods in the Low BE Group, compared to that of the High BE neighbourhoods. This can be seen by the negative Δr values in the top panel of each city, for travel time thresholds less than 30 minutes. Δr measures the difference

in assortativity between $G'_{BE}(c,t,High)$ and $G'_{BE}(c,t,Low)$, such that positive values imply larger assortativity coefficients for the High BE Group.

Notably, by combining the results from both panels, for each city, we can observe that within the 5 and 10 minute travel time threshold, differences in assortativity between BE groups tend to be minimal, while the assortativity values, themselves, are close to 1, revealing high assortativity. Thus, we note that within a 10 minute travel time, transit service expresses trends of socioeconomic homophily, regardless of BE characteristics. We hypothesise that high r values, yet low Δr estimates could be an artefact of spatial autocorrelation of socioeconomic status. That is, shorter travel time thresholds are reflective of not just transit service, but also the immediate surroundings of a given neighbourhood. This pattern can be observed in Table 6.5. The second column reflects average Δr values for travel time thresholds of ten minutes or less, denoted by $\overline{\Delta r_{t\leq 10}}$. This column conveys which cities have the least differences in network assortativity between BE groups (Δr), while still having high levels of socioeconomic homophily in transit service ($t \leq 10$). The third column in Table 6.5, however, shows the global spatial autocorrelation, considering tract-level median incomes for each city. Asterisks refer to the significance of Moran's I, compared to randomly shuffled median income levels. By comparing the second and third columns, we see initial signs that our hypothesis is valid, as areas with lower differences in network assortativity across BE groups and high levels of assortativity, tend to have higher magnitudes of spatial autocorrelation, as denoted by Moran's I. That is, less socio-spatial disparities in transit service for lower time thresholds are associated with higher socioeconomic spatial autocorrelation, revealing how network assortativity for shorter travel time thresholds are reflective of both transit service features and socioeconomic characteristics of nearby neighbourhoods.

Ultimately, this analysis considers the set of neighbourhoods that are reachable within a given time threshold, and reveals that socioeconomic homophily is present in transit service across all cities, with socio-spatial disparities between BE groups arising for five of the

eight cities. Consequently, these results emphasise how transit infrastructure not only has longer travel times for neighbourhoods in the Low BE group, but also provides access to neighbourhoods of similar economic backgrounds. Thus, we show how low-income neighbourhoods in the Low BE category face mobility constraints, in that transit does not provide options to travel to neighbourhoods of higher income, within a 30 minute journey. In the next section, we shift our focus from integral accessibility (in which we consider all possible destinations) and focus on system-facilitated accessibility in each neighbourhood.

6.4.3 System-Facilitated Accessibility

Having assessed differences across spatial demographics in how transit systems provide mobility opportunities to access areas of different socioeconomic backgrounds, we consider how transit provides service for different types of journeys. Thus, we build upon the spatial inequalities in transit systems that were explored in previous chapters, to examine how system-facilitated transit access levels vary across socio-spatial demographics, with respect to amenity visitations (Chapter 5) and commuting trips (Chapter 4). In this section, we include system-facilitated transit accessibility into our analysis by measuring how many amenities are accessible within a travel time threshold.

Previous studies that have established the concept of the 15-minute city, in which accessibility is derived as a measure of reachable amenities within a 15 minute walk, cycle, or transit trip [205, 305, 109]. Thus, we set a travel time threshold of 15 minutes. We focus on 'essential' amenities, as defined by Elldér et. al., who show how providing a basic supply of access to essential amenities incentivises more sustainable modes of travel [81]. Thus, essential amenities are comprised of grocery stores, schools, and pharmacies. It is crucial to note that amenity features are already incorporated into the BE index metric, in the form of amenity diversity. The distinction between amenity diversity and essential amenity access is that the former metric measures the variety of amenity types that are present in a

neighbourhood. Whether or not residents in a neighbourhood visit each amenity type is not of importance, as amenity diversity reflects the opportunity to satisfy a range of trip purposes. Meanwhile, the latter metric (essential amenity access) focuses on specific amenities, on which we assume the majority of the population depends. Moreover, by concentrating on the number of amenities, rather than how amenities are distributed over different categories, we incorporate the notion that amenities vary in quality. Thus, a larger number of amenities suggests that individuals have more choice when visiting essential amenities. OpenStreetMap provides amenity locations with respect to these three categories.

Furthermore, we implement an approach, similar to our mobility opportunity methodology, in which we retain the 100 randomly sampled origins for every census tract, but define destinations as the aforementioned amenities. For every origin point in a census tract we calculate the travel time to each point of interest in an amenity category:

$$J_{transit}(ct,a) = \frac{\sum_{i \in ct} J'_{transit}(i,a)}{100} \tag{6.7}$$

where $J_{transit}(ct,a)$ denotes the average time it takes to reach an amenity a from a census tract, ct, considering all 100 of the randomly sampled points within the tract. Considering all the amenities belonging to an amenity category (pharmacies, grocery stores, or schools), we can determine the number of amenities that are accessible within 15 minute transit journey from census tract, ct:

$$RA(ct,\mathbf{A}_{cat}) = \left| \left\{ J_{transit}(ct,a) \middle| \left(J_{transit}(ct,a) \leq 15 \right) \wedge \left(a \in \mathbf{A}_{cat} \right) \right\} \right| \tag{6.8}$$

where $\mathbf{A}_{cat}$ is the set of all amenities in an amenity category, cat. Thus, $RA(ct,\mathbf{A}_c)$ reflects the number of amenities of type cat, that are reachable within a 15 minute transit journey from tract ct.

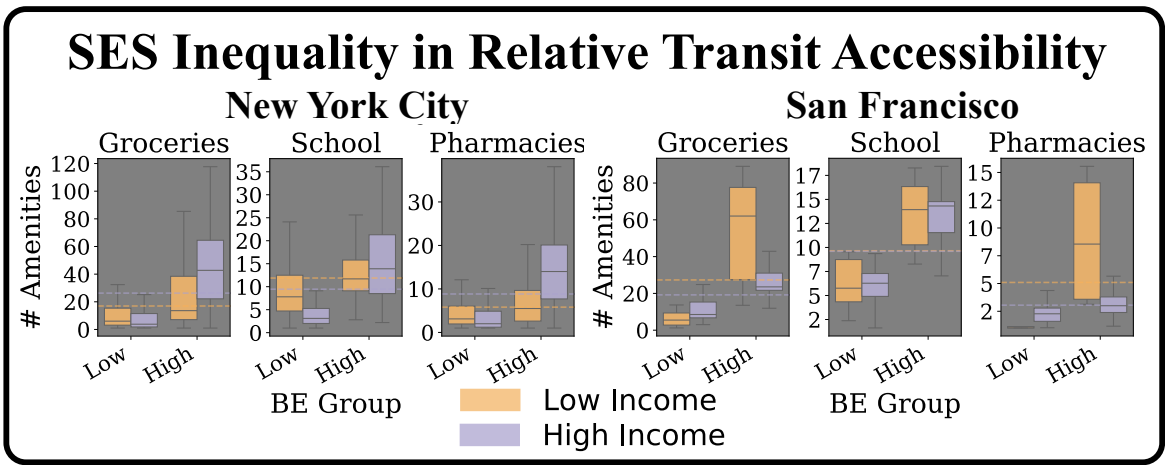

Fig. 6.5 **Socio-spatial inequalities in system-facilitated access to essential amenities reveals, for a given city, the type of built environment in which transit disparities arise.** From left to right, each panel depicts access to grocery stores, schools, and pharmacies, within a 15 minute transit ride, for New York City and San Francisco. The y-axis shows the number of amenities that are accessible for each amenity category. Socio-spatial demographics are reflected through the x-axis, which delineates the spatial aspect (BE groups) and the colour of the box plots, defining the socioeconomic component. Orange and purple box plots reflect the low and high-income group, respectively, for the given BE group.

Figure 6.5 uses New York City and San Francisco to highlight the importance of considering system-facilitated transit access with respect to BE characteristics. Figure C.8 in the Appendix shows the results for the other six cities. The x-axis captures the Low and High BE groups for both cities, while the orange and purple box plots reflect Low and High income groups, respectively. The y-axis shows the number of amenities that are accessible for each socio-spatial demographic. For both cities, regardless of which essential amenity we consider, we observe that the High BE group tends to have access to a larger number of amenities than the Low BE group does. The orange and purple horizontal lines depict the median number of amenity accessible within 15 minutes for the low and high income group, respectively. The horizontal lines disregard BE groups, only accounting for the socioeconomic composition of neighbourhoods.

In this manner, New York City serves as an example of a city in which the Low BE, Low income group tends to have access to more amenities than the Low BE, High income

group. This can be seen through the left two box plots in each plot. However, when we consider socioeconomic transit disparities for neighbourhoods with high mobility flexibility, we see that the lower income neighbourhoods have lower system-facilitated transit access than higher income neighbourhoods. Figure 6.5 elucidates that New York City transit service, in areas with higher degrees of design, density, and diversity, provides access to more critical amenities for higher income neighbourhoods. The socioeconomic disparities in accessing grocery stores and pharmacies, for low-income groups (shown by the dashed horizontal lines) reveal how transit inequality is not an artefact of inadequate transit service in neighbourhoods with low mobility flexibility. In this manner, transit planning can specifically target the spatial demographics in which socioeconomic inequalities in system-facilitated access occur. San Francisco depicts the opposite scenario, in which the higher-income neighbourhoods with less mobility flexibility have more access to amenities than their lower-income counterparts with similar BE features. Meanwhile, the High BE, Low income neighbourhoods have greater access to essential amenities than the High BE, High income census tracts. In this case, transport systems can improve equality in accessibility by focusing efforts on improving service for lower income neighbourhoods in areas with lower built environment features (mobility flexibility).

By comparing the socio-spatial disparities in system-facilitated access to the median socioeconomic disparities (horizontal, dashed lines), we highlight the need to consider BE features when analysing transit access. That is, in San Francisco, the high income group tends to have lower system-facilitated transit access to essential amenities, when we do not account for BE groups. Yet, the box plots reveal that lower income neighbourhoods in the Low BE groups have less amenity access. These results align with other studies that show how transit disparities arise for low income neighbourhoods that are further from the urban core [5]. When interpreting system-facilitated transit access to grocery stores and pharmacies, New York City appears to have less overall access for low-income neighbourhoods, regardless of

	Groceries		Schools		Pharmacies	
City	Low BE	High BE	Low BE	High BE	Low BE	High BE
Boston	63.0	227.5	**200.0****	250.5	60.0	104.0
Chicago	**1486.0****	**457.0*****	**3016.5*****	1818.0	**806.0*****	**667.0*****
Minneapolis	**265.0****	473.0	**195.5****	261.5	27.0	**267.5***
New York City	**21564.0****	**12877.0*****	**31664.5*****	**20489.5***	**19118.0****	**10684.5*****
Philadelphia	**1202.0*****	522.5	**1285.0*****	705.5	701.5	436.0
Portland	103.0	56.0	52.0	19.0	48.5	45.0
San Francisco	**101.0***	**132.0***	170.0	83.5	**35.0*****	**148.0****
Seattle	94.0	**20.0***	49.0	12.0	50.0	13.0

$*p < 0.05$; $**p < 0.01$; $***p < 0.001$

Table 6.6 Mann-Whitney U test statistic when comparing the distribution of transit times between the low and high income groups, for a given BE group and essential amenity. Bold values reflect instances in which the medians of compared distributions are not the same. Thus, bold values imply inequalities in essential amenity access via transit, for the respective socio-spatial demographic.

BE characteristics. Yet, we can observe that this disparity is largely due to transit inequalities within the High BE group. We confirm the observed differences in relative transit accessibility across socio-spatial demographics by using the Mann-Whitney U test. For a given city (row) in Table 6.6, we highlight the BE group for which socioeconomic disparities exist in essential amenity access. Although both the Mann-Whitney U test and the Kolmogorov-Smirnov test do not assume that samples follow a normal distribution, the Mann-Whitney U test is more powerful at comparing distributions with small sample sizes. While this was not relevant when comparing transit time distributions across spatial groups, focusing on socio-spatial groups leads to smaller sample sizes, as seen by Table 6.4. However, The Mann-Whitney U test focuses on comparing the median of the two samples, while the Kolmogorov-Smirnov test compares the shape and spread. The Mann-Whitney U test works by ranking the combined samples and then comparing the sum of the ranks in each sample. The bold values in Table 6.6 reflect BE groups for which we can reject the null hypothesis that the medians of each income group's distribution are not equal.

The findings from this section underscore how considering BE features can help pinpoint which type of neighbourhood suffer from transit disparities. Accordingly, transit planning efforts can be shaped around the observed socio-spatial disparities in transit service, for a given city. Having considered inequalities in system-facilitated and integral transit access, with respect to BE and socioeconomic features, we conclude this chapter by analysing socio-spatial transit disparities in accessing workplaces.

6.4.4 Socioeconomic Inequality in Commuting via Transit

Thus far, we have shown the importance of including BE features when analysing socioeconomic transit inequalities, with respect to integral access and system-facilitated access to essential amenities. Having highlighted how areas with higher mobility flexibility tend to receive faster transit service, this final section returns to the concept of the spatial mismatch hypothesis, which refers to a disconnect between residential and employment locations that negatively impacts more vulnerable demographics [143, 102]. In particular, we leverage socio-spatial demographics to assess the types of neighbourhoods in which lower income workers have poor access to their employment opportunities. In Chapter 4.6.2, we show how workers in neighbourhoods that are vulnerable to housing insecurity, generally have transit commute times of over half an hour to their workplaces. Furthermore, we highlight how vulnerable housing demographics would have even longer transit commutes to access jobs that are associated with better opportunities, with most of the commute times being over an hour. While this section does not consider housing demographics, we reveal socioeconomic inequalities in job accessibility, focusing on the spatial groups in which individuals from lower income neighbourhoods tend to face low employment accessibility.

Since 2010, the average travel time to work has increased from 25 to 28 minutes [41]. Furthermore, this report shows how, in 2019, the U.S. average travel time across all modes was 27.6 minutes. However, individuals who commuted using light rails, buses, or subway

systems reported average commuting times of 45.8 minutes, 46.6 minutes, and 48.8 minutes, respectively. With this in mind, we consider employment areas accessible if they can be reached within 50 minutes. Then, we calculate then percentage of workers from a given socio-spatial group that can access their workplaces within a 50 minute transit journey. Given the 2021 LODES data, we define the percentage of commuters in a particular socio-spatial demographic as:

$$CA^t(i,b) = \frac{|\mathbf{L}^t(i,b)|}{|\mathbf{L}(i,b)|} * 100 \tag{6.9}$$

where L reflects the tract-level commuting flows from the LODES data, for a given city. Moreover, L^t captures the commuting flows within a travel time threshold of t. Then, $\mathbf{L}^t(i,b)$ denotes the number of commuters, in a given city, from neighbourhoods in income group i and BE group b, that can reach their employment locations within a t-minute transit commute. $\mathbf{L}(i,b)$ measures the total number of commuters in a given socio-spatial demographic in the same city, regardless of how long their transit commutes take. Thus, $CA^t(i,b)$ indicates the percentage of commuters in a socio-spatial demographic that can reach their workplaces within a given transit time threshold, t, which we set to be 50 minutes.

We use $CA^{50}(i,b)$ to define income disparities in workplace access via transit across different BE groups. To accomplish this we calculate the percentage point difference (Δ pp) between high and low income neighbourhoods in comparable types of built environment. That is, we simply subtract $CA^{50}(Low,b)$ from $CA^{50}(High,b)$, to understand the percentage point change between workplace access, for a given built environment group b in a particular city.

Figure 6.6 visualises the change in percentage point, with respect to the percentage of commuters in a socio-spatial demographic that can access their workplace within a 50 minute commute. Each city is represented along the x-axis, and BE groups are reflected along the y-axis, with the socioeconomic workplace access disparity across neighbourhoods, regardless

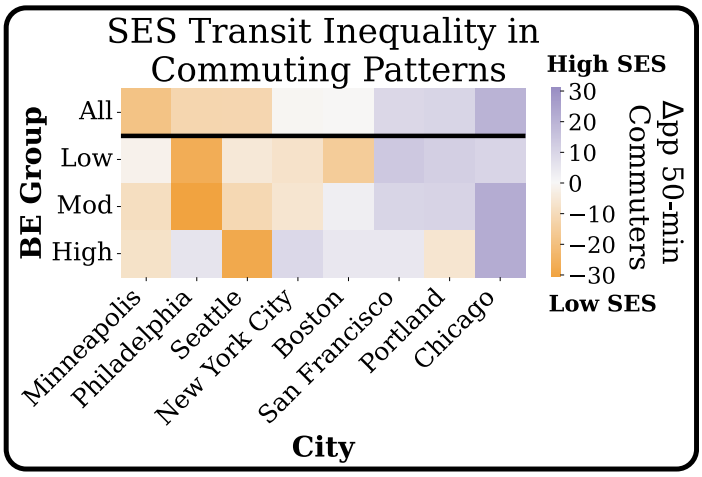

Fig. 6.6 **Combining socio-spatial demographics and commuting flows reveal the type of built environment in which employment access disparities (via transit) arise.** Each of the 8 cities are shown across the x-axis, with the top row conveying strictly socioeconomic disparities in workplace access for all neighbourhoods, ignoring BE features. Each cell in the bottom three rows reflects advantage for a socio-spatial demographic, with the y-axis denoting the spatial group and the hue symbolising the socioeconomic group that has more transit access. Purple hues represent socio-spatial in which a higher percentage of individuals from the High income group have access to their workplaces within a 50 minute transit commute, compared to workers in Low-income groups of the same spatial demographic.

of BE features, represented by the first row. In this manner, each cell reflects the percentage point difference between the percentage of high-income and low-income commuters that have a commute that is less than 50 minutes. Orange cells reflect spatial demographics in which lower income neighbourhoods have a higher percentage of workers that can access their workplace within the transit time threshold, compared to that of high income neighbourhoods in the same BE group. Table C.3 in the Appendix lists the percentages of commuters with workplace access within 50 minute transit journeys.

When we only consider socioeconomic transit disparities in workplace access (the top row of Figure 6.6), we observe that Minneapolis, Philadelphia, and Seattle have a larger percentage of workers in low-income neighbourhoods who have 50-minute workplace access, compared to commuters in the High income group. On the other hand, San Francisco, Portland, and Chicago (last three cells in the top row) convey that, when we disregard BE

features, higher-income neighbourhoods have a higher percentage of workers with access to their jobs than workers commuting from neighbourhoods in the Low income group. The last three rows in Figure 6.6 reveal socioeconomic disparities in employment access, with respect to BE features, for the Low, Moderate, and High BE Groups.

In Philadelphia, shown in the second column, neighbourhoods in the Low income, Low BE group tend to receive more access to employment opportunities within a 50 minute transit journey, compared to High income, Low BE neighbourhoods. However, when we compare employment access via transit for the High BE neighbourhoods in Philadelphia, we note that higher-income neighbourhoods have more access than lower-income tracts. This trend is present in New York City (fourth column) and Boston (fifth column), as well. In Minneapolis and Seattle, displayed in the first and third column respectively, we observe that lower-income neighbourhoods with low mobility flexibility (second row) tend to have less access to employment opportunities than their lower-income counterparts in the Moderate and High BE Group (third and fourth row, respectively). Portland depicts a contrasting case to Philadelphia, in which a larger fraction of workers in the High income, Low BE neighbourhoods can reach their workplaces within a 50 minute transit commute, compared to the Low income, Low BE tracts. Yet, lower income neighbourhoods with high mobility flexibility tend to provide their residents with greater employment access than compare to their higher income counterparts. In Chicago (last column), although low income neighbourhoods tend to have less workplace access than high income neighbourhoods, across all BE groups, we point out the larger transit access disparities occur in neighbourhoods with higher mobility flexibility. Ultimately, this section highlights how incorporating BE features when analysing commuting flows reveals disparities that are otherwise obscured when simply considering socioeconomic inequality in transit access to workplaces.

