University of California
Santa Barbara

# Sensing Human Activity and Interaction Patterns through Movement Observations

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Geography

by

Rongxiang Su

Committee in charge:

Professor Konstadinos G. Goulias, Co-chair
Associate Professor Somayeh Dodge, Co-chair
Professor Trisalyn Nelson

December 2023

The Dissertation of Rongxiang Su is approved.

---

Professor Trisalyn Nelson

---

Associate Professor Somayeh Dodge, Committee Co-Chair

---

Professor Konstadinos G. Goulias, Committee Co-Chair

December 2023

Sensing Human Activity and Interaction Patterns through Movement Observations

Copyright © 2023

by

Rongxiang Su

## Acknowledgements

When I began my PhD in the fall of 2019 at UCSB, I never expected to have the opportunity to work with two advisors. At this moment, I am deeply grateful for the privilege of working with Kostas and Somi, who are academic giants in transportation and GI-Science, respectively. While they always joked that I needed to write two dissertations since I have two advisors, I am well aware of the substantial support they have provided in shaping my dissertation and guiding every decision I've made. Their diverse perspectives continually prompt me to think how I can leverage the strengths of one domain to advance the other. Kostas always concludes our meetings with the words "have fun", a reminder not to overwork and to savor life, for which I am truly appreciative. He taught me not just to be a good researcher but a better person. Somi is my role model in every aspect — a superwoman capable of achieving anything. Despite her busy schedule, she always prioritized meetings with me. I am genuinely thankful for their extensive, insightful, and selfless guidance and recommendations for my research, doctoral milestones, and career plan. Special thanks also to my committee member Trisalyn, who provided very valuable suggestions during my qualifying exams and dissertation research.

My gratitude extends to my Geotrans and MOVE lab colleagues, including Elissa, Jingyi, Hui, Evgeny, Seonga, Zijian, Yifei, Danial, Mary, Rui, Ren, Mingshuang, and Leye. Working and having fun with all of you has been a wonderful experience. I'd also like to express my thanks to some important friends I've made here in Santa Barbara for their delightful companionship: Chongzheng, Marc, Huijie, Ling, Clare, Jiwon, Pinshu, Hao, Jiahua, Hsin-Ju, Ye, Ryoko, and Yanni. Studying abroad as an international student was not easy for me, but with all of you, it became a joyous journey. Last but not least, I want to acknowledge my partner Pu, my parents, my sister, and my brother. Completing my PhD would not have been possible without their unwavering support and love.

# Curriculum Vitæ

Rongxiang Su

## Education

| | |
|---|---|
| 2023 | Ph.D. in Geography, University of California, Santa Barbara |
| 2019 | M.S. in Cartography and Geographic Information System, Wuhan University, China |
| 2016 | B.S. in Geographic Information System, Hefei University of Technology, China |

## Publications

1. **Su, R.**, Goulias, K. (2023). Untangling the relationships among residential environment, destination choice, and daily walk accessibility. *Journal of Transport Geography*, 109, 103595. https://doi.org/10.1016/j.jtrangeo.2023.103595
2. **Su, R.**, Dodge, S., Goulias, K. (2022). A classification framework and computational methods for human interaction analysis using movement data. *Transactions in GIS*, 26(4), 1665-1682. https://doi.org/10.1111/tgis.12960
3. **Su, R.**, Dodge, S., Goulias, K. (2022). Understanding the impact of temporal scale on human movement analytics. *Journal of Geographical Systems*, 24(3), 353-388.**(JGS Editors' choice article)**https://doi.org/10.1007/s10109-021-00370-6
4. **Su, R.**, Dodge, S., Goulias, K. (2021). A time-geographic approach to quantify the duration of interaction in movement data. *In Proceedings of the 1st ACM SIGSPATIAL International Workshop on Animal Movement Ecology and Human Mobility* (pp. 18-26).**(Best Paper Award)** https://doi.org/10.1145/3486637.3489490
5. **Su, R.**, Xiao, J., McBride, E.C., Goulias, K. (2021). Understanding senior's daily mobility patterns in California using human mobility motifs. *Journal of Transport Geography*, 94, 103117. https://doi.org/10.1016/j.jtrangeo.2021.103117
6. **Su, R.**, McBride, E.C., Goulias, K. (2021). Unveiling daily activity pattern differences between telecommuters and commuters using human mobility motifs and sequence analysis. *Transportation Research Part A: Policy and Practice*, 147, 106-132. https://doi.org/10.1016/j.tra.2021.03.002
7. **Su, R.**, Goulias, K. (2021). Evolution of the Chinese Spring Festival Travel network during the COVID-19 early outbreak. *Transportation letters*, 13(5-6), 492-500. https://doi.org/10.1080/19427867.2021.1896065
8. **Su, R.**, McBride, E.C., Goulias, K. (2020). Pattern recognition of daily activity patterns using human mobility motifs and sequence analysis. *Transportation Research Part C: Emerging Technologies*, 120, 102796. https://doi.org/10.1016/j.trc.2020.102796

9. **Su, R.**, Fang, Z. (2019). A review of studies in taxi mobility and e-hailing taxi service. *Journal of Smart Cities*, 4(1), 1-6.
10. **Su, R.**, Fang, Z., Xu, H., Huang, L. (2018). Uncovering Spatial Inequality in Taxi Services in the Context of a Subsidy War among E-Hailing Apps. *ISPRS International Journal of Geo-Information*, 7(6), 230. https://doi.org/10.3390/ijgi7060230
11. **Su, R.**, Fang, Z., Luo, N., Zhu, J. (2018). Understanding the dynamics of the pick-up and drop-off locations of taxicabs in the context of a subsidy war among e-hailing apps. *Sustainability*, 10(4), 1256. https://doi.org/10.3390/su10041256
12. Shi, H., **Su, R.**, Goulias, K. (2023). Exploring the impact of COVID-19 pandemic on Americans time use related subjective well-being. *Wellbeing, Space and Society*, 100148. https://doi.org/10.1016/j.wss.2023.100148
13. Shi, H.,**Su, R.**, Xiao, J., Goulias, K. (2022). Spatiotemporal analysis of activity-travel fragmentation based on spatial clustering and sequence analysis. *Journal of Transport Geography*, 102, 103382. https://doi.org/10.1016/j.jtrangeo.2022.103382
14. Dodge, S., **Su, R.**, Johnson, J., Simcharoen, A., Goulias, K., Smith, J., Ahearn, S. (2021). ORTEGA: an object-oriented time-geographic analytical approach to trace space-time contact patterns in movement data. *Computers, Environment and Urban Systems*, 88, 101630. https://doi.org/10.1016/j.compenvurbsys.2021.101630
15. Goulias, K., **Su, R.**, McBride, E.C. (2020). Revisiting the Impact of Teleworking on Activity-Travel Behavior Using Recent Data and Sequence-Based Analytical Techniques. (Research report to Pacific Southwest Region 9 University Transportation Center)
16. Xiao, J., **Su, R.**, McBride, E.C., Goulias, K. (2020). Exploring the correlations between spatiotemporal daily activity-travel patterns and stated interest and perception of risk with self-driving cars. *AGILE: GIScience Series*, 1, 1-15. https://doi.org/10.5194/agile-giss-1-22-2020
17. Goulias, K., McBride, E.C., **Su, R.** (2020). Life cycle stages, daily contacts, and activity-travel time allocation for the benefit of self and others. In: *Scheiner, J. and Rau, H. (eds) Mobility Across the Life Course*, Publisher: Edward Elgar. (Book chapter)
18. Fang, Z., **Su, R.**, Huang, L. (2018). Understanding the effect of an E-hailing app subsidy war on taxicab operation zones. *Journal of Advanced Transportation*, 2018. https://doi.org/10.1155/2018/7687852

**Conference Presentations**

1. Spatiotemporal Dynamics of Racial-ethnic Diversity and Segregation: Insights from Mobile Phone Data. *ESRI Higher Education Webinar: Discover the Power of Spatial Data Science for Advancing Knowledge* (online). December 7, 2023. (Panelist)

2. Spatiotemporal Dynamics of Racial-ethnic Diversity and Segregation: Insights from Mobile Phone Data. *Spatial Hour at the Center for Spatial Studies and Data Science.* UCSB, California. November 29, 2023. (Oral presentation)
3. ORTEGA: An open-source Python package for context-aware interaction analysis based on movement data. *2023 GISS Specialty Group Student Paper Competition, American Association of Geographers Annual Meeting.* Denver, CO. March 24, 2023. (Full paper submission; Oral presentation; **Honorable Mention Award**)
4. Toward a taxonomy and computational methods for human interaction analysis using movement data. *2022 ESRI User Conference*. San Diego, California. July 12, 2022. (Oral presentation)
5. Heterogeneity in the relationship of vehicle miles traveled with the built environment: A quantile regression approach. *2022 UCGIS Symposium.* Syracuse, NY. June 7-9, 2022. (Extended abstract submission; Lightning talk, acceptance rate: 9/40)
6. Understanding the impact of temporal scale on human movement analytics. *2022 American Association of Geographers Annual Meeting* (online). February 26, 2022. (Oral presentation)
7. A time-geographic approach to quantify the duration of interaction in movement data. *1st ACM SIGSPATIAL International Workshop on Animal Movement Ecology and Human Mobility.* Beijing, China (online). November 2, 2021. (Full paper submission; Oral presentation; **Best Paper Award**)
8. Unveiling the taxonomy of daily travel and time use patterns using human mobility motifs and sequence analysis. *2021 American Association of Geographers Annual Meeting* (online). April 8, 2021. (Oral presentation)
9. An exploration of human mobility motifs in the California component of the 2017 National Household Travel survey. *hEART 2020: 9th Symposium of the European Association for Research in Transportation* (online). February 3-4, 2021. (Extended abstract submission; Recorded oral presentation)
10. Unveiling daily activity pattern differences between telecommuters and commuters using human mobility motifs and sequence analysis. *2021 Transportation Research Board Annual Meeting* (online). January 27, 2021. (Poster presentation)
11. Pattern recognition of daily activity patterns using human mobility motifs and sequence analysis. *2021 Transportation Research Board Annual Meeting* (online). January 26, 2021. (Poster presentation)
12. Unveiling daily activity pattern differences between telecommuters and commuters using human mobility motifs and sequence analysis. *Activity and Time-Use Patterns Subcommittee Meeting of 2021 Transportation Research Board Annual Meeting* (online). January 5, 2021. (Invited talk)

13. Accurate road anomaly detection by trajectory mining. *1st ACM SIGSPATIAL International Workshop on Ride-hailing Algorithms, Applications, and Systems*. Chicago, IL. November 5, 2019. (Full paper submission; Poster presentation)
14. Uncovering the changes of urban taxi travel demand in the context of a subsidy war among e-hailing apps. *China Geographic Information Science Theory and Methodology Annual Conference*. Taiyuan, China. November 2-4, 2018. (Full paper submission; Oral presentation; **Best Paper Award**)
15. Uncovering the changes of the pick-up and drop-off locations of taxicabs in the context of a subsidy war among e-hailing apps. *The 9th International Forum of Spatially Integrated Humanities and Social Sciences*. Shanghai, China. June 23-25, 2018. (Oral presentation)

**Teaching Experience**

| | |
|---|---|
| Winter 2021 | TA for GEOG 111B: Transportation Modeling and Simulation |
| Fall 2020 | TA for GEOG 111A: Transportation Planning and Modeling |
| Spring 2020 | TA for GEOG 111C: Smart Green Cities |
| Winter 2020 | TA for GEOG 111B: Transportation Modeling and Simulation |
| Fall 2019 | TA for GEOG 111A: Transportation Planning and Modeling |

**Services**

- Geography Events Committee (2020-2021)
- Geography Sport Committee (2019-2020)
- Volunteer for the Geography Peer Mentor Program (2022-2023)
- Volunteer for the department's open house for prospective graduate students (2023)
- Volunteer in graduate student panel for the department's "How-to-Apply" orientation for prospective graduate students (2021)
- Peer Reviewer for *Nature Human Behaviour*
- Peer Reviewer for International Journal of Geographical Information Science
- Peer Reviewer for *Transactions in GIS*
- Peer Reviewer for *Transportation Research Part A: Policy and Practice*
- Peer Reviewer for *Transportation Research Part C: Emerging Technologies*
- Peer Reviewer for *IEEE Transactions on Intelligent Transportation Systems*
- Peer Reviewer for *Journal of Spatial Information Science*
- Peer Reviewer for *Journal of Transport Geography*
- Peer Reviewer for *Travel Behaviour and Society*
- Peer Reviewer for *Transportation*
- Peer Reviewer for *GIScience&Remote Sensing*

- Peer Reviewer for *European Journal of Transport and Infrastructure Research*
- Peer Reviewer for *IEEE Access*
- Peer Reviewer for *Public Transport*
- Peer Reviewer for *Transportation Letters*
- Peer Reviewer for *Transportation Research Board Annual Meeting*
- Peer Reviewer for *Transportation Research Record*
- Peer Reviewer for *Computational Urban Science*
- Peer Reviewer for *Data Science for Transportation*

### Fellowships, Scholarships, and Awards

| | |
|---|---|
| 2023 | AAG-GISS Specialty Group Student Paper Competition Honorable Mention Award |
| 2022 | Jack and Laura Dangermond Fellowship |
| 2022 | UCSB Geography Department Excellence in Research Award |
| 2022 | Chinese Government Award for outstanding self-financed students abroad |
| 2022 | Chinese-American Engineers and Scientists Association of Southern California Scholarship |
| 2021 | UCSB Geography Department Summer Research Stipend Award |
| 2020 | UCSB Multidisciplinary Research on the Coronavirus and its Impacts Grant |
| 2019 | Outstanding Graduate Student Award, Wuhan University |
| 2018 | Chen-Yongling Scholarship, Wuhan University |
| 2018 | First-Class Scholarship, Wuhan University |
| 2016 | Distinguished College Graduate, Hefei University of Technology |
| 2014, 2015 | First-Class Scholarship, Hefei University of Technology |

### Travel Grants

| | |
|---|---|
| 2019-2023 | Jack and Laura Dangermond Travel Award (7 times) |
| 2022-2023 | UCSB Graduate Student Association Conference Travel Grant (twice) |
| 2023 | AAG Spatial Analytics and Modeling Speciality Group Student Travel Award |
| 2023 | AAG Transportation Geography Specialty Group Travel Award |
| 2022 | UCSB Doctoral Student Travel Grant |
| 2022 | UCGIS Symposium Travel Award |

**Abstract**

Sensing Human Activity and Interaction Patterns through Movement Observations

by

Rongxiang Su

Human beings have a fundamental need for social interaction according to Maslow's hierarchy of needs. Human interaction contributes to individual psychological development and well-being, and enhances social cohesion more broadly. Among others, two scientific fields that are interested in human interaction in space and time progressed in separate paths but emerged from similar inspirations and aim at similar end goals. The first is activity analysis with a focus on developing models of activity participation and travel behavior using travel surveys. These are commonly referred to as activity-based approaches in travel demand analysis. The second is the study of human movement in space and time using geometric relationships that evolve over time and space derived from the observations of moving individuals, a tradition followed mostly by geographic information scientists. Travel behavior researchers recognize human interaction as one of the most important determinants affecting people's daily time allocation and travel. However, inter-individual interactions, particularly inter-household ones, in travel behavior research have received less attention while these interactions contribute to a significant amount of travel. Traditional travel surveys are usually limited in sample size and observation period due to expensive costs. They also often lack sufficient information regarding inter-household interactions. Besides human interactions when individuals engage in activities at specific locations, it is also crucial to consider potential interactions in shared spaces, such as when their activity spaces or paths intersect during movement. In recent years, the increased availability of diverse and fine-grained movement data offers an un-

precedented opportunity to study human movement and interaction at an unparalleled level of fine granularity. However, there is a shortage of advanced computational methods to trace and understand human interactions present in fine-grained and large-volume movement data. It also remains unclear to what extent the knowledge gained from individual movement data can contribute to our understanding of collective behaviors across different socio-geographical contexts. The primary goal of this dissertation is to develop and evaluate new computational methods for identifying and quantifying spatiotemporal patterns of human daily activities and interactions, using both conventional travel diaries and massive movement data of individuals. By doing so, this dissertation explores important questions to understand heterogeneity in people's daily activity-travel patterns and interactions across different socio-geographical spaces.

# Contents

**Curriculum Vitae** **v**

**Abstract** **x**

**1 Introduction** **1**
1.1 Background and Motivation . . . . . 1
1.2 Research Questions . . . . . 6
1.3 Dissertation Synopsis . . . . . 9

**2 Unveiling Daily Activity Pattern Differences between Telecommuters and Commuters using Human Mobility Motifs and Sequence Analysis** **11**
2.1 Introduction . . . . . 13
2.2 Data . . . . . 17
2.3 Methodology: Motifs and Sequence Analysis . . . . . 22
2.4 Results . . . . . 31
2.5 Conclusions . . . . . 61

**3 A Time-geographic Approach for Human Interaction Analysis using Movement Data** **67**
3.1 Introduction . . . . . 69
3.2 Background . . . . . 71
3.3 A New Classification Framework for Human Interactions . . . . . 75
3.4 Methods: Identifying Interaction Patterns . . . . . 78
3.5 ORTEGA v1.0 Python Package . . . . . 84
3.6 Experiments . . . . . 93
3.7 Discussion . . . . . 109
3.8 Conclusions . . . . . 110

**4 Spatiotemporal Dynamics of Racial-ethnic Diversity and Segregation: Insights from Mobile Phone Data** **112**
4.1 Introduction . . . . . 113
4.2 Background . . . . . 116

4.3 Mobile Phone Location Data . . . . . 119
4.4 Methodology . . . . . 120
4.5 Results . . . . . 127
4.6 Discussion and Future Directions . . . . . 141

**5 Conclusions and Future Work** **144**
5.1 Research Contributions . . . . . 144
5.2 Discussions . . . . . 148
5.3 Limitations and Future Works . . . . . 153

**A** **158**
A.1 Determining the optimal number of clusters for motif groups . . . . . 158

**Bibliography** **161**

# Chapter 1

# Introduction

## 1.1 Background and Motivation

Human beings have a fundamental need for social interaction or social contact, according to Maslow's hierarchy of needs [1]. In practical terms, at the individual level, social contact significantly contributes to psychological development and subjective well-being [2]. At the community level, social interaction may strengthen social cohesion, which, in turn, leads to increased solidarity, trust, pro-social behavior, well-being, and reduced violence and crime in society [3, 4, 5, 6].

Among others, two scientific fields that are interested in human interaction in space and time progressed in separate paths but emerged from similar inspirations and aimed at similar end goals. The first is activity analysis with a focus on developing models of activity participation and travel behavior using travel surveys. Over time, considerations about the daily social interactions of people, time allocation to different activities, physical separation of places, and the roles people play as they progress through their life cycle stages led researchers to develop more complex models for planning and operations that are known as activity-based approaches for travel demand analysis [7, 8, 9]. The second

is the study of human movement in space and time using geometric relationships that evolve over time and space derived from the observations of moving individuals, which is a tradition that is followed mostly by geographic information scientists [10, 11, 12, 13]. With the increasing availability of fine-grained human movement data, the primary concern of this stream of research is to develop new computational approaches and modeling techniques for analyzing and understanding human interaction at more granular levels.

The conceptual framework for analyzing human interactions is motivated by Hägerstrand's time geography and its derivatives [14, 15, 16]. Hägerstrand was opposed to aggregate analyses of people's behavior and advocated microanalysis of human behavior. He defined many of the fundamental elements in time geography, such as space-time path tracing the movement of an individual in space and time, space-time prism for measuring individual activity space, space–time bundle (convergence of two or more paths for a shared activity), three types of constraints people face in their everyday life (i.e., capability, coupling, and authority constraints), activity motivated by basic human needs, and an evolutionary approach to all this [14]. Of these concepts, the space-time bundle is the most directly relevant for the study of human interaction. In time geography, a bundle refers to the grouping of activities or events (i.e., the convergence of two or more paths for some shared activity) within a defined time period or temporal window [14, 17]. A bundle can occur when individuals are stationary at the same location (i.e., stationary bundle) or during movement in a shared space (i.e., mobile bundle). The idea of space-time bundle lays the foundation for understanding complex spatiotemporal patterns of human activities and interactions [17, 18]. In this dissertation, I use the term *interaction* to explicitly discuss spatiotemporal co-location behaviors of different individuals at the same activity location or during movement in a shared space.

Travel behavior researchers recognize human interaction as one of the most important determinants affecting people's daily time allocation and travel [19, 9, 20, 7]. The

majority of daily human interactions happen within households (i.e., intra-household interactions). This is not only because household members live together but also because they share household responsibilities. Consequently, intra-household interactions are frequently incorporated into activity-travel analyses, and numerous researchers have highlighted their substantial roles in elucidating people's daily life [21, 22, 23, 24, 25]. Exploration of this topic has roots in Samuelson's work [26] in which he formulated household utility as a function of the utility of each individual household member. This approach was further developed by Townsend [27] and Koppelman [28], who established a theoretical framework for a household time allocation model, considering various types of interdependencies within a household. According to [29], intra-household relationships can be categorized into three types: substitution, complementarity, and companionship. In essence, researchers are interested in understanding the group decision-making process regarding how people allocate time to themselves and others in activity and travel (e.g., parents picking up the children from school, joint activities between household members such as having dinner together or engaging in the same social event) [21, 22, 27, 28, 29, 30, 31, 32, 33].

Nevertheless, inter-individual interactions, particularly inter-household ones in travel behavior research have received less attention while these interactions contribute to substantial daily travel and significantly impact the duration and start time of activities [33]. Inter-household interactions may not necessarily refer to social contact between familiar individuals (e.g., friends); they also include anonymized and casual encounters between strangers or acquaintances. This phenomenon is also referred to as the "familiar strangers" concept proposed by Milgram [34]. While such contact may seem insignificant, it serves as a precondition for more complex and meaningful social interactions that shape a healthy and lively environment [35, 36]. For instance, sharing rides with "familiar strangers" provides a sense of security without introducing additional obliga-

tions to engage in interpersonal interactions [37]. Familiar strangers have the potential to enhance more sophisticated social interactions, contributing to the formation of social identity and social networks among individuals [38, 39]. Traditional data sources supporting interaction analysis primarily consist of individual records of daily travel (e.g., travel diaries) or sometimes, in conjunction with GPS devices (e.g., GPS-enhanced travel surveys). Travel surveys usually collect people's individual and household level sociodemographic characteristics as well as detailed information about daily travel or activities such as each trip's departure and arrival times, trip purpose or activity type, activity location, and mode choice. In addition, GPS technology has been integrated into the travel survey collection process for various purposes, such as detecting travel modes [40, 41], identifying trip end or activity locations [42, 43], assessing transportation network conditions [44], and analyzing route choices [45, 46]. However, these data are usually limited in sample size and observation period due to expensive data collection costs. They also often lack sufficient information regarding inter-household interactions.

It is also crucial to take into account human interactions not only when individuals engage in activities together at specific locations but also potential interactions in shared spaces, such as when their activity spaces or paths intersect during movement [47, 48, 11, 10, 49, 50]. Geographic information scientists, also inspired by Hägerstrand's time geography framework [14], pursue a distinct research direction with objectives akin to those of travel behavior researchers. Their focus lies in unraveling the intricacies of human movement and interaction patterns by using the geometric relationships of their movement paths that evolve over time and space [51, 52, 47, 48, 11, 10, 49, 50]. Many of the existing tools for interaction analysis rely on the spatial proximity between two individuals and require predefined spatial and temporal thresholds [10, 53]. These proximity-based approaches can be limited when high-resolution data are not available or when moving individuals are not tracked simultaneously. To address this issue, recent

developments leveraged time geography [14] to enhance interaction analytics [50, 49, 48]. Specifically, these approaches model the accessible area of individuals given each pair of successive GPS fixes along their movement paths using the concept of *Potential Path Area (PPA)* [17, 16, 15]. Subsequently, the intersection of two PPAs of different individuals identifies the joint-accessibility space where potential interaction between individuals is possible. Existing efforts indicate that the PPA-based approaches can well identify potential interactions when individuals are not tracked simultaneously [50]. Moreover, the PPA-based approaches are capable of identifying delayed interactions when individuals visit the same locations asynchronously with a time lag [50, 49]. However, existing approaches fall short in their capacity to quantify the duration of dynamic interactions or frequency of contact between moving individuals, which are critical factors for differentiating different types of human interactions. Furthermore, a comprehensive taxonomy of the definition and classification of human movement interaction is also lacking, which is fundamental for the formalization of interaction patterns and their properties to inform subsequent more sophisticated analysis.

In recent years, the surge in location-aware technologies (LATs) has propelled the study of human activity and interaction patterns to unprecedented spatiotemporal granularities. This progress is attributed to the increased availability of multi-faceted and multi-sourced movement data [12], collected in larger volumes and varying formats across multiple temporal granularities in the so-called Big Data era [54]. Such data are sourced from various devices, including GPS data recorders, cell phone traces, Call Detail Records (CDR) from smartphones, Radio Frequency Identification (RFID) tags, Wi-Fi and Bluetooth sensors, as well as georeferenced social media applications [55]. Among them, the cellular networks operated by mobile phone providers function as ubiquitous sensors, offering an extensive amount of information about the movement patterns of nearly the entire population [56]. This offers an unprecedented opportunity to study human move-

ment and interaction at an unparalleled level of fine spatiotemporal resolution and a large geographic scale [57].

However, there is a shortage of advanced computational methods to trace and understand human interactions present in fine-grained and large-volume data sets of human movement. It also remains unclear to what extent the knowledge gained from the analysis of individual movement data can contribute to our understanding of collective behaviors across different socio-geographical contexts, such as social mixing and segregation among diverse social groups within metropolitan areas. The primary goal of this dissertation is to develop and evaluate new computational methods for identifying and quantifying spatiotemporal patterns of human daily activities and interactions, using both conventional travel diaries and massive movement data of individuals. By doing so, this dissertation explores important questions to understand heterogeneity in people's daily activity-travel patterns and interactions across different socio-geographical spaces.

## 1.2 Research Questions

My dissertation attempts to bridge the research gaps discussed above by addressing three individual yet interconnected key research questions.

**Research Question 1***: Are there significant differences in daily activity-travel patterns between different occupational groups, such as telecommuters and commuters?*

Unlike the conventional definition of telecommuting in the late 1970s, telecommuting nowadays is no longer a home-based work arrangement, but is multiple-place-based. With the advancement of information and telecommunication technologies (ICTs), the emergence of ubiquitous workplaces allows telecommuters to work not only at home but at many places with available internet connection such as a coffee shop, a bookstore, a restaurant, and so forth. This provides opportunities for human interactions between

physical and virtual spaces. However, it remains unclear if significant differences exist between telecommuters and customary commuters in their daily mobility and time allocation patterns. To answer RQ 1, I propose a new joint pattern recognition method that combines network motif-based analysis with an activity sequence alignment method in Chapter 2. The proposed method can identify individuals' variety-seeking behaviors in daily destination choice and distinct time allocation patterns for activity participation. The new method will be evaluated through a case study using the California component of the 2017 National Household Travel Survey, which consists of 4,168 telecommuters and 20,280 commuters. An empirical analysis focusing on time-of-day variation in activity participation and travel between telecommuters and commuters will be discussed.

**Research Question 2***: What can long-term movement data reveal about dynamic interactions between individuals?*

Besides human interactions when individuals engage in activities together at specific locations, it is also crucial to consider potential interactions in shared spaces, such as overlapped movement paths and activity spaces. While the majority of human interactions occur within households, inter-household interactions also contribute to a significant amount of travel. However, inter-household interactions have received less attention, primarily due to limited access to data capturing inter-household interactions. To understand different types of interactions between moving individuals such as intra-household and inter-household interactions, it is imperative to identify and quantify the spatiotemporal properties of these interactions. To achieve this goal, in Chapter 3, I utilize the time-geographic framework to enhance existing approaches for spatiotemporal analytics of human movement interactions. The proposed method can identify potential human interactions and quantify the duration or frequency of contact between moving individuals. In contrast to the investigation of RQ 1, which relies solely on conventional

travel diaries, I evaluate the proposed method leveraging fine-granularity, continuous, and long-term movement tracking data obtained from 986 participants using wearable GPS devices. This research further incorporates the travel diaries of the same 986 participants to validate the identified intra-household interactions from movement data. This multi-faceted approach ensures a comprehensive understanding of the spatiotemporal dynamics of intra-household and inter-household interactions in movement data.