6.5 Discussion

In this chapter, we propose a methodology for defining neighbourhoods based on their street network design, building density, and amenity diversity. By applying this framework to eight US cities, we define three types of neighbourhoods, which we term Low, Moderate, and High built environment (BE) groups to reflect the levels of mobility flexibility in a census tract. We find that High BE groups tend to receive better transit service. This is in line with other findings in the transport research field, and serve as a reasoning for the incorporating neighbourhood features into transit analysis [53]. We compare transit disadvantage between socio-spatial groups with respect to integral accessibility as well as system-facilitated accessibility to essential amenities and workplaces within a commuting time threshold. We underscore the relevance of incorporating BE characteristics when analysing socioeconomic transit inequality by revealing how only considering socioeconomic disparities obfuscates the types of neighbourhoods in which low accessibility measures arise. Specifically, we identify Seattle, Portland, Minneapolis, and Boston as having increased levels of socioeconomic homophily when considering integral access to all potential destinations. Moreover, we show socio-spatial transit inequalities when analysing system-facilitated access to essential amenities. For Low-Income neighbourhoods that have low BE features (which we identify in San Francisco, Boston and Chicago), this can contribute to exacerbated forms of social exclusion [5, 167]. Finally, we incorporating commuting patterns to show how employment accessibility changes across socio-spatial dynamics. We show how in San Francisco, Portland and Chicago, low income neighbourhoods with low mobility flexibility tend to have less access to their workplaces than their high income counterparts. In Boston, New York City, and Philadelphia we can see how simply considering socioeconomic features obscures inequalities in employment accessibility for low income neighbourhood with high mobility flexibility. Ultimately, by considering different forms of accessibility, we show how inequality can arise for different mobility purposes and how the inclusion of spatial

features helps to pinpoint what types are neighbourhoods are particularly vulnerable to transit disparities.

Limitations of this work include the use of household median incomes to describe the socioeconomic level of a census tract. This obscures the prevalence of individuals with extreme poverty or wealth that reside in a given neighbourhood. Specifically, in the context of transit poverty, the concentration of residents that earn below the poverty line could be analysed to assess disadvantage for vulnerable socioeconomic groups in particular. Furthermore, a high flow of travellers between two regions may not directly correspond with high public transit usage, if all travellers use cars as their transport mode. Thus, integrating rates of transit use within each tract would provide further clarity in assessing transit service and inequalities.

Moreover, the commuting data can be leveraged beyond empirical commuting times. For example, the accessibility of jobs, with respect to different industries or wages, can be further explored. For analysing inequalities with regards to the spatial mismatch hypothesis, introduced in Chapter 2.3.3, it would be useful to combine the identified socio-spatial disparities in employment access with proximity between residential and workplace locations. Additionally, our transit networks are reflective of transit service on a weekday morning. However, transit service is temporal in nature. While we focus on a static snapshot of a city, this work can be extended to explore how transit service inequalities fluctuate throughout the day, or throughout the weekends. Finally, the findings of this work can be reconciled with data that estimates the ability to accommodate more public transit. For instance, street frontage can be assessed to determine whether neighbourhoods that have poor transit service have the built environment characteristics to allow for bus stops or transit stations to be built.

In essence, this study aims to highlight the importance of the built environment when measuring transit inequality. By considering disparities in transit service across socio-spatial demographics, we show how built environment features are associated with varying levels

of transit service and should, therefore, be considered in the context of transport poverty. Neighbourhoods that have poor transit service, coupled with low levels of design, density, and diversity, are vulnerable not only in terms of their socioeconomic standing and the quality of their immediate environment, but also with respect to the mobility options provided by public transit. Thus, physical features should be accounted for when assessing transit disparities, in order to provide improved mobility options to vulnerable demographics and bolster accessibility.

Chapter 7

Conclusions

Inequality in urban areas has been a persistent trend throughout much of history [114]. While urban inequality has been conventionally framed in the context of residential segregation or income disparities, access to high resolution mobility data has allowed researchers to conceptualise a more nuanced understanding of how disparities are exhibited within urban spaces [256, 312, 90, 316]. Human mobility, however, does not exist in a vacuum. That is, urban and transport infrastructure can influence the areas to which individuals travel, the modes of transport they use, and the reasons for which they travel [206, 198]. A consistent finding across developed countries, is that lower income households tend to spend a larger portion of their earning on transport and are more vulnerable to forced car ownership [64]. Despite a handful of similarities, identified transport inequalities across these countries tend to contradict one another [177]. Moreover, inconsistencies across travel surveys and data sources pose obstacles in comparing results of transit analysis across different regions [2]. Thus, the importance of considering the geographical context when measuring transport inequality becomes clear, as each region is subject to its own set of economic, cultural, and social structures.

7.1 Residential-Workplace Dependencies

The work presented in this thesis provides a nuanced perspective on public transit service in the US, accounting for how transit intersects with various dimensions of urban inequality. In doing so, we explore how public transportation systems facilitate different types of journeys, from commutes to amenity visitations, using detailed pathfinding algorithms for transit schedules. The first urban inequality we analyse is residential-workplace dependencies. We build off of decades of research on the spatial mismatch hypothesis to reveal structural and experiential commuting disparities for vulnerable demographics.

7.1.1 Disparities in Diversity of Spatial Dependencies

First, we leverage network entropy to highlight how conventional segregation metrics, which focus on segregation within residential or employment areas, obscure the spatial dependencies that exist between these two urban dimensions. In doing so, we address our first research question regarding how commuting networks can be leveraged to inform complexities in residential-workplace dynamics. Furthermore, we find that areas in which lower income workers have a greater diversity of commuting destinations, tend to have a lower average commuting time, leading us to discuss whether shorter transit times are reflective of well planned transit service or of residential choices being constrained by employment opportunities. Thus, we highlight how diversity of commuting origins and destinations can have different implications, depending on which demographic group is being considered. These findings are in line with previous research that assesses how privileged demographics command space, while deprived demographics are constrained by it [119, 45, 28].

In Section 2.3.3 we discuss three approaches to addressing spatial mismatch in residential-workplace dependencies [129]. On one hand, local in-flow entropy can identify employment hubs and inform where to apply incentives to move individuals to jobs. On the other hand, local out-flow entropy reveals the diversity of workplaces for a given residential area,

informing the neighbourhoods where policy makers can invest in creating employment opportunities. Finally, combining transit service with commuting network entropy can address the last approach to spatial mismatch: connecting individuals to jobs. That is, local out-flow entropy can identify neighbourhoods that largely depend on a small subset of employment areas. Then, urban planners and policy makers can assess how improving transit service to bolster commutes to different employment areas may increase opportunities for individuals in a given neighbourhood.

7.1.2 Housing Insecurity

Then, we focus on the experiential aspect of commuting behaviour, paying particular attention to demographics that are vulnerable to housing insecurity. We address our second research question by developping a clustering framework to define housing demographics, with respect to a neighbourhood's level of housing affordability, quality, safety, and stability. We apply local spatial autocorrelation to identify employment hotspots based on where individuals from each housing demographic work. To address the role that transit service plays in providing access to better employment opportunities (our third research question) we build transit-pedestrian networks to estimate transit travel times. Then, we measure how average commuting times shift between empirical driving commutes, empirical transit commutes, and potential transit commutes to workplaces associated with better housing conditions. We unveil the inefficiency of US transit systems, compared to driving, when studying workplace accessibility.

In essence, transport infrastructure has the potential to provide accessibility, such that it accounts and, perhaps, even corrects for disparities in residential-workplace dependencies. Previous research indicates that increased workplace access improves employment probabilities and income levels, particularly for lower income demographics [253, 22]. However, our results highlight how transit service, in its current state, does not provide individuals who

live in areas that are already vulnerable to housing insecurity, with opportunities to commute to employment areas that are associated with better economic outcomes, within less than an hour.

The findings from this analysis can be applied to understand how zoning policies, which can impact the types of households that live in a neighbourhood [265], intersect with levels of housing insecurity. Approaches to enabling social mobility tend to focus on introducing housing policies, such as the Section 8 Housing Choice Vouchers in the US [257]. However, our work highlights how social mobility can be addressed from a transit perspective, by improving connections between neighbourhoods that are vulnerable to housing insecurity and workplaces that are associated with better housing conditions.

7.2 Urban Segregation

Having analysed transit systems based on how they connect residential and employment locations, we shift our focus to understand how transit inequalities can be shaped by disparities in residential-amenity dependencies. By combining amenity visitation data with transport routing tools, we analyse socioeconomic segregation at the residential, amenity, and public transit level. We use the Index of Concentration at the Extremes to address our fourth research question, showing how individuals experience lower levels of segregation at the amenities they visit than in their residential areas. This finding that is consistent with other research studies [206, 319, 1, 215, 67]. However, lower income neighbourhoods tend to decrease their levels of residential segregation to a larger extent than their high income counterparts. This is in line with previous findings that show how vulnerable demographics, such as immigrants, have more constrained activity spaces [120, 266].

We answer our fifth research question by, first, contrasting neighbourhood accessibility between transit systems and cars. In doing so, we reveal how travelling by car provides accessibility to a larger range of neighbourhoods, while transit service limits neighbourhood

access to areas with a similar socioeconomic composition as the origin neighbourhood. This supports previous findings that individuals travelling by car tend to have higher accessibility than those using transit [98].

Finally, we measure experienced segregation on public transit lines based on empirical mobility journeys, to respond to the second half of our fifth research question. In doing so, we show trends of experienced segregation in transit use, albeit at a smaller magnitude than residential segregation. Moreover we underscore how spatial inequalities in amenity distributions across a city generally lead to higher levels of experienced transit segregation. This work can be applied to understand which neighbourhoods segregated transit service arises.

7.3 Built Environment

The last section considers spatial inequalities with respect to the built environment. Considering transit planning is shaped around travel demand and urban landscapes, we analyse how transit service characteristics change across different types of neighbourhoods. Responding to our sixth research questions, we highlight how areas with lower mobility flexibility tend to face longer transit times. To address our final research question we show how areas with low built environment features lack access to neighbourhoods of different socioeconomic backgrounds, within a half hour journey. Furthermore, we reveal how overlooking built environment features can obscure what types of neighbourhoods face the largest disadvantages in terms of essential amenity and workplace access. Literature on transport poverty supports these findings, showing that low transit access and low income can limit activity spaces and employment outcomes [5, 176]. In doing so, resources to improve transit service can be focused on the types of neighbourhoods, in regards to the built environment, that socioeconomic transit disparities arise, to ensure that transport policies are not targeting the wrong spatial demographic [176].

7.4 Discussion

Urban life is composed of numerous dynamics, from commuting and shopping for groceries to exploring leisure activities and meeting friends for a meal. Thus, while conventional metrics, such as residential segregation, are useful for understanding constraints in housing affordability and availability, they fall short at understanding the dynamic nature of cities. Different facets or urban life can impact the magnitude of inequality in a particular region. These mechanisms include, but are not limited to, housing markets, employment opportunities, income inequality, and urban design.

The increasing availability of mobility and public transit data has provided researchers with the tools to evaluate how transportation infrastructure (i.e. public transit systems, street networks) bolsters or hinders individuals from travelling to different types of areas. Our work is novel in that it develops a multi-dimensional approach to measuring housing insecurity in the US. Moreover, it introduces a framework for estimating segregation in multiple urban dimensions, revealing segregation in transit service and showing how amenity landscapes can impact the level of experienced segregation while using public transit. Finally, our work underscores the importance of considering built environment features when using transit, as transit disadvantages may be prevalent in particular types of neighbourhoods. These three contributions, coupled with detailed transit modelling that estimates system-facilitated and integral transit accessibility, show the importance of incorporating the complexities of urban dynamics (i.e., housing markets, urban design) and transit systems (i.e., fluctuations of travel time based on departure time, time spent walking to or waiting for transit) to appropriately identify neighbourhoods that are vulnerable to transit poverty. Ultimately, the work presented in this thesis highlights how the coupled effects of spatial and transit inequalities can serve as barriers to accessibility for vulnerable demographics.

7.5 Limitations and Future Work

In this thesis, we leverage a wide variety of datasets and methodologies to provide a quantitative perspective of transport justice in urban areas. In doing so, we hope to motivate future work in developing a nuanced understanding of how vulnerable demographics may face transit-related social exclusion. Accordingly, this section discusses limitations in our work and how future work can be adapted to avoid such constraints.

First, our analysis is largely dependent on census surveys, provided by the US Census Bureau. As discussed in Chapter 3.1.1, survey samples are vulnerable to margins of error, especially surveys that have smaller sample sizes. Throughout our work, we prioritise the high temporal resolution of the American Community Survey over the larger data certainty that comes with the Decennial Census. Moreover, our use of SafeGraph amenity visitations requires making a few assumptions in mobility behaviour, due to the level of aggregation in the data. That is, given a certain flow of visitors from a census block group to an amenity, we assume that all residents in the block group are equally likely to visit the amenity. In doing so, we overlook differences in the types of amenities demographics may visit. Furthermore, to ensure privacy of users, SafeGraph reports amenities that receive between two to four visits from a census block group as four. In cases such as these we randomly sample the visitor count to be either two, three, or four visits. While most mobility datasets are subject to biases in demographic representation [166], high-resolution data that provides individual and anonymised trajectories, would eliminate the need for the aforementioned assumptions. Furthermore, individual-level mobility data would enable research that could explore time-dependent fluctuations of transit segregation levels or analyse how the amenity types that certain demographic groups tend to visit change throughout the day or week.

Second, the majority of the results presented in this thesis focus on socioeconomic disparities in US public transit systems. However, these frameworks can be applied to understand transit inequalities with respect to different demographic groups, similar to how

human mobility studies cover a range of demographic inequalities, from gender to race [3, 180, 68]. Furthermore, the LODES commuting dataset, used in Chapter 4, informs origin-destination commuting flows with respect to workers' ages, earnings, and the industries in which they work. In this manner, structural inequalities in residential-workplace dependencies can be explored with respect to age and industry. Additionally, the American Community Survey (ACS), from which we derive our segregation metric in Chapter 5, can inform a vast array of sociodemographic groups, including, but not limited to gender, race, disability status, age, and educational attainment. Thus, the methodologies presented in this thesis can be extended beyond the socioeconomic dimension.

In Chapter 4.4, we present a clustering approach for measuring housing insecurity. This approach quantifies a neighbourhood's vulnerability to housing insecurity with respect to housing affordability, quality, safety and stability. However, sociological definitions conceptualise housing insecurity to also include neighbourhood safety, neighbourhood quality and homelessness. The data sources for these additional dimensions were only available for all cities at a spatial resolution that was lower than the census tract-level that we used in our analysis. Moreover, the measures of amenity diversity defined in Chapter 6 could be incorporated into the definition of neighbourhood quality. Neighbourhood quality could also account for exposure to pollution. Meanwhile, neighbourhood safety could include features of essential amenity accessibility, used in Chapter 6.4.3. Crime and road accidents data could also be used to inform neighbourhood safety. Ultimately, future work can leverage these potential data sources to define vulnerability to housing insecurity with respect to all seven dimensions. Finally, the classification approach, outlined in Chapter 4.4, provides a fundamental building block to quantify housing insecurity in a comprehensive manner. However, it falls short, in that it considers all neighbourhoods in the most vulnerable demographic as equally burdened by housing insecurity, when, in reality, they may be

burdened to different degrees and in different dimensions. Thus, future work could focus on developing a continuous metric for vulnerability to housing insecurity.

Fourth, we calculate transit travel times using transit schedules for a weekday morning. However, public transit service changes throughout the day, as well as between weekdays and weekends. Thus, future work could address how integral and system-facilitated transit accessibility changes over time, with respect to different types of trip purposes. Doing so would shine light on what types of neighbourhoods, employment areas, and amenities are accessible during and outside of conventional working hours. Furthermore, incorporating high-resolution transit ridership data could reveal demographic-level relationships between temporal transit service and usage.

Finally, the analysis in this thesis focuses on transit systems in the United States of America. Future work could expand this to other countries in the Global North, to compare spatial inequalities in transit systems across regions with similar traits. Previous findings point to how, even in the context of Global North countries, demographic inequalities in public transit vary greatly [177]. Measures of housing insecurity and the built environment must be translated to a Global South context with caution, as cultural and economic differences may require different data sources in order to adequately quantify these urban mechanisms [251]. Ultimately, by analysing transit with respect to the spatial inequalities explored in this thesis, future work can explore whether these contrasting findings result from disparities in transit service or in spatial landscapes.

References

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

[2] Akkermans, L., Maerivoet, S., Ahern, A., Redelbach, M., Schulz, A., and Vannacci, L. (2013). Harmonisation of national travel statistics in europe. *Publications Office of the European Union, Luxembourg City, Luxembourg*.

[3] Alessandretti, L., Aslak, U., and Lehmann, S. (2020). The scales of human mobility. *Nature*, 587(7834):402–407.

[4] Alessandretti, L. and Szell, M. (2022). Urban mobility. *arXiv preprint arXiv:2211.00355*.

[5] Allen, J. and Farber, S. (2019). Sizing up transport poverty: A national scale accounting of low-income households suffering from inaccessibility in canada, and what to do about it. *Transport policy*, 74:214–223.

[6] Alonso, W. (1964). *Location and land use: toward a general theory of land rent*. Harvard university press.

[7] Aman, J. J. C. and Smith-Colin, J. (2020). Transit deserts: Equity analysis of public transit accessibility. *Journal of Transport Geography*, 89:102869.

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

[9] Anselin, L. (2019). The moran scatterplot as an esda tool to assess local instability in spatial association. In *Spatial analytical perspectives on GIS*, pages 111–126. Routledge.

[10] Arambepola, R., Schaber, K. L., Schluth, C., Huang, A. T., Labrique, A. B., Mehta, S. H., Solomon, S. S., Cummings, D. A., and Wesolowski, A. (2023). Fine scale human mobility changes within 26 us cities in 2020 in response to the covid-19 pandemic were associated with distance and income. *PLOS Global Public Health*, 3(7):e0002151.

[11] Asabor, E. N., Warren, J. L., and Cohen, T. (2022). Racial/ethnic segregation and access to covid-19 testing: spatial distribution of covid-19 testing sites in the four largest highly segregated cities in the united states. *American Journal of Public Health*, 112(3):518–526.

[12] Atkinson, A. B. (2015). *Inequality: What can be done?*, pages 9–14. Harvard University Press.

[13] Aurand, A., Emmanuel, D., Threet, D., Rafi, I., and Yentel, D. (2019). The gap: A shortage of affordable rental homes. *National Low-Income Housing Coalition. Available from tinyurl. com/rz2n2c2*.

[14] Banister, D. (2008). The sustainable mobility paradigm. *Transport policy*, 15(2):73–80.

[15] Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., and Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports*, 734:1–74.

[16] Barbosa, H., de Lima-Neto, F. B., Evsukoff, A., and Menezes, R. (2015). The effect of recency to human mobility. *EPJ Data Science*, 4:1–14.

[17] Barbosa, H., Hazarie, S., Dickinson, B., Bassolas, A., Frank, A., Kautz, H., Sadilek, A., Ramasco, J. J., and Ghoshal, G. (2021). Uncovering the socioeconomic facets of human mobility. *Scientific reports*, 11(1):8616.

[18] Barrington-Leigh, C. and Millard-Ball, A. (2019). A global assessment of street-network sprawl. *PloS one*, 14(11):e0223078.

[19] Barthélemy, M. (2011). Spatial networks. *Physics reports*, 499(1-3):1–101.

[20] Bast, H. (2009). Car or public transport—two worlds. *Efficient algorithms: Essays dedicated to kurt mehlhorn on the occasion of his 60th birthday*, pages 355–367.

[21] Bast, H., Delling, D., Goldberg, A., Müller-Hannemann, M., Pajor, T., Sanders, P., Wagner, D., and Werneck, R. F. (2016). Route planning in transportation networks. *Algorithm engineering: Selected results and surveys*, pages 19–80.

[22] Bastiaanssen, J., Johnson, D., and Lucas, K. (2022). Does better job accessibility help people gain employment? the role of public transport in great britain. *Urban studies*, 59(2):301–322.

[23] Batty, M. (2004). Distance in space syntax.

[24] Bauer, R., Columbus, T., Katz, B., Krug, M., and Wagner, D. (2010). Preprocessing speed-up techniques is hard. In *International Conference on Algorithms and Complexity*, pages 359–370. Springer.

[25] Been, V. and Glashausser, A. (2009). Tenants: innocent victims of the nation's foreclosure crisis. *Alb. Gov't L. Rev.*, 2:1.

[26] Benfer, E., Robinson, D. B., Butler, S., Edmonds, L., Gilman, S., Mckay, K. L., Neumann, Z., Owens, L., Steinkamp, N., and Yentel, D. (2020). Covid-19 eviction crisis: An estimated 30-40 million people in america are at risk.

[27] Berube, A. (2019). Segregation, suburbs, and the future of fair housing. In *The Dream Revisited: Contemporary Debates About Housing, Segregation, and Opportunity*, pages 66–68. Columbia University Press.

[28] Bhat, C., Handy, S., Kockelman, K., Mahmassani, H., Chen, Q., Weston, L., et al. (2000). Development of an urban accessibility index: Literature review.

[29] Bishop-Royse, J., Lange-Maia, B., Murray, L., Shah, R., and DeMaio, F. (2021). Structural racism, socio-economic marginalization, and infant mortality. *Public Health*, 190:55–61.

[30] Blake, K. S., Kellerson, R. L., and Simic, A. (2007). Measuring overcrowding in housing.

[31] Blanchard, S. D. and Waddell, P. (2017). Urbanaccess: generalized methodology for measuring regional accessibility with an integrated pedestrian and transit network. *Transportation research record*, 2653(1):35–44.

[32] Boeing, G. (2019). Urban spatial order: Street network orientation, configuration, and entropy. *Applied Network Science*, 4(1):1–19.

[33] Boeing, G. (2020). Online rental housing market representation and the digital reproduction of urban inequality. *Environment and Planning A: Economy and Space*, 52(2):449–468.

[34] Boeing, G. (2021). Off the grid... and back again? the recent evolution of american street network planning and design. *Journal of the American Planning Association*, 87(1):123–137.

[35] Bokányi, E., Juhász, S., Karsai, M., and Lengyel, B. (2021). Universal patterns of long-distance commuting and social assortativity in cities. *Scientific reports*, 11(1):1–10.

[36] Bordagaray, M., dell'Olio, L., Ibeas, A., and Cecín, P. (2014). Modelling user perception of bus transit quality considering user and service heterogeneity. *Transportmetrica A: Transport Science*, 10(8):705–721.

[37] Bradshaw, J., Kemp, P., Baldwin, S., and Rowe, A. (2004). The drivers of social exclusion. *London: Social Exclusion Unit*.

[38] Brock, D. W. (1973). Recent work in utilitarianism. *American Philosophical Quarterly*, 10(4):241–276.

[39] Brockmann, D. and Theis, F. (2008). Money circulation, trackable items, and the emergence of universal human mobility patterns. *IEEE Pervasive Computing*, 7(4):28–35.

[40] Bruce Newbold, K. (2021). The urban geography of segregation. *The Economic Geography of Cross-Border Migration*, pages 293–306.

[41] Burd, C., Burrows, M., and McKenzie, B. (2021). Travel time to work in the united states: 2019. *American Community Survey Reports, United States Census Bureau*, 2.

[42] BUREAU, U. (2022). Design and methodology, american community survey. Technical report, Technical report, US Government Printing Office, Washington, DC.

[43] Bureau, U. C. (2022a). 2019 american community survey 5-year estimates, table b19001,b02001. accessed on 15/7/2022.

[44] Bureau, U. C. (2022b). Lehd origin-destination employment statistics data (2002-2019). accessed on 15/12/2022, Longitudinal-Employer Household Dynamics Program, LODES 7.5.

[45] Burns, L. D. (1980). Transportation, temporal, and spatial components of accessibility.

[46] Calafiore, A., Samardzhiev, K., Rowe, F., Fleishmann, M., and Arribas-Bel, D. (2023). Inequalities in experiencing urban functions. an exploration of human digital (geo-) footprints. *Environment and Planning B: Urban Analytics and City Science*, pages 1–15.

[47] Calthorpe, P. (1993). *The next American metropolis: Ecology, community, and the American dream*. Princeton architectural press.

[48] Carleton, P., Hoover, S., Fields, B., Barnes, M., and Porter, J. D. (2019). Gtfs-ride: Unifying standard for fixed-route ridership data. *Transportation Research Record*, 2673(12):173–181.

[49] Cervero, R. and Kockelman, K. (1997). Travel demand and the 3ds: Density, diversity, and design. *Transportation research part D: Transport and environment*, 2(3):199–219.

[50] Chambers, B. D., Baer, R. J., McLemore, M. R., and Jelliffe-Pawlowski, L. L. (2019). Using index of concentration at the extremes as indicators of structural racism to evaluate the association with preterm birth and infant mortality—california, 2011–2012. *Journal of Urban Health*, 96:159–170.

[51] Chen, C., Feng, T., Ding, C., Yu, B., and Yao, B. (2021). Examining the spatial-temporal relationship between urban built environment and taxi ridership: Results of a semi-parametric gwpr model. *Journal of Transport Geography*, 96:103172.

[52] Chen, Y., Wang, S., and Wu, X. (2022). Exploring perceived transportation disadvantages: Distribution, disparities, and associations with the built environment. *Transportation Research Part D: Transport and Environment*, 112:103497.

[53] Cheng, J., Yan, R., and Gao, Y. (2020). Exploring spatial heterogeneity in accessibility and transit mode choice. *Transportation Research Part D: Transport and Environment*, 87:102521.

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

[55] Chiu, M., Shah, B. R., Maclagan, L. C., Rezai, M.-R., Austin, P. C., and Tu, J. V. (2015). Walk score® and the prevalence of utilitarian walking and obesity among ontario adults: a cross-sectional study. *Health Reports*, 26(7):3.

[56] Cingano, F. (2014). Trends in income inequality and its impact on economic growth.

[57] Clark, W. A. (2013). The aftermath of the general financial crisis for the ownership society: What happened to low-income homeowners in the us? *International Journal of Housing Policy*, 13(3):227–246.

[58] Clark, W. A., Duque-Calvache, R., and Palomares-Linares, I. (2017). Place attachment and the decision to stay in the neighbourhood. *Population, space and place*, 23(2):e2001.

[59] Conway, M. W., Byrd, A., and van der Linden, M. (2017). Evidence-based transit and land use sketch planning using interactive accessibility methods on combined schedule and headway-based networks. *Transportation Research Record*, 2653(1):45–53.

[60] Cook, K. S. and Hegtvedt, K. A. (1983). Distributive justice, equity, and equality. *Annual review of sociology*, 9(1):217–241.

[61] Coulter, R. and Scott, J. (2015). What motivates residential mobility? re-examining self-reported reasons for desiring and making residential moves. *Population, Space and Place*, 21(4):354–371.

[62] Cowell, F. A. (2011). *Measuring inequality*. Oxford University Press.

[63] Cox, R., Henwood, B., Rodnyansky, S., Rice, E., and Wenzel, S. (2019). Road map to a unified measure of housing insecurity. *Cityscape*, 21(2):93–128.

[64] Currie, G. and Delbosc, A. (2013). Exploring trends in forced car ownership in melbourne. In *Proceedings of the 36th Australasian Transport Research Forum (ATRF), Brisbane, Queensland, Australia*, pages 1–9. Department of Infrastructure and Regional Development Canberra, Australia.

[65] Delling, D., Pajor, T., and Werneck, R. F. (2015). Round-based public transit routing. *Transportation Science*, 49(3):591–604.

[66] Delling, D., Sanders, P., Schultes, D., and Wagner, D. (2009). Engineering route planning algorithms. In *Algorithmics of large and complex networks: design, analysis, and simulation*, pages 117–139. Springer.

[67] Delmelle, E., Nilsson, I., and Adu, P. (2021). Poverty suburbanization, job accessibility, and employment outcomes. *Social Inclusion*, 9(2):166.

[68] Deng, H., Aldrich, D. P., Danziger, M. M., Gao, J., Phillips, N. E., Cornelius, S. P., and Wang, Q. R. (2021). High-resolution human mobility data reveal race and wealth disparities in disaster evacuation patterns. *Humanities and Social Sciences Communications*, 8(1):1–8.

[69] Denmark, D. (1998). The outsiders: Planning and transport disadvantage. *Journal of Planning Education and Research*, 17(3):231–245.

[70] Desmond, M. (2012). Eviction and the reproduction of urban poverty. *American journal of sociology*, 118(1):88–133.

[71] Desmond, M. and Gershenson, C. (2016). Housing and employment insecurity among the working poor. *Social Problems*, 63(1):46–67.

[72] Desmond, M., Gromis, A., Edmonds, L., Hendrickson, J., Krywokulski, K., Leung, L., and Porton, A. (2018). Eviction lab national database: Version 1.0. *Online database, Princeton University*.

[73] Desmond, M. and Kimbro, R. T. (2015). Eviction's fallout: housing, hardship, and health. *Social forces*, 94(1):295–324.

[74] Desmond, M. and Shollenberger, T. (2015). Forced displacement from rental housing: Prevalence and neighborhood consequences. *Demography*, 52(5):1751–1772.

[75] Diaz, R., Garrido, N., and Vargas, M. (2021). Segregation of high-skilled workers and the productivity of cities. *Regional Science Policy & Practice*, 13(5):1460–1478.

[76] Dibbelt, J., Pajor, T., Strasser, B., and Wagner, D. (2013). Intriguingly simple and fast transit routing. In *International Symposium on Experimental Algorithms*, pages 43–54. Springer.

[77] Dijkstra, E. W. (2022). A note on two problems in connexion with graphs. In *Edsger Wybe Dijkstra: His Life, Work, and Legacy*, pages 287–290.

[78] Doi, K., Kii, M., and Nakanishi, H. (2008). An integrated evaluation method of accessibility, quality of life, and social interaction. *Environment and Planning B: Planning and Design*, 35(6):1098–1116.

[79] Easthope, H. (2004). A place called home. *Housing, theory and society*, 21(3):128–138.

[80] El-Geneidy, A., Levinson, D., Diab, E., Boisjoly, G., Verbich, D., and Loong, C. (2016). The cost of equity: Assessing transit accessibility and social disparity using total travel cost. *Transportation Research Part A: Policy and Practice*, 91:302–316.

[81] Elldér, E., Haugen, K., and Vilhelmson, B. (2022). When local access matters: A detailed analysis of place, neighbourhood amenities and travel choice. *Urban Studies*, 59(1):120–139.

[82] Ellis, M., Wright, R., and Parks, V. (2004). Work together, live apart? geographies of racial and ethnic segregation at home and at work. *Annals of the Association of American Geographers*, 94(3):620–637.

[83] Erhardt, G. D., Hoque, J. M., Goyal, V., Berrebi, S., Brakewood, C., and Watkins, K. E. (2022). Why has public transit ridership declined in the united states? *Transportation research part A: policy and practice*, 161:68–87.

[84] Ermagun, A., Janatabadi, F., and Maharjan, S. (2023). Inequity analysis of spatial mismatch for low-income socially vulnerable populations across america. *Transportation research part D: transport and environment*, 118:103692.

[85] Ewing, R. and Cervero, R. (2010). Travel and the built environment: A meta-analysis. *Journal of the American planning association*, 76(3):265–294.

[86] Ewing, R., Schieber, R. A., and Zegeer, C. V. (2003). Urban sprawl as a risk factor in motor vehicle occupant and pedestrian fatalities. *American journal of public health*, 93(9):1541–1545.

[87] Fan, Z., Su, T., Sun, M., Noyman, A., Zhang, F., Pentland, A., and Moro, E. (2023). Diversity beyond density: Experienced social mixing of urban streets. *PNAS nexus*, 2(4):pgad077.

[88] Farber, S., Bartholomew, K., Li, X., Páez, A., and Habib, K. M. N. (2014). Assessing social equity in distance based transit fares using a model of travel behavior. *Transportation Research Part A: Policy and Practice*, 67:291–303.

[89] Farber, S., Neutens, T., Miller, H. J., and Li, X. (2013). 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.

[90] Farber, S., O'Kelly, M., Miller, H. J., and Neutens, T. (2015). Measuring segregation using patterns of daily travel behavior: A social interaction based model of exposure. *Journal of transport geography*, 49:26–38.

[91] Feng, Z. and Boyle, P. (2014). Do long journeys to work have adverse effects on mental health? *Environment and Behavior*, 46(5):609–625.

[92] Foda, M. A. and Osman, A. O. (2010). Using gis for measuring transit stop accessibility considering actual pedestrian road network. *Journal of Public Transportation*, 13(4):23–40.

[93] for Geoinformation Technology (HeiGIT), H. I. Openrouteservice. https://openrouteservice.org. (Accessed on March 2023).

[94] Foti, F., Waddell, P., and Luxen, D. (2012). A generalized computational framework for accessibility: from the pedestrian to the metropolitan scale. https://github.com/UDST/pandana. (Data accessed on November 2022).

[95] Franklin, J. H. (1956). History of racial segregation in the united states. *The Annals of the American Academy of Political and Social Science*, 304(1):1–9.

[96] Galster, G. (2001). On the nature of neighbourhood. *Urban studies*, 38(12):2111–2124.

[97] Gan, Z., Yang, M., Feng, T., and Timmermans, H. J. (2020). Examining the relationship between built environment and metro ridership at station-to-station level. *Transportation Research Part D: Transport and Environment*, 82:102332.

[98] Gao, Q.-L., Yue, Y., Zhong, C., Cao, J., Tu, W., and Li, Q.-Q. (2022). Revealing transport inequality from an activity space perspective: A study based on human mobility data. *Cities*, 131:104036.

[99] Geisberger, R., Sanders, P., Schultes, D., and Delling, D. (2008). Contraction hierarchies: Faster and simpler hierarchical routing in road networks. In *Experimental Algorithms: 7th International Workshop, WEA 2008 Provincetown, MA, USA, May 30-June 1, 2008 Proceedings 7*, pages 319–333. Springer.

[100] Getis, A. (2009). Spatial autocorrelation. In *Handbook of applied spatial analysis: Software tools, methods and applications*, pages 255–278. Springer.

[101] Giuliano, G. (2005). Low income, public transit, and mobility. *Transportation Research Record*, 1927(1):63–70.

[102] Gobillon, L., Selod, H., and Zenou, Y. (2007). The mechanisms of spatial mismatch. *Urban studies*, 44(12):2401–2427.

[103] Goetz, S. J., Han, Y., Findeis, J. L., and Brasier, K. J. (2010). Us commuting networks and economic growth: Measurement and implications for spatial policy. *Growth and Change*, 41(2):276–302.

[104] Gogia, N. (2006). Unpacking corporeal mobilities: The global voyages of labour and leisure. *Environment and Planning A*, 38(2):359–375.

[105] Goldblum, C. and Wong, T.-C. (2000). Growth, crisis and spatial change: a study of haphazard urbanisation in jakarta, indonesia. *Land Use Policy*, 17(1):29–37.

[106] Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782.

[107] Gordon, I. (2008). Density and the built environment. *Energy Policy*, 36(12):4652–4656.

[108] Gordon, I. R. and McCann, P. (2000). Industrial clusters: complexes, agglomeration and/or social networks? *Urban studies*, 37(3):513–532.

[109] Graells-Garrido, E., Serra-Burriel, F., Rowe, F., Cucchietti, F. M., and Reyes, P. (2021). A city of cities: Measuring how 15-minutes urban accessibility shapes human mobility in barcelona. *PloS one*, 16(5):e0250080.

[110] Griffin, G. P. and Sener, I. N. (2016). Public transit equity analysis at metropolitan and local scales: A focus on nine large cities in the us. *Journal of public transportation*, 19(4):126–143.

[111] Hackl, A. (2018). Mobility equity in a globalized world: Reducing inequalities in the sustainable development agenda. *World development*, 112:150–162.

[112] Haklay, M. and Weber, P. (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive computing*, 7(4):12–18.

[113] Hall, M., Iceland, J., and Yi, Y. (2019). Racial separation at home and work: Segregation in residential and workplace settings. *Population Research and Policy Review*, 38:671–694.

[114] Hamnett, C. (2019). 16. urban inequality. *Handbook of urban geography*, page 242.

[115] Handy, S. L., Boarnet, M. G., Ewing, R., and Killingsworth, R. E. (2002). How the built environment affects physical activity: views from urban planning. *American journal of preventive medicine*, 23(2):64–73.

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

[117] Hansson, E., Mattisson, K., Björk, J., Östergren, P.-O., and Jakobsson, K. (2011). Relationship between commuting and health outcomes in a cross-sectional population survey in southern sweden. *BMC public health*, 11:1–14.

[118] Hartman, C. and Robinson, D. (2003). Evictions: The hidden housing problem. *Housing Policy Debate*, 14(4):461–501.

[119] Harvey, D. (1973). Social justice and the city. University of Georgia Press.

[120] Haupt, A. and Ebner, C. (2020). Occupations and inequality: Theoretical perspectives and mechanisms. *Kölner Zeitschrift für Soziologie und Sozialpsychologie (KZfSS)*, 72.

[121] Hedman, L., Kadarik, K., Andersson, R., and Östh, J. (2021). Daily mobility patterns: Reducing or reproducing inequalities and segregation? *Social Inclusion*, 9(2):208–221.

[122] Hellerstein, J. K., McInerney, M., and Neumark, D. (2011). Neighbors and coworkers: The importance of residential labor market networks. *Journal of Labor Economics*, 29(4):659–695.

[123] Hepburn, P., Louis, R., Fish, J., Lemmerman, E., Alexander, A. K., Thomas, T. A., Koehler, R., Benfer, E., and Desmond, M. (2021). Us eviction filing patterns in 2020. *Socius*, 7:23780231211009983.

[124] Hepburn, P., Rutan, D. Q., and Desmond, M. (2023). Beyond urban displacement: Suburban poverty and eviction. *Urban Affairs Review*, 59(3):759–792.

[125] Hickman, R., Seaborn, C., Headicar, P., and Banister, D. (2010). Planning for sustainable travel. *Integrated Transport: from policy to practice*, page 33.

[126] Hine, J. (2009). Transport and social exclusion. In Kitchin, R. and Thrift, N., editors, *International Encyclopedia of Human Geography*, pages 429–434. Elsevier, Oxford.