**Research Question 3***: To what degree can the analysis of individual movement data contribute to our understanding of collective behaviors between communities, such as social mixing and segregation between diverse groups?*

Chapter 2 investigates RQ1 by relying on a limited-size self-reported travel diary collected across the entire state of California, which provides information solely on the origin and destination of each trip but lacks details about the movement paths between these locations. In contrast, to answer RQ2, Chapter 3 utilizes continuous and long-term movement data collected using GPS devices from 986 individuals to identify intra-household and inter-household interactions during movement and uses travel diaries of the same participants to validate intra-household interactions. However, it remains unclear to what extent the analysis of individual movement data can inform our understanding of human collective behaviors between communities, such as social mixing and segregation among diverse social groups. To answer this question, in Chapter 4, I develop a dynamic measure of the diversity of human interactions among different social groups at the finest spatial scale, specifically at the Points of Interest (POI) level and various temporal contexts. Leveraging one month of privacy-enhanced mobile phone location data of 283,556 anonymized individuals collected in Los Angeles County along with the POI data, a case study explores the racial diversity and segregation simultaneously in people's daily interactions at various types of places.

## 1.3 Dissertation Synopsis

This dissertation is organized around the integration of four individual yet interconnected articles, presented in Chapters 2, 3, and 4, respectively.

Chapter 2, titled "Unveiling Daily Activity Pattern Differences between Telecommuters and Commuters using Human Mobility Motifs and Sequence Analysis", is a revised version based on the published journal article below.

- Su, R., McBride, E.C., Goulias, K. (2021). Unveiling daily activity pattern differences between telecommuters and commuters using human mobility motifs and sequence analysis. *Transportation Research Part A: Policy and Practice*, 147, 106-132. https://doi.org/10.1016/j.tra.2021.03.002

Chapter 3, titled "A Time-geographic Approach for Human Interaction Analysis using Movement Data", is a synthesis of two published journal articles.

- Su, R., Dodge, S., Goulias, K. (2022). A classification framework and computational methods for human interaction analysis using movement data. *Transactions in GIS*, 26(4), 1665-1682. https://doi.org/10.1111/tgis.12960
- Su, R., Dodge, S., Goulias, K. (2022). Understanding the impact of temporal scale on human movement analytics. *Journal of Geographical Systems*, 24(3), 353-388.https://doi.org/10.1007/s10109-021-00370-6

The content of Chapter 4, titled "Spatiotemporal Dynamics of Racial-ethnic Diversity and Segregation: Insights from Mobile Phone Data", is derived from a journal article currently under review.

- Su, R., Dodge, S. (2023). Spatiotemporal dynamics of racial-ethnic diversity and segregation: Insights from mobile phone data. (in review)

Finally, Chapter 5 provides a summary of significant contributions made and insights into the three research questions. The chapter also delves into the limitations of this dissertation and suggests potential avenues for future work.

# Chapter 2

# Unveiling Daily Activity Pattern Differences between Telecommuters and Commuters using Human Mobility Motifs and Sequence Analysis

| Peer-reviewed Publication | |
|---|---|
| Title | Unveiling daily activity pattern differences between telecommuters and commuters using human mobility motifs and sequence analysis |
| Authors | Rongxiang Su, Elizabeth McBride, Konstadinos Goulias |
| Venue | Transportation Research Part A: Policy and Practice |
| Editors | Elisabetta Cherchi |
| Publisher | Elsevier |
| Pages | Volume 147, 106-132 |
| Submitted Date | 6 August 2020 |
| Accepted Date | 2 March 2021 |
| Publication Date | 16 March 2021 |
| Copyright | Reprinted with permission from Elsevier |
| DOI | https://doi.org/10.1016/j.tra.2021.03.002 |


**Abstract** This chapter demonstrates the use of motif and sequence analysis in tandem to analyze differences and commonalities between telecommuters and usual commuters. In terms of substantive findings, telecommuters are by far more diverse in their allocation of time to places, activities, and travel. Approximately 20% of telecommuters stay at home all day during a workday, while only 8% of commuters do. Telecommuters that have at least one trip during their workday accrue more vehicle miles traveled and number of trips than their commuter counterparts. However, they travel less driving alone and tend to have more complex schedules visiting more locations. Within telecommuters and commuters, however, we have substantial variation in activity participation and travel captured by the combination of motifs and sequence analysis. As expected, a substantial proportion of commuters display morning and afternoon peaks of arriving at and departing from work, and telecommuters do not show this pattern. In addition, telecommuters do not only perform work tasks from home. Instead, during the day a high percentage travel to a variety of locations to either visit customers and/or use their spatiotemporal schedule flexibility to perform work tasks from locations other than home. In contrast, more than 80% of commuters perform work at their workplace. In addition,

a slightly higher proportion of telecommuters function as designated drivers escorting other people to their activity locations.

## 2.1 Introduction

Telecommuting and telework is the use of information and telecommunication technology (ICT) to replace the more traditional working at workplaces and traveling to work. In the 1970s telecommuting was envisioned as a policy tool in a Travel Demand Management (TDM) toolkit to help decrease congestion, air pollution, and waste of resources. Legislation and planning at many levels of government support telecommuting as a measure to benefit the public, employers, and employees. Paradoxical findings in the literature of telework is one motivation of the research reported here [58, 59, 60]. These paradoxes are also amplified by telework market penetration that is small considering the forces involved. On the one hand, we see a push for more teleworking in public agencies (see [61]), and the Telework Enhancement Act of 2010 [62], evidence of substantial benefits for agencies and employees with estimates reaching $11,000 per person year for a business, $2,000 to $7,000 savings per year per employee, substantial greenhouse gas emission reduction and energy consumption [63] and $700 Billion a year national savings if eligible workers would work half the time from home [64, 61]. On the other hand, only 7% of the US-based firms make telework available, the Bureau of Labor Statistics reports a decline in working at home from the past [65], and major corporations as reported by news media are cutting down on telecommuting programs. However, due to the recent pandemic telework might become a more popular option with employees. According to Global Workplace Analytics this can reach 25–30% of the workforce in 2021[1] providing a second important motivation of our research here.

[1] `https://globalworkplaceanalytics.com/work-at-home-after-covid-19-our-forecast`

Often these paradoxes are not paradoxes at all. They can be explained by carefully studying the context, data definitions, and related analysis [66]. The context side of this story needs careful scrutiny because we need to discern differences and commonalities among settings of where telework is adopted [67], demographics [68], technologies used [69, 70, 71], supervisory roles [72], and perspectives of adopter vs non-adopter and manager vs employee [73]. Similarly, context can also be extracted from databases with documented data collection settings, data collection processes, and wording of questions (the Bureau of Labor Statistics is asking if a person works full-time or part-time at home and the US Census American Community Survey is asking persons 16 and over if they work from home instead of traveling to work in the week before the Census day). All these are important details that first need to be harmonized in terms of basic definitions of work, telework, and telecommuting and then contextual analysis can reveal the underlying correlations of the propensity to stay at home and work instead of traveling to work. We will return to this point in the conclusions.

Research on telecommuting in California has a long history starting as early as the 1970s [74], tested in different places [75], and experienced a major push by the State of California Telecommuting Pilot Project and other planning initiatives [76, 77]. Another push forward was the 1999 National Telecommuting and Air Quality Act's establishment of pilot programs in five large Metropolitan Statistical Areas (MSAs), of which one was Los Angeles. This was an impetus for developing a market-based emission-credit-trading program encouraging telecommuting. However, it did not create a mechanism to track telecommuting impacts over time. Early research studies point out the need for careful telecommuting ontological analysis [78], the possibility that telecommuting may not be as desirable to a worker as planners think [79], and telecommuting may not have a positive impact on non-work travel [80]. On the more positive side, telecommuting is associated with positive perceived autonomy, job satisfaction, turnover intent, and posi-

tive impacts on work-family conflict, but may lead to negative impacts on relationships at work [81]. For transportation policy and planning, telecommuting appeared as an important component of Transportation Demand Management (TDM) as a promising option to decrease trip generation, vehicle miles traveled, and peak-period trips, and possibly gains in energy consumption [82, 83]. This is challenged by research that accounts for energy use at work and at home when telecommuting participation is low in frequency [84]. More recent data analyses also challenge the claim that telecommuting leads to travel decrease. In fact, Zhu et al. [85], in their comparison of the impact of telecommuting among different Metropolitan Statistical Areas (MSA) using data from the National Household Travel Survey of 2001 and 2009 conclude that policies that encourage telecommuting indeed increase travel demand rather than decrease. All these studies offer evidence that the travel behavior impact of telecommuting is evolving and in more recent years we see that telecommuting today is different than the telecommuting we knew in the late 1980s and early 1990s. However, all this analysis about telework and mobility is lacking in fundamental aspects of behavioral measurement.

First, the effects of individual's situational constraints on telecommuting are not accounted for in the literature. Internal personal barriers and motivation to adopt telecommuting are analyzed only tangentially through correlations with person and household characteristics and analyses accounting for endogeneity [86]. Instead, context needs to also be explored based on commitments people have (e.g., the need to escort children to school). To understand this type of context, we need to statistically examine daily patterns at fine-grained time intervals and use a classification of activities that emphasize the need to escort people to places. Second, studying aggregates such as daily averages may mask a variety of relationships. In fact, we need to be cautious with past analyses because all these studies used average daily summaries such as the number of trips to work per day, vehicle miles traveled per day, and so forth. Behind averages, we find

hidden variability in behavior (heterogeneity) that should have been explored in more detail.

In this chapter, we demonstrate a new analytical method that captures heterogeneity within and between telecommuters and customary commuters. To do this we use a fine-grained time-of-day allocation of time to activities and travel and patterns of destinations and trips in between them. We develop a joint pattern recognition method that combines network-based motif analysis with activity-based sequence analysis on telecommuters' and commuters' activities and travel. The combination of building a taxonomy of daily mobility patterns using the concept of motif with the taxonomy of minute-by-minute daily time allocation patterns derived from sequence analysis enables the study of interdependency in space and time of multiple facets of behavior that are fundamental in building activity-based models without imposing a priori behavioral assumptions. The concept of human mobility motif introduced by [87] refers to a directed network of nodes and links of daily patterns among destinations. Schneider et al. [87] identified 17 unique motifs that capture approximately 90% of daily mobility patterns of the total population in travel survey data and GPS traces data. Recently, another application by [88] showed that 10 unique location-based motifs and 10 unique activity-based motifs can represent 99.35% and 98.46% of human mobility patterns of the total population respectively using mobile phone positioning data collected in Shenzhen, China. The advantages of the motif approach in the studies of human mobility pattern analysis are twofold. First, the motif approach takes into account the interconnection of visited locations at the individual level compared to traditional trip-based models (referred to as "4-step" models, see explanation in [89]) that only use the individual person trip as the basic unit of analysis. Second, the motif model can be used directly in simplifying modeling and simulation of daily activity scheduling by decreasing the number of possible combinations of the many variables used in activity microsimulation. The analysis is followed by a comparison of

telecommuters' and commuters' daily patterns to uncover commonalities and differences across these two groups but also different daily patterns within each group.

## 2.2 Data

The data used in this study comes from the California component of the 2017 National Household Travel Survey (California-NHTS). NHTS collects social and demographic data about each participating household and every individual in the household. For each person, NHTS also collects a one-day travel diary. The data were collected between April 24th, 2016 and April 24th, 2017 with the one-day travel diary on an assigned day. The assigned day attempts to provide a uniform assignment throughout a complete year and the diary day for each household can be any weekday, weekend day, or holiday. For this analysis, we use person and household characteristics such as age, sex, educational background, job category, household income, full-time or part-time working status, retirement status, and homeworker status. From the one-day diary, we use the trip start time and end time, origin and destination location types, and mode choice. The one-day travel diary contains all reported trips between 4:00 AM on the survey day and 3:59 AM on the following day.

In total, in the NHTS California component, we have 24,448 part-time or full-time workers (4,168 telecommuters plus 20,280 commuters). In this study, we only focus on records of workers during non-holiday weekdays (we will use the term "workdays" throughout this chapter), since human mobility patterns have distinct differences between workdays and non-workdays [90]. An immediate next step is to compare the mobility patterns of telecommuters and commuters between workdays and non-workdays. We will get back to this point in the conclusion section. As we aim to explore the difference in mobility patterns between telecommuters and commuters, the records of 15,045

respondents (2,236 telecommuters plus 12,809 commuters) who made at least one trip on a workday are used in this analysis. Telecommuters are the persons in NHTS that answered the question "Do you usually work from home?" The possible responses are "Yes", "No", "Refuse to answer", or "Not applicable". Among the selected sample, we find that 15.8% of commuters and 26.43% of telecommuters spend no time at work on non-holiday weekdays. This observation indicates that given they are employed, many of them shifted their work to another day or just happened to take a day off work.

In this study, we explore the patterns of activity and travel during what is considered a typical weekday in regional simulation models and this includes people who perform work tasks but also people who take a day off for personal vacation, illness, or other reasons. As we will see in the analysis these workers are active outside their home even on days during which they do not perform any work tasks. The excluded 9,403 workers are people with diary days on non-workdays, or staying at the same location (mostly home) the entire diary day during workdays and have no trips to report (i.e. missing data in the one-day travel diaries). Of the 9,403 excluded workers, 1,932 are telecommuters and 7,471 are commuters. Among the excluded telecommuters, 624 people reporting on a workday did not make any trips. This means that among the 2,860 telecommuters interviewed on a workday, 21.82% stay at home all day. In contrast, among the 13,918 commuters that were interviewed on a workday, 1,109 (7.97%) stayed at home all day and did not make any trips on workdays. Presumably due to personal vacation, sick leave, or to care for another person, etc. However, we do not have the information needed to verify this. Exclusion from the analysis of telecommuters and commuters staying at home the entire diary day is explained later and complemented by a comparison between the excluded and the included samples in Section 2.4.4.

Table 2.1 shows the descriptive statistics for the 15,045 persons who made at least one trip on the interview workday. Male respondents account for 49.11% of the sample

Table 2.1: Descriptive statistics for telecommuters and commuters.

| Variable | Definition | Subgroup | Telecommuter (n=2,236) | Commuter (n=12,809) |
|---|---|---|---|---|
| *Individual characteristics* | | | | |
| Sex | Respondent's sex: female or male | Female | 50.81% | 48.58% |
| Age group | Respondent's age group | Under 26 | 2.86% | 8.85% |
| | | 26-35 | 11.58% | 19.34% |
| | | 36-50 | 25.58% | 30.26% |
| | | 51-65 | 40.38% | 34.12 |
| | | Above 65 | 19.54% | 7.29% |
| Education attainment | Respondent's education attainment | Below bachelor's degree | 9.79% | 15.01% |
| | | Some college or associate's degree | 25.00% | 29.84% |
| | | Bachelor's degree or above | 65.16% | 55.11% |
| Full-time or part-time | Full-time or part-time worker | Full-time | 56.17% | 80.12% |
| Job category | Respondent's job category | Sales or service | 27.55% | 21.21% |
| | | Clerical or administrative support | 6.40% | 11.61% |
| | | Manufacturing, construction, maintenance, farming | 8.59% | 10.84% |
| | | Professional, managerial, or technical | 57.07% | 56.13% |
| *Household-level characteristics* | | | | |
| Household annual income | Respondent's household annual income level | Less than $24,999 | 9.97% | 7.43% |
| | | $25,000 to $99,999 | 39.89% | 44.90% |
| | | $100,000 to $199,999 | 31.62% | 33.90% |
| | | $200,000 or more | 16.14% | 11.94% |
| Household structure | Respondent's household structure | One adult | 15.16% | 14.75% |
| | | 2+ adults | 35.51% | 37.33% |
| | | One adult, retired | 2.55% | 0.49% |
| | | 2+ adults, retired | 17.71% | 11.02% |
| | | One adult living with people younger than 21 | 2.95% | 3.04% |
| | | 2+ adults living with people younger than 21 | 26.12% | 33.37% |
| Residential setting | Household residential location types: urban or rural | Urban | 87.79% | 91.01% |

Notes: The percentage of the alternative option for some binary variables (e.g., sex, full-time or part-time worker, residential setting) are omitted. The total percentage of some variables is not exactly 100% because very few of the respondents refused to answer the survey questions (mostly below 1%).

of telecommuters. This is slightly lower than the male percentage of commuters which is 51.36%. When we look at the age groups, it turns out that there is a lower percentage of people under 50 belonging to telecommuters and a higher percentage of people over 50 belonging to telecommuters. Perhaps more senior workers achieve the freedom to choose to work from home and some of them might be part-time home workers. In terms of employment, 56.17% of telecommuters are full-time workers, while 80.12% of commuters are full-time workers. 27.55% of telecommuters work in sales or service industries, which is higher than the 21.21% for commuters. The percentage of clerical or administrative support workers among telecommuters is 5% less than that of commuters. This is not surprising because these occupations usually require face-to-face communication. In terms of household income level, the percentage of telecommuters with household income less than $24,999 is 2% more than the one of commuters. Presumably, as many telecommuters are part-time workers, they earn less money. However, the percentage of telecommuters with household income of more than $200,000 is 5% greater than that of commuters, indicating that many telecommuters are high-income individuals. In addition, more than 35% of the telecommuters and commuters are from households with two adults, 17.71% of telecommuters are from households with two retirees, while the corresponding percentage for commuters is only 11.02%, and 33.37% of commuters are living with another adult and people younger than 21 while only 26.12% of telecommuters do. The percentages of people living in urban or rural environments show that apparently, a higher percentage of telecommuters are living in rural areas compared to commuters, which aligns with the generally lower accessibility in rural areas. Presumably, people tend to work from home more to save commuting time. The differences in personal characteristics between telecommuters and commuters are consistent with other past analyses using data from NHTS [91, 92].

According to the 2017 NHTS-California codebook, people's travel mode choices are

classified into seven categories including walking, biking, transit, car as passengers, car drive alone, car drive someone else, and one category labeled "other" comprising all other modes. Table 2.2 summarizes the mean and standard deviation of vehicle miles traveled (VMT), person miles traveled (PMT) as well as the number of trips by different modes. We apply the Mann-Whitney U test to examine whether the average VMT, PMT, and number of trips by different modes of telecommuters are significantly different from commuters. The results indicate that there is no significant difference in PMT for biking and the number of bicycle trips, while the other types of VMT, PMT, and the number of trips of telecommuters are significantly different from commuters at the 99.9% confidence level. It is also noteworthy that for telecommuters, VMT total, VMT driving someone else, VMT as passengers, and PMT as walking are significantly higher than the corresponding quantity of commuters, while VMT driving alone, and PMT using transit of telecommuters are significantly lower than commuters.

This observation is also consistent with the number of trips by different modes. On the whole, Table 2.2 tells us that on average, telecommuters make 1.37 more VMT in a day compared to commuters as well as 0.53 more trips in a day. Specifically, telecommuters drive alone 3.76 miles fewer, but drive others 2.71 miles more than commuters, and they carpool 2.42 miles more than commuters daily. In addition, telecommuters walk 0.16 miles more in a day. These results defy the expectations that telecommuters travel less on workdays. When we examine people who made at least one trip in a day, telecommuters travel more than commuters on workdays. In contrast, if we include in these statistics the persons who did not make any trips on workdays (624 telecommuters and 1,109 commuters), the derived daily total VMT of telecommuters becomes 29.16 miles per day compared to 33.06 miles per day for commuters. It is the inclusion in statistics of people who do not travel at all that show on average less travel by telecommuters than commuters. In the subsequent analysis of this study, we first explore further the patterns

Table 2.2: Statistics for VMT, PMT and number of trips.

| Variables | Telecommuter (n=2,236) Mean | SD | Commuter (n=12,809) Mean | SD | Difference |
|---|---|---|---|---|---|
| *VMT/PMT statistics* | | | | | |
| VMT total | 37.29 | 59.3 | 35.92 | 47.14 | 1.37*** |
| VMT drive alone | 20.3 | 40.64 | 24.06 | 34.54 | -3.76*** |
| VMT drive someone else | 10.79 | 34.96 | 8.08 | 28.93 | 2.71*** |
| VMT as passengers | 6.21 | 33.04 | 3.79 | 22.81 | 2.42*** |
| PMT as biking | 0.17 | 1.29 | 0.18 | 1.54 | -0.01 |
| PMT as walking | 0.42 | 1.07 | 0.26 | 0.86 | 0.16*** |
| PMT as transit | 1.56 | 17.65 | 1.71 | 10.24 | -0.15*** |
| *Number of trips statistics* | | | | | |
| Total number of trips | 4.77 | 2.74 | 4.23 | 2.43 | 0.54*** |
| Drive alone trips | 2.4 | 2.3 | 2.45 | 2.03 | -0.05 |
| Drive someone else trips | 1.15 | 1.95 | 0.85 | 1.61 | 0.3*** |
| Passenger trips | 0.43 | 1.2 | 0.32 | 0.99 | 0.11*** |
| Bike trips | 0.07 | 0.45 | 0.06 | 0.43 | 0.01 |
| Walk trips | 0.65 | 1.37 | 0.43 | 1.08 | 0.22*** |
| Transit trips | 0.06 | 0.33 | 0.1 | 0.46 | -0.04*** |

Note: *$p$ <0.1; **$p$ <0.05; ***$p$ <0.01.

of telecommuters and commuters who made at least one trip in a day and then turn to the no-travel portion of the analysis.

## 2.3 Methodology: Motifs and Sequence Analysis

The method used here combines motif analysis that identifies patterns of locations visited in a day with one-way or two-way trips in between these distinct locations with sequence analysis of activities and trips applied to minute-by-minute diary records.

### 2.3.1 Construction of motifs

A motif is a directed network, in which nodes represent visited locations and directed edges (links) represent trips between locations [87]. The 2017 California-NHTS data keeps a one-day travel diary recording trips for everyone on the day of the interview. The one-day travel diary contains an anonymous person identifier and household identifier, trip start time and end time, origin and destination unique identifier, and corresponding location types. To construct motif representations, we use the origin and destination locations as nodes and connect nodes with a directed edge if there is a trip between them. An example is shown in Figure 2.1. The person has two commute trips but also visits another destination labeled "Other". Therefore, this person has a motif with three nodes representing home, work, and other location and four directed edges between home and workplace, and workplace and other location (the trips). A human mobility motif was constructed in this way for every respondent in our data.

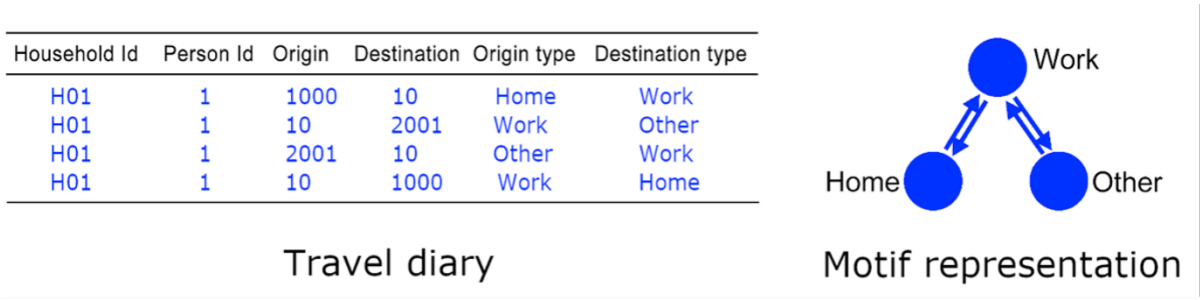



Figure 2.1: Example of constructing a motif from a trip diary.

According to Schneider et al. [87], for a given network size $N$, a possible number of edge combinations $N_p$ can be computed by the following equation.

$$N_p = 2^{N^2-N} \tag{2.1}$$

These are the counts of the feasible number of daily motif networks. Schneider et al.

[87] used two constraints to derive a lower number of feasible patterns with the reasons that people need to sleep in a 24-hour period and trips should not have unreasonable patterns. The need for sleep requires that people come back to their first origin location, mostly home, after one day's activity. The consistency of trips ensures that we do not have trips ending in a node and never coming out of the node. For example, in Figure 2.1 we cannot have a trip going to other from work and a trip going to other from home and no trips leaving from other. In data using diaries and in the California-NHTS, the origin location is not always home and some persons do not return home during the observation period. This means the Schneider et al. [87] constraints when imposed on the data, will eliminate legitimate daily patterns that contain important information about daily behavior, but we retain these patterns in the analysis here. Also, motifs with just one node can be the null motif (stay at a location all day) and a motif that is the base for one or more trips that start at a location and end at the same location (e.g., walking the dog, going out for jogging, strolling an infant). A trip like this is recorded in NHTS as a loop trip, and we include this pattern in the analysis here.

For motifs that have more than one node, each of the nodes can also include a loop trip. Distinguishing motifs with the same nodes but different patterns of loop trips depends on the content in the data. Recent diary survey data have a record of loop trips, but mobile phone data use a range of points and time intervals that extract stay points (activity locations) by setting spatial and temporal constraints. One example is by Cao et al. [88] who clustered as stay points several consecutive GPS points that are within 500 m range and persist for longer than 60 min. This type of point classification would miss a home-based jog. In California-NHTS workday data, 2.42% of trips are loop trips and 84.27% of the loop trips are home-based trips. Descriptive statistics show that the mean duration of home-based loop trips is 39.37 min and 88.27% of home-based loop trips are less than 60 min. The majority of people spent less than an hour on

their home-based loop trips, with many of these trips for exercise. In motif classification patterns that have the same number of nodes, patterns with and without a loop trip are classified in the same group. The loop trip information, however, is retained and used in the daily time allocation pattern investigation as explained later. From a pattern recognition standpoint, the primitives (elemental pieces of information used in pattern classification) of motifs are the stops (locations) and the movement between two locations (the trip).

### 2.3.2 Sequence analysis to identify activity-travel patterns

#### Sequence of individual daily activities and travel

We first introduce the concept of sequence. A sequence is a series of time points at which a subject can move from one discrete "state" to another. The states are defined by the types of places people visit and stay during their diary day and according to the 2017 California-NHTS, they are Home with unspecified activity (*Home*); Work from home as in telecommuting and home stay combined (*Home&Work*); Work at a workplace or other places (*Work*); Education at the school location (*School*); Drop off or Pick up someone (*DropPickup*); Change type of transportation (*ChangeTrans*); Purchase goods such as groceries, clothes, appliances, gas (*BuyGoods*); Purchase services such as dry cleaners, banking, service a car, pet care (*BuyService*); Go out for a meal, snack, carry-out (*BuyMeal*); Run other general errands such as post office (*ShopServ*); Engage in recreational activities such as visit parks, movies, bars (*Recreation*); Exercise (*Exercise*); Visit friends and/or relatives (*VisitFrsRls*); and all *Other*. Travel between these places is also considered a "state" noted as *Trip* (and distinguishable from *DropPickup* and *ChangeTrans*). The daily sequence has a length of 1440 which is the total minutes in an assigned survey day. Every minute of the day contains one of the 15 distinct states for

each person. An example of a sequence with a length of ten minutes can be noted as $x = \{Home, Home, Home, Trip, Trip, Trip, Trip, Trip, Work, Work\}$, that is staying at home for 3 minutes, then travel for 5 minutes and then staying at the workplace for 2 minutes.

### Complexity of individual daily activity schedule

We use an indicator of *Complexity* to measure the variability of individual daily activity schedules. The Complexity indicator is based on the concept of *Entropy* which is a measurement of "prediction uncertainty" [93] and transitions between distinct states within a sequence $x$. The explanation follows [94, 95] closely. Entropy and Complexity can be calculated as follows.

$$h(x) = h(\pi_1, ..., \pi_s) = -\sum_{i=1}^{s} \pi_i log(\pi_i) \tag{2.2}$$

$$C(x) = \sqrt{\frac{nt(x)}{(l(x)-1)} \frac{h(x)}{h_{max}}} \tag{2.3}$$

where $h(x)$ is the entropy function of $x$ which represents the sequence, $s$ is number of potential states and $\pi_i$ is proportion of occurrences of the $i$th state in the considered sequence. Since this measure does not account for the number of state changes and contiguity of states, the indicator of Complexity is introduced that includes entropy and the number of transitions $nt(x)$ in a sequence $x$, normalized by the maximum theoretical entropy $h_{max}$ and the length of the sequence $l(x)$. This indicator takes a value between 0 and 1, with 0 corresponding to entropy zero and no transitions (e.g., staying at a single place for the entire day of observation).

**Sequence alignment method**

Using these sequences we seek to build a taxonomy of spatiotemporal daily activity and travel. NHTS collects data using a travel diary with information about the place and activity at the origin and destination of each trip and the time of day for departure and arrival for each trip. For each one-day travel record, we insert activities before the first trip, after the last trip, and in between trips based on origin–destination place, the timing of trips, and information about activity/purpose at the origin and destination of each trip. In this way, for each person, we have a sequence of place-activity-travel as a series of 1440 bins—one for each minute of the survey day starting at 4:00 AM and ending the next day at 3:59 AM. Each bin is assigned to a category by a word (i.e., *Home, Home&Work, Work, School, DropPickup, ChangeTrans, BuyGoods, BuyService, BuyMeal, ShopServ, Recreation, Exercise, VisitFrsRls, Other, and Trip*).

To identify similar sequences among the person-sequences of the sample data, we use methods from sequence alignment techniques. In the sequence alignment literature, the measurement of dissimilarity and the number of operations needed to make two sequences the same is called a "distance". Typically, the distance between two sequences is the minimum combination of substitution and indel (i.e., insertion and deletion) [96]. As [93] stated, favoring indel reduces the importance of time shifts in the comparison while favoring substitutions gives more importance to position-wise similarities. Sequential order is of primary concern in daily activity and time allocation pattern recognition because consecutive activities are likely to affect one another [97]. For this reason, we only consider substitution when computing the dissimilarity between two sequences. Hamming distance is the most appropriate dissimilarity measure, as it measures the minimum cost of substitutions required to change one sequence into another. For ease of interpretation, in this chapter, we call the minimal cost of transforming one sequence to another the

"dissimilarity score" between the two sequences.

To compute the dissimilarity score between two sequences, we need to define the cost for substitution. According to [93], the substitution cost between states $S_p$ and $S_q$ can be obtained with the following formula.

$$SC(S_p, S_q) = 2 - P(S_p|S_q) - P(S_q|S_p) \tag{2.4}$$

where $SC(S_p, S_q)$ is the substitution cost between states $S_p$ and $S_q$ with a value between 0 and 2; $P(S_p|S_q)$, the transition rate from state $S_q$ to state $S_p$, is the probability of observing state $S_p$ at time $t+1$ given that state $S_q$ has been observed at time $t$. If the transition rate from state $S_q$ to state $S_p$ has a value close to 1, it means that a person in a given state $S_q$ at time $t$ has a great probability to transition to state $S_p$ at time $t+1$. Notice that $P(S_p|S_q)$ is not equal to $P(S_q|S_p)$. The idea is to set a high cost when changes between $S_p$ and $S_q$ are not observed often and lower cost when they are frequent.

A substitution-cost matrix can be derived by computing $SC(S_p, S_q)$ for each pair of distinct states. The substitution-cost matrix is a square symmetrical matrix of dimension $s$ by $s$, where $s$ is the number of distinct states, in our case 15. Then we can compute the Hamming distance between each two sequences based on the substitution-cost matrix. The output of the Hamming distance computation among all the sequences is a matrix of dissimilarity scores. The dissimilarity matrix contains $N$ by $N$ ($N$ is the count of respondents) cells of dissimilarity scores among sequences for each person-day in our sample. The matrix is symmetric with zeros along the diagonal.

$$D\_Matrix = \begin{bmatrix} 0 & d_{x_1,x_2} & ... & d_{x_1,x_N} \\ d_{x_2,x_1} & 0 & ... & d_{x_2,x_N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{x_N,x_1} & d_{x_N,x_2} & ... & 0 \end{bmatrix} (d_{x_i,x_j} = d_{x_j,x_i} (i, j \in [1, N])) \tag{2.5}$$

where $d_{x_i,x_j}$ is the dissimilarity score between two sequences $x_i$ and $x_j$ using Hamming distance as the measurement; $D\_Matrix$ is the derived $N$ by $N$ symmetric matrix of dissimilarity scores.

### Agglomerative nesting clustering

In the next step, we derive homogeneous groups of sequences using cluster analysis on the dissimilarity matrix ($D\_Matrix$) among sequences of 1440 min in a day, and each minute is classified in one of the 15 categories mentioned above. We use the agglomerative nesting clustering method (AGNES) which is a type of hierarchical clustering. However, other clustering methods such as Partitioning Around Medoids (PAM) and Divisive Analysis Clustering (DIANA) could also be used to identify distinct groups of sequences with similar patterns [93, 96]. Compared to AGNES which uses a bottom-up approach, DIANA follows the top-down hierarchical clustering approach. However, either bottom-up or top-down fashion will yield the same dendrogram, which means no difference in choosing AGNES or DIANA. PAM is not considered here because this algorithm works effectively for small data sets but does not scale well for large data sets due to computational complexity [98]

To implement AGNES, we start with all sequences and group pairs of sequences based on their dissimilarity scores. Then, we compare all the cluster dissimilarity score averages to each other and lump together clusters with smaller dissimilarity scores. We proceed

until all observations are in one cluster [96]. This process includes a dendrogram that at each stage contains different combinations of sequences in groups. Different groupings can be compared in terms of the differences in the variance of sequences within the groups and across the groups. We aim to maximize the variance across groups and minimize the variance within groups to obtain an optimal number of clusters. This optimal number of clusters is determined using as rule the minimization of the within-cluster sum of squares, average silhouette (an indicator of dissimilarity), and a daily behavior representation conforming to travel behavior literature. We provide additional explanation on this later in this chapter.

**Example of the identified daily time allocation pattern**

Figure 2.2 shows an example of the daily time allocation pattern of a group of person-days. The group is one of the daily pattern clusters we will also discuss later in this chapter and we use here as an example. There are 15 states of activity in this study. The x axis represents the time of day in minutes starting at 4:00 am and ending at 3:59 am the next day. The y axis is the relative frequency of people who are doing a certain type of activity in the cluster in a stacked bar format. It shows that most persons in this group started their day from home and they spend most of their time either at home or working at their workplace. These persons spend a small amount of time on trips, buying goods, buying meals, recreation, drop and pick up, and visiting friends or relatives. The first peak for buying meals is noon and the second peak is after work but for a small percentage of the people in this group. The overall pattern we see in this figure is a typical commuting pattern in the United States.

From a pattern recognition standpoint, the primitives of sequence analysis are the minutes in a day (1440), and their assignment to one of the 15 categories of activity or trip type. Using this representation and information we derive daily patterns such as

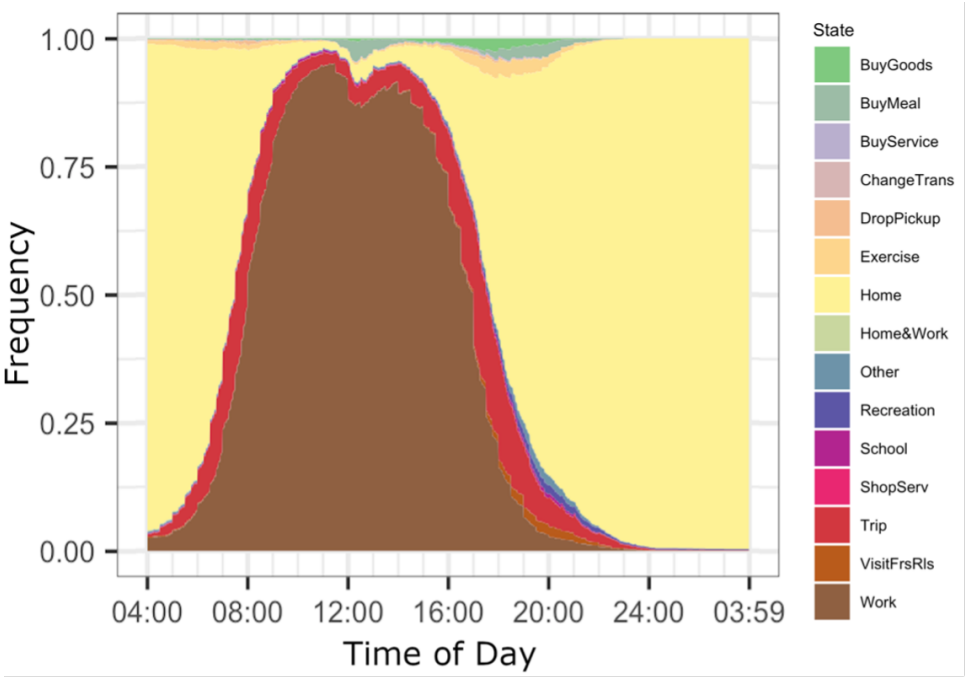



Figure 2.2: An example of daily pattern of min-by-min activity sequences.

typical workday, typical school day, and mostly at home day with running some errands.

## 2.4 Results

### 2.4.1 Composition of the motifs

Analysis of the daily motif patterns in the 2017 California-NHTS workday data shows 218 and 107 distinct types of motifs for telecommuters and commuters respectively. Telecommuters have a smaller sample size, yet more diverse location visiting and travel patterns than commuters. This counterintuitive result tells us that telecommuters do not obey uniform travel patterns, and substantial heterogeneity exists among individuals. Figure 2.3 shows that only 17 unique motifs from telecommuters and 13 motifs from commuters are above 1% of their respective samples. Taken together the 17 motifs of telecommuters capture mobility patterns of approximately 79.96% of telecommuters'

daily patterns and the 13 motifs of commuters capture 81.93% of commuters' daily patterns. Figure 2.3 enumerates each of the most popular motifs, the structure of each motif, and its percentage. The motifs are grouped according to the number of nodes separated by dashed lines. One advantage of using motifs is our ability to distinguish multi-tour daily patterns from single-tour patterns. In this study, a tour (also known as a trip chain) is made up of individual trips and all the stops a person made along the way at distinct locations. Some motifs have multiple tours in a day. For example, compare motifs 16, 17, and 18. Motif 16 is a single tour with a shorter round trip to a location not in the tour, motif 17 contains one tour, and motif 18 contains two tours. It should be noted that each of the remaining motifs of telecommuters and commuters not included in Figure 2.3 were found in less than 1% of total respondents of telecommuters or commuters, and they are labeled as all other below.

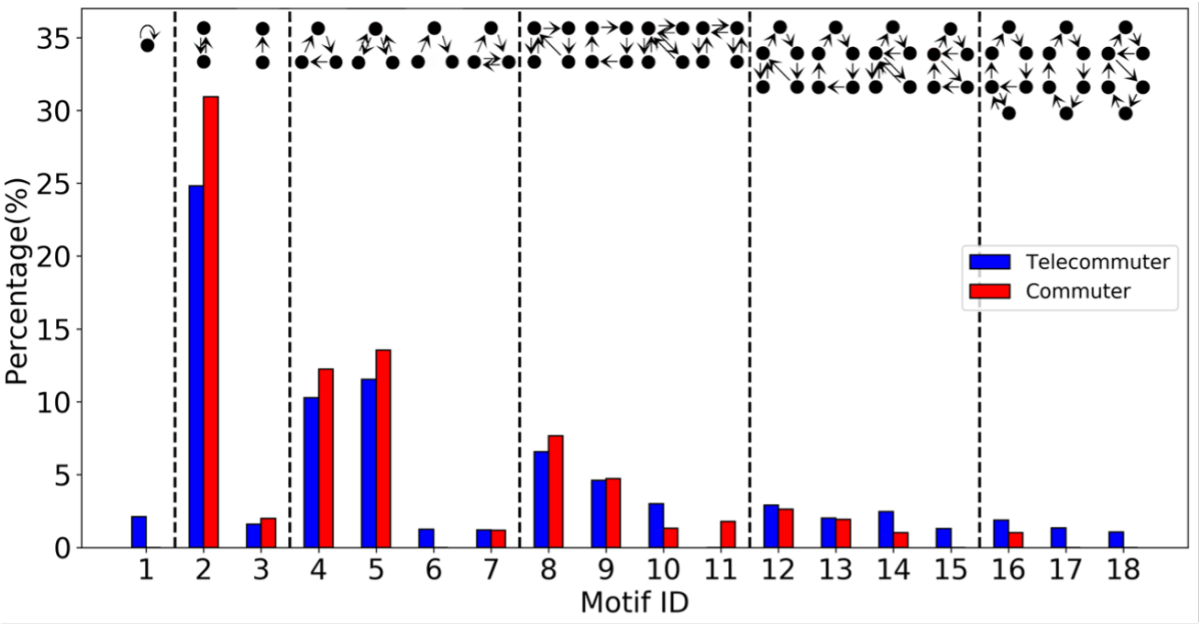



Figure 2.3: Percentage of human daily mobility motif patterns.

Motif 1 contains respondents whose trips all have the same origin and destination (loop trips). Around 2% of telecommuters and less than 1% of commuters perform motif

1. Motif 2 (with two locations visited and at least a round trip between them in a day) accounts for the highest percentages for both categories. This is consistent with the location-based motif in Schneider et al. [87] and Cao et al. [88], given that we only focus on the telecommuters and commuters during workdays in this chapter instead of the whole sample. 30% of commuters do motif 2, while telecommuters have a lower percentage of motif 2 compared to commuters, presumably because telecommuters are less constrained by a fixed work location. It is also noteworthy that commuters are more concentrated in less complex motifs (with less than five nodes) while telecommuters are more spread out in terms of the percentage of each distinct motif. The substantial heterogeneity in human mobility patterns, especially for telecommuters, points to location-based differences in travel patterns that we analyze in more depth in the next sub-section.

For each of these motifs, we compute the average characteristics of the members, including daily travel behavior characteristics. Table 2.3 provides the list of the telecommuters' 17 motifs, plus one category containing all the other motifs. Table 2.4 provides the same characteristics for commuters' motifs. For each motif, the tables provide the within-motif average characteristics of time allocation by the survey respondents.

Tables 2.3 and 2.4 show that the complexity of daily schedules increases with the number of combinations of nodes and edges in the motifs. This is expected because people tend to participate in different activities when they visit different places. The same is found for both telecommuters and commuters despite their differences in work styles. In general, given the same structure of motifs, commuters tend to spend more time at work as well as outside the home. Specifically, commuters using motif 3 spend the highest amount of time, 734.2 min, at work and 959.1 min outside home on average. When we look at the location type, it shows that 75.09% of the trips made by commuters using motif 3 are work-related trips (either origin or destination is workplace). Even though similar to commuters using motif 3, telecommuters using the same motif 3 spend

Table 2.3: Motifs and time-based behavioral indicators for telecommuters (including size and sample percentage of motif membership).

| Motif | Complexity | Time at Work | Time Outside Home | Time Traveling | Travel Time Ratio | Person in Sample | Percent in Sample |
|---|---|---|---|---|---|---|---|
|  1 | 0.019 | 81.1 | 288.7 | 82.7 | 0.733 | 47 | 2.10% |
|  2 | 0.026 | 236.9 | 400.0 | 65.9 | 0.323 | 555 | 24.82% |
|  3 | 0.018 | 193.3 | 882.1 | 183.6 | 0.262 | 36 | 1.61% |
|  4 | 0.033 | 228.6 | 467.0 | 80.0 | 0.308 | 230 | 10.29% |
|  5 | 0.043 | 177.9 | 416.8 | 91.6 | 0.342 | 258 | 11.54% |
|  6 | 0.029 | 305.9 | 977.2 | 224.0 | 0.238 | 28 | 1.25% |
|  7 | 0.046 | 157.8 | 355.1 | 135.9 | 0.504 | 27 | 1.21% |
|  8 | 0.050 | 145.2 | 447.5 | 113.0 | 0.321 | 147 | 6.57% |
|  9 | 0.040 | 93.9 | 315.6 | 96.4 | 0.372 | 103 | 4.61% |
|  10 | 0.057 | 137.0 | 424.1 | 127.5 | 0.368 | 67 | 3.00% |
|  12 | 0.056 | 203.0 | 513.0 | 118.2 | 0.295 | 65 | 2.91% |
|  13 | 0.047 | 157.7 | 420.7 | 116.8 | 0.325 | 45 | 2.01% |
|  14 | 0.063 | 94.2 | 392.5 | 120.9 | 0.337 | 55 | 2.46% |
|  15 | 0.057 | 141.1 | 412.6 | 113.9 | 0.358 | 29 | 1.30% |
|  16 | 0.062 | 125.3 | 460.5 | 153.4 | 0.353 | 42 | 1.88% |
|  17 | 0.052 | 157.2 | 481.7 | 143.5 | 0.358 | 30 | 1.34% |
|  18 | 0.061 | 108.4 | 432.9 | 131.8 | 0.334 | 24 | 1.07% |
| All other | 0.067 | 153.8 | 573.7 | 176.8 | 0.367 | 448 | 20.04% |

Note: Time is measured in minutes per day. The values of complexity, time at work, time outside home, time traveling, and travel time ratio are the average values of groups of people belonging to each motif. The background color in a gradient is according to the data in each column and the same for the rest of the summary tables in this section. Motif 11 is missing in this table because it is below 1% of the included sample of telecommuters. Same for Tables 2.5a and 2.6.

Table 2.4: Motifs and time-based behavioral indicators for commuters.

| Motif | Complexity | Time at Work | Time Outside Home | Time Traveling | Travel Time Ratio | Person in Sample | Percent in Sample |
|---|---|---|---|---|---|---|---|
|  2 | 0.028 | 463.4 | 574.6 | 68.1 | 0.158 | 3962 | 30.93% |
|  3 | 0.017 | 734.2 | 959.1 | 66.2 | 0.127 | 255 | 1.99% |
|  4 | 0.036 | 430.8 | 598.2 | 83.3 | 0.171 | 1568 | 12.24% |
|  5 | 0.044 | 426.6 | 608.7 | 94.7 | 0.18 | 1734 | 13.54% |
|  7 | 0.046 | 385.8 | 536.0 | 81.2 | 0.203 | 150 | 1.17% |
|  8 | 0.050 | 404.5 | 621.5 | 105.9 | 0.197 | 983 | 7.67% |
|  9 | 0.043 | 401.3 | 612.0 | 102.0 | 0.195 | 604 | 4.72% |
|  10 | 0.055 | 331.2 | 582.1 | 114.7 | 0.234 | 170 | 1.33% |
|  11 | 0.058 | 443.3 | 668.6 | 131.3 | 0.204 | 228 | 1.78% |
|  12 | 0.057 | 376.6 | 642.2 | 125.1 | 0.213 | 336 | 2.62% |
|  13 | 0.049 | 373.1 | 603.8 | 112.3 | 0.215 | 246 | 1.92% |
|  14 | 0.062 | 299.2 | 596.2 | 132.8 | 0.243 | 130 | 1.01% |
|  16 | 0.062 | 313.8 | 613.2 | 136.0 | 0.235 | 129 | 1.01% |
| All other | 0.062 | 323.3 | 675.4 | 165.2 | 0.272 | 2314 | 18.07% |

Note: Motifs 1, 6, 15, 17, and 18 are missing in this table because they are below 1% of the included sample of commuters. Same for Tables 2.5b and 2.7.

a very high amount of time 882.1 min outside home compared to other motifs, the time at work is only 193.3 min for these telecommuters. None of the trips they made are to or depart from workplaces, which indicates that telecommuters using motif 3 go to places labeled "other" to do work for a short time. Except for motif 3, commuters using motif 2 representing the popular typical daily commuting pattern spend the highest amount of time at work (463.4 min). However, telecommuters using the same motif2 only spend 236.9 min at work. When we look at the location type, only 0.95% of trips made by telecommuters using motif 2 are work-related while the percentage for commuters is

83.43%. These findings are indicative of the changing nature of workplaces and the flexibility of performing work tasks away from traditional work locations hinted by the literature in the introduction.

Furthermore, in general, time spent traveling also increases from simple motifs to more complex motifs. This is expected because people spend more time traveling when they visit more and different places. However, motifs 3 and 6 of telecommuters are an exception because they contain only unidirectional trips. It is also noteworthy that given the same structure of motif 3 for telecommuters and commuters, telecommuters spend 183.6 min traveling while commuters only use 66.2 min on average. When we look at the location type, it turns out that 97.22% of telecommuters using motif 3 are either leaving home to a place labeled "other" or coming back home from a place labeled "other". However, for commuters, this percentage only accounts for 17.74%. Given the same motif structure, telecommuters using motif 3 are people who leave home to a place far away from their home or come back home from a distant place while for commuters, the origin and destination types are more diverse, and on average they travel relatively shorter time compared to telecommuters. Tables 2.3 and 2.4 include another synoptic indicator of daily travel called *Travel Time Ratio* (TTR) [99], defined as the total travel time in a day divided by the sum of the total time in activities outside the home plus the total travel time in a day. Higher values of TTR indicate a higher proportion of out-of-home time spent traveling. In general, telecommuters have higher values of TTR on workdays compared to commuters which means telecommuters spend more time traveling and less time in longer out-of-home activities (e.g., working at a workplace), but in shorter activities such as running errands or picking up and dropping off children at places.

Table 2.5 shows person's characteristics for each motif. In general, women and men are quite evenly distributed across these distinct motifs except in motif 3 for telecommuters and commuters and motif 13 for commuters. Women are less likely to use motif

Table 2.5: Motifs and respondents' characteristics (percent of persons within each motif group) for (a) telecommuters and (b) commuters.

(a) Telecommuters

| Motif | Woman | Full-time Employee | Retired |
|---|---|---|---|
| 1 | 57.45% | 76.60% | 6.38% |
| 2 | 47.21% | 61.80% | 8.11% |
| 3 | 33.33% | 55.56% | 11.11% |
| 4 | 53.04% | 53.91% | 10.43% |
| 5 | 51.94% | 58.91% | 8.91% |
| 6 | 50.00% | 64.29% | 7.14% |
| 7 | 55.56% | 55.56% | 11.11% |
| 8 | 51.70% | 55.78% | 9.52% |
| 9 | 57.28% | 44.66% | 13.59% |
| 10 | 47.76% | 55.22% | 7.46% |
| 12 | 46.15% | 58.46% | 10.77% |
| 13 | 48.89% | 42.22% | 33.33% |
| 14 | 49.09% | 47.27% | 12.73% |
| 15 | 48.28% | 65.52% | 10.34% |
| 16 | 57.14% | 61.90% | 9.52% |
| 17 | 50.00% | 50.00% | 20.00% |
| 18 | 45.83% | 33.33% | 16.67% |
| All other | 53.57% | 51.79% | 8.04% |

(b) Commuters

| Motif | Woman | Full-time Employee | Retired |
|---|---|---|---|
| 2 | 44.93% | 81.52% | 1.56% |
| 3 | 38.43% | 86.67% | 1.18% |
| 4 | 50.26% | 80.23% | 1.91% |
| 5 | 46.54% | 81.31% | 1.61% |
| 7 | 54.00% | 78.67% | 0.00% |
| 8 | 50.46% | 80.67% | 1.53% |
| 9 | 54.47% | 80.30% | 1.66% |
| 10 | 50.59% | 74.71% | 1.76% |
| 11 | 47.37% | 91.23% | 0.88% |
| 12 | 52.98% | 77.08% | 1.49% |
| 13 | 60.57% | 76.42% | 2.03% |
| 14 | 49.23% | 76.15% | 1.54% |
| 16 | 51.94% | 72.09% | 3.10% |
| All other | 51.47% | 76.62% | 2.81% |

Note: The percentages of part-time employees are omitted.

3 (the one-way trip to elsewhere), presumably because they need to take on household responsibilities ([100, 95]), which makes them less likely to leave home and not return in the survey day. As Table 1 shows, 43.83% of telecommuters are part-time employees while only 19.88% of commuters are part-time employees. This explains the very high

percentage of full-time employees of different motifs among commuters while the percentages of full-time and part-time workers in telecommuters are more balanced. Specifically, a high percentage (76.6%) of telecommuters belonging to motif 1 are full-time employees since telecommuters do not need to commute to a workplace and many of them tend to work from home. Motif 18, which includes two tours, for telecommuters has the highest percentage of part-time employees which could be a motif related to some part-time workers who work in multiple jobs. The percentages of full-time employees in the commuters group are quite spread out throughout different motifs except motifs 3 and 11. The part-time employee percentages of motifs 3 and 11 of commuters are substantially below the overall average percentage of 19.88%. Note also that 75.29% of commuters in motif 3 are either leaving home to work or returning home from work. Intuitively, part-time workers are less likely to leave home for work and do not come back after work since they are usually required to work for shorter hours compared to full-time workers. This could explain the very low percentage of part-time workers in motif 3 among commuters. Motif 11 of commuters is a very special mobility pattern accounting for 1.78% (228) of total commuters that are people who need to run errands or give rides as a daily routine before going to work (e.g. escort kids to school)and they also go out for lunch during work. The percentages of full-time (91.23%) and part-time (8.77%) commuters of motif 11 show that full-time commuters tend to be more likely to use motif 11. In terms of retirement status, it is not surprising that we have very few retirees throughout different motifs in commuters since most employed retirees do not commute to work. It is possible for retired people to get a job after retirement and leave their home during the workday. Retirees who telecommute tend to be concentrated in complex motifs such as motifs 13, 17, and 18. This shows that many retired persons have quite diverse mobility patterns and do not just stay at home. Note also that retirees can work as part-time workers and many part-time workers tend to telecommute as we discussed before.

Table 2.6: Motifs and respondents' travel mode choice for telecommuters.

| Motif | Walk | Bike | Transit | Passenger | Drive Alone | Drive Someone Else | Other Mode | Gini |
|---|---|---|---|---|---|---|---|---|
|  1 | 84.04% | 3.19% | 0.00% | 0.00% | 10.64% | 2.13% | 0.00% | 0.011 |
|  2 | 13.35% | 2.15% | 2.01% | 12.13% | 54.08% | 15.92% | 0.36% | 0.068 |
|  3 | 8.33% | 0.00% | 0.00% | 27.78% | 38.89% | 22.22% | 2.78% | 0.000 |
|  4 | 10.10% | 0.29% | 1.01% | 11.23% | 55.76% | 21.02% | 0.58% | 0.111 |
|  5 | 16.70% | 2.85% | 2.23% | 8.32% | 48.11% | 21.79% | 0.00% | 0.308 |
|  6 | 1.79% | 0.00% | 3.57% | 23.21% | 41.07% | 30.36% | 0.00% | 0.143 |
|  7 | 7.16% | 3.70% | 0.00% | 2.10% | 43.09% | 43.95% | 0.00% | 0.254 |
|  8 | 13.61% | 2.43% | 2.18% | 9.63% | 51.28% | 20.60% | 0.27% | 0.294 |
|  9 | 6.99% | 1.13% | 0.97% | 12.91% | 58.46% | 19.53% | 0.00% | 0.118 |
|  10 | 20.21% | 1.85% | 1.99% | 5.53% | 48.96% | 21.46% | 0.00% | 0.413 |
|  12 | 10.52% | 0.51% | 2.05% | 5.46% | 58.00% | 23.45% | 0.00% | 0.307 |
|  13 | 3.43% | 0.00% | 0.00% | 8.89% | 76.57% | 11.11% | 0.00% | 0.079 |
|  14 | 19.22% | 0.18% | 1.04% | 8.40% | 40.75% | 30.41% | 0.00% | 0.383 |
|  15 | 11.68% | 3.63% | 1.72% | 9.20% | 42.02% | 31.75% | 0.00% | 0.267 |
|  16 | 12.20% | 0.00% | 0.68% | 8.30% | 57.95% | 20.52% | 0.34% | 0.317 |
|  17 | 5.14% | 0.00% | 1.11% | 13.33% | 65.42% | 15.00% | 0.00% | 0.073 |
|  18 | 7.25% | 0.00% | 0.00% | 6.61% | 51.61% | 34.52% | 0.00% | 0.364 |
| All other | 11.37% | 1.34% | 0.66% | 9.71% | 48.36% | 28.39% | 0.19% | 0.346 |

Note: The first seven columns (all except the Gini column) are the group average percentages of each mode used on the survey day. The percentage values of the seven modes in each row sum up to 1. Same for Table 2.7.