[127] Hochstenbach, C. and Musterd, S. (2021). A regional geography of gentrification, displacement, and the suburbanisation of poverty: Towards an extended research agenda. *Area*, 53(3):481–491.

[128] Holliday, A. L. and Dwyer, R. E. (2009). Suburban neighborhood poverty in us metropolitan areas in 2000. *City & Community*, 8(2):155–176.

[129] Hoover, E. M. (1941). Interstate redistribution of population, 1850–1940. *The Journal of Economic History*, 1(2):199–205.

[130] Hu, S. and Chen, P. (2021). Who left riding transit? examining socioeconomic disparities in the impact of covid-19 on ridership. *Transportation Research Part D: Transport and Environment*, 90:102654.

[131] Ihlanfeldt, K. R. and Sjoquist, D. L. (1998). The spatial mismatch hypothesis: A review of recent studies and their implications for welfare reform. *Housing policy debate*, 9(4):849–892.

[132] Ingold, T. (1994). Companion encyclopedia of anthropology. pages 1010–1011.

[133] Ingram, D. R. (1971). The concept of accessibility: A search for an operational form. *Regional studies*, 5(2):101–107.

[134] Iyer, N., Menezes, R., and Barbosa, H. (2023). Network entropy as a measure of socioeconomic segregation in residential and employment landscapes. *Complex Networks XIV. CompleNet 2023. Springer Proceedings in Complexity.*

[135] Iyer, N., Menezes, R., and Barbosa, H. (To appear April 2024). Public transit inequality in the context of the built environment. *Complex Networks XV. CompleNet 2024. Springer Proceedings in Complexity.*

[136] Iyer, N., Menezes, R., and Barbosa, H. (Under Review). The role of transport systems in housing insecurity: a mobility-based analysis. *EPJ Data Science*.

[137] Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York: Random House.

[138] Jiang, S., Yang, Y., Gupta, S., Veneziano, D., Athavale, S., and González, M. C. (2016). The timegeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences*, 113(37):E5370–E5378.

[139] Johnson, D., Ercolani, M., and Mackie, P. (2017). Econometric analysis of the link between public transport accessibility and employment. *Transport Policy*, 60:1–9.

[140] Johnson, R. and Cureton, A. (2004). Kant's moral philosophy.

[141] Johnston, R., Harris, R., Jones, K., Manley, D., Wang, W. W., and Wolf, L. (2020). Quantitative methods ii: How we moved on–decades of change in philosophy, focus and methods. *Progress in Human Geography*, 44(5):959–971.

[142] Jones, C. and Richardson, H. W. (2014). Housing markets and policy in the uk and the usa: A review of the differential impact of the global housing crisis. *International Journal of Housing Markets and Analysis*.

[143] Kain, J. F. (1968). Housing segregation, negro employment, and metropolitan decentralization. *The quarterly journal of economics*, 82(2):175–197.

[144] Kamruzzaman, M., Baker, D., Washington, S., and Turrell, G. (2014). Advance transit oriented development typology: case study in brisbane, australia. *Journal of transport geography*, 34:54–70.

[145] Kang, S. (2016). Inequality and crime revisited: effects of local inequality and economic segregation on crime. *Journal of Population Economics*, 29:593–626.

[146] Karimi, F. and Oliveira, M. (2023). On the inadequacy of nominal assortativity for assessing homophily in networks. *Scientific Reports*, 13(1):21053.

[147] Kattiyapornpong, U. and Miller, K. E. (2009). Socio-demographic constraints to travel behavior. *International Journal of Culture, Tourism and Hospitality Research*, 3(1):81–94.

[148] Kaufman, A. (2018). *Rawls's egalitarianism*, pages 6–8. Cambridge University Press.

[149] Kaufmann, V., Bergman, M. M., and Joye, D. (2004). Motility: Mobility as capital. *International journal of urban and regional research*, 28(4):745–756.

[150] King, A. T. and Mieszkowski, P. (1973). Racial discrimination, segregation, and the price of housing. *Journal of political economy*, 81(3):590–606.

[151] Kingsley, G. T., Smith, R., and Price, D. (2009). *The Impacts of foreclosures on families and communuities*. Urban Institute Washington, DC.

[152] Klaesson, J. and Öner, Ö. (2021). Ethnic enclaves and segregation—self-employment and employment patterns among forced migrants. *Small Business Economics*, 56:985–1006.

[153] Knaap, E., Wolf, L., Rey, S., Kang, W., and Han, S. (2019). The dynamics of urban neighborhoods: a survey of approaches for modeling socio-spatial structure.

[154] Kneebone, E. and Holmes, N. (2015). The growing distance between people and jobs in metropolitan america. *Brook Inst March*.

[155] Krupka, D. J. (2007). Are big cities more segregated? neighbourhood scale and the measurement of segregation. *Urban Studies*, 44(1):187–197.

[156] Kullback, S. (1997). *Information theory and statistics*. Courier Corporation.

[157] Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review*, 45(1):1–28.

[158] Kvålseth, T. O. (1995). Comment on the coefficient of ordinal variation. *Perceptual and motor skills*, 81(2):621–622.

[159] Lambert, S. J., Haley-Lock, A., and Henly, J. R. (2012). Schedule flexibility in hourly jobs: Unanticipated consequences and promising directions. *Community, Work & Family*, 15(3):293–315.

[160] Larrabee Sonderlund, A., Charifson, M., Schoenthaler, A., Carson, T., and Williams, N. J. (2022). Racialized economic segregation and health outcomes: a systematic review of studies that use the index of concentration at the extremes for race, income, and their interaction. *PloS one*, 17(1):e0262962.

[161] Law, M. (1991). The environment: A focus for occupational therapy.

[162] Lenormand, M., Arias, J. M., San Miguel, M., and Ramasco, J. J. (2020a). On the importance of trip destination for modelling individual human mobility patterns. *Journal of The Royal Society Interface*, 17(171):20200673.

[163] Lenormand, M., Samaniego, H., Chaves, J. C., da Fonseca Vieira, V., da Silva, M. A. H. B., and Evsukoff, A. G. (2020b). Entropy as a measure of attractiveness and socioeconomic complexity in rio de janeiro metropolitan area. *Entropy*, 22(3):368.

[164] Leopold, J., Cunningham, M., Posey, L., and Manuel, T. (2016). Improving measures of housing insecurity: A path forward. *The Urban Institute*, 1.

[165] Levinson, D. and King, D. (2020). Transport access manual: A guide for measuring connection between people and places.

[166] Li, Z., Ning, H., Jing, F., and Lessani, M. N. (2024). Understanding the bias of mobile location data across spatial scales and over time: a comprehensive analysis of safegraph data in the united states. *Plos one*, 19(1):e0294430.

[167] Liu, C. and Bardaka, E. (2021). The suburbanization of poverty and changes in access to public transportation in the triangle region, nc. *Journal of transport geography*, 90:102930.

[168] Liu, D. and Kwan, M.-P. (2020). Measuring spatial mismatch and job access inequity based on transit-based job accessibility for poor job seekers. *Travel Behaviour and Society*, 19:184–193.

[169] Liu, E.-J. and Yan, X.-Y. (2020). A universal opportunity model for human mobility. *Scientific Reports*, 10(1).

[170] Liu, Y., Bunker, J., and Ferreira, L. (2010). Transit users' route-choice modelling in transit assignment: A review. *Transport Reviews*, 30(6):753–769.

[171] Lo, H. K., Yip, C.-W., and Wan, Q. K. (2004). Modeling competitive multi-modal transit services: a nested logit approach. *Transportation Research Part C: Emerging Technologies*, 12(3-4):251–272.

[172] Logan, J. R., Spielman, S., Xu, H., and Klein, P. N. (2011). Identifying and bounding ethnic neighborhoods. *Urban Geography*, 32(3):334–359.

[173] Louail, T., Lenormand, M., Picornell, M., Garcia Cantu, O., Herranz, R., Frias-Martinez, E., Ramasco, J. J., and Barthelemy, M. (2015). Uncovering the spatial structure of mobility networks. *Nature communications*, 6(1):1–8.

[174] Lowry, I. S. (1964). *A Model of Metropolis*. RAND Corporation, Santa Monica, CA.

[175] Lu, S., Fang, Z., Zhang, X., Shaw, S.-L., Yin, L., Zhao, Z., and Yang, X. (2017). Understanding the representativeness of mobile phone location data in characterizing human mobility indicators. *ISPRS International Journal of Geo-Information*, 6(1):7.

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

[177] Lucas, K., Mattioli, G., Verlinghieri, E., and Guzman, A. (2016). Transport poverty and its adverse social consequences. In *Proceedings of the institution of civil engineers-transport*, volume 169, pages 353–365. Thomas Telford Ltd.

[178] Lucas, K., Tyler, S., Cervero, R., and Orfeuil, J. (2006). Moving from welfare to work: the role of transport.

[179] Luxen, D. and Vetter, C. (2011). Real-time routing with openstreetmap data. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '11, pages 513–516, New York, NY, USA. ACM.

[180] Macedo, M., Lotero, L., Cardillo, A., Barbosa, H., and Menezes, R. (2020). Gender patterns of human mobility in colombia: Reexamining ravenstein's laws of migration. In *Complex Networks XI*, pages 269–281. Springer.

[181] Macedo, M., Lotero, L., Cardillo, A., Menezes, R., and Barbosa, H. (2022). Differences in the spatial landscape of urban mobility: gender and socioeconomic perspectives. *Plos one*, 17(3):e0260874.

[182] Maciag, M. (2014). Public transportation's demographic divide. *Governing*, 25.

[183] Mackett, R. (2014). Has the policy of concessionary bus travel for older people in britain been successful? *Case studies on transport policy*, 2(2):81–88.

[184] Mahendra Dev, S. (2018). Inequality, employment and public policy. *The Indian Journal of Labour Economics*, 61:1–42.

[185] Malekzadeh, A. and Chung, E. (2020). A review of transit accessibility models: Challenges in developing transit accessibility models. *International journal of sustainable transportation*, 14(10):733–748.

[186] Mansfeld, Y. (1992). From motivation to actual travel. *Annals of tourism research*, 19(3):399–419.

[187] Marin, V., Molinero, C., and Arcaute, E. (2022). Uncovering structural diversity in commuting networks: global and local entropy. *Scientific Reports*, 12(1):1–13.

[188] Marshall, S. (2004). *Streets and patterns*, pages 29–34. Routledge.

[189] Martens, K., Golub, A., and Robinson, G. (2012). A justice-theoretic approach to the distribution of transportation benefits: Implications for transportation planning practice in the united states. *Transportation research part A: policy and practice*, 46(4):684–695.

[190] Massey, D. S. (2001). The prodigal paradigm returns: ecology comes back to sociology. *Does it take a village*, pages 41–48.

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

[192] Mattioli, G. (2017). " forced car ownership" in the uk and germany: socio-spatial patterns and potential economic stress impacts. *Social Inclusion*, 5(4):147–160.

[193] Mavoa, S., Witten, K., McCreanor, T., and O'sullivan, D. (2012). Gis based destination accessibility via public transit and walking in auckland, new zealand. *Journal of transport geography*, 20(1):15–22.

[194] Mazziotta, M. and Pareto, A. (2016). Methods for constructing non-compensatory composite indices: a comparative study. In *Forum for Social Economics*, volume 45, pages 213–229. Taylor & Francis.

[195] McHugh, B. (2013). Pioneering open data standards: The gtfs story. *Beyond transparency: open data and the future of civic innovation*, pages 125–135.

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

[197] Merriman, P., Jones, R., Cresswell, T., Divall, C., Mom, G., Sheller, M., and Urry, J. (2013). Mobility: Geographies, histories, sociologies. *Transfers*, 3(1):147–165.

[198] Millard-Ball, A., West, J., Rezaei, N., and Desai, G. (2022). What do residential lotteries show us about transportation choices? *Urban Studies*, 59(2):434–452.

[199] Minocha, I., Sriraj, P., Metaxatos, P., and Thakuriah, P. (2008). Analysis of transit quality of service and employment accessibility for the greater chicago, illinois, region. *Transportation Research Record*, 2042(1):20–29.

[200] Miranda, A. S. (2020). The shape of segregation: The role of urban form in immigrant assimilation. *Cities*, 106:102852.

[201] MobilityData (2023). The mobility database. https://database.mobilitydata.org/. Accessed: 2022-06-11.

[202] Mokhtarian, P. L., Salomon, I., and Singer, M. E. (2015). What moves us? an interdisciplinary exploration of reasons for traveling. *Transport reviews*, 35(3):250–274.

[203] Molz, J. G. (2006). 'watch us wander': mobile surveillance and the surveillance of mobility. *Environment and Planning A*, 38(2):377–393.

[204] Montroll, E. W. and Weiss, G. H. (1965). Random walks on lattices. ii. *Journal of Mathematical Physics*, 6(2):167–181.

[205] Moreno, C., Allam, Z., Chabaud, D., Gall, C., and Pratlong, F. (2021). Introducing the "15-minute city": Sustainability, resilience and place identity in future post-pandemic cities. *Smart Cities*, 4(1):93–111.

[206] Moro, E., Calacci, D., Dong, X., and Pentland, A. (2021). Mobility patterns are associated with experienced income segregation in large us cities. *Nature communications*, 12(1):4633.

[207] Moroni, S. (2016). Urban density after jane jacobs: the crucial role of diversity and emergence. *City, Territory and Architecture*, 3:1–8.

[208] Mullen, C., Tight, M., Whiteing, A., and Jopson, A. (2014). Knowing their place on the roads: What would equality mean for walking and cycling? *Transportation research part A: policy and practice*, 61:238–248.

[209] Muntaner, C., Li, Y., Ng, E., Benach, J., and Chung, H. (2011). Work or place? assessing the concurrent effects of workplace exploitation and area-of-residence economic inequality on individual health. *International Journal of Health Services*, 41(1):27–50.

[210] Nassir, N., Hickman, M., Malekzadeh, A., and Irannezhad, E. (2016). A utility-based travel impedance measure for public transit network accessibility. *Transportation Research Part A: Policy and Practice*, 88:26–39.

[211] Nawakitphaitoon, K. and Ormiston, R. (2015). Occupational human capital and earnings losses of displaced workers: does the degree of similarity between pre-and post-displacement occupations matter? *Journal for Labour Market Research*, 48(1):57–73.

[212] Newman, M. E. (2003). Mixing patterns in networks. *Physical review E*, 67(2):026126.

[213] Nicoletti, L., Sirenko, M., and Verma, T. (2022). Disadvantaged communities have lower access to urban infrastructure. *Environment and Planning B: Urban Analytics and City Science*, page 23998083221131044.

[214] Nieuwenhuis, J., Tammaru, T., Van Ham, M., Hedman, L., and Manley, D. (2020). Does segregation reduce socio-spatial mobility? evidence from four european countries with different inequality and segregation contexts. *Urban Studies*, 57(1):176–197.

[215] Nieuwenhuis, J. and Xu, J. (2021). Residential segregation and unequal access to schools. *Social Inclusion*, 9(2):142–153.

[216] Nilsson, I. and Delmelle, E. C. (2020). On the link between rail transit and spatial income segregation. *Applied Geography*, 125:102364.

[217] Norton, P. D. (2007). Street rivals: Jaywalking and the invention of the motor age street. *Technology and culture*, 48(2):331–359.

[218] Nozick, R. (2013). Distributive justice. In *Modern Understandings of Liberty and Property*, pages 179–260. Routledge.

[219] on Physical Activity, N. R. C. U. C., Transportation, Use, L., Board, T. R., and of Medicine, I. (2005). *Does the Built Environment Influence Physical Activity?: Examining the Evidence–Special Report 282*, volume 282, pages 134–137. Transportation Research Board.

[220] Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and techniques in modern geography*.

[221] OpenStreetMap contributors (2017). Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org.

[222] Ord, J. K. and Getis, A. (2001). Testing for local spatial autocorrelation in the presence of global autocorrelation. *Journal of regional science*, 41(3):411–432.

[223] Oviedo, D. and Guzman, L. A. (2020). Revisiting accessibility in a context of sustainable transport: Capabilities and inequalities in bogota. *Sustainability*, 12(11):4464.

[224] Özer, S. and Jones, A. (2022). Changing socio-spatial definitions of sufficiency of home: evidence from london (uk) before and during the covid-19 stay-at-home restrictions. *International Journal of Housing Policy*, pages 1–26.

[225] O'Hare, W. P., Robinson, J. G., West, K., and Mule, T. (2016). Comparing the us decennial census coverage estimates for children from demographic analysis and coverage measurement surveys. *Population Research and Policy Review*, 35:685–704.

[226] Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., and Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature communications*, 6(1):8166.

[227] Paul, J. and Taylor, B. D. (2021). Who lives in transit-friendly neighborhoods? an analysis of california neighborhoods over time. *Transportation research interdisciplinary perspectives*, 10:100341.

[228] Peach, C. (1996). Good segregation, bad segregation. *Planning perspectives*, 11(4):379–398.

[229] Pereira, R. H. (2019). Future accessibility impacts of transport policy scenarios: Equity and sensitivity to travel time thresholds for bus rapid transit expansion in rio de janeiro. *Journal of Transport Geography*, 74:321–332.

[230] Pereira, R. H., Saraiva, M., Herszenhut, D., Braga, C. K. V., and Conway, M. W. (2021). r5r: rapid realistic routing on multimodal transport networks with r 5 in r. *Findings*.

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

[232] Peters, T. and Halleran, A. (2020). How our homes impact our health: using a covid-19 informed approach to examine urban apartment housing. *Archnet-IJAR: International Journal of Architectural Research*.

[233] Petersen, J., Gibin, M., Longley, P., Mateos, P., Atkinson, P., and Ashby, D. (2011). Geodemographics as a tool for targeting neighbourhoods in public health campaigns. *Journal of Geographical Systems*, 13:173–192.

[234] Piketty, T. (2014). *Capital in the twenty-first century*. Harvard University Press.

[235] Pirie, G. (2009). Distance. Elsevier.

[236] Porta, S., Crucitti, P., and Latora, V. (2006). The network analysis of urban streets: A dual approach. *Physica A: Statistical Mechanics and its Applications*, 369(2):853–866.

[237] Porton, A., Gromis, A., and Desmond, M. (2021). Inaccuracies in eviction records: Implications for renters and researchers. *Housing Policy Debate*, 31(3-5):377–394.

[238] Pyrga, E., Schulz, F., Wagner, D., and Zaroliagis, C. (2008). Efficient models for timetable information in public transportation systems. *Journal of Experimental Algorithmics (JEA)*, 12:1–39.

[239] Qian, H. (2013). Diversity versus tolerance: the social drivers of innovation and entrepreneurship in us cities. *Urban Studies*, 50(13):2718–2735.

[240] Ramakrishnan, K., Champion, E., Gallagher, M., and Fudge, K. (2021). Why housing matters for upward mobility: Evidence and indicators for practitioners and policymakers.

[241] Ranjan, G., Zang, H., Zhang, Z.-L., and Bolot, J. (2012). Are call detail records biased for sampling human mobility? *ACM SIGMOBILE Mobile Computing and Communications Review*, 16(3):33–44.

[242] Rao, J. (2003). Some new developments in small area estimation.

[243] Raphael, S. and Stoll, M. A. (2010). Job sprawl and the suburbanization of poverty. Technical report.

[244] Rawls, J. (1971). *A Theory of Justice: Original Edition*. Harvard University Press.

[245] Reardon, S. F. and Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological methodology*, 32(1):33–67.