Tables 2.6 and 2.7 show respondents' travel mode choice for each motif. The percentage of each mode used in the survey day for every respondent can be easily computed. The percentage values of each mode in Tables 2.6 and 2.7 are the group average of people belonging to each motif and each row except for the Gini index sums up to one. The Gini index quantifies the daily variation of mode choices and is computed as 1 minus

the summation of the squared value of the percentage of each mode choice. The Gini index of mode choices takes values between 0 and 1, with 0 implying that a person only uses one mode to travel throughout the survey day. From Tables 2.6 and 2.7 we can observe that bike, transit, and other mode (which contains all unaccounted-for modes including air) are the least favorite modes among all people. As expected, the vehicle-based modes including car as passenger, drive alone, and drive someone else are the most popular modes and they span from 75.95% in motif 10 to 96.57% in motif 13 except motif 1 among telecommuters. Commuters' percentages span from 73.26% in motif 11 to 94.12% in motif 3. Telecommuters who use motif 1 tend to walk most and rarely use other modes, yielding a very low Gini, which complements our previous analysis on this motif of home-based trips in which people tend to exercise or take a walk. In terms of the typical commuting pattern motif 2, 71.91% of commuters drive alone, while only 54.08% of telecommuters driving alone. There is also a higher percentage of passenger trips and drive someone else trips among telecommuters compared to commuters. This is consistent with our previous finding in Table 2.2. We also notice that the loop trip patterns with more than three locations (motifs 9, 13, and 17) all have more than 90% of vehicle-based modes and the lowest Gini index compared to other motifs of the same number of locations. This tells us that when people need to make a loop trip among four or more locations, the vehicle is the most preferable and dominant mode choice. People who are included in motifs 10, 14, and 18 of telecommuters and motifs 10, 11, and 14 of commuters have a quite large Gini index implying a higher combination of modes used in a day.

Table 2.7: Motifs and respondents' travel mode choice for commuters.

| Motif | Walk | Bike | Transit | Passenger | Drive Alone | Drive Someone Else | Other Mode | Gini |
|---|---|---|---|---|---|---|---|---|
|  2 | 4.68% | 1.43% | 5.21% | 6.08% | 71.91% | 10.25% | 0.43% | 0.043 |
|  3 | 1.37% | 2.16% | 1.96% | 10.20% | 67.65% | 16.27% | 0.39% | 0.006 |
|  4 | 4.04% | 0.75% | 3.08% | 6.36% | 70.52% | 15.18% | 0.06% | 0.122 |
|  5 | 13.59% | 2.64% | 3.03% | 8.32% | 54.34% | 17.88% | 0.20% | 0.302 |
|  7 | 7.16% | 4.04% | 0.80% | 5.66% | 60.91% | 21.32% | 0.11% | 0.234 |
|  8 | 11.54% | 1.10% | 2.27% | 8.63% | 55.93% | 20.01% | 0.52% | 0.318 |
|  9 | 3.61% | 1.38% | 2.26% | 8.09% | 65.00% | 19.49% | 0.17% | 0.150 |
|  10 | 19.85% | 2.80% | 1.48% | 7.15% | 47.20% | 21.33% | 0.20% | 0.394 |
|  11 | 20.61% | 2.05% | 3.94% | 6.65% | 42.78% | 23.83% | 0.15% | 0.441 |
|  12 | 8.89% | 0.74% | 2.52% | 9.12% | 55.81% | 22.87% | 0.05% | 0.313 |
|  13 | 4.31% | 0.65% | 2.03% | 7.07% | 60.69% | 25.25% | 0.00% | 0.154 |
|  14 | 18.57% | 2.77% | 1.95% | 6.86% | 47.78% | 22.07% | 0.00% | 0.392 |
|  16 | 11.23% | 0.78% | 1.88% | 7.52% | 54.55% | 22.50% | 1.55% | 0.300 |
| All other | 10.71% | 1.21% | 1.28% | 8.98% | 51.82% | 25.37% | 0.63% | 0.319 |

### 2.4.2 Time allocation pattern recognition using sequence analysis

The analysis in the previous section demonstrates a strong relationship between network motifs, daily time allocation, person characteristics, and travel mode choice to activities and travel in a day. In this section, we explore this further using sequence analysis as introduced in Section 2.3 to identify patterns of time allocation to activities and travel as well as to compare the differences in the time allocation patterns between telecommuters and commuters. To begin with, instead of repeating sequence analysis on the 17 motifs plus one category of all other motifs for telecommuters, and 13 motifs plus one category of all other motifs for commuters respectively, and further compare the

derived time allocation patterns, we merge motifs into five groups based on the number of nodes, which represents the number of distinct locations visited in the survey day. Another reason we cannot directly apply sequence analysis on persons using each of the distinct motifs is that the percentages for some motifs are very low (approximately 1%), hence, it is not feasible to implement sequence analysis and clustering on very small samples. Furthermore, as we conclude from Tables 2.3 and 2.4, the complexity of daily schedules increases with the number of combinations of nodes and edges in the motifs. Therefore, it is more intuitive to aggregate motifs by the number of nodes than any other strategy. The derived five groups of motifs also reflect five levels of complexity in daily schedules.

To distinguish telecommuters and commuters, we use the name "Group T" for telecommuters and "Group C" for commuters to use as labels for each group of patterns. Group T.I and Group C.I are both motif1 with only one node (recall these are persons that had a trip with the same origin and destination – the loop trip). Group T.II and Group C.II include motifs 2 and 3 with two nodes. Group T.III and Group C.III include motifs with three nodes. Group T.IV and Group C.IV include motifs with four nodes. All the other less frequent motifs with five and more nodes are classified as Group T.V and Group C.V. In the telecommuters' group, there are 47 persons (2.1%) in Group T.I, 591 persons (26.43%) in Group T.II, 557 persons (24.91%) in Group T.III, 405 persons (18.11%) in Group T.IV, and 636 persons (28.44%) in Group T.V. The composition by five groups of motifs for commuters is that 25 persons (0.20%) in Group C.I, 4,217 persons (32.92%) in Group C.II, 3,654 persons (28.53%) in Group C.III, 2,287 persons (17.85%) in Group C.IV, and 2,626 persons (20.50%) in Group C.V.

Following the procedure of sequence analysis, we construct a daily activity-travel sequence consisting of 15 distinct states with a length of 1440 (total minutes in a survey day) for each respondent in the five motif Groups. The procedure is followed by a

computation of the dissimilarity matrix among sequences of 1440 min in a day and implementing AGNES on the dissimilarity matrix to identify distinct clusters of sequences with similar patterns. Our objective is not only to build clusters with as similar sequences as possible but also to derive clusters in which their most representative "sequence" is as different as possible across clusters. To do this, we use the clustering solution that combines the WSS and Silhouette as introduced above. The usual procedure is to compute WSS and Silhouette for different numbers of clusters and then plot them as in Figure 2.4. The well-known elbow method in the clustering literature refers to a rapid drop in WSS and then slow reach to an asymptote of the graph in the right-hand side of Figure 2.4. It shows that selecting 4 clusters, or 5 clusters will lead to similar WSS, so a solution with more than 5 clusters would not be considered. The Silhouette graph shows the best solution for the separation among clusters is for 4 clusters that has the highest average silhouette coefficient. These two indicators show that 4 clusters would be an optimal solution for motif Group T.II. We apply the same strategy to determine the optimal number of clusters for the other motif groups (see Appendix A.1).

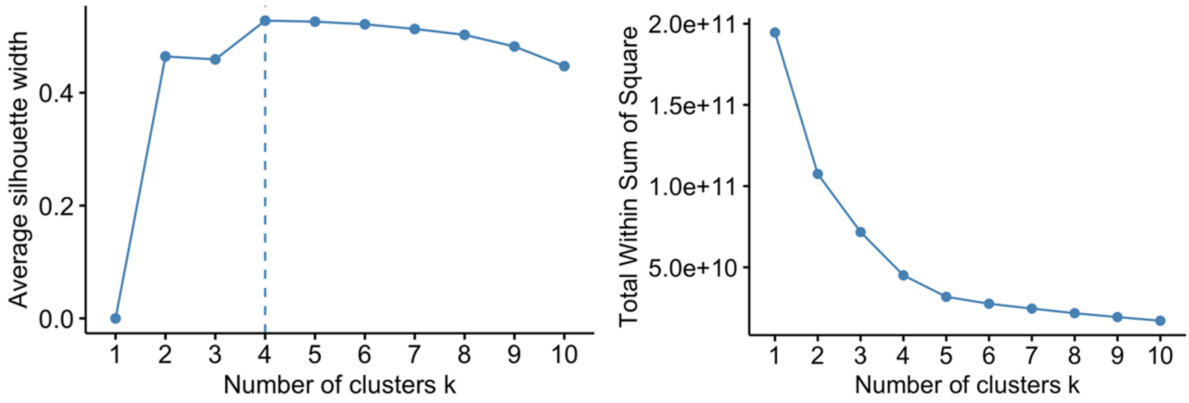



Figure 2.4: Illustration of the results of Silhouette (left) and WSS elbow (right) for motif Group T.II.

Recalling that within each motif group, we identify several daily activity-travel cluster sequences that represent different behaviors, we will use a numbering convention that

shows motif group and sequence cluster within each group. To distinguish them, we number them by assigning to the motif group a Roman numeral (I, II, III, IV, and V) and the sequence cluster an Arabic numeral (1,2,3,4,5). For example, in Group T.II, we will have more than one distinct type of sequence cluster labeled as II.1, II.2, II.3, and II.4.

Figures 2.5-2.12 show the distinct daily time of day patterns for each group of motifs for telecommuters and commuters respectively. The daily cluster names are based on the daily behavior each cluster represents. Group T.I and Group C.I both have a very small sample size, and running a cluster analysis on them yields some clusters with only one observation. This implies that a clustering analysis does not apply to these two small groups. Thus, the clustering results for Group T.I and Group C.I are not reported. On the other hand, all the other four groups of telecommuters have four distinct activity sequence patterns, and Group C.II and Group C.V have four distinct patterns as well while Group C.III and Group C.IV have five distinct patterns. Table 2.8 shows the statistics for the number of people, average entropy, average complexity, and Gini index of mode choice for each cluster in each group. In general, the average complexity within each group of motifs increases with the number of locations visited in a day. Groups T.I and T.II have the lowest Gini index implying less diverse mode choices compared to other groups of telecommuters. Similarly, Groups C.I and C.II have the lowest Gini index. Also, within each motif group, we find substantial heterogeneity of daily activity-travel patterns with different compositions of activities, timing of activities and travel, complexity, and Gini index indicating different transitions and variation of mode choice among activities.

All four motifs groups of telecommuters contain four similar major patterns, that is Workand Run ErrandsDay, Mostly Out of Home Day, Work from Home Day, and Long Work from Home Day. Some telecommuters in Work and Run Errands Day still

Table 2.8: Statistics for the number of people, entropy, complexity, Gini index for each cluster in each group.

| Group | Cluster | Pattern | No. of people | Ave. Entropy | Ave. Complexity | Ave. Gini |
|---|---|---|---|---|---|---|
| ***Telecommuter*** | | | | | | |
| T.I | / | / | 47 | 0.546 | 0.0192 | 0.0106 |
| T.II | T.II.1 | Work and Run Errands Day | 307 | 0.635 | 0.0255 | 0.0709 |
| | T.II.2 | Mostly Out of Home Day | 78 | 0.6 | 0.0231 | 0.0599 |
| | T.II.3 | Work from Home Day | 115 | 0.906 | 0.0313 | 0.0541 |
| | T.II.4 | Long Work from Home Day | 91 | 0.404 | 0.0194 | 0.0565 |
| T.III | T.III.1 | Work and Run Errands Day | 247 | 0.764 | 0.0364 | 0.191 |
| | T.III.2 | Mostly Out of Home Day | 83 | 0.749 | 0.0338 | 0.195 |
| | T.III.3 | Work from Home Day | 144 | 1.07 | 0.0477 | 0.277 |
| | T.III.4 | Long Work from Home Day | 83 | 0.593 | 0.0314 | 0.206 |
| T.IV | T.IV.1 | Work and Run Errands Day | 237 | 0.901 | 0.0487 | 0.319 |
| | T.IV.2 | Mostly Out of Home Day | 39 | 1.05 | 0.0495 | 0.327 |
| | T.IV.3 | Work from Home Day | 90 | 1.2 | 0.0551 | 0.25 |
| | T.IV.4 | Long Work from Home Day | 39 | 0.741 | 0.0417 | 0.166 |
| T.V | T.V.1 | Work and Run Errands Day | 422 | 1.06 | 0.0648 | 0.296 |
| | T.V.2 | Mostly Out of Home Day | 50 | 1.12 | 0.0623 | 0.281 |
| | T.V.3 | Work from Home Day | 128 | 1.3 | 0.0697 | 0.369 |
| | T.V.4 | Long Work from Home Day | 36 | 0.89 | 0.0552 | 0.178 |
| ***Commuter*** | | | | | | |
| C.I | / | / | 25 | 0.488 | 0.017 | 0 |
| C.II | C.II.1 | Typical Workday | 3039 | 0.81 | 0.0293 | 0.0396 |
| | C.II.2 | Very Late Workday | 243 | 0.763 | 0.0273 | 0.0348 |
| | C.II.3 | Leave Home Day | 270 | 0.606 | 0.0209 | 0.0203 |
| | C.II.4 | Mostly Home Day | 665 | 0.517 | 0.0221 | 0.0582 |
| C.III | C.III.1 | Typical Workday | 2628 | 0.939 | 0.0415 | 0.22 |
| | C.III.2 | Very Late Workday | 264 | 0.861 | 0.0391 | 0.19 |
| | C.III.3 | Leave Home Day | 101 | 1.08 | 0.0424 | 0.225 |
| | C.III.4 | Mostly Home Day | 552 | 0.668 | 0.0339 | 0.188 |
| | C.III.5 | Mixed Workday | 109 | 1 | 0.0415 | 0.229 |
| C.IV | C.IV.1 | Typical Workday | 1325 | 1.04 | 0.0516 | 0.303 |
| | C.IV.2 | Very Late Workday | 106 | 0.987 | 0.0487 | 0.222 |
| | C.IV.3 | Leave Home Day | 95 | 1.16 | 0.0513 | 0.276 |
| | C.IV.4 | Mostly Home Day | 364 | 0.732 | 0.042 | 0.253 |
| | C.IV.5 | Early Workday | 397 | 1.01 | 0.05 | 0.274 |
| C.V | C.V.1 | Typical Workday | 1138 | 1.1 | 0.0641 | 0.326 |
| | C.V.2 | Late Workday | 575 | 1.21 | 0.0677 | 0.363 |
| | C.V.3 | Leave Home Day | 161 | 1.25 | 0.0678 | 0.321 |
| | C.V.4 | Mostly Home Day | 752 | 0.988 | 0.0609 | 0.299 |

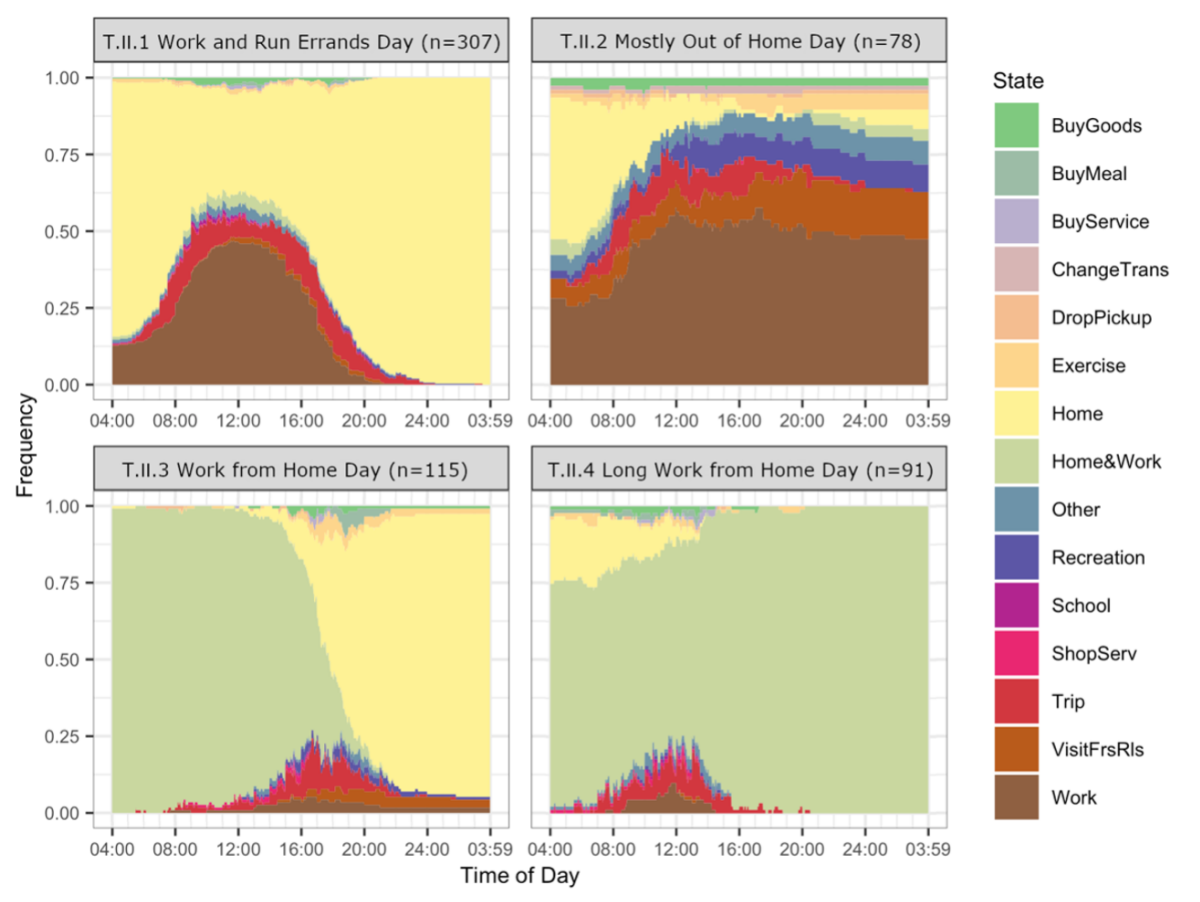



Figure 2.5: Daily time of day patterns of activity sequences for telecommuters Group II.

go to work and they also spend time on recreational activities, buying goods, and other activities. It might be counterintuitive at first glance that telecommuters still go to workplaces. However, the destination types of these trips show that 15.71% of destinations of the total 1457 trips going to work were at home, a very small percentage (1.65%) were at workplaces, and the majority (82.57%) were at places labeled as other, and only one trip was to a school location. As expected, most telecommuters do not commute to workplaces. Instead, they go to a variety of other locations including coffee shops, bookstores, customers' locations, or elsewhere to work. This confirms our introductory comments about the increases in work location flexibility and the multiple different locations that

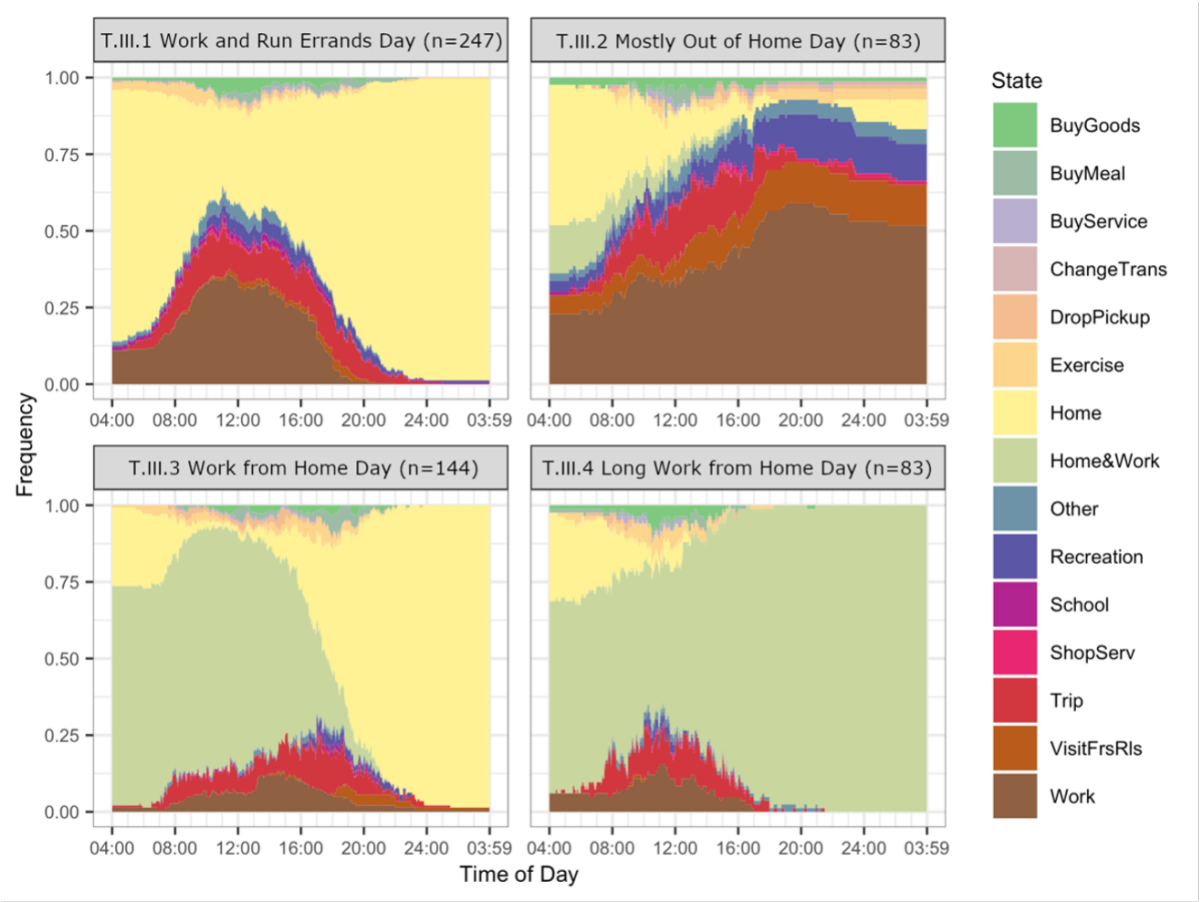



Figure 2.6: Daily time of day patterns of activity sequences for telecommuters Group III.

function as places where work is performed. In contrast, among the commuters' 15571 work trips, 80.69% are to a workplace, 18.38% are to a place labeled as other, 0.89% are to home, and 0.04% to school. The presence of a substantial number of places labeled as Other is notable because it parallels the telecommuters' tendency of either needing to visit a client's site and/or using some of the flexibility with technology to perform work tasks from anywhere.

The Mostly Out of Home Day pattern includes telecommuters who spend most of their time working, visiting friends and relatives, performing recreational activities, buying goods, and doing other activities, and they rarely return home in the evening. The Work from Home Day and the Long Work from Home Day are both typical telecommuting

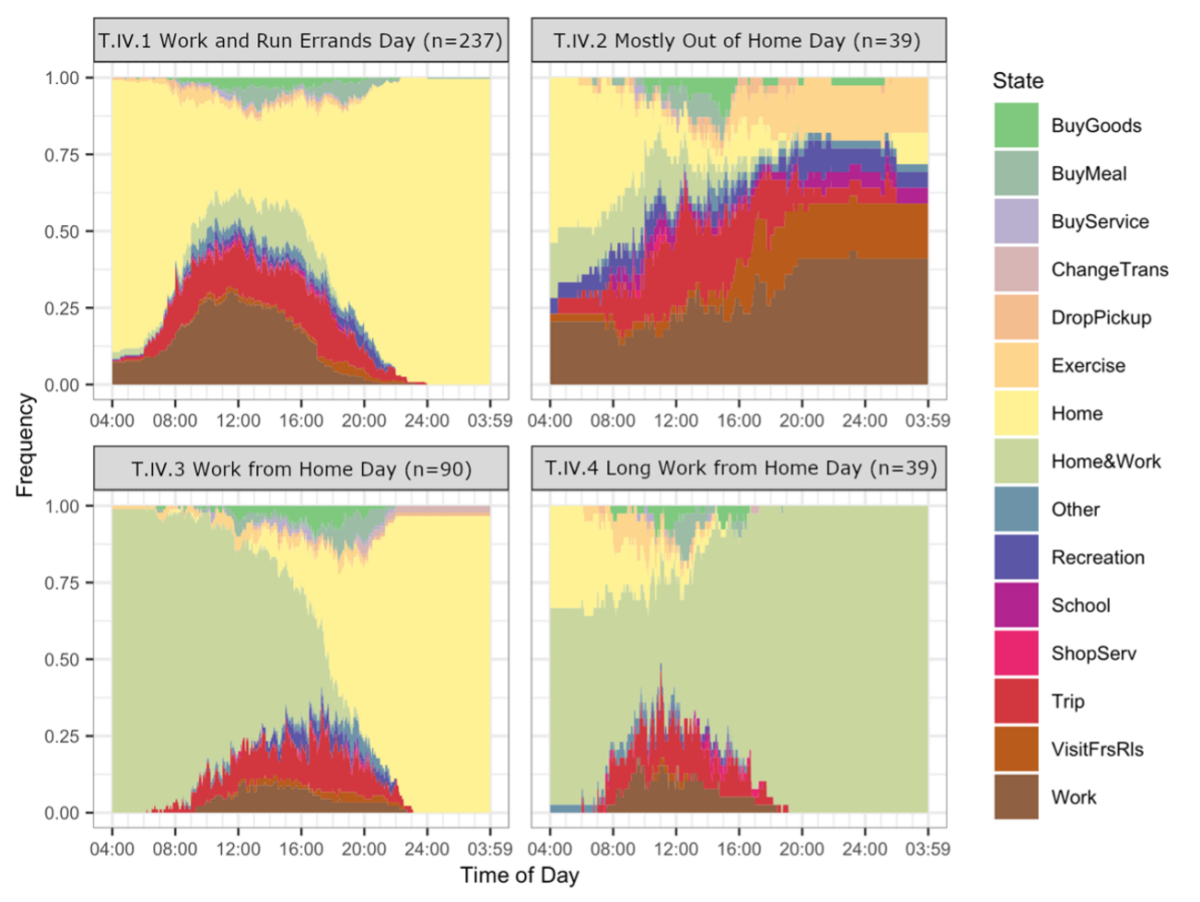



Figure 2.7: Daily time of day patterns of activity sequences for telecommuters Group IV.

daily routines. These two patterns share similar characteristics: telecommuters start to work from home in the very early morning and work most of the day. Beyond that, a few of them still go to a workplace, travel, buy goods, and have fun. The difference is that telecommuters in the Long Work from Home Day pattern work from home until late. In conclusion, telecommuters present very diverse daily activity-travel patterns within each motif group, but we still find four representative daily patterns within each group with similarities across groups.

We turn now to the daily patterns of commuters. All four groups of commuters contain the Typical Workday pattern. This pattern is a typical commuting routine. People leave home to work in the morning and return home after work. Patterns C.V.2 and C.IV.5

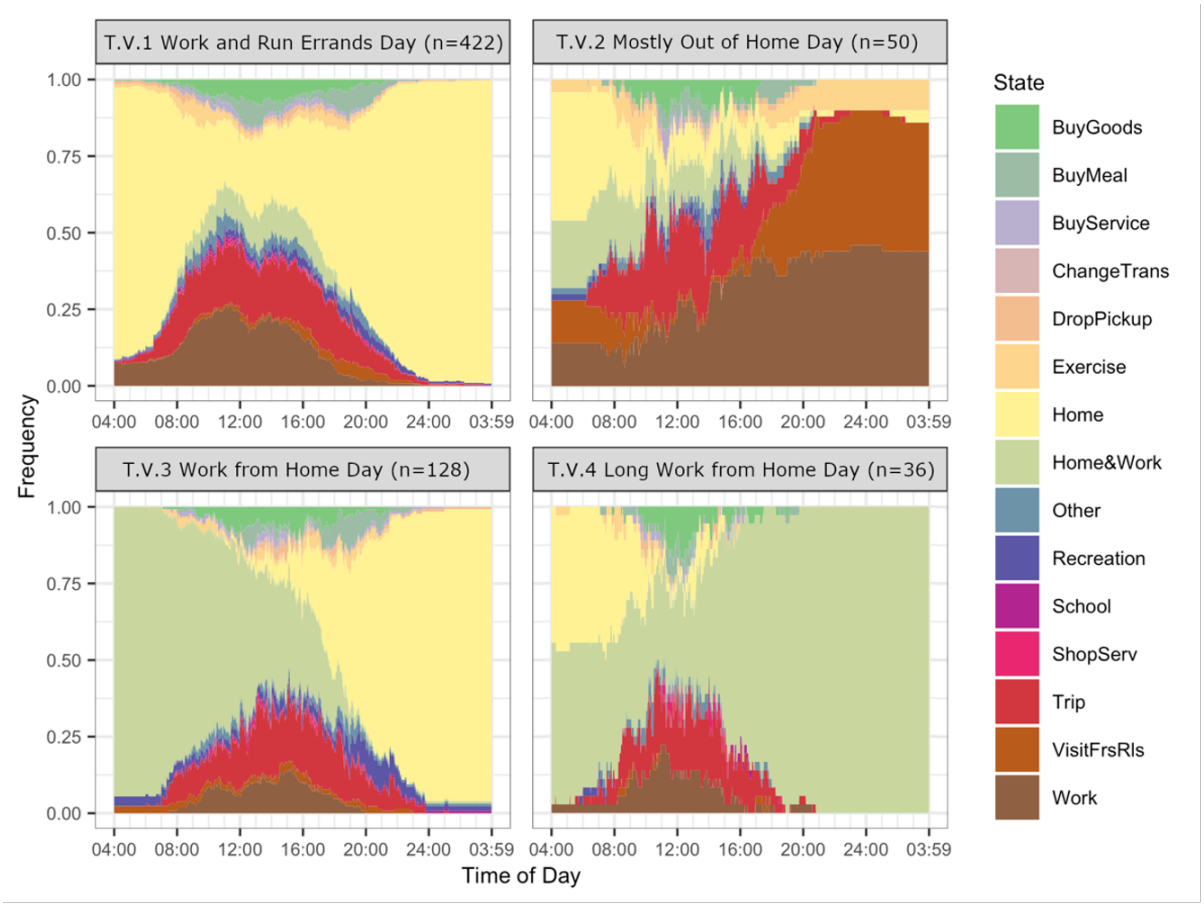



Figure 2.8: Daily time of day patterns of activity sequences for telecommuters Group V.

are similar to this pattern. However, an obvious difference is that commuters in pattern C.V.2 leave home to work later than the Typical Workday pattern, and commuters in pattern C.IV.5 leave home to work earlier than people in the Typical Workday pattern. Thus, we name pattern C.V.2 Late Workday and pattern C.IV.5 as Early Workday. In addition to these typical commute patterns, patterns C.II.2, C.III.2, and C.IV.2 are very similar and possibly contain people who work night shifts, as they start work in the afternoon and end at midnight. Leave Home Day is also a common pattern among these four groups of commuters. Most commuters in these groups did not return home in the evening. The Leave Home Day patterns C.III.3 and C.IV.3 share similar characteristics: people go to work in the morning and have fun after work. Many people in the Leave

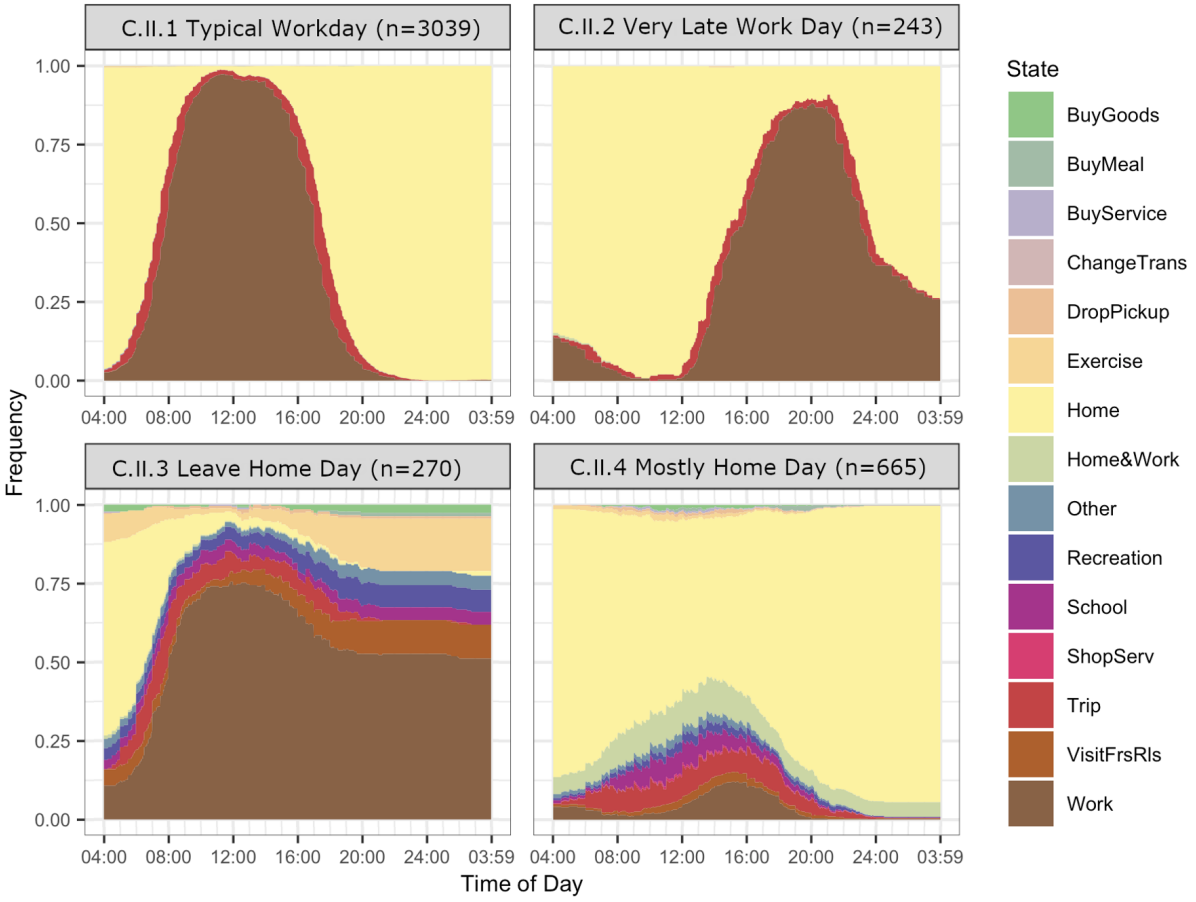



Figure 2.9: Daily time of day patterns of activity sequences for commuters Group II.

Home Day pattern C.II.3 did not return home because of working late. The Leave Home Day pattern C.V.3 has both late-night workers and people having fun after their workday. The common Mostly Home Day patterns within the four groups also share similar characteristics that include long periods at home, going out to buy goods, and a small fraction of time at work and traveling. Pattern C.III.5 is not a typical pattern. People in this group tend to work from home during the daytime and some of them also commute to work. Therefore, we call this pattern Mixed Workday.

Visual comparison of the time of day patterns between telecommuters and commuters for the same number of locations visited reveals that we have substantially different time

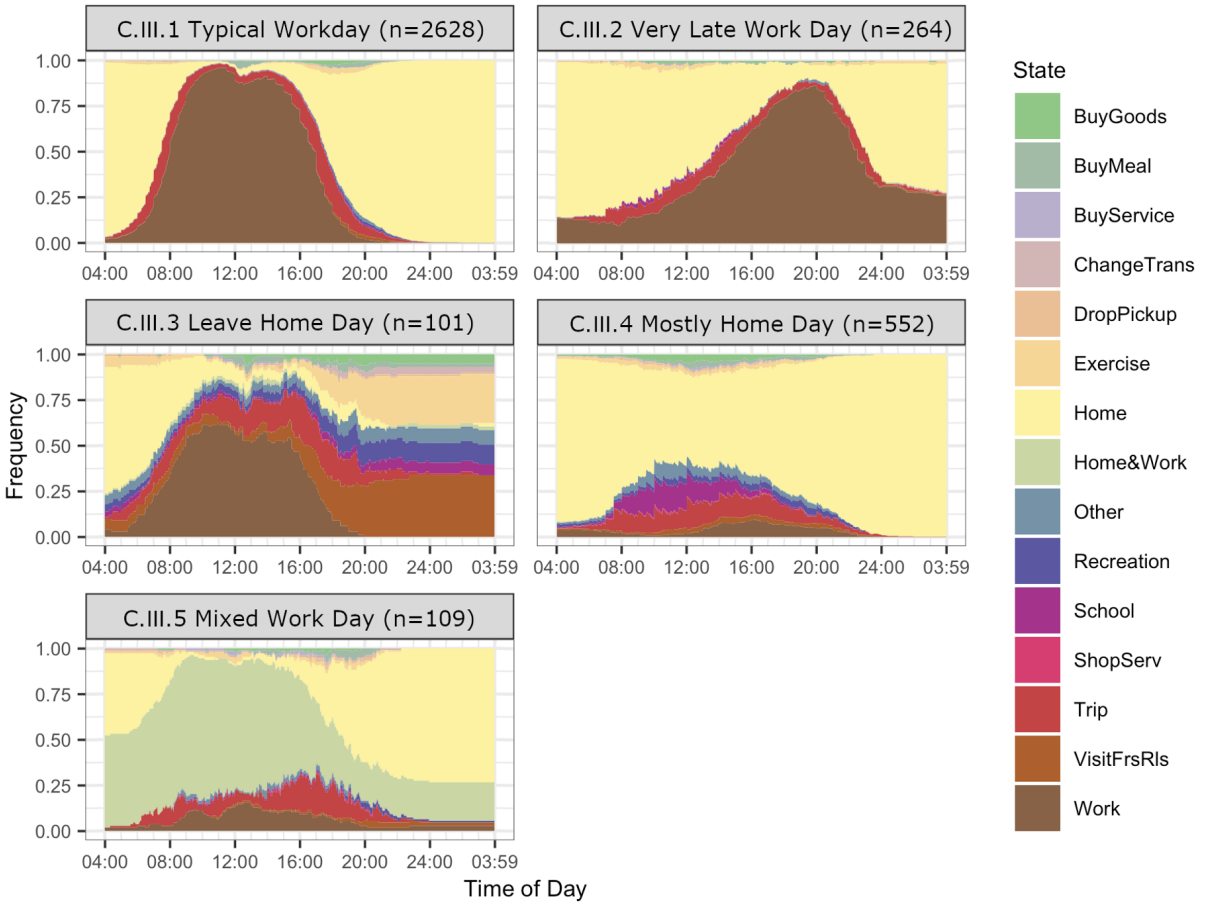



Figure 2.10: Daily time of day patterns of activity sequences for commuters Group III.

of day allocation to activities and travel. For example, the peaking of arrival time to and departure time from work that we see in the commuter patterns is mostly absent from the telecommuter groups, although those peaks are present for a small proportion of respondents. Also, an obvious difference as we would expect is the substantial number of telecommuter respondents that work at home for a long amount of time (T.II.3, T.II.4, T.III.3, T.III.4, T.IV.3, T.IV.4, T.V.3, and T.V.4), and only one pattern among the commuters (C.III.5) has some work at home.

Comparing the percentage of people who made dropping off/picking up trips at least once in each pattern, we find that a relatively higher percentage occurs in patterns

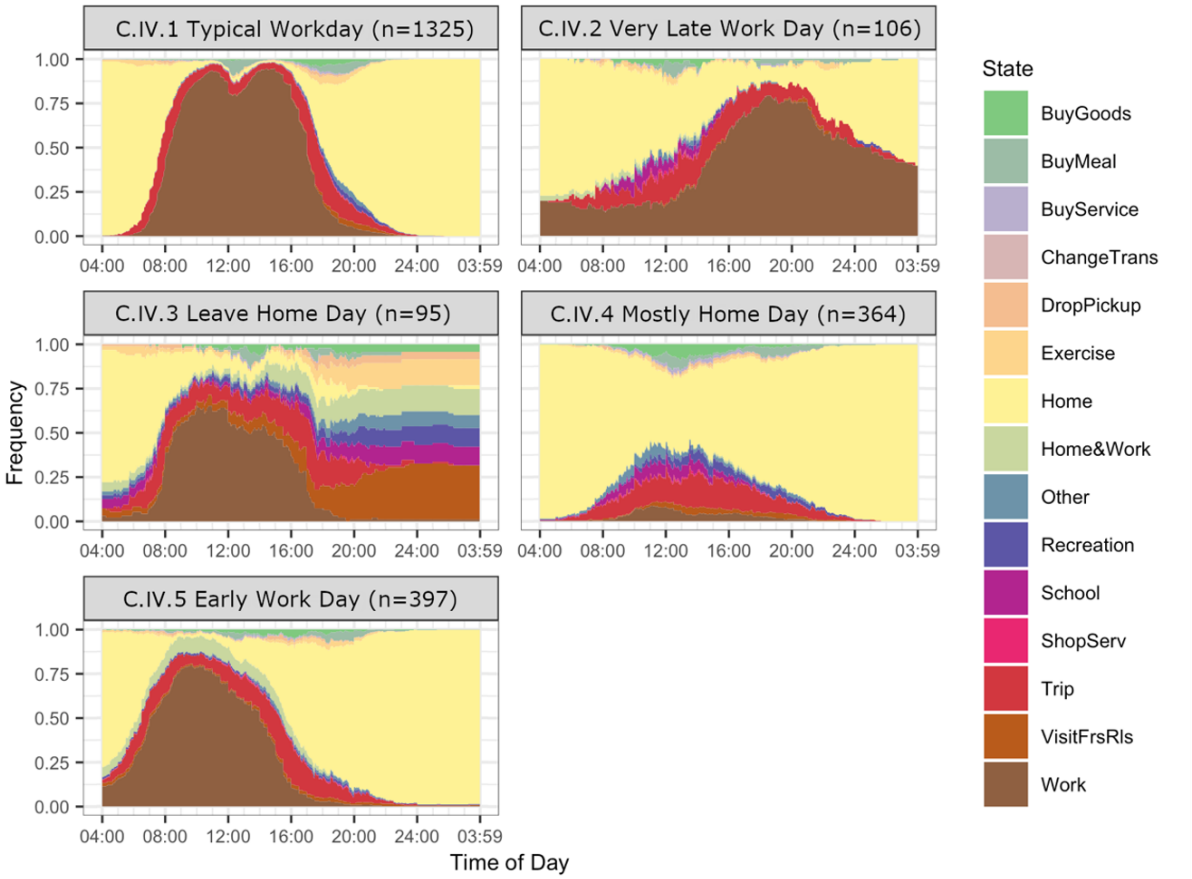



Figure 2.11: Daily time of day patterns of activity sequences for commuters Group IV.

T.IV.1 (16.54%) and T.V.1 (22.80%) for telecommuters, and in patterns C.III.1 (11.36%), C.IV.1 (15.09%), and C.V.1 (16.49%) for commuters. This shows a higher percentage of telecommuters in the Work and Run Errands Day pattern with 4 or more visited locations, as well as commuters in the Typical Workday pattern with 3 or more visited locations, are the designated drivers to escort people (e.g., taking children to school). In addition, telecommuters in patterns T.IV.1 and T.V.1 have a higher percentage of people driving others when compared to commuters in patterns C.III.1, C.IV.1, and C.V.1. Overall, 20.71% of telecommuters made at least one trip for dropping off/picking up people and only 16.52% of commuters did that. All these support our expectation that more persons from the telecommuters group are the designated drivers. However, a

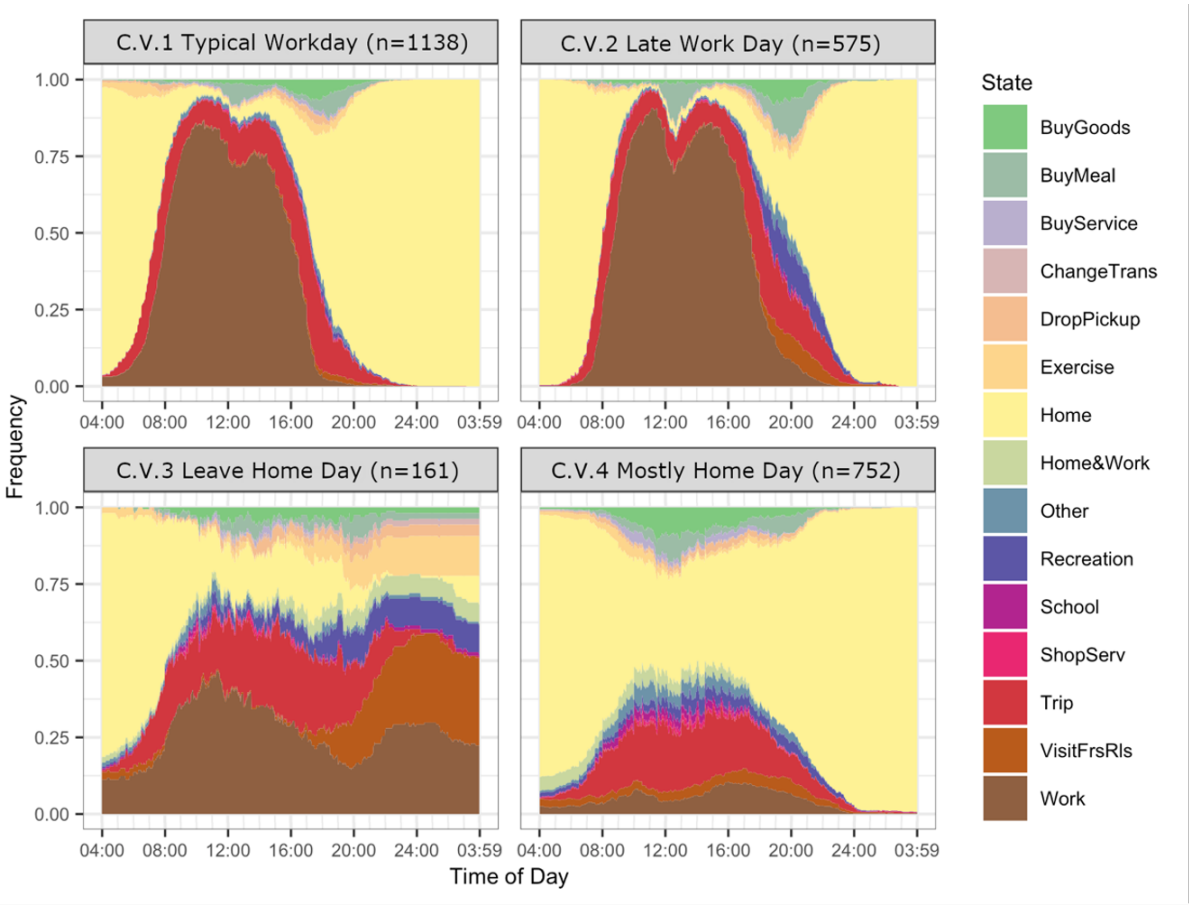



Figure 2.12: Daily time of day patterns of activity sequences for commuters Group V.

substantial number of regular commuters are also performing this task.

### 2.4.3 Intra-household interaction analysis

[101] studied the intra-household interactions of dual-earning household members and found a very strong association between telecommuting frequency and time allocation. With our techniques, we explored these associations without imposing the strong assumptions required in regression-like statistical models. We use 8,385 workers in 4,328 dual-earner households and 1,618 workers in 566 three-earner households (including workers making no trips) to investigate the combination of daily time of day patterns of activities

and travel for the dual-earner and three-earner households.

The frequency of different combinations of telecommuters and commuters for the dual-earner and three-earner households are listed in Table 2.9. It shows that 2 commuters is the most popular combination (67.24%) in dual-earner households; likewise, 3 commuters is the dominant combination for three-earner households. The second most popular combinations for dual-earner and three-earner households are combinations that have 1 telecommuter plus 1 commuter for dual-earner households (21.05% of the sample of dual-earner households) and 1 telecommuter plus 2 commuters (22.30% of the sample of three-earner households). All the other less frequent combinations are below 6%. We further look into the female percentage in the 1 telecommuter plus 1 commuter households and 1 telecommuter plus 2 commuters households. The statistics show that 54.67% of telecommuters are women while 44.68% of commuters are women in dual-earner households, which demonstrates that a higher percentage of women are more likely to be telecommuters in dual-earner households. However, in the three-earner households, only 49.19% of telecommuters are women and 50.81% of commuters are women which implies that men and women are almost equally likely to work as commuters or telecommuters in three-earner households.

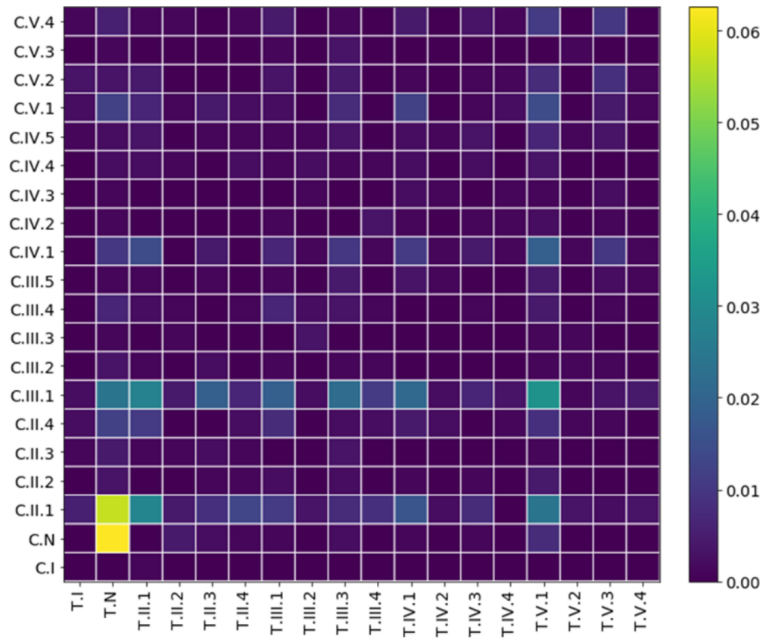
We further investigate the combination of daily time-of-day patterns of activities and travel for the dual-earner and three-earner households. We select three categories in Table 2.9 including 1 telecommuter plus 1 commuter, 2 telecommuters, and 1 telecommuter plus 2 commuters households as case studies. We aim to use these three groups containing telecommuters to explore the situation when there are commuters in a household how the other telecommuter's daily schedule looks and when both of the household members are telecommuters how their daily time allocation patterns look like on workdays. The results show that there are 205 distinct combinations for 1 telecommuter plus 1 commuter household, 92 distinct combinations for 2 telecommuter households, and 102 distinct

Table 2.9: Combinations of telecommuters and commuters for the dual-earner and three-earner households.

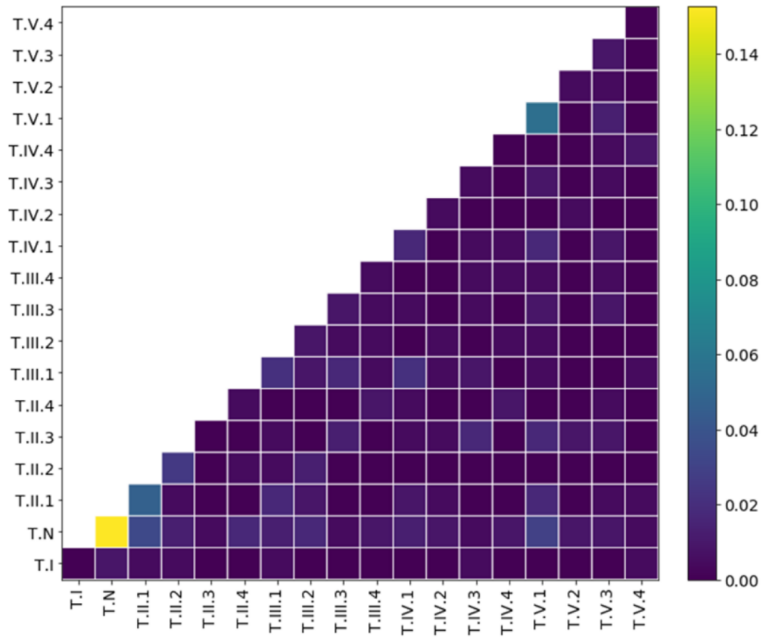
| Category | Number of households (Percentage) |
|---|---|
| *Dual-earner household (n=4328)* | |
| 2 commuters | 2910 (67.24%) |
| 1 telecommuter + 1 commuter | 911 (21.05%) |
| 2 telecommuters | 236 (5.45%) |
| 1 commuter | 229 (5.29%) |
| 1 telecommuter | 42 (0.97%) |
| *Three-earner household (n=566)* | |
| 3 commuters | 351 (63.13%) |
| 1 telecommuter + 2 commuters | 124 (22.30%) |
| 2 commuters | 33 (5.94%) |
| 2 telecommuters + 1 commuter | 30 (5.40%) |
| 1 telecommuter + 1 commuter | 6 (1.08%) |
| 3 telecommuters | 5 (0.90%) |
| 1 commuter | 4 (0.72%) |
| 2 telecommuters | 3 (0.54%) |

Note: If a category contains only one worker in a dual-earner household, the other worker is not accounted for in the NHTS data because they may be located outside California. Similarly, this holds for three-earner households.

combinations for 1 telecommuter plus 2 commuters households. Figure 2.13 shows the percentages of all possible combinations of daily time of day patterns of activities and travel for the 1 telecommuter plus 1 commuter, and 2 telecommuters households respectively. Note that C.N and T.N indicate commuters making no trips and telecommuters making no trips respectively. It is noticeable that in the 1 telecommuter plus 1 commuter households, a combination of a commuter making no trip associated with a telecommuter making no trip accounts for the highest percentage (6.26%). Likewise, in the 2 telecommuters households, the percentage of T.N together with T.N is also the highest (15.25%). A possible explanation could be that people from these two combinations coordinate their schedules to be together at home for family events or other situations that require every household member to be involved at home. In addition, C.II.1 and C.III.1 associated with




(a) 1 telecommuter plus 1 commuter households




(b) 2 telecommuters households

Figure 2.13: Percentages of various combinations of daily time-of-day patterns for the (a) 1 telecommuter plus 1 commuter and (b) 2 telecommuters households. Figure (b) is symmetric, so we only visualize half of the matrix of percentages.

other telecommuter patterns also have a relatively high percentage, which implies that given a household member works as a commuter, the other household member working as a telecommuter has much more diversity in the daily time of day patterns and takes on more household responsibilities such as escorting kids to school, grocery shopping, and so forth. This is even more obvious by some frequent telecommuter patterns including T.II.1, T.IV.1, and T.V.1 indicating that telecommuters tend to perform a Work and Run Errands Day pattern. In the 2 telecommuters households, the highest percentages occur in the diagonal pattern combinations including T.II.1 + T.II.1 (11 households, 4.66%) and T.V.1 + T.V.1 (13 households, 5.51%). The travel diaries of the household members following these two diagonal combinations show that 5 out of the 11 households doing T.II.1 + T.II.1 have the same daily schedule. The same happens in 5 out of the 13 households doing the combination T.V.1 + T.V.1. This proves that many people in 2 telecommuters households doing the pattern combination of T.II.1 + T. II.1 and T.V.1 + T.V.1 tend to do activities jointly through the day.