[246] Reichelt, M. and Abraham, M. (2017). Occupational and regional mobility as substitutes: A new approach to understanding job changes and wage inequality. *Social Forces*, 95(4):1399–1426.

[247] Resch, B. and Szell, M. (2019). Human-centric data science for urban studies.

[248] Rey, S. J., Kang, W., and Wolf, L. (2016). The properties of tests for spatial effects in discrete markov chain models of regional income distribution dynamics. *Journal of Geographical Systems*, 18:377–398.

[249] Rosen, H. S. (1985). Housing behavior and the experimental housing-allowance program: What have we learned? In *Social experimentation*, pages 55–94. University of Chicago Press.

[250] Roy, A. (2005). Urban informality: Toward an epistemology of planning. *Journal of the american planning association*, 71(2):147–158.

[251] Roy, D., Bernal, D., and Lees, M. (2020). An exploratory factor analysis model for slum severity index in mexico city. *Urban studies*, 57(4):789–805.

[252] SafeGraph (2021). Weekly patterns schemma. accessed on 20/3/2021.

[253] Saif, M. A., Zefreh, M. M., and Torok, A. (2019). Public transport accessibility: A literature review. *Periodica Polytechnica Transportation Engineering*, 47(1):36–43.

[254] Schleimer, J. P., Buggs, S. A., McCort, C. D., Pear, V. A., Biasi, A. D., Tomsich, E., Shev, A. B., Laqueur, H. S., and Wintemute, G. J. (2022). Neighborhood racial and economic segregation and disparities in violence during the covid-19 pandemic. *American journal of public health*, 112(1):144–153.

[255] Schmidt, S. and Németh, J. (2010). Space, place and the city: Emerging research on public space design and planning. *Journal of Urban Design*, 15(4):453–457.

[256] Schnell, I. and Yoav, B. (2001). The sociospatial isolation of agents in everyday life spaces as an aspect of segregation. *Annals of the Association of American Geographers*, 91(4):622–636.

[257] Schwartz, A. F. (2021). *Housing policy in the United States*. Routledge.

[258] Scott, J. L., Lee-Johnson, N. M., and Danos, D. (2022). Place, race, and case: Examining racialized economic segregation and covid-19 in louisiana. *Journal of Racial and Ethnic Health Disparities*, pages 1–13.

[259] Sen, A. (1993). Capability and well-being. *The quality of life*, 30:270–293.

[260] Severen, C. (2023). Commuting, labor, and housing market effects of mass transportation: Welfare and identification. *Review of Economics and Statistics*, 105(5):1073–1091.

[261] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

[262] Sheller, M. (2018). Theorising mobility justice. *Tempo Social*, 30:17–34.

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

[264] Shen, Q. (1998). Location characteristics of inner-city neighborhoods and employment accessibility of low-wage workers. *Environment and planning B: Planning and Design*, 25(3):345–365.

[265] Shertzer, A., Twinam, T., and Walsh, R. P. (2022). Zoning and segregation in urban economic history. *Regional Science and Urban Economics*, 94:103652.

[266] Silm, S., Mooses, V., Puura, A., Masso, A., Tominga, A., and Saluveer, E. (2021). The relationship between ethno-linguistic composition of social networks and activity space: A study using mobile phone data. *Social Inclusion*, 9(2):192–207.

[267] Simini, F., González, M. C., Maritan, A., and Barabási, A.-L. (2012a). A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100.

[268] Simini, F., González, M. C., Maritan, A., and Barabási, A.-L. (2012b). A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100.

[269] Simpson, W. (1980). A simultaneous model of workplace and residential location incorporating job search. *Journal of Urban Economics*, 8(3):330–349.

[270] Singh, V. K., Bozkaya, B., and Pentland, A. (2015). Money walks: implicit mobility behavior and financial well-being. *PloS one*, 10(8):e0136628.

[271] Smolak, K., Siła-Nowicka, K., Delvenne, J.-C., Wierzbiński, M., and Rohm, W. (2021). The impact of human mobility data scales and processing on movement predictability. *Scientific Reports*, 11(1):15177.

[272] Song, C., Koren, T., Wang, P., and Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature physics*, 6(10):818–823.

[273] Song, X., Kanasugi, H., and Shibasaki, R. (2016). Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level.

[274] Southworth, M. and Ben-Joseph, E. (2004). Reconsidering the cul-de-sac. *Access Magazine*, 1(24):28–33.

[275] Spielman, S. E., Folch, D., and Nagle, N. (2014). Patterns and causes of uncertainty in the american community survey. *Applied geography*, 46:147–157.

[276] Spielman, S. E. and Logan, J. R. (2013). Using high-resolution population data to identify neighborhoods and establish their boundaries. *Annals of the Association of American Geographers*, 103(1):67–84.

[277] Spielman, S. E. and Yoo, E.-h. (2009). The spatial dimensions of neighborhood effects. *Social science & medicine*, 68(6):1098–1105.

[278] Spielman, S. E., Yoo, E.-H., and Linkletter, C. (2013). Neighborhood contexts, health, and behavior: understanding the role of scale and residential sorting. *Environment and Planning B: Planning and Design*, 40(3):489–506.

[279] Steffen, B., Carter, G. R., Martin, M., Pelletiere, D., Vandenbroucke, D. A., and Yao, Y.-G. D. (2015). Worst case housing needs: 2015 report to congress. *Available at SSRN 3055230*.

[280] Stein, M. S. (2008). *Distributive justice and disability: Utilitarianism against egalitarianism*, pages 1–2. Yale University Press.

[281] Stiglitz, J. E. (2012). *The price of inequality: How today's divided society endangers our future*. WW Norton & Company.

[282] Strömgren, M., Tammaru, T., Danzer, A. M., van Ham, M., Marcińczak, S., Stjernström, O., and Lindgren, U. (2014). Factors shaping workplace segregation between natives and immigrants. *Demography*, 51(2):645–671.

[283] Tasci, A. D. and Ko, Y. J. (2017). Travel needs revisited. *Journal of Vacation Marketing*, 23(1):20–36.

[284] Tetteh, J. D., Templeton, M. R., Cavanaugh, A., Bixby, H., Owusu, G., Yidana, S. M., Moulds, S., Robinson, B., Baumgartner, J., Annim, S. K., et al. (2022). Spatial heterogeneity in drinking water sources in the greater accra metropolitan area (gama), ghana. *Population and Environment*, 44(1-2):46–76.

[285] Theil, H. (1972). Statistical decomposition analysis: With applications in the social and administrative sciences. *(No Title)*.

[286] Theil, H. and Finizza, A. J. (1971). A note on the measurement of racial integration of schools by means of informational concepts.

[287] Tiznado-Aitken, I., Lucas, K., Munoz, J. C., and Hurtubia, R. (2022). Freedom of choice? social and spatial disparities on combined housing and transport affordability. *Transport Policy*, 122:39–53.

[288] Toch, E., Lerner, B., Ben-Zion, E., and Ben-Gal, I. (2019). Analyzing large-scale human mobility data: a survey of machine learning methods and applications. *Knowledge and Information Systems*, 58:501–523.

[289] Tribby, C. P. and Zandbergen, P. A. (2012). High-resolution spatio-temporal modeling of public transit accessibility. *Applied Geography*, 34:345–355.

[290] Tu, W., Cao, R., Yue, Y., Zhou, B., Li, Q., and Li, Q. (2018). Spatial variations in urban public ridership derived from gps trajectories and smart card data. *Journal of Transport Geography*, 69:45–57.

[291] Vansnick, J.-C. (1990). Measurement theory and decision aid. In *Readings in multiple criteria decision aid*, pages 81–100. Springer.

[292] Veneri, P., Comandon, A., Garcia-López, M.-À., and Daams, M. N. (2021). What do divided cities have in common? an international comparison of income segregation. *Journal of Regional Science*, 61(1):162–188.

[293] Verbich, D., Badami, M. G., and El-Geneidy, A. M. (2017). Bang for the buck: Toward a rapid assessment of urban public transit from multiple perspectives in north america. *Transport Policy*, 55:51–61.

[294] Wachs, M. and Kumagai, T. G. (1973). Physical accessibility as a social indicator. *Socio-Economic Planning Sciences*, 7(5):437–456.

[295] Waddell, P. (2002). Urbansim: Modeling urban development for land use, transportation, and environmental planning. *Journal of the American planning association*, 68(3):297–314.

[296] Wagner, D. and Zündorf, T. (2017). Public transit routing with unrestricted walking. In *17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[297] Wallace, M. E., Crear-Perry, J., Green, C., Felker-Kantor, E., and Theall, K. (2019). Privilege and deprivation in detroit: infant mortality and the index of concentration at the extremes. *International Journal of Epidemiology*, 48(1):207–216.

[298] Walzer, M. (2008). *Spheres of justice: A defense of pluralism and equality*. Basic books.

[299] Wang, W. L. and van de Lindt, J. W. (2021). Quantitative modeling of residential building disaster recovery and effects of pre-and post-event policies. *International Journal of Disaster Risk Reduction*, 59:102259.

[300] Wang, Y., Zhang, Z., Zhu, M., and Wang, H. (2020). The impact of service quality and customer satisfaction on reuse intention in urban rail transit in tianjin, china. *Sage Open*, 10(1):2158244019898803.

[301] Wardrip, K., Williams, L., and Hague, S. (2011). The role of affordable housing in creating jobs and stimulating local economic development. *Journal of Planning Literature*, 21(4):371–385.

[302] Webster, G. R. and Leib, J. (2010). Living on the grid: The us rectangular public land survey system and the engineering of the american landscape. In *Engineering Earth: The Impacts of Megaengineering Projects*, pages 2123–2138. Springer.

[303] Wei, Z. and Mukherjee, S. (2023). Examining income segregation within activity spaces under natural disaster using dynamic mobility network. *Sustainable Cities and Society*, 91:104408.

[304] Weinberg, D. H. (1979). The determinants of intra-urban household mobility. *Regional Science and Urban Economics*, 9(2-3):219–246.

[305] Weng, M., Ding, N., Li, J., Jin, X., Xiao, H., He, Z., and Su, S. (2019). The 15-minute walkable neighborhoods: Measurement, social inequalities and implications for building healthy communities in urban china. *Journal of Transport & Health*, 13:259–273.

[306] Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W., and Buckee, C. O. (2013). The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of the Royal Society Interface*, 10(81):20120986.

[307] Wetzstein, S. (2017). The global urban housing affordability crisis. *Urban Studies*, 54(14):3159–3177.

[308] Williams, N. E., Thomas, T. A., Dunbar, M., Eagle, N., and Dobra, A. (2015). Measures of human mobility using mobile phone records enhanced with gis data. *PloS one*, 10(7):e0133630.

[309] Wise, A., Kianian, B., Chang, H. H., Linton, S., Wolfe, M. E., Smith, J., Tempalski, B., Des Jarlais, D., Ross, Z., Semaan, S., et al. (2023). Socioeconomic and racial/ethnic spatial polarization and incarceration among people who inject drugs in 19 us metropolitan areas, 2015. *SSM-Population Health*, 23:101486.

[310] Wissink, B., Schwanen, T., and Van Kempen, R. (2016). Beyond residential segregation: Introduction.

[311] Wong, D. W. (1999). Geostatistics as measures of spatial segregation. *Urban geography*, 20(7):635–647.

[312] Wong, D. W. and Shaw, S.-L. (2011). Measuring segregation: An activity space approach. *Journal of geographical systems*, 13:127–145.

[313] Wu, J., Du, B., Gong, Z., Wu, Q., Shen, J., Zhou, L., and Cai, C. (2023). A gtfs data acquisition and processing framework and its application to train delay prediction. *International Journal of Transportation Science and Technology*, 12(1):201–216.

[314] Xiaohong, C., Jihao, D., and Peng, C. (2020). Measuring transit equity for people living in affordable housing: A spatial principal component analysis. In *International Conference on Transportation and Development 2020*, pages 36–49. American Society of Civil Engineers Reston, VA.

[315] Xu, Y., Belyi, A., Santi, P., and Ratti, C. (2019). Quantifying segregation in an integrated urban physical-social space. *Journal of the Royal Society Interface*, 16(160):20190536.

[316] Yabe, T. and Ukkusuri, S. V. (2020). Effects of income inequality on evacuation, reentry and segregation after disasters. *Transportation research part D: transport and environment*, 82:102260.

[317] Yao, J., Wong, D. W., Bailey, N., and Minton, J. (2019). Spatial segregation measures: A methodological review. *Tijdschrift voor economische en sociale geografie*, 110(3):235–250.

[318] Yao, S., Luo, D., and Wang, J. (2014). Housing development and urbanisation in china. *The World Economy*, 37(3):481–500.

[319] Zhang, Y., Cheng, S., Li, Z., and Jiang, W. (2023). Human mobility patterns are associated with experienced partisan segregation in us metropolitan areas. *Scientific Reports*, 13(1):9768.

[320] Zhao, Z., Shaw, S.-L., Xu, Y., Lu, F., Chen, J., and Yin, L. (2016). Understanding the bias of call detail records in human mobility research. *International Journal of Geographical Information Science*, 30(9):1738–1762.

[321] Zielstra, D. and Hochmair, H. H. (2011). Comparative study of pedestrian accessibility to transit stations using free and proprietary network data. *Transportation Research Record*, 2217(1):145–152.

[322] Zipf, G. K. (1946). The p1 p2/d hypothesis: On the intercity movement of persons. *American Sociological Review*, 11(6):677.

Appendix A

Details on Residential-Workplace Disparities

A.1 Spectral Clustering

Chapter 4.4.1 provides details regarding how we choose the number of clusters to create housing demographics for each city. Considering that each city has its own housing characteristics, we apply choose the initial number of clusters with respect to Graph Laplacian properties, for each city. Specifically, given the sorted eigenvalues derived from a city's Graph Laplacian, we set the number of clusters to reflect the largest gap in the first fifteen eigenvalues. Furthermore, we do not consider gaps for the first two eigenvalues, as we need a minimum of three clusters to derive the less vulnerable, mildly vulnerable, and most vulnerable demographics.

Figure A.1 shows the spectral gaps for each city, reflected by the vertical dashed line. Then, we apply K-means clustering to housing features, setting the number of clusters, $|\mathbf{SC}|$, to to reflect the spectral gap. Given that each census tract belongs to a cluster in $\mathbf{SC}$, we rank each K-Means cluster with respect to its housing features (discussed in detail in chapter 4.4.1). Then, we define the housing demographics, such that number of k-means clusters in

City	# Census Tracts $\sum_i^{CT} P_{ij} = 0$	# Census Tracts Total	H_{GN}^{in} Original	H_{GN}^{in} Adjusted	% Change
San Francisco	1	196	0.7573	0.7580	0.097
New York City	13	2164	0.8336	0.8342	0.078
Charleston	0	85	0.8808	0.8808	0.000
Bridgeport	0	210	0.9023	0.9023	0.000
Las Vegas	0	487	0.7869	0.7869	0.000
Albuquerque	0	153	0.8308	0.8308	0.000
Jacksonville	0	173	0.8342	0.8342	0.000
Chicago	3	1318	0.7982	0.7984	0.032
Detroit	2	1163	0.8653	0.8655	0.024
Philadelphia	1	384	0.8140	0.8144	0.044
Cincinnati	0	222	0.8702	0.8702	0.000
Milwaukee	0	297	0.8500	0.8500	0.000
New Orleans	1	176	0.7116	0.7124	0.110
Boston	1	204	0.7588	0.7595	0.092
Cleveland	0	446	0.8408	0.8408	0.000
Columbus	0	347	0.8541	0.8541	0.000
Dallas	0	529	0.8229	0.8229	0.000
Fort Worth	0	357	0.8616	0.8616	0.000
Gainesville	0	56	0.8127	0.8127	0.000
Greenville	0	111	0.8674	0.8674	0.000
Hartford	0	224	0.8671	0.8671	0.000
Houston	0	921	0.8728	0.8728	0.000
Indiannapolis	0	224	0.8093	0.8093	0.000
Kansas City	1	283	0.8887	0.8892	0.063
Memphis	0	221	0.8116	0.8116	0.000

Table A.1 The occurrences of zeros in the in-flow distribution has a minimal impact on global in-flow entropy measures. The second column reflects the number of census tracts in which the incoming node strength is zero. The third column lists the total number of census tracts in the given city. The fourth and fifth columns present the global in-flow entropy shown in Figure 4.2 and the global in-flow entropy, when adjusting for zeros. The final column shows the percent change between the original and adjusted global in-flow entropy in the previous two columns.

City	# Census Tracts $\sum_j^{CT} P_{ij} = 0$	# Census Tracts Total	H_{GN}^{out} Original	H_{GN}^{out} Adjusted	% Change
San Francisco	0	196	0.9824	0.9824	0.000
New York City	0	2164	0.9815	0.9815	0.000
Charleston	0	85	0.9405	0.9405	0.000
Bridgeport	0	210	0.9843	0.9843	0.000
Las Vegas	0	487	0.9809	0.9809	0.000
Albuquerque	0	153	0.9827	0.9827	0.000
Jacksonville	0	173	0.9761	0.9761	0.000
Chicago	0	1318	0.9838	0.9838	0.000
Detroit	0	1163	0.9805	0.9805	0.000
Philadelphia	0	384	0.9835	0.9835	0.000
Cincinnati	0	222	0.9772	0.9772	0.000
Milwaukee	0	297	0.9830	0.9830	0.000
New Orleans	0	176	0.9787	0.9787	0.000
Boston	1	204	0.9754	0.9763	0.092
Cleveland	0	446	0.9759	0.9759	0.000
Columbus	0	347	0.9765	0.9765	0.000
Dallas	0	529	0.9854	0.9854	0.000
Fort Worth	0	357	0.9784	0.9784	0.000
Gainesville	0	56	0.9520	0.9520	0.000
Greenville	0	111	0.9715	0.9715	0.000
Hartford	0	224	0.9847	0.9847	0.000
Houston	0	921	0.9693	0.9693	0.000
Indiannapolis	0	224	0.9772	0.9772	0.000
Kansas City	0	283	0.9734	0.9734	0.000
Memphis	0	221	0.9647	0.9647	0.000

Table A.2 The occurrences of zeros in the out-flow distribution has a minimal impact on global out-flow entropy measures. The second column reflects the number of census tracts in which the outgoing node strength is zero. The third column lists the total number of census tracts in the given city. The fourth and fifth columns present the global out-flow entropy shown in Figure 4.2 and the global out-flow entropy, when adjusting for zeros. The final column shows the percent change between the original and adjusted global out-flow entropy in the previous two columns.