We next turn to the 1 telecommuter plus 2 commuters households. Given that there are 102 pattern combinations from the 124 households with 1 telecommuter plus 2 commuters when we look at the pattern distribution of the telecommuters, it shows many telecommuters are concentrated in patterns T.N (23.39%), T.II.1 (16.13%), T.V.1 (15.32%), T.IV.1 (8.06%), only 0.81% of telecommuters do T.I, and the rest are all below 5%. This tells us that given that many telecommuters from 1 telecommuter plus 2 commuters households do not make any trips on workdays (23.39% of T.N), a substantial number of telecommuters still visit 2, 4, and 5 locations to perform Work and Run Errands Day. This is consistent with dual-earner households where when the other household members are commuting, the telecommuter takes on household responsibility activities.

### 2.4.4 Comparison of stay-at-home persons

As mentioned in Section 2.2, of the total 24,448 workers (4,168 telecommuters and 20,280 commuters) in the 2017 California-NHTS, we only keep the workday subset comprising 2,236 telecommuters and 12,809 commuters for the motif and sequence analysis. The 9,403 excluded workers are people who were interviewed on non-workdays or stayed at the same location (mostly home) all day during workdays with no trips to report. To understand the differences between people who stayed at home all day and people who made at least one trip, two binary logistic regression models are estimated for telecommuters and commuters respectively. We use individual-level and household-level characteristics as explanatory variables and test their significance on telecommuters' and commuters' decisions on making trips. The final specification of the model was obtained by a systematic process of eliminating insignificant variables.

Table 2.10 summarizes the results of the two estimated binary logistic regression models. Among the set of individual sociodemographic variables, we find that commuters aged 66 and over are more likely to leave home and make trips compared to the rest age groups. Surprisingly, commuters younger than 26 are the age group that is least likely to leave home and make trips. Presumably, many of them are college students working as part-time commuters and their job duties are flexible which allows them to ask for leave and stay at home the entire survey day. However, none of the coefficients of the four age groups are significant for telecommuters, which indicates that different age groups do not have significant impacts on telecommuters' decisions to make trips. In addition, both telecommuters and commuters who attained a bachelor's degree or above are more likely to leave home and make trips compared to people who have less than a bachelor's degree. Telecommuters who work as part-time workers are more likely to leave home and make trips compared to their non-part-time worker counterparts. Nevertheless, the

Table 2.10: Estimated parameters of binary logistic regression models for telecommuters and commuters.

| Variables | Model 1: Telecommuter | | Model 2: Commuter | |
|---|---|---|---|---|
| | Estimate | z-value | Estimate | z-value |
| **Individual characteristics** | | | | |
| Age (base: Under 26) | | | | |
| 26-35 | 0.262 | 1.039 | 0.268 | 2.716*** |
| 36-50 | 0.197 | 0.834 | 0.334 | 3.564*** |
| 51-65 | 0.369 | 1.585 | 0.297 | 3.271*** |
| Above 65 | 0.222 | 0.901 | 0.397 | 3.064*** |
| Education attainment (base: Below bachelor's degree) | | | | |
| Some college or associate's degrees | 0.118 | 0.86 | 0.077 | 1.024 |
| Bachelor's degree or above | 0.505 | 3.829*** | 0.35 | 4.602*** |
| Part-time employee (base: Non-part-time) | 0.237 | 2.635*** | -0.115 | -1.726* |
| Travel day (base: Sunday) | | | | |
| Monday | -0.054 | -0.36 | 0.972 | 10.847*** |
| Tuesday | 0.305 | 1.926* | 1.363 | 13.525*** |
| Wednesday | 0.023 | 0.154 | 1.291 | 13.118*** |
| Thursday | 0.372 | 2.430** | 1.281 | 13.115*** |
| Friday | 0.012 | 0.084 | 1.184 | 12.484*** |
| Saturday | 0.151 | 1.001 | 0.165 | 2.234** |
| **Household characteristics** | | | | |
| Household annual income (base: Less than $24,999) | | | | |
| $25,000 to $99,999 | 0.187 | 1.443 | 0.2 | 2.337** |
| $100,000 to $199,999 | 0.128 | 0.925 | 0.284 | 3.046*** |
| $200,000 or more | 0.518 | 3.085*** | 0.283 | 2.449** |
| Household structure (base: 2+ adults, retired) | | | | |
| 1 adult | 0.734 | 4.591*** | 0.481 | 4.590*** |
| 2+ adults | 0.052 | 0.458 | 0.273 | 3.252*** |
| 1 adult, retired | 0.717 | 2.113** | 0.108 | 0.317 |
| 1 adult with people under 21 | 0.901 | 2.839*** | 0.584 | 3.301*** |
| 2+ adults with people under 21 | 0.484 | 3.484*** | 0.376 | 4.186*** |
| Residential setting (base: Rural) | | | | |
| Urban | 0.471 | 4.278*** | 0.293 | 3.642*** |
| Constant | -0.163 | -0.534 | 0.398 | 2.498** |
| **Model statistics** | | | | |
| Number of observations | 4168 | | 20280 | |
| Log-likelihood Restricted | -1976.5 | | -5790.6 | |
| Log-likelihood Unrestricted | -1911.3 | | -5476.6 | |
| Log-likelihood Ratio | 130.4***(df=22) | | 627.96***(df=22) | |
| McFadden's Pseudo $R^2$ | 0.033 | | 0.054 | |

Note: *$p$ <0.1; **$p$ <0.05; ***$p$ <0.01.

situation is the opposite for commuters. In terms of travel day (survey day), commuters are found more likely to leave home and make trips during weekdays as opposed to weekends. Meanwhile, commuters are most likely to stay at home on Sundays. When it comes to telecommuters, we do not observe the same trend and most of the coefficients are not significant indicating that telecommuters' decisions on making trips are largely not affected by the day of the week. This is reasonable as we discussed before that telecommuters do not need to commute to work resulting in a flexible daily schedule.

We turn now to the effects of household characteristics on telecommuters' and commuters' decisions on making trips. In general, higher-income commuters are more likely to make trips compared to lower-income commuters. Telecommuters with annual household income more than $200k are most likely to make trips compared to the other telecommuters with lower household incomes. As expected, both telecommuters and commuters who live in urban areas are more likely to leave home and make trips compared to people living in rural areas (presumably due to higher accessibility to opportunities but we turn to this in the conclusions). The effect of household structure indicates that compared to two-retiree households, telecommuters or commuters from a household with one adult and people under 21 are most likely to make trips, which is reasonable because they need to run errands for the younger members of their households. We also find that regardless of being telecommuters or commuters, single non-retired persons, and people from a household with two or more adults with people under 21 are more likely to leave home and make trips compared to two-retiree households. The differences between telecommuters and commuters are that telecommuters from a household with two adults are as likely as telecommuters from two-retiree households to make trips; however, commuters from a household with two adults are more likely to make trips compared to two-retiree households. Similarly, single retirees working as telecommuters are more likely to make trips compared to two-retiree households; while there is no significant difference between

commuters from single-retiree households and commuters from two-retiree households in terms of the propensity of making trips. The findings here point to the need for more in-depth household structure and task allocation information to tease out the motivation underlying the behavior described above.

## 2.5 Conclusions

### 2.5.1 Main findings

In this study, we show that using motif and sequence analysis in combination provides more insights into differences and commonalities between telecommuters and usual commuters. This study proposes to add these new methods to the toolbox of travel behavior analysis. In terms of substantive findings, we see here that telecommuters are by far more diverse in their time of day allocation of time to places, activities, and travel. First approximately 20% of the telecommuters stay at home all day during a workday, while only 8% of the commuters do the same. Telecommuters that have at least a trip during their workday travel more (in VMT and trips) than their counterpart commuters but travel less driving alone. In addition, telecommuters have by far greater variety in their motifs with more complex tour formation and a distinct time of day allocation to activities and travel from the commuters. Within telecommuters and commuters, however, we have substantial variation in activity participation and travel. Commuters show a substantial proportion in routine morning and afternoon peaks of arriving at and departing from work. In addition, telecommuters do not perform work tasks only from home. Instead, during the day they visit a variety of locations to visit customers and/or use their spatiotemporal flexibility to perform work tasks anywhere they want. In addition, a higher proportion of telecommuters function as designated drivers escorting other people

to their activity locations.

In the comparison between people who stay at the same place (mostly home) all day and people who make at least a trip, we find that telecommuters and commuters who have higher educational attainment (bachelor's degree or above), higher household income, living in urban areas, or living with people younger than 21 are less likely to be homebound. Commuters under 26 years old are least likely to leave home and make trips compared to the other age groups, while age group membership does not significantly impact telecommuters' decision to leave home on the interview day. The bottom line is that commuters are more likely to make trips on weekdays and stay at home on weekends, while telecommuters do not have such constraints and are more flexible.

### 2.5.2 Policy implications

Telecommuting is an official policy in California and appears in many regional plans and sustainability strategies as defined by California legislation. Unlike the conventional definition of telecommuting in the late 1970s, telecommuting nowadays as verified in this study is no longer a home-based work arrangement, but is multiple-place-based. With the advancement of ICTs, the emergence of ubiquitous workplaces allows telecommuters to work not only at home but at many places with available internet connection such as a coffee shop, a bookstore, a restaurant, and so forth. As confirmed in this chapter using the 2017 California-NHTS as a case study, telecommuting has already a positive impact on transportation with a lower number of trips and number of miles driving alone, less travel during peak hours, and enables escorting people to places of activity. We also find substantial heterogeneity in daily mobility patterns and time allocation to activities and travel among telecommuters. We can view the higher variety of scheduling arrangements and visiting places among telecommuters as a positive impact because the internet is

ubiquitous and enables visiting more diverse places. But we can also view this as a not-so-positive impact when the added freedom of telecommuters may motivate them to circulate for a longer time in the network and possibly contribute to emissions. It is unknown, however, if they do this because the type of vehicle used plays a major role in estimating emissions. Moreover, also unknown for the telecommuters who did not leave home at all during the day of the interview whether this was done voluntarily or by necessity.

The mobility patterns and daily schedules will be most likely dissimilar in different settings due to national, cultural, policy, and infrastructure differences. For example, as [102] discussed in the late 90s, this is due to evolving telework needs supportive social norms and/or legal protection of the workers. One could imagine many potential differences in telecommuters' daily schedules between those from well-developed Western countries where telecommuting has been a mature and relatively popular practice and those from developing countries where telecommuting is still new and less popular. Transportation infrastructure also has important impacts on telecommuters' daily mobility patterns as well as daily schedules of activities and travel. People living in California mostly rely on automobiles than other transportation modes, which is confirmed in Table 2.2 with both telecommuters and commuters driving alone the most and telecommuters who made at least one trip travel 1.37 more VMT as well as 0.53 more trips in a day compared to commuters. However, we anticipate the results will be different in other cities where public transit is more developed, such as Washington D.C., Boston, and Chicago. In addition, it is also important to take into account land-use types, community design, and infrastructure accessibility in investigations of the daily mobility of telecommuters. Telecommuters living in neighborhoods where grocery stores, restaurants, gyms, and other types of activity opportunities are easy to access are presumably less likely to be selected over driving alone to activity opportunities with longer

distances. Telecommuters living in rural areas with low accessibility to places are more likely to visit multiple places to fulfill their daily needs and drive longer distances. All in all, when one studies telecommuters' daily mobility patterns and daily schedules should always take into account the local cultural, policy, and infrastructure characteristics in the study area to better explain the observed behavior and data analysis findings. A promising future work for this study is to correlate the various mobility patterns and time allocation patterns with built environment characteristics.

The unexpected global event of the COVID-19 pandemic forced people to stay at home and reduce traveling to contain the spread of the virus. People adjusted to the new norm of daily life with COVID-19 [103]. In fact, we expect a substantial increase in the adoption of telecommuting by companies as it is happening at the time of writing this article and we also anticipate a large number of telecommuters will switch to motif Group 1 pattern, which is a single location-based (mostly home) loop trips pattern. Telecommuters using motif Group 1 will most likely work from home but take a walk during the workdays during the COVID-19 pandemic. In addition, we will see a substantial decrease in the telecommuters' patterns comprising a large period being present at workplaces, such as telecommuters' Work and Run Errands Day, and Mostly out of Home Day. Meanwhile, telecommuters will be less likely to use complex motif groups such as motif Groups 4 and 5 as they work from home most of the time and have less time left for other activities and also out of the concerns of reducing the risk of infection by traveling less. We also anticipate that different sociodemographic groups will react in different ways and occupation that already plays a major role in defining daily commuting schedules will continue driving COVID-19-induced telecommuting. However, longitudinal data analysis as in MOBIS-COVID19 [2] shows that we may experience a return to the use of the private automobile and maybe a tendency to return to pre-COVID patterns.

[2] https://ivtmobis.ethz.ch/mobis/covid19/en/

### 2.5.3 Contributions and limitations

Our contributions are fourfold. First, our study advances people's understanding of telecommuter's travel behavior nowadays. For example, contrary to conventional belief, telecommuters in California who have at least a trip during their workdays travel 1.37 more VMT and 0.53 more trips than their counterpart commuters. In addition, telecommuters tend to use carpooling more than driving alone. Second, this study reveals the substantial heterogeneity in daily mobility patterns and time allocation patterns for telecommuters and commuters. Third, armed with this new method combining motif and sequence analysis, the self-selection bias discussed in the literature [86, 104, 66] can be handled using the patterns described here in a way that accounts for behavioral heterogeneity in a more insightful and behaviorally informative way. Fourth, we challenge the crisp distinction between commuters and telecommuters that should be replaced by different ways of work from many different places accounting for the work task flexibility enabled by information and communication technology. This complements the [105] and [106] different ways of telecommuting, but now it can be done using observed patterns that classify respondents in statistically derived groups based on time allocated to work, home, household responsibilities, shopping and serving activities, and travel.

Several limitations need further investigation. First, unlike a weekly travel diary, the data used in this study, 2017 California-NHTS, contains only a one-day travel diary for a respondent. Some of the respondents might not follow their regular workday schedule due to unexpected reasons such as sick leave, business travel, personal vacation, emergency, etc. Granted that there is little to do about this bias, a promising future study can apply a longitudinal survey to study the recurrent mobility patterns and to further detect anomaly travel patterns at the individual level. Second, an immediate next step is to apply the motif and sequence analysis to the excluded non-workday samples to fur-

ther compare the differences in daily mobility patterns of telecommuters and commuters between workdays and non-workdays. Third, to extend the intra-household analysis in more depth, we plan to leverage the more detailed information in the survey regarding the intra-household interactions of daily activity to understand how a household member's daily mobility pattern can influence other household members' mobility on the same day. We also plan to look into the large percentage of employees who do not go to work or work from home during non-holiday weekdays and to study how this correlates with the activity and travel patterns of other people in the same household.

# Chapter 3

# A Time-geographic Approach for Human Interaction Analysis using Movement Data

| Peer-reviewed Publication | |
|---|---|
| Title | A classification framework and computational methods for human interaction analysis using movement data |
| Authors | Rongxiang Su, Somayeh Dodge, Konstadinos Goulias |
| Venue | Transactions in GIS |
| Editors | John P. Wilson |
| Publisher | Wiley |
| Pages | Volume 26, Issue 4, 1665-1682 |
| Submitted Date | 17 January 2022 |
| Accepted Date | 9 May 2022 |
| Publication Date | 30 May 2022 |
| Copyright | Reprinted with permission from Wiley |
| DOI | `https://doi.org/10.1111/tgis.12960` |

| Peer-reviewed Publication | |
|---|---|
| Title | Understanding the impact of temporal scale on human movement analytics |
| Authors | Rongxiang Su, Somayeh Dodge, Konstadinos Goulias |
| Venue | Journal of Geographical Systems |
| Guest editors | Michael F. Goodchild, Daoqin Tong, WenWen Li |
| Publisher | Springer Nature |
| Pages | 24, 353–388 |
| Submitted Date | 23 April 2021 |
| Accepted Date | 21 December 2021 |
| Publication Date | 08 February 2022 |
| Copyright | Reprinted with permission from Springer Nature |
| DOI | https://doi.org/10.1007/s10109-021-00370-6 |


**Abstract** Interaction between moving individuals is a critical factor in shaping urban dynamics and human social networks. Recent advancements in trajectory analytics have resulted in promising methods to identify and extract spatio-temporal patterns of interaction using movement tracking data. However, methodologies to quantify the duration of interaction remain limited. In the present work, we advance the existing time-geographic-based approach that relies on potential path area computation and polygon intersection to quantify the duration of potential concurrent interactions between mobile individuals. In addition, this research is the first attempt toward developing a taxonomy of human interaction in space and time that are observable from movement data. In the proposed framework, interaction between moving individuals is divided into three categories based on the time when it occurs and its duration, including encounter, concurrent interaction, and delayed interaction. This article incorporates an evaluation of the proposed time-geographic interaction analysis method to quantify the duration of interaction between moving individuals using human-generated GPS tracking data in California. The evaluation involves a comparison with the traditional proximity-based approach, commonly utilized in digital contact tracing technologies, which relies on the intersection of spatial and temporal buffers of two individuals. The results show that

the proposed time-geographic-based approach can measure the duration of interaction more accurately, whereas the proximity-based approach underestimates the duration of interactions. The chapter provides further insights into the impact of temporal scale on human interaction analysis through a systematic evaluative experiment.

## 3.1 Introduction

Interaction between moving individuals is a critical factor in shaping urban dynamics and human social networks [50, 11, 10, 48]. Identifying and quantifying dynamic interactions between moving individuals attracts scholars from GIScience, computer science, movement ecology, epidemiology, and related disciplines [10, 107]. Recent advancements in trajectory analytics have created methods to identify and extract spatiotemporal patterns of interaction using movement tracking data [108]. Many of the existing tools for interaction analysis, including Bluetooth-based digital contact tracing technologies, rely on the spatial proximity between two individuals and require predefined spatial and temporal thresholds [10, 53]. These proximity-based approaches can be limited when high-resolution data are not available or when moving individuals are not tracked simultaneously. Simultaneous tracking of multiple individuals usually is hard to achieve due to limited battery life and capacity concerns [109].

In addition to proximity-based approaches, several studies leveraged time geography framework [14] to enhance interaction analytics [50, 49, 48]. Specifically, these approaches model the accessible area of individuals giving each pair of successive GPS fixes along their movement paths using the concept of *Potential Path Area* (PPA)[17, 16, 15]. Subsequently, the intersection of two PPAs of different individuals identifies the joint-accessibility space where potential interaction between individuals is possible. ORTEGA (Object-oRiented TimE-Geographic Analytical approach) is one of these newly emerged

time-geographic-based approaches (or called PPA-based approaches) for identifying various types of interactions at a reasonable computation cost [50]. Existing efforts indicate that ORTEGA or in general, the PPA-based approaches can well identify potential interactions when individuals are not tracked simultaneously [50]. Moreover, the PPA-based approaches are capable of identifying delayed interactions when individuals visit the same locations asynchronously with a time lag [50, 49].

While existing approaches have advanced our ability to identify and trace patterns of interaction in movement data, they fall short in estimating the duration of interaction. Duration of interaction is a critical factor in differentiating different types of human interactions. For instance, in the context of contact tracing for COVID-19, it is necessary to distinguish between brief-duration interactions and long-duration ones to enable different levels of sensitivity and specificity in terms of detection of contacts [110, 111]. In addition, definitions and classifications of interaction between moving individuals are isolated (see our literature review later). Existing classifications mainly consider the occurrence of interaction in space and time [112, 50, 49] but do not incorporate the role of duration of interaction and the context in which interaction happens.

This research addresses these gaps by first introducing a new technique to estimate the duration of interaction using movement trajectories of interacting individuals. The proposed method builds upon the theoretical foundations of ORTEGA, initially presented in [50]. Specifically, our work advances the time-geographic-based interaction analytical approaches to trace and compute the duration of concurrent contacts (i.e. synchronous interaction in space and time) between mobile individuals. Secondly, this chapter establishes a new classification framework for human interaction, highlighting the crucial role of interaction duration and the spatiotemporal properties inherent in human interactions. The proposed framework classifies human interactions based on the *concurrence* of overlapping or intersecting individuals' movement paths and their *dura-*

*tions*, which lays the foundation for building a more comprehensive taxonomy of human movement interactions. The proposed method extending based on ORTEGA for quantifying the concurrency, delay, and duration of interaction facilitates the differentiation among different types of interactions introduced in the proposed classification framework. Human-generated GPS tracking data in California are used to assess the efficacy of the new method for the quantification and identification of different types of interactions based on the proposed classification framework.

## 3.2 Background

### 3.2.1 Analytical Methods for Human Interaction

In time geography, the activity space of a moving entity can be measured by a space-time prism [14] which is shaped by a pair of origin and destination locations (denoted by $P_i(x_i, y_i, t_i)$ and $P_j(x_j, y_j, t_j)$), a time budget $(t_j - t_i)$, and the maximum speed capacity $(v_{max})$, as shown in Figure 3.1. The projection of a space-time prism onto a two-dimensional Euclidean space is called the potential path area. PPA delimits the accessible locations to the moving entity over a given time budget (for details about the definition and formalization of PPA, the readers are referred to [113, 17]). When applied on consecutive GPS observations along the trajectory of an object, the PPA models a potential area that the entity could reach over the duration between the GPS recordings. Hence, it can consider the uncertainty of positioning and gaps in movement data for interaction analysis.

In this chapter, the term *interaction* refers to potential contacts between individuals in space and time. Such contact does not necessarily indicate the occurrence of physical or social interaction, but it may lead to exposure to risk factors or opportunities

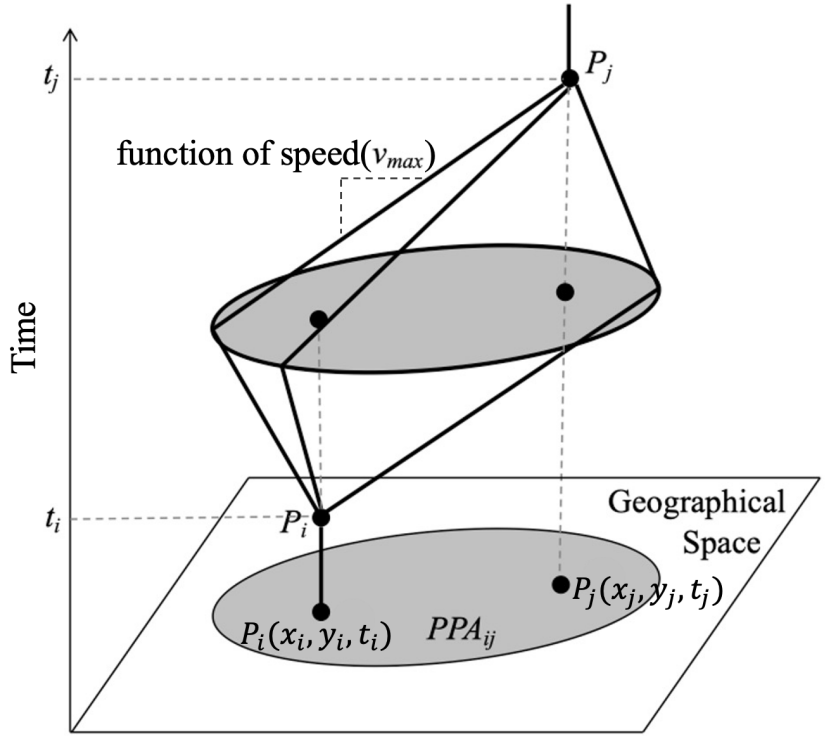



Figure 3.1: Illustration of the space–time prism and potential path area in a 3D space. $P_i(x_i, y_i, t_i)$ and $P_j(x_j, y_j, t_j)$ are the two consecutive tracking points. $v_{max}$ is the maximum speed capacity of the moving entity at the time interval $[t_i, t_j]$ given the two fixed locations $P_i$ and $P_j$. $PPA_{ij}$ denotes the potential path area in a 2D Euclidean space between $P_i$ and $P_j$ (modified from [17]).

for social interaction. Recent studies suggest that the potential interactions between moving entities can be identified by intersecting their PPA ellipses and/or space-time prisms along their trajectories [47, 48, 49, 50]. In general, existing time-geographic-based interaction analytical approaches can be classified as either deterministic or probabilistic approaches. Deterministic time geography assumes movement probabilities are uniform within a PPA or a space-time prism. This assumption makes computation easy, but it may not be reliable when fine-resolution tracking data are not available. It is not plausible to assume a moving entity will present at any location uniformly in a PPA/space-time prism if the movement was tracked at a coarse temporal resolution. This is because the derived PPA/space-time prism may be very large, and intersecting them directly

will over-estimate interaction [114]. Due to its computation efficiency and simplicity, researchers have developed several deterministic approaches for interaction analysis between multiple moving entities. For example, [51, 52] proposed alibi query algorithms based on intersecting network-based space-time prisms to determine whether two moving entities could have possibly met. [48] developed a joint Potential Path Area (jPPA) approach for mapping areas of spatial-temporal overlap from wildlife tracking data. [49] extended the jPPA approach to capture potential delayed interactions that occur between moving entities with a time lag.

Compared with deterministic time-geographic-based interaction analysis, probabilistic interaction analysis is less evolved in the literature. Probabilistic interaction analysis relies on probabilistic time geography [115]. In essence, probabilistic time geography models visit probabilities within a PPA/space-time prism based on various simulation models, such as Random Walk, Brownian Bridges, and kernel density estimation [115, 116, 117, 118]. The application of probabilistic time geography on interaction analysis is limited. [47] developed a voxel-based probabilistic interaction analytical approach by intersecting discretized space-time prisms of two or more entities. However, this approach still requires heavy computation and the predicted interaction probabilities strongly rely on many user-defined parameters such as the spatial resolution of the voxel.

Despite the extensive work on computational methods for movement interaction analysis as described in this section, there remains a research gap to quantify the duration of interaction between moving individuals. The present research fills this gap by advancing ORTEGA to estimate the duration of potential concurrent interactions between mobile individuals.

### 3.2.2 Taxonomy of Human Interactions

It is essential to establish a comprehensive taxonomy encompassing the definition and classification of human interaction, as it is fundamental for formalization of the interaction patterns and their properties to inform future machine learning and movement analytics techniques. In the movement ecology domain, [112] argues that interaction can be *static* or *dynamic* where *static interaction* indicates that individuals' home ranges (or activity space utilization) intersect spatially but not necessarily temporally while *dynamic interaction* occurs when animals move in close spatial proximity at the same time.

In human movement, dynamic interaction plays a key role in forming a wide range of phenomena like human social behavior, urban land use dynamics, virus transmissions through human contact, to name a few [50, 11, 119]. However, dynamic interaction is usually represented by aggregate metrics (e.g., average distance over time, or the fraction of proximal locations) in the existing literature to reflect association or the interdependency in the movement of two individuals [10, 107, 53] and this classification (i.e., static versus dynamic) is limited in capturing temporal characteristics of interaction, such as the duration or lag of interaction. To reflect this property, others suggest that human interaction can be classified as *concurrent (or direct/synchronous) interaction* or *delayed (or indirect/asynchronous) interaction* [50]. Aside from dynamic interaction, this classification focuses on how paths of individuals intersect in space and time which is a driving factor for social activities and urban dynamics [120, 18, 121, 122]. Concurrent interaction indicates that individuals move in spatial proximity of each other in a shared space and at the same time, while delayed interaction occurs when individuals visit the same locations at different times. This classification captures the spatiotemporal properties of human interaction but does not reflect the duration or nature of interaction, which is

critical information for quantifying and contextualizing human interaction. Duration of interaction is a critical factor in differentiating different types of human interactions. For instance, in the context of contact tracing for COVID-19, it is necessary to distinguish between brief-duration interactions and long-duration ones to enable different levels of sensitivity and specificity in terms of detection of contacts [110, 111]. A recent study on wildlife interaction analysis suggests considering interactions between individuals not only as isolated spatiotemporal co-occurrence behavior but as continuous events, which they term contact phases, as interactions are often clustered in time [108]. However, the idea has not been fully incorporated in human movement analysis literature and techniques to quantify the duration of interaction are limited.

To sum up, existing classifications mainly consider the occurrence of interaction in space and time but do not incorporate the role of duration of interaction and the context in which interaction happens. The present research fills this gap by developing a classification framework of human interaction that emphasizes the critical role of duration of interaction. The main contribution of this article is a new classification framework for interaction between moving individuals in human movement data, which lays the foundation for building a more comprehensive taxonomy of human movement interaction.