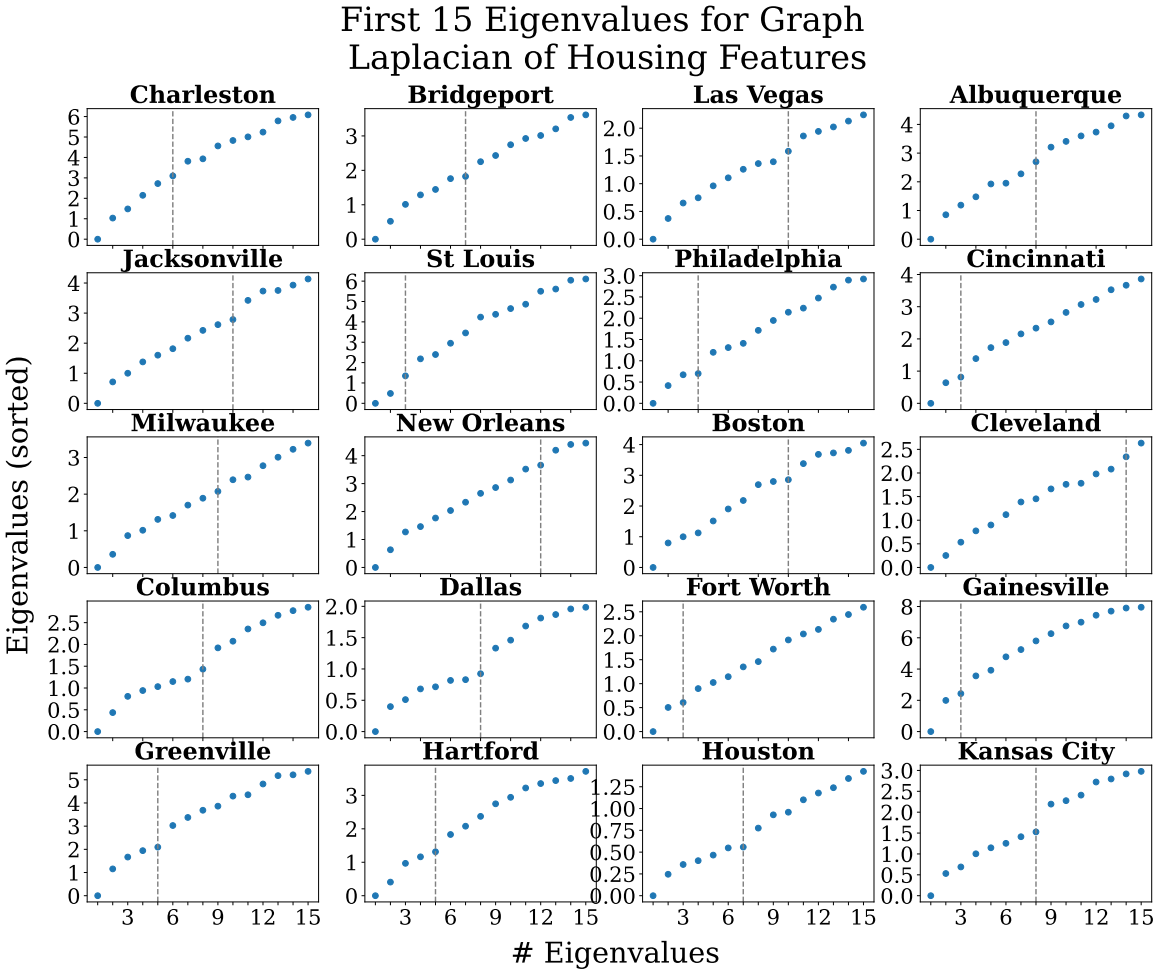

Fig. A.1 Sorted eigenvalues of the Graph Laplacian, for housing features in each city. The x-axis reflect the number of eigenvalues, while the y-axis shows the sorted eigenvalues themselves. The vertical dashed line denotes the number of clusters for which the spectral gap occurs.

the less and most vulnerable demographic are $\lfloor |\mathbf{SC}|/3 \rfloor$. Meanwhile, the number of clusters in the mildly vulnerable housing demographic is $\lfloor |\mathbf{SC}|/3 \rfloor + |\mathbf{SC}| \pmod 3$.

A.2 Socioeconomic Characteristics of Housing Demographics

In chapter 4.4.1, we introduce a spectral clustering approach for defining census tracts based on their vulnerability to housing insecurity. This consists of measuring housing affordability (i.e. rent burden, mortgage, housing density), quality (kitchen, plumbing, and phone facilities), and stability (i.e. evictions, overcrowding). In doing so we define three types of housing demographics: most vulnerable, mildly vulnerable, and less vulnerable. Figure A.2 displays the population distribution across each of the housing demographics. The dark red bar denotes the number of residents in census tracts that are the most vulnerable to housing insecurity. In this manner, we can observe that Fort Worth, Cleveland and Gainesville have the largest number of individuals living in vulnerable housing areas. Meanwhile, St. Louis Bridgeport, and Columbus represent cities in which less than ten percent of the population lives in neighbourhoods that are vulnerable to housing insecurity. In line with the capabilities approach to distributive justice, it is still crucial to consider accessibility for vulnerable individuals, regardless of how representative of the entire population they are.

The following following tables compares statistical features for each housing demographic in a given city, with respect to different sociodemographic characteristics. Each row in the Tables A.3 to A.8 reflects one of the 20 analysed cities. Columns one and two reflects the mean of the sociodemographic characteristic being considered for the less vulnerable and most vulnerable demographic, respectively. Columns three and four denote the median value, while the last two columns list the standard deviations.

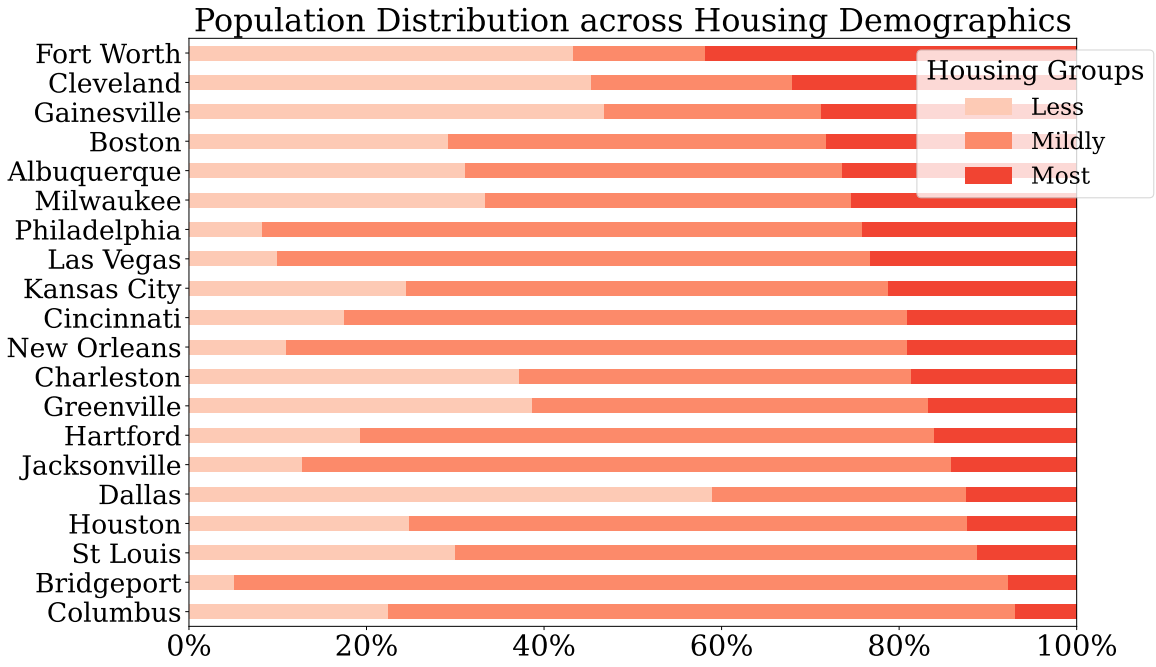

Fig. A.2 Population distribution across each housing demographic. The x-axis reflects the percentage of the population that lives in a particular housing demographic, denoted by the colour of the bar. Lighter hues convey individuals living in neighbourhoods that are less vulnerable to housing insecurity. The y-axis specifies which city is being considered.

Table A.3 Statistical properties of educational attainment levels in the census tracts that are less and most vulnerable to housing insecurity. Generally, the most vulnerable tracts tend to have a lower fraction of their population holding an Associate's Degree or higher.

% holding an Associate's Degree or higher						
Housing Vuln	Mean		Median		Std. Dev.	
City	*Less*	*Most*	*Less*	*Most*	*Less*	*Most*
Philadelphia	80.119	23.396	80.749	19.360	9.610	14.604
Milwaukee	55.803	18.191	56.072	17.002	17.464	9.302
Boston	58.571	35.095	58.734	28.301	17.983	16.826
San Francisco	80.389	45.850	82.193	45.643	7.345	16.154
Cincinnati	72.222	31.664	75.347	26.736	12.531	18.211
Cleveland	53.462	27.862	54.572	24.836	17.920	14.763
Jacksonville	65.295	22.711	64.205	19.490	7.747	10.979
New Orleans	73.245	26.684	75.641	24.974	9.834	13.116
Albuquerque	45.388	41.100	43.834	37.610	20.369	19.230
Las Vegas	50.848	19.058	49.053	16.347	12.000	10.010
Houston	56.746	24.969	60.206	16.701	18.521	19.410
Gainesville	57.918	57.915	63.921	58.528	18.829	19.306
Columbus	71.569	23.727	72.214	18.338	13.078	14.697
Charleston	66.389	34.246	64.399	30.604	12.650	17.964
Kansas City	59.111	24.743	60.354	23.201	15.818	12.798
Greenville	60.974	33.848	61.077	29.318	15.321	14.950
Fort Worth	49.099	25.094	47.810	21.889	17.381	14.777
Bridgeport	87.083	20.912	86.994	15.518	5.616	13.968
Hartford	69.616	22.199	72.019	19.743	11.235	14.066
Dallas	41.521	41.458	38.129	30.318	24.376	30.201

Table A.4 Statistical properties of median income levels in the census tracts that are less and most vulnerable to housing insecurity. Generally, the most vulnerable tracts tend to have a lower household incomes than the less vulnerable tracts.

Median Household Income						
Housing Vuln	Mean		Median		Std. Dev.	
City	*Less*	*Most*	*Less*	*Most*	*Less*	*Most*
Philadelphia	92314	30770	95430	29656	30282	11558
Milwaukee	78811	30854	71563	30051	27030	10087
Boston	102007	44970	101769	48055	31691	15349
San Francisco	165592	46839	163226	44504	32195	21257
Cincinnati	102469	37120	100419	32658	27703	18578
Cleveland	80359	37918	73854	32944	34264	18610
Jacksonville	95083	34874	93378	34149	20081	11993
New Orleans	97843	30681	93193	26728	33092	13186
Albuquerque	68917	52552	58308	48651	31672	20258
Las Vegas	99783	42305	91823	39302	30376	16136
Houston	103621	51605	102798	41707	34873	31687
Gainesville	62339	33310	62327	30355	28860	13363
Columbus	116719	33665	110059	35250	32699	9550
Charleston	99792	39785	97486	33933	36397	12531
Kansas City	101569	40974	95282	38789	31163	12126
Greenville	86751	44149	82284	45844	26603	10843
Fort Worth	90964	55642	80820	52006	38736	21510
Bridgeport	244254	39408	250001	39862	17490	12716
Hartford	126334	36465	124215	33518	30665	13081
Dallas	77307	87544	67630	48750	38263	75188

Table A.3 displays statistic properties (i.e. mean, median, standard deviation) for educational attainment in neighbourhoods that are less vulnerable and most vulnerable to housing insecurity. Barring Gainesville and Dallas, we highlight that Table A.3 reveals much larger percentage of individuals in less vulnerable neighbourhoods hold an associate's degree or higher, compared to individuals in the most vulnerable census tracts. Comparing their median values, which is more robust to extreme values, we observe that even the less vulnerable demographics in Gainesville and Dallas have a higher fraction of residents with more educational attainment.

Table A.5 Statistical properties of unemployment levels in the census tracts that are less and most vulnerable to housing insecurity. Generally, the most vulnerable tracts tend to have a higher unemployment rates, with the two exceptions being Albuquerque and Boston.

% Unemployed						
Housing Vuln	Mean		Median		Std. Dev.	
City	*Less*	*Most*	*Less*	*Most*	*Less*	*Most*
Philadelphia	2.020	6.743	1.745	6.332	1.458	3.658
Milwaukee	2.103	5.489	2.048	5.000	1.241	3.054
Boston	3.587	6.206	2.959	5.965	3.092	2.637
San Francisco	2.643	4.910	2.729	4.977	1.095	2.182
Cincinnati	1.854	6.423	1.626	5.695	1.200	4.165
Cleveland	3.014	6.620	2.524	6.067	1.991	4.438
Jacksonville	1.801	5.091	1.915	4.894	0.971	2.537
New Orleans	2.672	7.779	1.968	6.926	2.135	5.950
Albuquerque	2.639	3.968	2.511	3.526	1.803	2.632
Las Vegas	2.959	5.285	2.674	4.928	2.032	2.540
Houston	3.197	5.056	2.797	4.702	1.967	3.179
Gainesville	3.317	3.460	2.582	2.641	2.761	2.204
Columbus	1.769	6.514	1.433	5.466	1.321	3.618
Charleston	1.894	4.041	1.428	3.830	1.515	2.515
Kansas City	2.309	4.453	1.931	3.865	1.571	3.068
Greenville	1.887	4.002	1.504	3.254	1.523	3.096
Fort Worth	2.727	3.887	2.422	3.624	1.799	2.152
Bridgeport	2.704	9.015	2.534	7.837	1.926	3.441
Hartford	2.170	6.715	1.896	6.501	1.257	3.080
Dallas	3.046	3.361	2.848	3.072	1.888	2.015

Table A.6 Statistical properties of poverty levels in the census tracts that are less and most vulnerable to housing insecurity. The most vulnerable tracts tend to have much higher rates of poverty than the less vulnerable tracts.

% Households in Poverty						
Housing Vuln	Mean		Median		Std. Dev.	
City	*Less*	*Most*	*Less*	*Most*	*Less*	*Most*
Philadelphia	12.820	36.373	8.201	35.345	13.306	11.983
Milwaukee	7.187	37.104	6.240	35.231	4.120	13.678
Boston	9.832	27.269	7.946	25.214	7.270	11.130
San Francisco	6.073	21.947	5.568	21.908	3.258	8.956
Cincinnati	4.854	33.685	4.551	32.708	2.972	16.573
Cleveland	7.838	27.423	5.958	27.612	6.693	12.705
Jacksonville	6.626	28.542	6.006	27.585	4.006	13.677
New Orleans	7.425	32.906	6.820	30.253	3.968	14.370
Albuquerque	13.406	17.642	10.991	17.431	9.214	8.978
Las Vegas	5.568	24.521	5.172	25.010	3.381	10.614
Houston	6.942	24.196	5.148	25.927	5.992	13.353
Gainesville	15.440	34.209	7.347	32.663	17.478	14.196
Columbus	3.467	35.648	2.541	35.802	2.470	11.374
Charleston	5.904	26.539	5.569	27.570	2.893	10.405
Kansas City	3.685	23.448	3.114	22.395	3.167	10.727
Greenville	6.067	17.952	4.240	14.480	4.541	8.252
Fort Worth	7.475	18.609	6.355	17.460	5.376	11.507
Bridgeport	2.691	31.134	2.430	28.311	1.748	10.849
Hartford	3.349	29.652	3.045	27.418	2.067	11.725
Dallas	12.004	15.867	9.694	15.005	9.150	11.072

Table A.7 Statistical properties of transit-based commuting levels in the census tracts that are less and most vulnerable to housing insecurity. The most vulnerable tracts have higher rates of transit-based commutes than the less vulnerable tracts.

% Commuting via Public Transit						
Housing Vuln	Mean		Median		Std. Dev.	
City	*Less*	*Most*	*Less*	*Most*	*Less*	*Most*
Philadelphia	17.101	35.238	17.137	33.522	4.676	10.905
Milwaukee	2.176	11.393	1.245	9.453	2.591	9.159
Boston	26.087	34.146	26.303	33.251	10.222	7.438
San Francisco	30.810	29.142	30.237	27.606	8.647	11.926
Cincinnati	1.183	11.053	0.703	9.987	1.424	7.867
Cleveland	2.018	9.853	1.014	6.870	3.138	9.375
Jacksonville	0.182	6.122	0.000	4.241	0.302	6.156
New Orleans	3.461	11.331	2.590	8.401	3.809	10.271
Albuquerque	1.758	1.137	0.587	0.486	2.489	1.695
Las Vegas	0.731	8.027	0.000	6.478	1.449	6.872
Houston	1.820	3.460	1.144	1.997	1.981	4.357
Gainesville	1.997	7.977	0.679	6.787	2.664	7.744
Columbus	0.828	6.982	0.256	7.556	1.363	4.251
Charleston	0.960	4.607	0.506	1.955	1.466	6.912
Kansas City	0.676	5.433	0.000	3.854	1.239	6.112
Greenville	0.407	0.640	0.000	0.000	0.951	1.179
Fort Worth	0.420	0.757	0.000	0.339	0.755	1.280
Bridgeport	25.208	14.798	25.975	12.354	6.591	8.226
Hartford	1.092	11.899	0.842	8.482	1.149	10.187
Dallas	2.113	3.212	1.315	1.607	3.108	4.509

Table A.8 Statistical properties of commuting times in the census tracts that are less and most vulnerable to housing insecurity. This value includes all commuters, regardless of what mode of transit they use for commuting. The distinction between housing demographics based on commuting times tends to be less clear than other sociodemographic indicators.

% Commuting over an Hour						
Housing Vuln	Mean		Median		Std. Dev.	
City	*Less*	*Most*	*Less*	*Most*	*Less*	*Most*
Philadelphia	8.451	16.599	7.138	14.732	4.743	7.555
Milwaukee	2.949	5.597	2.690	4.492	1.761	4.214
Boston	11.518	15.397	10.362	13.412	5.490	7.756
San Francisco	13.594	8.530	13.026	7.996	4.267	3.645
Cincinnati	2.332	6.745	1.981	5.824	1.839	4.811
Cleveland	3.776	6.352	3.292	4.762	3.331	5.511
Jacksonville	3.273	5.983	3.039	5.570	2.268	3.777
New Orleans	3.963	6.697	3.325	5.523	2.493	5.357
Albuquerque	5.888	4.345	4.779	3.231	4.991	4.033
Las Vegas	3.021	7.154	2.463	6.356	2.522	4.773
Houston	8.545	9.977	8.177	9.090	5.031	5.583
Gainesville	4.064	3.039	3.547	1.842	2.945	2.710
Columbus	3.101	6.598	2.540	6.429	2.460	4.540
Charleston	3.936	5.073	2.469	3.600	3.918	6.056
Kansas City	3.333	3.494	2.743	2.979	3.354	3.461
Greenville	3.261	3.663	2.363	3.404	3.093	2.763
Fort Worth	6.637	7.723	6.057	7.162	3.773	3.673
Bridgeport	35.148	14.149	36.190	12.915	6.890	6.525
Hartford	5.847	5.608	5.914	4.832	2.344	4.052
Dallas	7.117	7.185	6.195	7.031	4.973	4.784

A.4 Residential Proximity to Transit Core

Table A.9 Mean distance (km) from tracts to transit core, with respect to housing vulnerability.

City	Most Vulnerable	Mildly Vulnerable	Less Vulnerable
Philadelphia	6.267	6.729	8.512
Milwaukee	5.003	6.889	9.650
Boston	5.007	5.571	4.085
San Francisco	3.959	4.454	3.065
Cincinnati	5.355	9.800	13.344
Cleveland	7.063	11.800	16.417
Jacksonville	10.702	10.226	15.331
New Orleans	5.193	5.218	6.116
Albuquerque	8.530	9.107	11.138
Las Vegas	8.333	18.854	17.406
Houston	19.868	22.007	21.452
Gainesville	4.160	10.509	8.510
Columbus	8.514	13.957	17.122
Charleston	8.338	12.028	10.397
Kansas City	8.594	17.018	22.131
Greenville	13.575	8.748	10.541
Fort Worth	13.530	15.031	17.993
Bridgeport	11.411	18.355	22.971
Hartford	6.749	14.600	13.577
Dallas	15.034	14.182	13.770

A.3 Transit Efficiency for All Cities

In chapter 4.5.2, we highlight three cities (Cleveland, Albuquerque, and Bridgeport) to show distinct trends in transit systems, when considering how transit efficiency corresponds with the length of a particular transit journey. Figure A.3 expands on this idea, showing transit efficiency as a function of distance for all 20 cities in our analysis.