## 3.3 A New Classification Framework for Human Interactions

Although the taxonomy of animal movement interactions has been discussed widely among researchers (see for example in [10, 107, 53]), many of these ideas may not be interchangeable for human movement interaction analysis. For instance, interaction analysis for animals does not necessarily require animals to be present in the same space because

they can sense each other at a distance or avoid one another. However, the majority of meaningful human interactions through movement occur when individuals are present in a shared space at the same or different times. It is noteworthy that the focus of this study is on human interactions in shared physical spaces that are observable using movement data. In this section, we create a classification framework for human interactions through movement.

Human interaction may occur when individuals stay close in space and time, or when their movement paths intersect, but not necessarily at the same time. This space-time proximity may lead to an intentional interaction (social or not), or it can simply be an incidental occurrence happening by chance and without any intention. While movement data help in tracing proximity of individuals in space and time, the spatial proximity of recorded locations may not be sufficient to identify whether the interaction is intentional (social, confrontational, etc.) or non-intentional. In addition to location, time is a critical factor in further identifying the type of interaction. For instance, brief duration interactions such as encounters can reflect different natures of human behavior compared to long duration interactions which often involve shared activities. Therefore, the classification incorporates information on the lag or concurrency of overlapping or intersecting people's movement paths and the duration over which individuals present in shared spaces synchronously or asynchronously. Although some interactions may occur when individuals stay apart from each other (e.g., synchronous movements of two individuals in different spaces can reflect the association between their movements), it can be difficult to explain the meaning and to measure the duration of this type of interaction simply by relying on movement data without any ancillary or contextual information. In the proposed classification, we use both spatial and temporal information to distinguish between different types of interactions that are identifiable in movement data: *encounter, concurrent interaction, and delayed interaction.*

- An *encounter* occurs when individuals stay in spatial proximity in space and time for a brief duration (which can be specified depending on different applications and activities).

- A *concurrent interaction* occurs when individuals move synchronously in spatial proximity of each other in a shared space and at the same time with a longer duration compared to encounters.

- A *delayed interaction* happens when individuals visit the same locations in space asynchronously with a time lag.

Human interactions can be either intentional or incidental. An intentional interaction indicates that interacting individuals know each other (i.e., within their social networks) or a planned interaction involving strangers or individuals with a certain functional connection (i.e. teacher/student, co-workers, etc.); otherwise, it is an incidental interaction occurring by chance. Although the duration and frequency of contacts between people computed from movement data help to distinguish incidental versus intentional interactions, the identification can be further enhanced using auxiliary information. By incorporating contextual variables such as individual sociodemographic characteristics, movement mode, behavior, and geographic and environmental variables, it may be possible to further distinguish between intentional and incidental interactions. Unlike intentional interactions that can be verifiable by incorporating auxiliary data, incidental interactions are more difficult to verify as individuals do not realize they interact with each other (e.g., strangers pass each other on the street or use a shared space) and almost no data can certainly verify this type of interaction.

Examples of encounters can be a person picking up an order from a coffee shop, a person passing a stranger on the street, and strangers waiting for buses at a bus stop, just to name a few. Encounters can play a crucial role in disease transmission such as

COVID-19 which spreads via airborne particles and droplets that can stay in the air for some time [123]. Existing studies leverage movement data collected by location awareness technologies such as mobile phones to establish social networks of human interactions or contacts in the context of COVID-19 spread [124]. Researchers have focused on human movement interactions at different spatial and/or temporal granularity ranging from close face-to-face encounters [125], interactions within an accessible geographical area [50], to aggregate mobility flows or metrics at local, regional, or national levels [126, 127, 128, 129].

In terms of concurrent interactions, examples can be people traveling together or people engaging in the same activities for a certain amount of time, strangers moving together on the same road incidentally, etc. On the other hand, delayed interactions can reflect the behavior of people who visit the same place or are exposed to the same space at different times. This type of interaction may reveal a similar spatial behavior between individuals. For example, an intentional delayed interaction can be a marathon athlete left behind by another athlete for 30 min, or a person leaving their shift at a workspace and a colleague takes on the next shift a few minutes later. If two strangers visit the same place (e.g., a grocery store or a gym) or share a common movement path at different times (within a reasonable time lag) unintentionally, it is not considered as an intentional delayed interaction but rather an incidental delayed interaction. These types of delayed interactions may reveal similarities in spatial behavior of people or homophily in certain neighborhoods and activities.

## 3.4 Methods: Identifying Interaction Patterns

Following [130], we change the space-time prism by varying $v_{max}$ at different time intervals instead of using a constant value for the whole trajectory. Specifically, the

parameter $v_{max}$ for each time interval is the floating average of speed over an exponential kernel [131].

As shown in Figure 3.2, the intersection of two PPAs along trajectories of different individuals identifies the joint-accessibility space where potential interaction between individuals is possible [16, 50, 49, 48]. It is noteworthy that the area of intersections between two PPAs delimits locations for possible interactions but it does not certify the occurrence of interactions. Formally, a trajectory of a moving entity $e$ can be formalized as a series of $T_e = \{(x_0^e, y_0^e, t_0^e), (x_1^e, y_1^e, t_1^e), ..., (x_i^e, y_i^e, t_i^e), ..., (x_n^e, y_n^e, t_n^e)\}$, where $x_i^e$ and $y_i^e$ denote the geographical location of $e$ at the time $t_i^e$. Subsequently, the PPA between each pair of consecutive tracking points $(x_i^e, y_i^e, t_i^e)$ and $(x_{i+1}^e, y_{i+1}^e, t_{i+1}^e)$ of moving entity $e$ can be denoted by $PPA_{i,i+1}^e$.

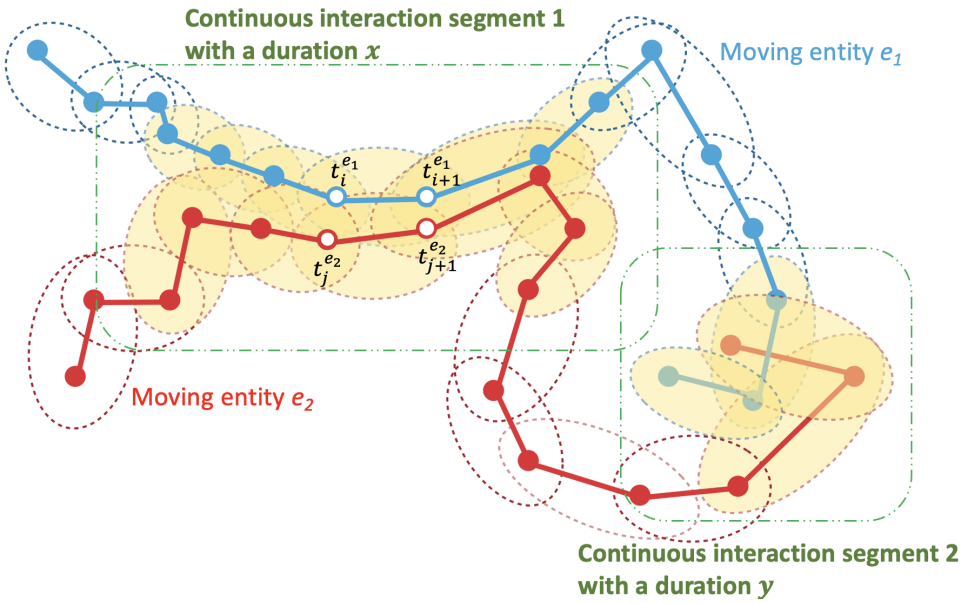



Figure 3.2: Illustration of PPA intersections for identifying potential interaction between two moving entities. The intersecting PPAs are shaded in yellow. The two green dashed rectangles highlight two continuous interaction segments.

Let $e_1$ and $e_2$ denote two moving entities and $PPA_{i,i+1}^{e_1}$ and $PPA_{j,j+1}^{e_2}$ denote the PPA of each of the two moving entities respectively. The potential interaction between $e_1$ and

$e_2$ can be formalized as follows (revised from [50]):

1. If $PPA^{e_1}_{i,i+1} \cap PPA^{e_2}_{j,j+1} = \phi$ (i.e., the two PPAs spatially not intersect), there is no potential interaction between the two moving entities.

2. If $[t^{e_1}_i - \tau, t^{e_1}_{i+1} + \tau] \cap [t^{e_2}_j, t^{e_2}_{j+1}] \neq \phi$ or $[t^{e_1}_i, t^{e_1}_{i+1}] \cap [t^{e_2}_j - \tau, t^{e_2}_{j+1} + \tau] \neq \phi$ (i.e., the two PPAs temporally overlap) and $PPA^{e_1}_{i,i+1} \cap PPA^{e_2}_{j,j+1} \neq \phi$ (the two PPAs spatially intersect), a potential encounter or a concurrent interaction can be identified. $\tau$ is a predefined allowable short *time window* that relaxes the requirement for exactly synchronous start and end times of the two PPAs for intersections.

3. If $[t^{e_1}_i - \tau^a, t^{e_1}_{i+1} + \tau^a] \cap [t^{e_2}_j, t^{e_2}_{j+1}] \neq \phi$ or $[t^{e_1}_i, t^{e_1}_{i+1}] \cap [t^{e_2}_j - \tau^a, t^{e_2}_{j+1} + \tau^a] \neq \phi$ (the two PPAs temporally overlap with a lag of $\tau^a$), and $[t^{e_1}_i - \tau^b, t^{e_1}_{i+1} + \tau^b] \cap [t^{e_2}_j, t^{e_2}_{j+1}] \neq \phi$ or $[t^{e_1}_i, t^{e_1}_{i+1}] \cap [t^{e_2}_j - \tau^b, t^{e_2}_{j+1} + \tau^b] \neq \phi$ (the two PPAs temporally overlap with a lag of $\tau^b$, where $\tau^b > \tau^a > \tau$), and $PPA^{e_1}_{i,i+1} \cap PPA^{e_2}_{j,j+1} \neq \phi$ (the two PPAs spatially intersect), a potential delayed interaction that occur with a time lag falling within the specified *time interval* $[\tau^a, \tau^b]$ is identified.

The equations above indicate that if two PPAs intersect spatially and their time intervals overlap within the allowable small *time window* $\tau$, potential encounters or concurrent interactions may occur between the two moving entities (we will discuss later how to distinguish between encounters and concurrent interactions). Algorithm 1 presents the details for identifying concurrently intersecting PPAs between two moving individuals. The parameter $\tau$ is essential because, in reality, the movement tracking of two individuals is often unsynchronized or collected at different sampling rates. The parameter is domain-specific and usually can be the same as the average temporal resolution of the input movement data. It can also be set to zero for a strict synchronous alignment, especially when high-resolution data are available. As an example, for human interaction

analysis using movement data collected at 1 min sampling interval, a time window of 5 min may be reasonable to identify potential concurrent interactions as 5 min might not be long enough for an interaction to be considered delayed [50]. However, if the goal is to identify encounters, the time window parameter should be rather small as encounters can only happen when individuals interact synchronously in time.

To detect delayed interactions, the user needs to specify a *time interval* $[\tau^a, \tau^b]$ so that interactions that occur with a time lag falling within the specified range of $\tau^a$ to $\tau^b$ can be identified. Both $\tau^a$ and $\tau^b$ should be greater than $\tau$ and can be determined according to the study's objectives. The algorithm for identifying intersecting PPAs with a time lag (i.e. delayed interactions) is not presented here, as it closely resembles Algorithm 1. The only modification required is to replace the conditional statement that assesses whether two PPAs intersect with a time lag.

**Algorithm 1:** Identifying concurrently intersecting PPAs between two moving entities

**Input:** $PPA^{e_1}$,$PPA^{e_2}$: two lists of PPAs of two moving entities $e_1$ and $e_2$;
$\tau > 0$: allowable short time window parameter.
**Output:** $PPA_{intersect}$: a list of spatially and temporally intersecting PPAs between $e_1$ and $e_2$ within a short time window $\tau$.

```
L^{e1} ← length of the PPA^{e1} list;
L^{e2} ← length of the PPA^{e2} list;
for i = 0 to L^{e1} − 1 do
    for j = 0 to L^{e2} − 1 do
        if [t_i^{e1} − τ, t_{i+1}^{e1} + τ] ∩ [t_j^{e2}, t_{j+1}^{e2}] ≠ φ or [t_i^{e1}, t_{i+1}^{e1}] ∩ [t_j^{e2} − τ, t_{j+1}^{e2} + τ] ≠ φ then
            if PPA_{i,i+1}^{e1} ∩ PPA_{j,j+1}^{e2} ≠ φ then
                PPA_intersect.insert([PPA_{i,i+1}^{e1}, PPA_{j,j+1}^{e2}])
            end
        end
    end
end
return L
```

Encounters are similar to concurrent interactions as they both happen between moving individuals synchronously in time. While encounters only last for a short duration

(e.g. less than a few minutes), concurrent interactions usually last longer. As such, it is necessary to quantify the duration of interactions to distinguish between encounters and concurrent interactions. To do this, we apply a new approach, to quantify the duration of interactions using a time-geographic-based interaction analysis method ("the PPA-based approach" herein) as shown in Algorithm 2.

**Algorithm 2:** Quantifying the duration of concurrent interactions

**Input:** $PPA_{intersect}$: a list of spatially and temporally intersecting PPAs between $e_1$ and $e_2$ within a time window $\tau$.
**Output:** $\Psi^{e_1,e_2}$: a list of the start time, end time, and duration of the continuous interaction segments between $e_1$ and $e_2$.
Let $\Theta^{e_1,e_2}$ and $\Psi^{e_1,e_2}$ be new arrays
$l \leftarrow$ length of the $PPA_{intersect}$ list
**for** $k = 0$ *to* $l - 1$ **do**
  **if** $t_S^{e_1}(k) == t_S^{e_1}(k+1)$ **then**
    $\Theta^{e_1,e_2}.insert[t_S^{e_1}(k), t_E^{e_1}(k), t_S^{e_1}(k+1), t_S^{e_2}(k), t_E^{e_2}(k), t_S^{e_2}(k+1), t_E^{e_2}(k+1)]$
  **end**
  **else if** $t_E^{e_1}(k) == t_S^{e_1}(k+1)$ **then**
    $\Theta^{e_1,e_2}.insert[t_S^{e_1}(k), t_E^{e_1}(k), t_S^{e_1}(k+1), t_S^{e_2}(k), t_E^{e_2}(k), t_S^{e_2}(k+1), t_E^{e_2}(k+1)]$
  **end**
  **else**
    **if** $|\Theta^{e_1,e_2}| == 0$ **then**
      $\Theta^{e_1,e_2}.insert[t_S^{e_1}(k), t_E^{e_1}(k), t_S^{e_2}(k), t_E^{e_2}(k)]$
    **end**
    $starttime \leftarrow min\{\Theta^{e_1,e_2}\}$
    $endtime \leftarrow max\{\Theta^{e_1,e_2}\}$
    $duration \leftarrow max\{\Theta^{e_1,e_2}\} - min\{\Theta^{e_1,e_2}\}$
    $\Psi^{e_1,e_2}.insert([starttime, endtime, duration])$
    $\Theta^{e_1,e_2} \leftarrow$ empty list
  **end**
**end**
**if** $|\Theta^{e_1,e_2}| != 0$ **then**
  $starttime \leftarrow min\{\Theta^{e_1,e_2}\}$
  $endtime \leftarrow max\{\Theta^{e_1,e_2}\}$
  $duration \leftarrow max\{\Theta^{e_1,e_2}\} - min\{\Theta^{e_1,e_2}\}$
  $\Psi^{e_1,e_2}.insert([starttime, endtime, duration])$
**end**
**return** $\Psi^{e_1,e_2}$

Specifically, after applying PPA intersection along the movement trajectories of the individuals, a list of spatially and temporally intersecting PPAs (e.g., all yellow ellipses

in Figure 3.2) denoted by $PPA_{intersect} = \{[PPA^{e_1}(1), PPA^{e_2}(1)], [PPA^{e_1}(2), PPA^{e_2}(2)],$ ...,$[PPA^{e_1}(k), PPA^{e_2}(k)]$, ..., $[PPA^{e_1}(l), PPA^{e_2}(l)]\}$ will be identified following the above specifications. Each element in $PPA_{intersect}$ indicates a pair of intersecting PPAs between $e_1$ and $e_2$. Note that the index $k$ here denotes the sequential order in the list of intersecting PPAs, but not the sequential order in the original PPA sequence of an individual. Let $[t_S^{e_1}(k), t_E^{e_1}(k)]$ and $[t_S^{e_2}(k), t_E^{e_2}(k)]$ denote the time intervals of the two PPAs of the $k$th intersection between $e_1$ and $e_2$ (i.e., $PPA^{e_1}(k)$ and $PPA^{e_2}(k)$).

By traversing every pair of intersecting PPAs based on their sequential order in the list of $PPA_{intersect}$, we can identify subsequences of continuously intersecting PPAs between the two moving entities (e.g., the portion or PPA intersections highlighted by the two green dashed rectangles in Figure 3.2) and their durations. We term this subsequence as a *continuous interaction segment*. It is also termed as *contact phases*[108]. The interaction duration is computed by taking the difference between a continuous interaction segment's start and end times. Eventually, for each continuous interaction segment, if the duration is smaller than $\delta_t$, where $\delta_t$ is a user-defined threshold parameter, an encounter is identified. Otherwise, it is classified as a concurrent interaction of a duration equal to the difference between the start and end times of the continuous intersected PPA sequence. The concept of duration may not apply to delayed interactions, as individuals are present in the same space but at different times.

The parameter $\delta_t$ must be small, as encounters usually refer to brief-duration close contacts. However, the parameter may be varied depending on different application scenarios. For instance, in Figure 3.2, the duration of the continuous interaction segment 1 exceeds the threshold value ($x > \delta_t$), thereby classifying it as a concurrent interaction. In contrast, segment 2, with a duration ($y$) shorter than $\delta_t$, is categorized as an encounter. Users should be aware of the impact of temporal granularity of the tracking data when deciding the parameter $\delta_t$. Setting the parameter to a time value lower than the tracking

frequency may overlook many potential encounters.

## 3.5 ORTEGA v1.0 Python Package

To advance the reproducibility and replicability of this research, we developed an open-source Python package named ORTEGA v1.0[1,2](from here on ORTEGA refers to ORTEGA version 1) for implementing the PPA-based method for identifying interaction patterns using movement data, as described in Section 3.4. The Python package ORTEGA is built on top of the existing theoretical developments of ORTEGA, originally introduced in [50]. ORTEGA is built upon existing Python's matplotlib [132], numpy [133], shapely [134], and pandas libraries. Table 3.1 describes ORTEGA's main functions. This section introduces the ORTEGA v1.0 Python package and provides step-by-step implementation instructions.

The ORTEGA package contributes two significant capabilities: 1) the functions to identify potential interactions (e.g., brief encounters, concurrent interactions, delayed interactions) from movement data of two or more entities using a time-geographic-based approach; and 2) the capacity to compute attributes of potential interaction events including start time, end time, interaction duration, and difference in movement parameters such as speed and moving direction, and also contextualize the identified potential interaction events. To the best of our knowledge, the present ORTEGA is the first open-source Python package that offers a time-geographic-based interaction analytical tool for movement interaction research.

[1] `https://github.com/move-ucsb/ORTEGA`
[2] `https://pypi.org/project/ortega`

Table 3.1: Description of ORTEGA's functions

| Function | Description |
|---|---|
| interaction_analysis() | Identifies spatiotemporal intersecting PPAs after the user specifies the short time window (for identifying concurrent interaction) or time interval (for delayed interaction) during the initialization step; computes the difference in speeds and movement directions for each intersecting PPA pair; and extracts continuously intersecting PPA segments (i.e., interaction events) if any exists. |
| compute_interaction_duration() | Computes the duration of each concurrent interaction event. This function only applies to concurrent interaction. |
| compute_ppa_perimeter() | Descriptive statistics of PPA ellipse's perimeter for each moving individual |
| compute_ppa_area() | Descriptive statistics of PPA ellipse's area for each moving individual |
| compute_ppa_interval() | Descriptive statistics of PPA time interval for each moving individual |
| compute_ppa_speed() | Descriptive statistics of PPA speed for each moving individual |
| attach_attributes() | Attaches attributes such as contextual variables to the dataframe of intersecting PPAs (derived from the function *interaction_analysis()*) for further correlation analysis |

### 3.5.1 Input data

ORTEGA v1.0 accepts conventional movement tracking data of a set of moving individuals, such as GPS or telemetry data, in the form of comma-separated values (CSV) files. The movement data should at least include a unique identifier for each individual (or entity), a set of coordinates (geographic coordinates in floating-point format, provided in separate columns) representing the movement observations of each entity, and the corresponding timestamp of each observation (in Python "datetime" format).

### 3.5.2 Setting up ORTEGA

As illustrated in Figure 3.3, ORTEGA first imports trajectories of two individuals, as provided by the user. To conduct interaction analysis for more than two individuals, the user will need to employ *For* loops, by considering a moving entity as a reference at each loop and running the analysis in conjunction with all other individuals in the data.

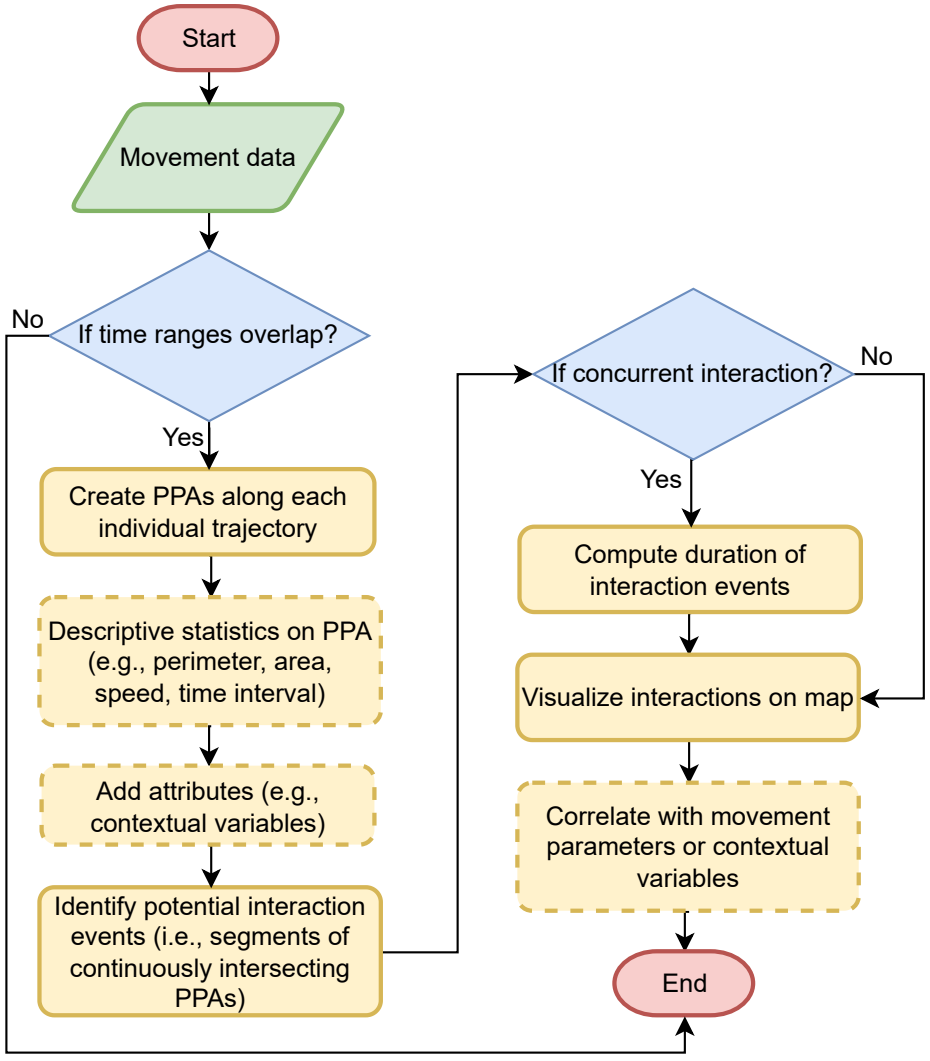



Figure 3.3: ORTEGA's workflow for analyzing interactions between a pair of moving individuals using movement tracking data (dashed polygons indicate optional steps).

As shown in the below code snippet in Figure 3.4, in the first step when initializing

an ORTEGA object, the user is required to specify the column names for the identifier, longitude and latitude coordinates, as well as the timestamp. As previously mentioned in Section 3.4, the user is also required to define a brief time window denoted as $\tau$ to detect concurrent interactions, or a time interval $[\tau^a, \tau^b]$ to identify desired delayed interactions. As an example, the code snippet in Figure 3.4 provides instructions on initializing ORTEGA to identify concurrent interaction. When the user provides a short time window (in minutes) as the value of the *minute_max_delay* parameter (i.e., in this case, *minute_max_delay* is considered as $\tau$), ORTEGA returns the results of potential concurrent interaction within that given brief time window. Alternatively, when both *minute_min_delay* and *minute_max_delay* are given (i.e., the user specifies a time interval $[\tau^a, \tau^b]$), ORTEGA detects delayed interactions that occur with a time lag within the specified interval (see the code in Figure 3.6). Otherwise, the program will raise an error to remind the user to specify at least the parameter *minute_max_delay*.

During the initialization phase, ORTEGA will check whether the temporal spans of the two movement trajectories intersect. Should at least a partial overlap occur, the program will proceed to compute and generate PPAs along each individual's trajectory; otherwise, it terminates with an error message. The output of the initialization phase is an instance of the ORTEGA class, which includes the original movement data and two lists containing the constructed PPAs of two moving individuals. Subsequently, the user can conduct descriptive statistics on PPA's perimeter, area, speed, and time interval using ORTEGA's descriptive statistics functions. Examples of implementing such statistics are shown in the same code snippet (see Figure 3.4).

```
import pandas as pd
import ortega

# Consider a scenario where the movement data is gathered at a 60 min
    interval for each GPS fix.
data1 = pd.read_csv('example.csv', sep=',',header=0)
# An example of a 60-minute time window to detect concurrent
    interactions. The brief allowable time window for interaction to be
    considered concurrent is normally set as the temporal resolution of
    the data (in this case, 60 min).
con_interaction = ortega.ORTEGA(data=data1,
                                minute_max_delay=60,
                                latitude_field = 'latitude_column',
                                longitude_field = 'longitude_column',
                                time_field='time_cloumn',
                                id_field='id_column')

# Should the time ranges of the two individuals not overlap, the user
    will encounter a ValueError: "Skipping pair {self.id1} and {self.id2
    } due to time lag greater than {self.minute_max_delay}!"

con_interaction.compute_ppa_perimeter()
con_interaction.compute_ppa_area()
con_interaction.compute_ppa_interval()
con_interaction.compute_ppa_speed()
```

Figure 3.4: Initialize ORTEGA and conduct descriptive statistics

### 3.5.3 Identify concurrent interaction

Next, the user can identify segments of continuously intersecting PPAs as potential interaction events using the function *interaction_analysis*(), as demonstrated in the code snippet in Figure 3.5. The outcome is a dataframe including the start and end times of each intersecting segment, along with the time difference between the two intersecting segments. The column of time difference indicates the time lag (if exists) between the two start points of the two intersecting trajectory segments. This information is especially helpful when the tracking of two individuals is not perfectly synchronized. It enables precisely pinpointing the location where the initial intersection or overlap of two trajectories occurred.

```
con_results = con_interaction.interaction_analysis()
print(con_results.df_interaction_events)
# Dataframe df_interaction_events includes the start and end times of
    each intersecting segment and the time difference between the two
    intersecting segments
con_results.compute_interaction_duration()
print(con_results.df_interaction_events)
# After running compute_interaction_duration(), df_interaction_events
    has one more column named 'duration' indicating the duration of the
    interaction event
```

Figure 3.5: Implement concurrent interaction analysis

The user can use *compute_interaction_duration*() subsequently to compute the duration of each interaction event, as presented in Figure 3.5. Recall that in Section 3.4, it is essential to consider the parameter $\delta_t$ for discerning between brief encounters and long-duration concurrent interactions. When the duration of an interaction event is shorter than $\delta_t$, it signifies an encounter, whereas a duration exceeding $\delta_t$ indicates a concurrent interaction. The outputs of interaction events are saved in a dataframe called *df_interaction_events* as an attribute of *con_results*. Each row in this dataframe indicates an interaction event. The user may take advantage of this dataframe to further interpret various types of identified interactions. In the provided example, where movement data is gathered every 60 minutes, a $\delta_t$ value of 60 minutes is sufficient to differentiate encounters from concurrent interactions.