Figure 4.11 in chapter 4.5.3 highlights how the neighbourhoods that are most vulnerable to housing security tend to live closer to the transit system's centre of mass. Table A.9 expands on this, listing the average distance (km) between tracts in a housing demographic and the transit core.

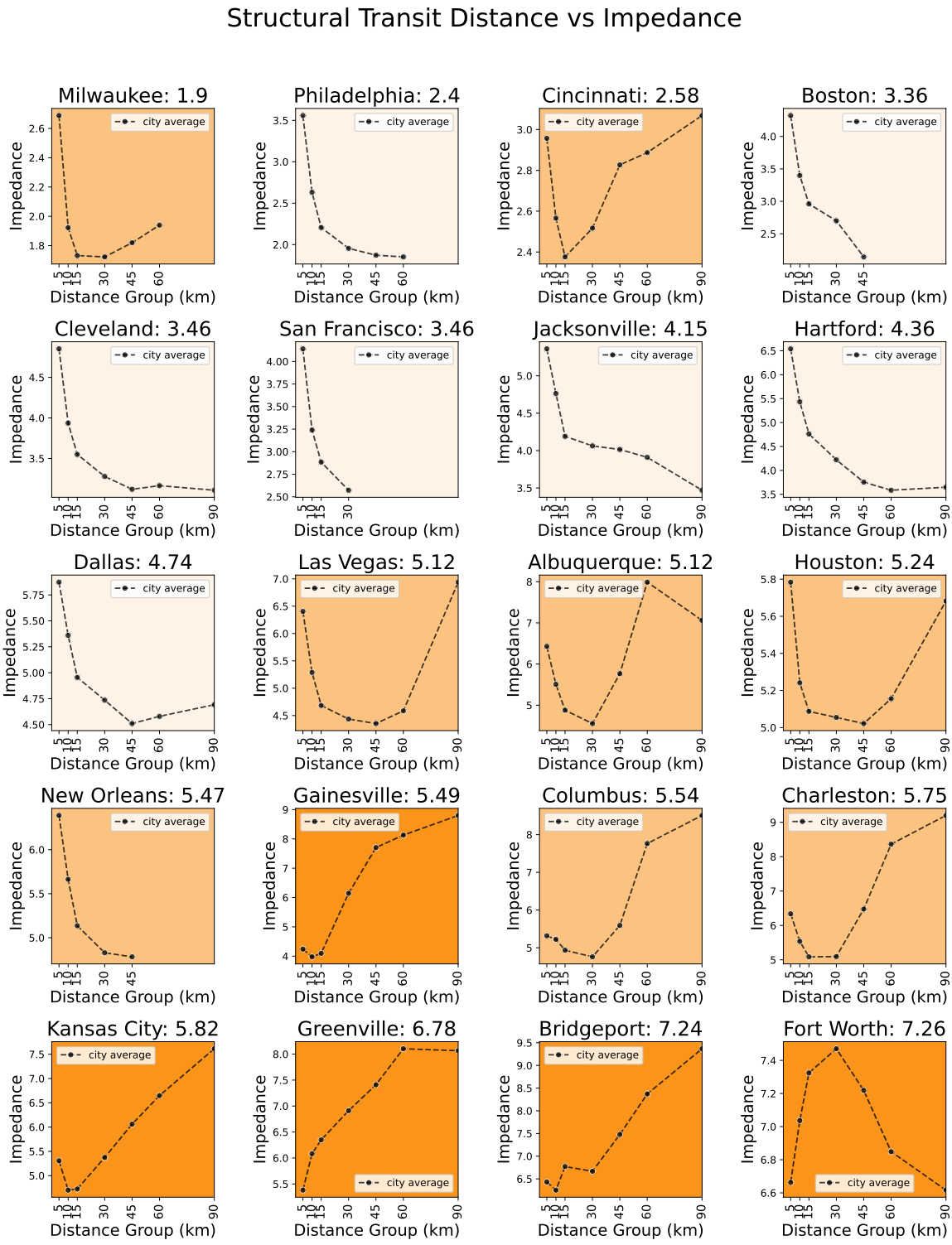

Fig. A.3 Transit efficiency as a function of trip distance for 20 US cities. Lighter orange plots reflect efficient systems in which travel impedance decreases as trip distance increases (Philadelphia, San Francisco, Boston, Cleveland, Jacksonville, Hartford, Dallas, and New Orleans). Dark orange plots indicate inefficient transit systems, with increasing travel impedance as trip distance increases (Gainesville, Kansas City, Greenville, Bridgeport, Fort Worth). The remaining orange plots capture cities with moderately efficient transit service, in which the relationship between travel impedance and trip distance switches from negative to positive at a given trip distance threshold (Milwaukee, Cincinnati, Albuquerque, Las Vegas, Charleston, Houston, Columbus).

Appendix B

Details on Transit Segregation

B.1 Segregation across Multiple Urban Dimensions in the Null Model

This section serves to complement Figure 5.9 in this main manuscript. Figure B.1 shows how measures of segregation change across urban dimensions, using amenity visitations that are generated for the null model rather than empirical mobility patterns. We observe that when spatial disparities in amenity segregation are removed, segregation in the transit dimension (middle axis) continues to exist. Furthermore, by building a null model in which we define mobility destinations as amenities that are sampled from uniform spatial distribution, segregation at the destination level (right axis) converges to one value. Moreover, destination segregation converges to values that are reflective of the city's overall socioeconomic composition.

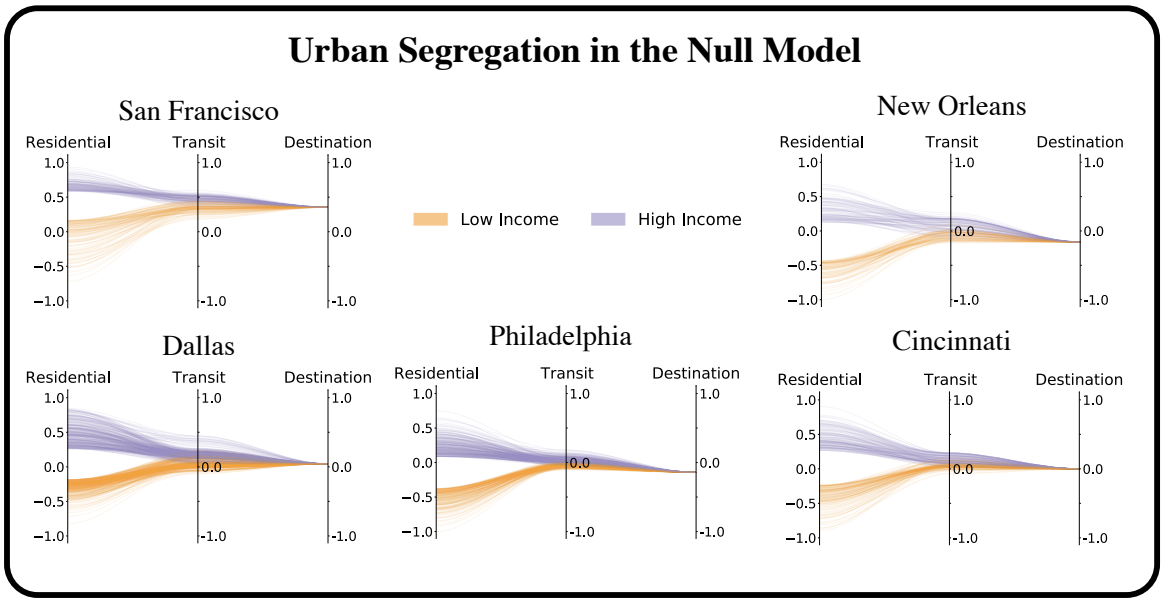

Fig. B.1 Changes in segregation between the residential, transit, and mobility domains, using mobility journeys generated for the null model comparison. Shown for 5 US Cities, for which the null model simulates mobility patterns to amenities that are uniformly distributed across a given city. Orange reflects the HS-Lo segregation group and purple represents the HS-Hi segregation group.

Appendix C

Details on the Built Environment Analysis

C.1 Built Environment Index: Robustness Check

Our approach to calculating the built environment (BE) index consists of calculating the geometric mean of its three components: street griddedness, building density, and amenity diversity. We choose the geometric mean as it is less sensitive to outliers, which is particularly given that some of the components' distributions are skewed. Figure C.1 shows the correlation coefficient between each of the BE components and the BE index. Darker colours denote larger correlation coefficients. Non-significant correlations, with p-values greater than or equal to 0.05, are indicated by the grey cells. Design refers to the grid index which measures the griddedness of a census tract's street layout. Diversity reflects the distribution of amenity types in a neighbourhoods, while density conveys the fraction of land area that consists of buildings. By looking at the last columns (or row) of each city's heat map, we can see that each component tends to have a stronger correlation with the BE index, than with the other components. Thus, we observe that each component captures different aspects of the built environment. We note how diversity tends to be the least correlated with the BE index, having non-significant correlations in Portland and Seattle.

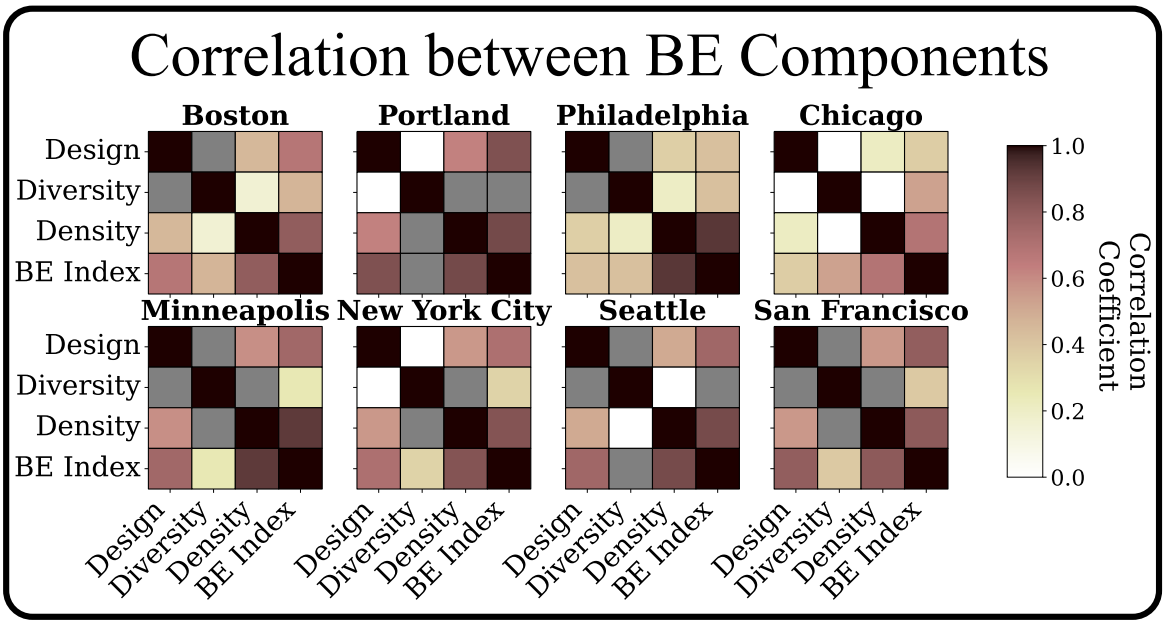

Fig. C.1 Correlation of BE components and the BE Index. Darker colours imply larger, significant correlations. Grey cells indicate non-significant correlations.

Since each BE component has different distributions, we consider different approaches to deriving the BE index, similar to the robustness tests for the grid index [34]. Figure C.2 shows the results of the validation, with the left most panels reflecting the distributions of BE features that we use for our analysis in chapter 6. The black line shows the distribution of the BE index, while the dashed lines show the spread of the components used to build the BE index. The orange, green, and yellow lines denote the distribution of grid indices, building density, and amenity diversity, respectively. The two alternative approaches we use involve mean normalisation and quantile normalisation of the components. The first involves standardising each component using mean normalisation, before scaling each dimension to range from 0 to 1, using min-max normalisation. Then, the BE index is calculated in a similar manner, taking the cube root of the three components. This can be seen through the centre plots for each city. By comparing the geometric mean (left most panels) to the mean normalised geometric mean (centre panels), we see that the shape of BE index distribution has minimal changes. Thus, standardising each component to have a mean of 0 and standard

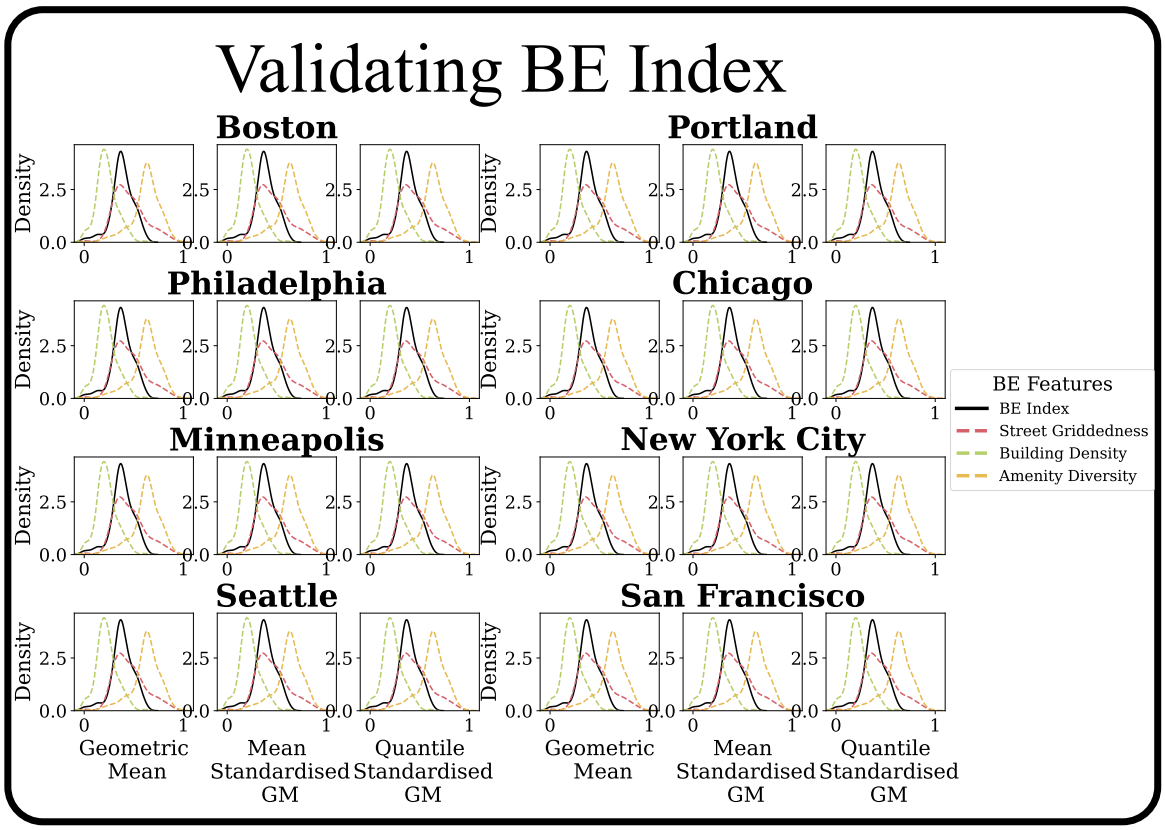

Fig. C.2 Comparing the distribution of BE components for various approaches to measuring the BE index. Pink, green, and orange dashed lines correspond to the street grid index, building density, and amenity diversity respectively. The black, solid line reflects the final BE index measured using the three components. The left most panel shows distributions for the geometric mean (GM), while the centre and right most panel convey the distribution for the mean-normalised GM and the quantile-normalised GM, respectively.

deviation of 1, impacts the distribution of each component, but does not lead to any drastic changes in the final spread of the BE index.

The second approach involved applying quantile standardisation to all the components, then applying min-max normalisation. This approach assumes all three components have a similar distribution and limits the effects of outliers. The effects quantile normalisation has on the BE components and index are shown in the right most panel for each city. While the distribution is more even across the entire range, the skews of the distributions themselves are not retained. Ultimately, we choose to retain the non-standardisation approach (left-most panels) as it allows for comparability across cities. Furthermore, standardisation allows

Table C.1 Correlation between BE index and transit characteristics, for different approaches of BE feature standardisation.

	Geometric Mean (GM)		**Mean-Standardised GM**		**Quantile-Standardised GM**	
City	Opportunity	Commute	Opportunity	Commute	Opportunity	Commute
Chicago	-0.063	**-0.232*****	-0.03	**-0.192*****	0.05	**-0.102****
New York City	**-0.261*****	**-0.064****	**-0.261*****	**-0.065****	**-0.239*****	0.023
Philadelphia	**-0.316*****	**-0.488*****	**-0.311*****	**-0.484*****	**-0.435*****	**-0.528*****
San Francisco	**-0.467*****	**-0.503*****	**-0.375*****	**-0.42*****	**-0.373*****	**-0.372*****
Boston	**-0.485*****	**-0.65*****	**-0.483*****	**-0.649*****	**-0.476*****	**-0.611*****
Seattle	**-0.498*****	-0.16	**-0.434*****	-0.186	**-0.431*****	-0.143
Portland	**-0.607*****	**-0.208***	**-0.515*****	-0.175	**-0.554*****	-0.157
Minneapolis	**-0.783*****	**-0.19****	**-0.784*****	**-0.205*****	**-0.679*****	**-0.168****

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

for less interpretability of each of the components. Table C.1 shows how the correlations between the BE index and transit characteristics changes based on the chosen standardisation approach. The first two columns of Table C.1 reflect the chosen approach, used in the formal analysis. These columns are the same as the last two columns in Table 6.2.

We can observe how the mean-standardised and quantile-standardised approaches result in fewer cities having significant correlations between their built environment and transit commuting times. Meanwhile, we see that for each approach, neighbourhoods with higher mobility flexibility, conveyed through the BE index, tend to have shorter commute times. This holds true for all cities except Chicago. As a final justification, we compare the spatial distribution of BE indices for each approach in Figure C.3, using New York City as an example. We observe that the distribution between the geometric mean (GM) and mean-standardised GM are similar. This can be confirmed through the left two plots in Figure C.2 for New York City, as the shape of the distribution remains the same, but the mean-normalised index spans a greater range. Thus, we would expect to see a larger range of colours in the mean-normalised spatial distribution, despite the two version having a similar spread. We

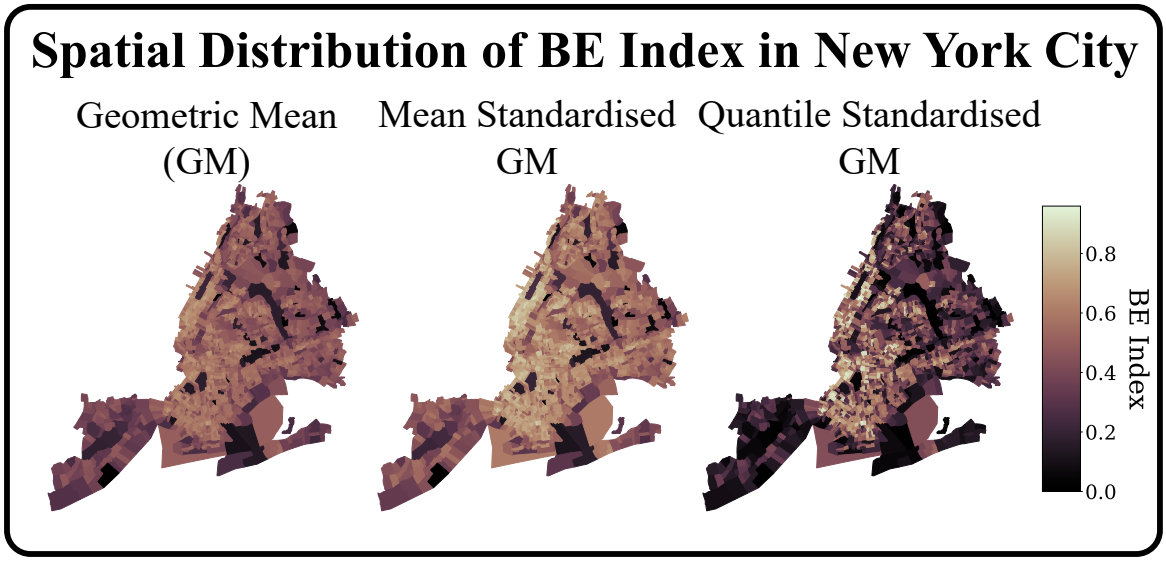

Fig. C.3 Comparing the spatial distribution of different BE index approaches for New York City. Lighter hues convey a larger BE index and, in turn, higher mobility flexibility for a given census tract.

choose the GM method over the mean-normalised, to ensure a BE index in one city captures the same value as a BE index in another.

The quantile-normalised version reflects a much starker difference and, most notably, does not identify high built environment areas in Queens (middle right of NYC) or the Bronx. Thus, to ensure explainability of the BE Index and comparability across cities, we retain the original approach, in which we simply take the cube root of the street griddedness, building density, and amenity diversity.

C.2 BE Group Characteristics

This section provides additional details regarding the characteristics of the built environment (BE) groups, discussed in section 6.2. Figure C.4 shows the distribution of different BE components across each BE tercile. The top panel highlights the distribution of the grid index, which reflects street design, for each BE group, across all of the eight US cities shown

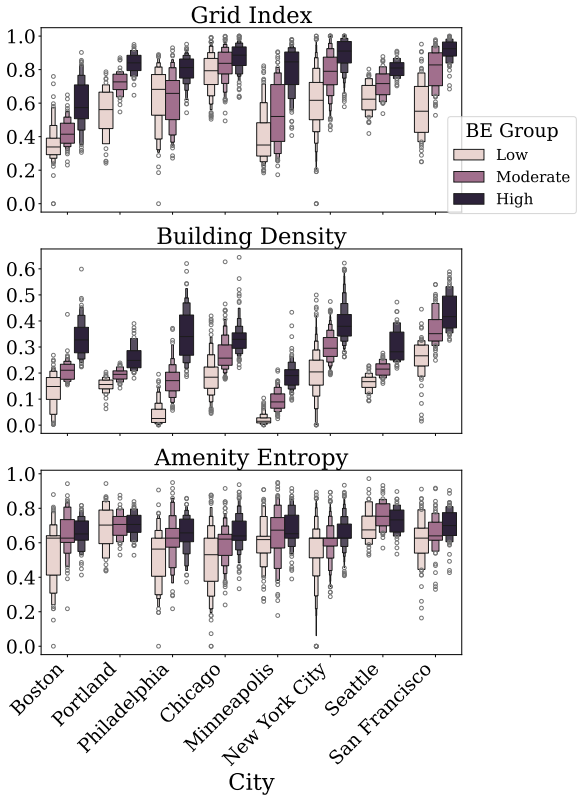

Fig. C.4 Distribution of BE features across the three BE groups. The x-axis refers to the city being analysed. The top panel reflects levels of street design, while the bottom two panels convey building density and amenity diversity, respectively. The y-axis shows the values for each of the features. Finally, lighter shades of purple reflect the distribution of BE characteristics for tracts in the Low BE group, while the darkest purple shows the features for neighbourhoods in the High BE Group.

on the x axis. The centre and bottom panels convey building density and amenity entropy, which represent the density and diversity components of the built environment, respectively.

Table C.2 conveys statistically significant distribution across Low and Moderate BE groups, as well as Moderate and High BE groups, that are shown in Figure 6.2

C.3 Integral Accessibility

Section 6.4.2 explores how socioeconomic homophily in transit service changes across different travel time thresholds. To do so, we compare levels of network assortativity in

	Opportunity		**Commute**	
City	Low-Mod	Mod-High	Low-Mod	Mod-High
Boston	**0.382*****	**0.471*****	**0.41*****	**0.565*****
Portland	**0.335***	**0.595*****	0.298	**0.411****
Philadelphia	0.086	**0.406*****	0.165	**0.45*****
Chicago	0.052	0.1	**0.143***	**0.189*****
Minneapolis	**0.457*****	**0.517*****	**0.211***	**0.382*****
New York City	**0.19*****	0.067	**0.102****	**0.116*****
Seattle	**0.407****	**0.348***	0.269	0.308
San Francisco	**0.297****	**0.328****	0.255	**0.313***

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Table C.2 Kolmogorov-Smirnov test statistic when comparing the distribution of transit times between the Low and Moderate BE group (columns two and four) as well as between the Moderate and High BE Group (columns three and five). We consider differences in distributions with respect to mobility opportunity transit times (second and third columns) and transit commuting times (last two columns). Bold values reflect instances in which the compared distributions are not the same.

transit access networks. Furthermore, we discuss the need to ensure that the increasing socioeconomic homophily associated with longer travel time thresholds is not an artefact of unequal socioeconomic demographic representation or the number of components in the network. Figures C.5, C.6, and C.7 reflect how our empirical transit access networks compare to configuration models, when considering all neighbourhoods, low BE neighbourhoods, and high BE neighbourhoods, respectively.

C.4 Relative Accessibility

Section 6.4.3 uses New York City and San Francisco as examples to underscore the importance of considering BE features when analysing transit inequality. Figure C.8 shows relative accessibility across socio-spatial groups for the remaining six cities.

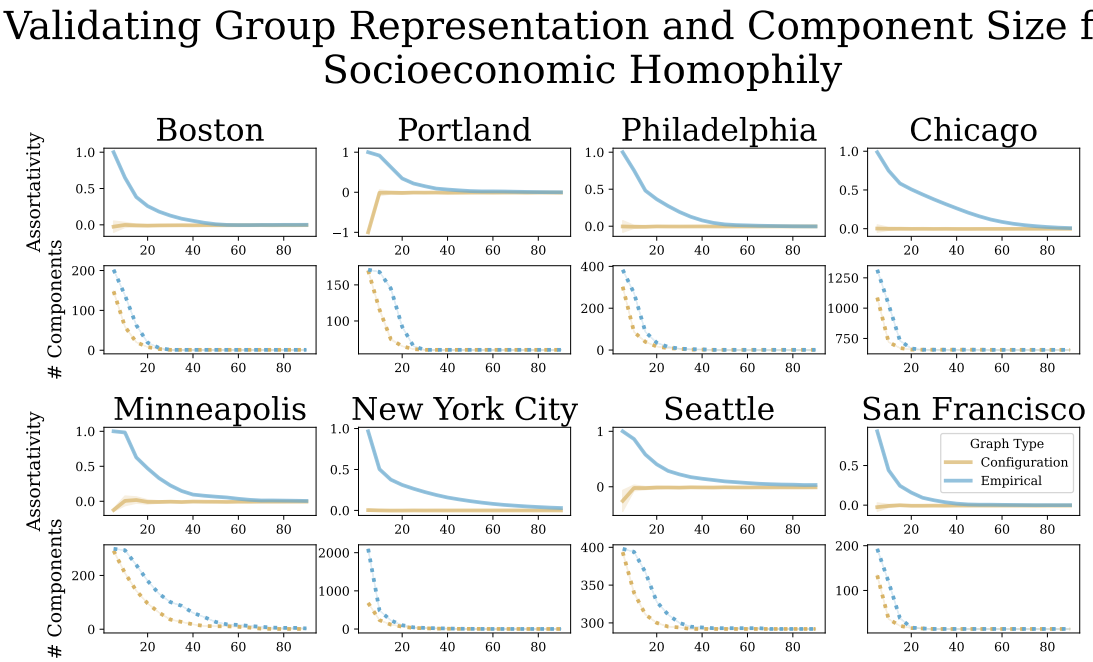

Fig. C.5 Comparing empirical transit access networks to their respective configuration models. For each city, the solid blue and yellow lines (top panels) show the coefficients of network assortativity (y-axis) for the empirical network and configuration models, respectively, across different travel time thresholds (x-axis). Meanwhile the dotted blue and yellow lines convey how the number of components (y-axis) changes for empirical and configuration models, respectively, as travel time thresholds increase (x-axis).

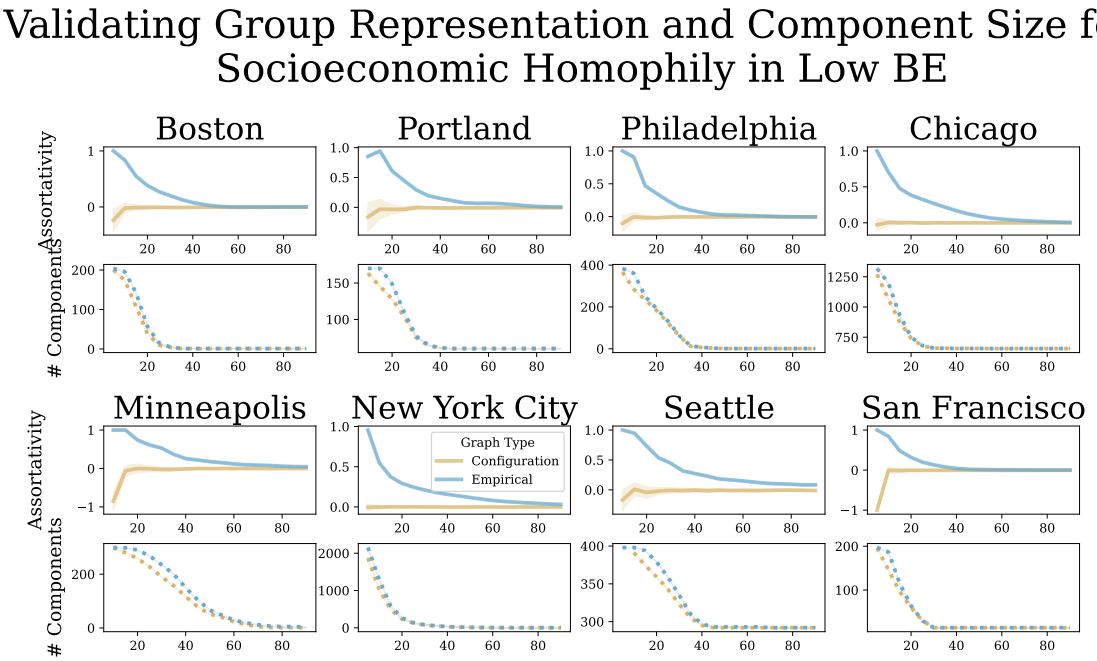

Fig. C.6 Comparing empirical transit access networks from Low BE neighbourhoods to their respective configuration models. For each city, the solid blue and yellow lines (top panels) show the coefficients of network assortativity (y-axis) for the empirical Low BE network and configuration models, respectively, across different travel time thresholds (x-axis). Meanwhile the dotted blue and yellow lines convey how the number of components (y-axis) changes for Low BE empirical and configuration models, respectively, as travel time thresholds increase (x-axis).

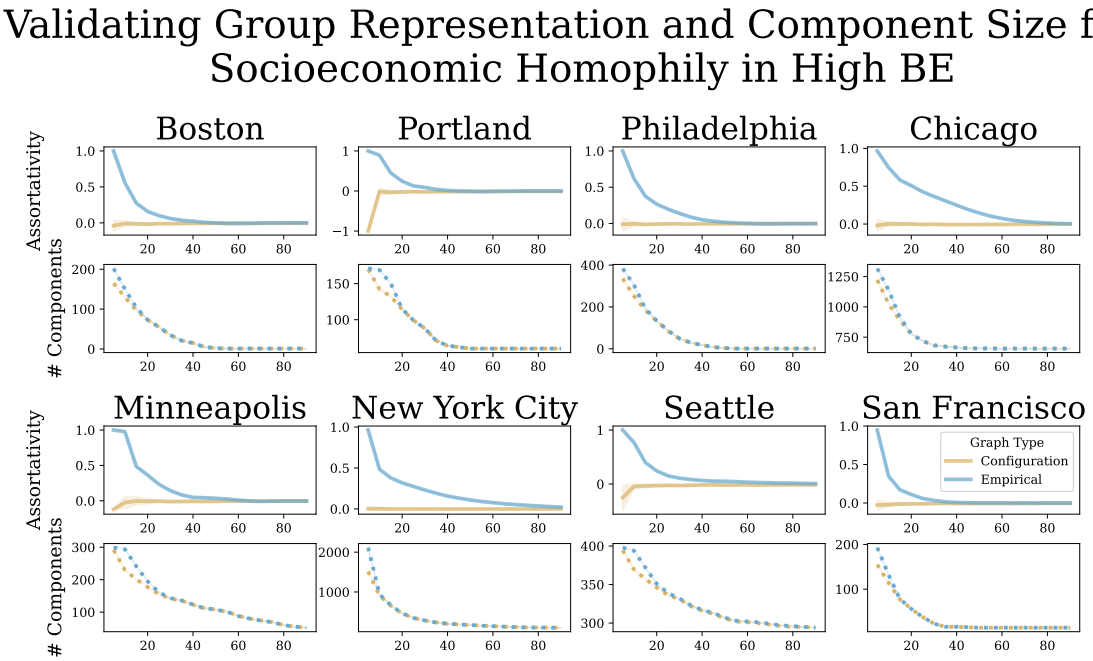

Fig. C.7 Comparing empirical transit access networks from High BE neighbourhoods to their respective configuration models. For each city, the solid blue and yellow lines (top panels) show the coefficients of network assortativity (y-axis) for the empirical High BE network and configuration models, respectively, across different travel time thresholds (x-axis). Meanwhile the dotted blue and yellow lines convey how the number of components (y-axis) changes for High BE empirical and configuration models, respectively, as travel time thresholds increase (x-axis).

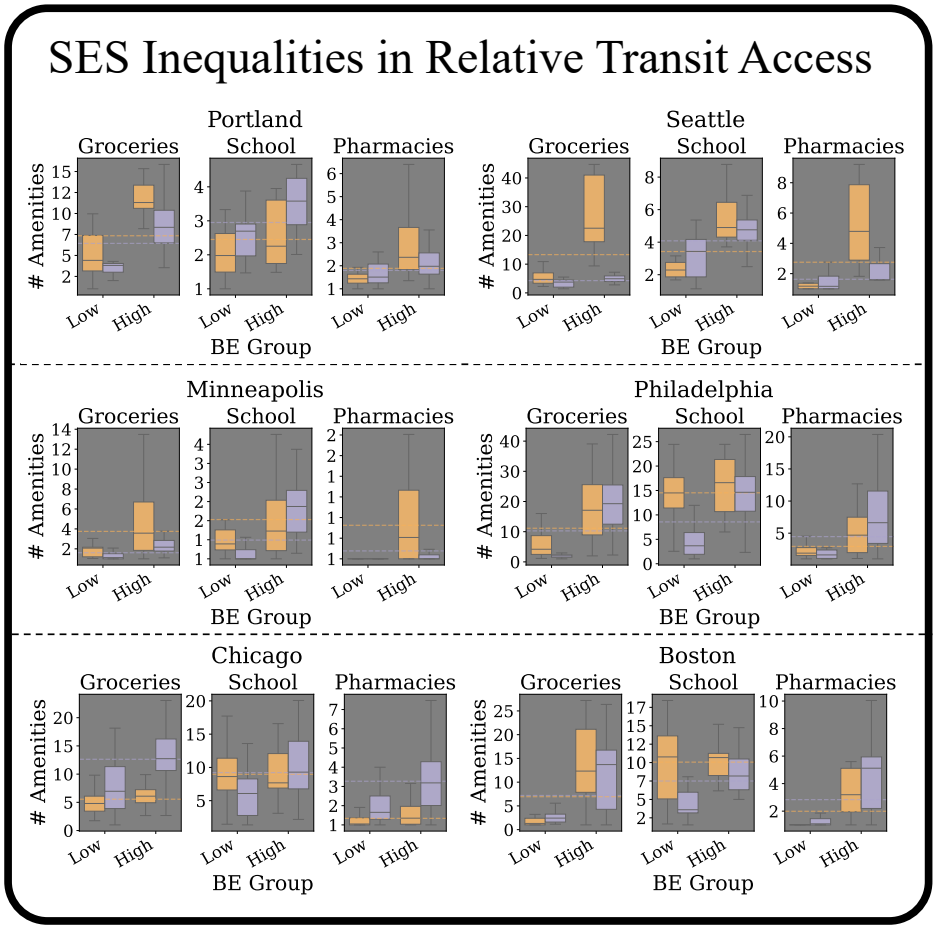

Fig. C.8 Socio-spatial inequalities in relative transit accessibility to essential amenities. The x-axis denotes the spatial demographic (BE Group), while the orange and purple reflect the low and high income socioeconomic groups, respectively. The orange and purple horizontal lines convey the accessibility of low and high socioeconomic groups, when the BE is not considered.

C.5 Workplace Access

In chapter 6.4.4 we discuss accessibility differences across socioeconomic groups in similar built environments, with respect to workplace accessibility within a 50 minute transit commute. Figure 6.6 shows the difference between the percentage of workers in high and low income neighbourhoods, in similar built environments, who have workplace accessibility. Table C.3 expands on this.

Table C.3 Percent of individuals in each socio-spatial demographic that can commute to their employment locations within a 50 minute transit journey. This table accompanies Figure 6.6, which compares percentage point differences in the Low and High income group across all BE groups.

		BE Group			
City	Income Group	*Low*	*Moderate*	*High*	*All*
Boston	Low	71.36	74.66	87.74	77.81
	Middle	48.4	67.86	87.84	68.38
	High	55.55	77.33	92.35	77.52
Chicago	Low	34.06	33.40	33.44	33.77
	Middle	35.22	34.75	34.85	34.90
	High	44.32	55.47	55.26	53.48
Minneapolis	Low	13.41	33.44	46.21	39.16
	Middle	11.76	25.83	42.84	25.87
	High	11.34	24.26	38.31	20.14
New York City	Low	22.58	27.28	29.24	26.38
	Middle	18.27	22.42	27.46	22.94
	High	15.04	20.69	38.29	25.61
Philadelphia	Low	70.98	74.67	79.04	73.84
	Middle	56.2	58.42	67.08	60.84
	High	44.04	43.96	84.76	61.61
Portland	Low	35.33	42.99	61.69	42.16
	Middle	34.50	43.23	58.05	45.56
	High	47.05	53.54	55.03	52.23
San Francisco	Low	76.98	82.81	90.98	83.25
	Middle	83.17	81.80	92.51	86.24
	High	90.49	92.75	95.63	92.44
Seattle	Low	13.66	20.99	44.56	21.89
	Middle	7.37	11.86	22.74	13.73
	High	8.34	9.48	16.19	10.10