### 3.5.4 Identify delayed interaction

Figure 3.6 illustrates an example of identifying delayed interactions that occur with a time lag between 120 to 360 minutes. Similar to identifying concurrent interaction, the user needs to run *interaction_analysis*() after initializing an ORTEGA object. The outcome of this step is the same as the concurrent interaction, namely a dataframe called *df_interaction_events* including the start and end times of each intersecting segment and

the time difference between the two intersecting segments. In essence, the column of time difference in the case of delayed interaction shows the time lag of the delayed interaction event or the time lag between the two intersecting segments.

```
# An example of using [120 min, 360 min] to detect delayed interactions
    that occur with a time lag between 2 to 6 hours.
delayed_interaction = ortega.ORTEGA(data=data1,
                                    minute_min_delay=120,
                                    minute_max_delay=360,
                                    latitude_field = 'latitude_column',
                                    longitude_field = 'longitude_column',
                                    time_field='time_cloumn',
                                    id_field='id_column')

delayed_results = delayed_interaction.interaction_analysis()
print(delayed_results.df_interaction_events)
```

Figure 3.6: Initialize ORTEGA and conduct descriptive statistics

### 3.5.5 Contextualizing the identified interactions

Movement parameters such as speed and direction can be directly computed from GPS coordinates. Given a pair of consecutive GPS points, the speed of the corresponding PPA can be easily computed by dividing the Euclidean distance between the two given locations by the time interval. Subsequently, the *ratio of the difference in speed* (represented as $r_v$) between the two entities when potential interactions occur can be formalized as the absolute difference between the speeds of the two intersecting PPAs, divided by half of the sum of the speeds of the two intersecting PPAs. The value of $r_v$ ranges from 0 to 2, with 0 indicating identical speeds and a larger value indicating greater disparities in speeds between two interacting entities. Movement direction ranges from -180 to 180 degrees. *Similarity of movement direction* (represented as $r_\theta$) can be measured by the cosine value of the difference between the movement directions of the two intersecting PPAs [10]. This value equals 1 when movement segments have

the same orientation, 0 when they are perpendicular, and -1 when they move in opposite directions. The ratio of the difference in speed and similarity of movement direction are automatically computed when running *interaction_analysis*(). The user can further use this information to examine the correlation between movement parameters and potential interactions.

In circumstances when contextual variables are available in addition to movement data, the user may leverage ORTEGA's context-aware functionality to further inform interaction analysis. As Figure 3.7 shows, the user may specify which contextual variables they want to include in the ORTEGA object during the initialization stage by specifying the $attr\_field$ parameter. Since each PPA consists of two GPS locations of an individual trajectory, each pair of intersecting PPAs includes contextual information of four GPS points of the corresponding PPAs of the two entities. ORTEGA offers a *attach_attributes*() function to aggregate these values at each potential interaction area (i.e., a PPA intersection). Currently, ORTEGA calculates the average of the values from the two GPS points that constitute each PPA. Subsequently, it computes either the average or the difference between the two mean values derived from the intersecting PPAs of the two individuals. This can be done by specifying the parameter *method* as 'mean' or 'difference'. Ultimately, every pair of intersecting PPAs is associated with one value of a specific contextual variable. The user may correlate the identified potential interactions with movement parameters such as speed and movement direction, or with contextual factors. This correlation analysis can help us understand the relationships between potential interactions and individuals' movement behaviors, as well as the surrounding environments where interactions occur.

```
con_interaction = ortega.ORTEGA(data=data1,
                                minute_max_delay=60,
                                latitude_field = 'latitude_column',
                                longitude_field = 'longitude_column',
                                time_field='time_cloumn',
                                id_field='id_column',
                                attr_fields=['attr1','attr2','attr3'])

results = con_interaction.interaction_analysis()
# calculate mean values of attr1 and attr2 of two PPAs, respectively
results.attach_attributes(col=['attr1','attr2'], method='mean')
# calculate the difference of attr3 values of two PPAs
results.attach_attributes(col='attr3', method='difference')
print(results.df_all_intersection_pairs)
```

Figure 3.7: Contextualize identified interactions

### 3.5.6 Mapping interactions

Lastly, ORTEGA can visualize the detected interaction events and their associated intersecting PPAs alongside the original movement data for better interpretation. The tool provides two visualization choices: 1) *plot_original_tracks*(): This option allows the user to visualize the original movement trajectories of both individuals, with PPAs depicted in red and blue (by default) for each individual, respectively. 2) *plot_interaction*(): Selecting this option showcases the original trajectories while highlighting the overlapping segments in yellow by default, enhancing the clarity of the segments indicating potential interactions. The user may replace the parameter *colors* to customize the color code using common color names or hexadecimal color codes, especially if they wish to map the interaction between more than two individuals.

```
# 1. Only show original movement trajectories
ortega.visualization.plot_original_tracks(con_interaction, colors=['red
    ', 'blue'])
# 2. Show original movement trajectories and interaction segments
ortega.visualization.plot_interaction(con_interaction, results.
    intersection_ellipse_pair, colors=['red', 'blue', 'yellow'])
```

Figure 3.8: Mapping interaction events

## 3.6 Experiments

### 3.6.1 Data description

This section assesses the usability of the proposed classification framework and methods for the identification of human interaction patterns using a subset of the 2010-2012 California Household Travel Survey (CHTS) [135]. The data consists of a single-day travel diary and the same day of GPS tracking from 986 participants of 447 households. The one-day travel diaries record every participant's household information by assigning a unique identifier, and for each individual's trip, start time and end time, travel mode, origin and destination (in longitude and latitude format), and corresponding location types (e.g., home, work, school, other) were collected. The GPS tracking data were collected using a wearable device that can be worn on the waist or clipped to a purse or backpack. The device only tracks people's movements that are faster than one mile per hour to safeguard the battery and storage capacity. Therefore, in this study, stationary activities are not included in the interaction analysis. The sampling rate of the tracking data is one minute per GPS fix.

### 3.6.2 Case studies to identify different types of interaction between two individuals

#### Encounters

The objective of the first case study is to illustrate encounters between moving individuals. Figure 3.9 shows an encounter between two individuals from different households (i.e., outside their social networks). The short time window parameter $\tau$ is set to one minute to identify brief-duration encounters. As shown in the inset map, only two PPAs from the traces of person A in red were found to intersect spatially and temporally with

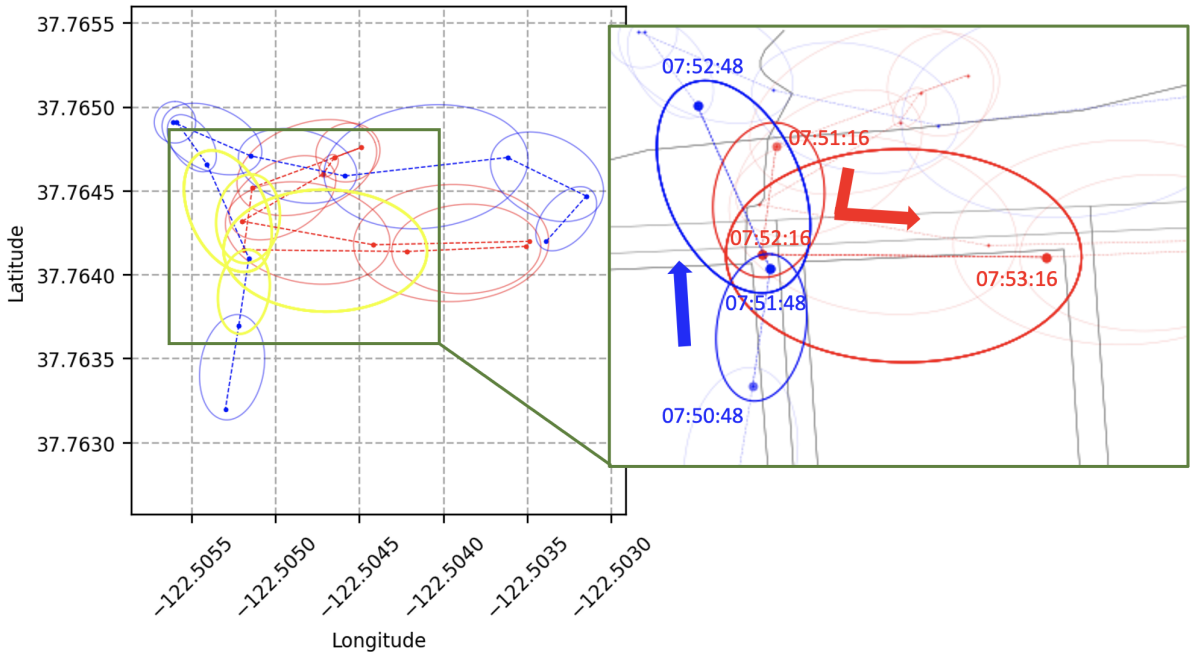



Figure 3.9: An encounter between two individuals from different households. The PPAs are presented in red (person A) and blue (person B) respectively for each individual. Intersecting PPAs are highlighted in yellow ellipses in the main map. The inset map shows the details of the intersecting PPAs between the two individuals with the annotation of the start time and end time of each intersecting PPA. The arrows indicate the moving direction at the selected time interval. The gray lines in the inset map represent the road network retrieved by OSMnx [136]. The base map is hidden in the main map to protect participants' privacy.

the other two PPAs from the traces of person B in blue. We can infer based on their opposite moving direction that this brief interaction most likely indicates an incidental encounter of two individuals who passed by each other on the street rather than an intentional encounter. Due to data limitations, the method does not reveal any brief PPA intersection between individuals of the same or different households which could be interpreted as an intentional encounter. To identify intentional encounters using PPA intersection, we would need to incorporate additional context information from auxiliary data like travel surveys or land use data (see the explanation in Section 3.3). Most encounters were detected among individuals from different households and on the road, and therefore they do not qualify to be classified as intentional encounters.

### Concurrent interactions

The second case study focuses on concurrent interactions occurring between moving individuals. The short time window parameter $\tau$ is set to five minutes to identify concurrent interactions. Figure 3.10 illustrates the one-day GPS tracking of two individuals from the same household in the main map. The result of interaction analysis shows that 100% of their PPAs intersect, implying highly overlapped movement paths and potential joint trips by these two individuals. A portion of the PPA intersections are highlighted using yellow ellipses as shown in the inset map in Figure 3.10. Table 3.2 summarizes the six identified continuous interaction events (i.e., continuous interaction segments) between the two moving individuals during the survey day, while Table 3.3 presents their self-reported travel diary.

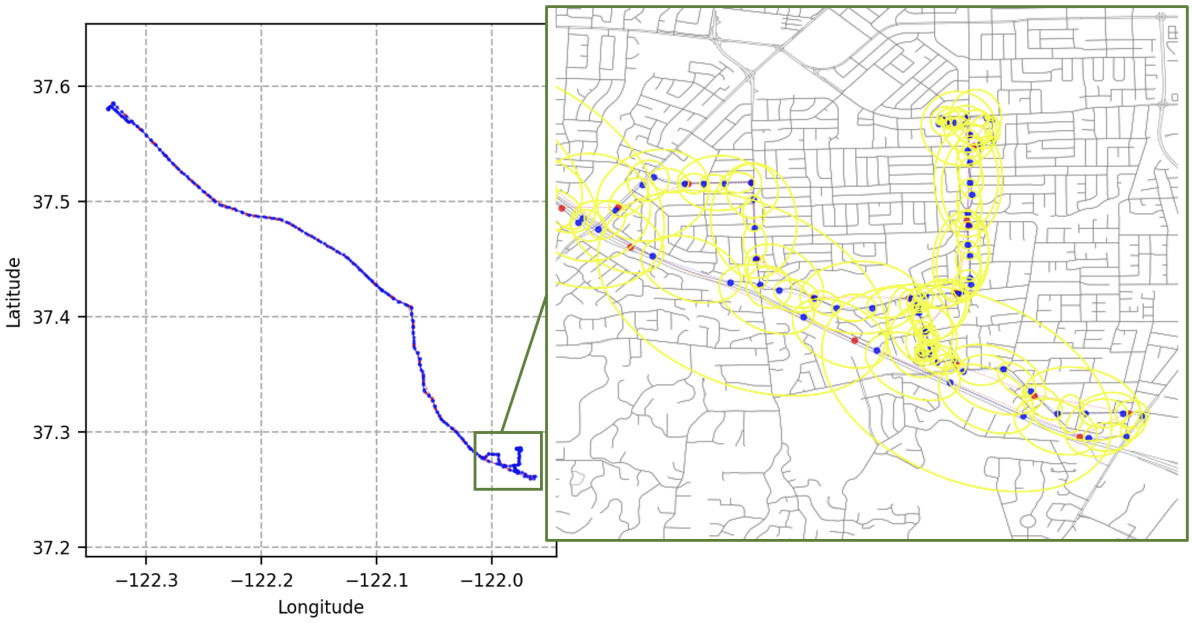



Figure 3.10: Concurrent interactions between two individuals from the same household. The traces are presented in red and blue dash-dotted lines respectively in the main figure (since their tracking points are very dense and could be unreadable if PPAs are visualized). The inset map highlights a subset of the intersecting PPAs between the two individuals using yellow ellipses.

Table 3.2: The identified continuous interaction events between individuals from the same households using the PPA-based approach.

| Event # | Start time | End time | Duration (min) |
|---|---|---|---|
| 1 | 14:16:29 | 14:52:57 | 36.5 |
| 2 | 15:00:58 | 15:04:46 | 3.8 |
| 3 | 15:47:49 | 16:29:39 | 41.8 |
| 4 | 18:16:25 | 18:23:07 | 6.7 |
| 5 | 18:34:11 | 18:37:16 | 3.1 |
| 6 | 18:44:53 | 18:50:05 | 5.2 |

Table 3.3: Travel diary reported by the two individuals in Figure 3.10

| Trip # | Start time | End time | Duration (min) | Origin | Destination |
|---|---|---|---|---|---|
| 1 | 14:17:00 | 14:52:00 | 35 | Home | Liquor store |
| 2 | 15:02:00 | 15:04:00 | 2 | Liquor store | Parent's home |
| 3 | 15:48:00 | 16:29:00 | 41 | Parent's home | Home |
| 4 | 18:16:00 | 18:23:00 | 7 | Home | Restaurant A |
| 5 | 18:34:00 | 18:35:00 | 1 | Restaurant A | Restaurant B |
| 6 | 18:37:00 | 18:45:00 | 8 | Restaurant B | Home |

To assess the validity of the detected interaction events, here we compare the time intervals of the joint trips identified by interaction analysis of GPS data and by reports from the individuals using the travel survey. According to the travel survey, these two individuals have six common trips reported during the survey day. The six concurrent interaction events identified by the PPA-based approach match the travel diary almost perfectly. The total absolute difference between the duration of each interaction event in the travel diary and the identified results is 20.05 min (the average absolute difference is 3.34 min). The slight difference is reasonable as self-reported travel diary data usually cannot reflect precisely the timing of each activity a person conducted in a day. Therefore, all six interaction events are classified as concurrent interactions. In general, the PPA-based approach can identify accurately the interaction events between these two individuals of the same household. We will further evaluate the efficacy of this approach for measuring the duration of concurrent interactions in Section 3.6.3 using a large sample.

Figure 3.11 presents another example of concurrent interactions but is between individuals from different households (i.e., outside their social networks). The two individuals were moving in the same direction. The identified start time and end time of this concurrent interaction are 16:34:17 and 16:40:46. The total duration of this interaction is about 6.48 min. According to their travel diaries, the two individuals were making two different trips during the selected concurrent interaction. Therefore, this implies an incidental concurrent interaction of two persons who move in close spatial proximity on the road by chance.

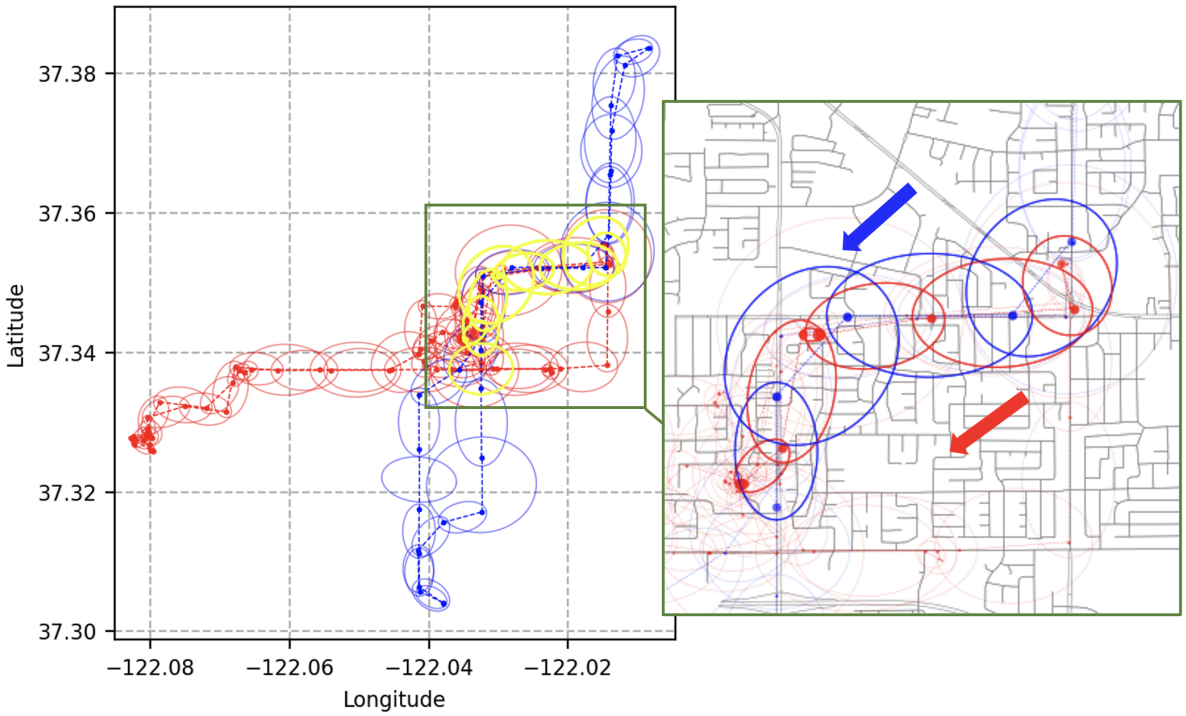



Figure 3.11: A concurrent interaction between two individuals from different households. The PPAs are presented in red and blue respectively for each individual. The intersecting PPAs are highlighted in yellow ellipses in the main map. The inset map shows the details of the continuous intersecting PPAs within 5 min time window. They were moving in the same direction over the selected concurrent interaction.

### Delayed interactions

To showcase the application, as an example, we arbitrarily set the time interval parameter $[\tau^a, \tau^b]$ to [25 min, 30 min] to identify several delayed interactions (25 to 30 min time lag can be considered long enough for delayed interactions between humans). It means that delayed interactions can be possible if PPAs of different individuals intersect spatially and temporally with a time lag falling within the specified range of 25 to 30 min. Figure 3.12a presents a delayed interaction between two individuals from the same household. The inset map shows one segment of the continuously intersecting PPAs with 25 to 30 min delay. The start time and end time of the selected delayed interaction of the individual in red color are 11:00:03 and 11:04:28. For the individual in blue, the time interval is 10:32:29 to 10:34:55. In addition, the two individuals were moving in opposite directions. Mapping the traces to the road networks shows that these segments overlap well with existing streets. Based on these observations, this delayed interaction very likely indicates two individuals who passed by the same road segment at different times with a lag of 25 to 30 min. Moreover, according to the travel survey, the two individuals were on different trips when this delayed interaction occurred. Hence, this delayed interaction occurrence is very likely to be incidental rather than intentional interaction even though these two individuals are related. However, this delayed interaction occurred near their household home location. Based on the inset map, presumably, the delayed interaction happened on the road segment connecting their home and the nearby freeway entrance. This delayed interaction, although is very likely to be incidental, can still reflect a social connection between the two individuals from the same household.

Figure 3.12b illustrates a delayed interaction between individuals from different households (i.e., outside their social networks). The start time and end time of the selected delayed interaction of the individual in red are 07:32:21 and 07:42:28. For the individual

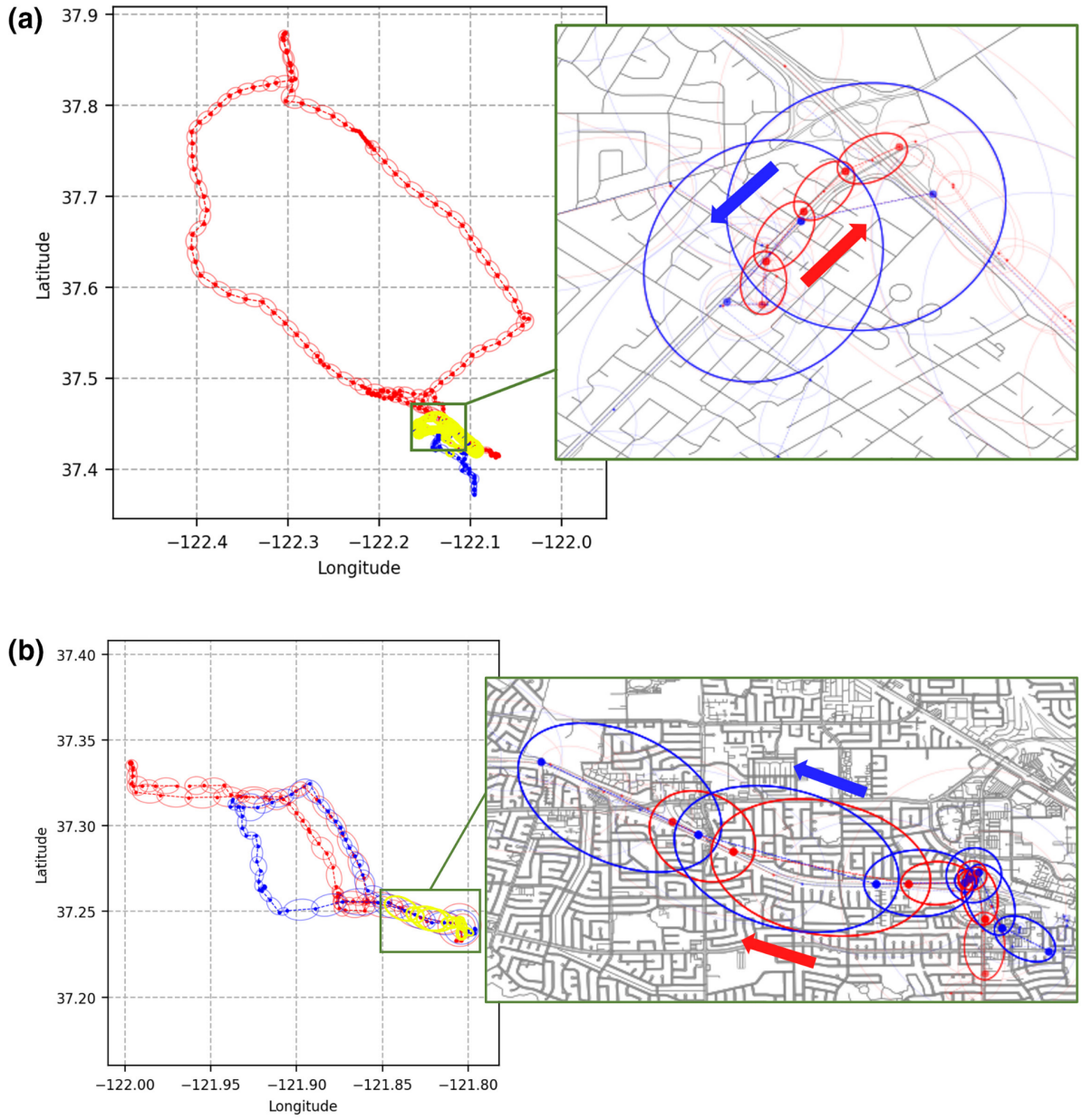



Figure 3.12: Delayed interactions between two individuals from the (a) same household or (b) different households. The inset map in each figure shows the details of one of the continuous intersecting PPAs within 25 min to 30 min time lag between the two individuals. The arrows indicate the moving directions of the two individuals over the time interval of the selected delayed interaction.

in blue, the time interval is 07:04:30 to 07:12:32. The two individuals were moving in the same direction. According to the inset map, this delayed interaction very likely implies

two individuals who passed several same road segments by chance in the morning with a time lag of 25 to 30 min.

### 3.6.3 Case studies to compute duration of interaction in a large network

Duration of interactions plays a critical role in the proposed classification of interactions in human movement data. It is an essential parameter to enable distinguishing between brief-duration encounters and long-duration interactions. This case study evaluates the PPA-based approach for quantifying the duration of interactions, mainly for concurrent interactions. We conduct three evaluative experiments to (1) evaluate the efficacy of the PPA-based interaction analysis technique using the tracking data of a large network of mobile individuals; (2) assess the impact of observational scale on the performance of the PPA-based method for duration computation; and (3) compare the efficacy of the PPA-based approach to a proximity-based approach.

Here, we evaluate the PPA-based approach comprehensively using a larger sample of the 2010-2012 CHTS data, including 986 participants. Only within-household interactions are used here for evaluation because individuals from the same household are very likely to travel together if they report common trips in their travel survey. To investigate the impact of different temporal observational scales on the efficacy of the PPA-based approach in quantifying the duration of interactions, the original tracking data are manually re-scaled to three levels, namely, 1 min, 5 min, and 10 min per GPS fix intervals. This is performed by eliminating tracking points between the required intervals along the trajectories, while the first and final GPS tracking points of each trip are preserved.

For the comparative study, the PPA-based approach is evaluated against the commonly used proximity-based approach based on spatial and temporal buffer(s). In a

proximity-based approach, a potential concurrent interaction can be identified by intersecting spatial buffers centered on synchronous GPS points of two individuals. Here, we use a spatial buffer size of 100 m, as a reasonable movement distance over 1 min (the resolution of the original tracking data used in this study). The bigger the spatial buffer, the higher the possibility of intersection between tracks, especially when synchronous or high resolution tracking data are not available [50]. A parameter $\hat{\tau}$ is used to reflect the allowable short time window between the GPS points of two individuals when judging if their spatial buffers intersect. To make the proximity-based approaches comparable to the PPA-based approach, the same 5 min threshold is applied for both approaches to identify concurrent interaction events (i.e. $\hat{\tau} = \tau = 5$ min).

Two common metrics are used to measure the accuracy of the estimated total duration of potential interactions, that is *Mean Absolute Error (MAE)* and *Root Mean Squared Error (RMSE)*. The smaller the value of the two metrics, the more accurate the estimated total duration of interactions is. Specifically, the total duration of interactions between each pair of individuals from the travel diary is used as the actual value to measure the accuracy of the estimated total duration of potential concurrent interactions.

The box plots in Figure 3.13 illustrate the distribution of the total duration of potential concurrent interactions using the two approaches and the reported values computed based on the travel diary. Figure 3.13a suggests that the PPA-based approach can better estimate the total duration when using 1 min per fix sample ($MAE \approx 17.28$ min, $RMSE \approx 29.61$ min), and it substantially outperforms the traditional proximity-based approach ($MAE \approx 32.84$ min, $RMSE \approx 58.3$ min). A Wilcoxon test on the results of the total duration of interactions using the PPA approach and the reported value from the travel diary indicates no significant difference between them at 0.1 level ($p \approx 0.0544$). As the temporal resolution of the tracking data decreases (see Figure 3.13b-c), the PPA approach tends to overestimate the total duration of potential interactions, while the

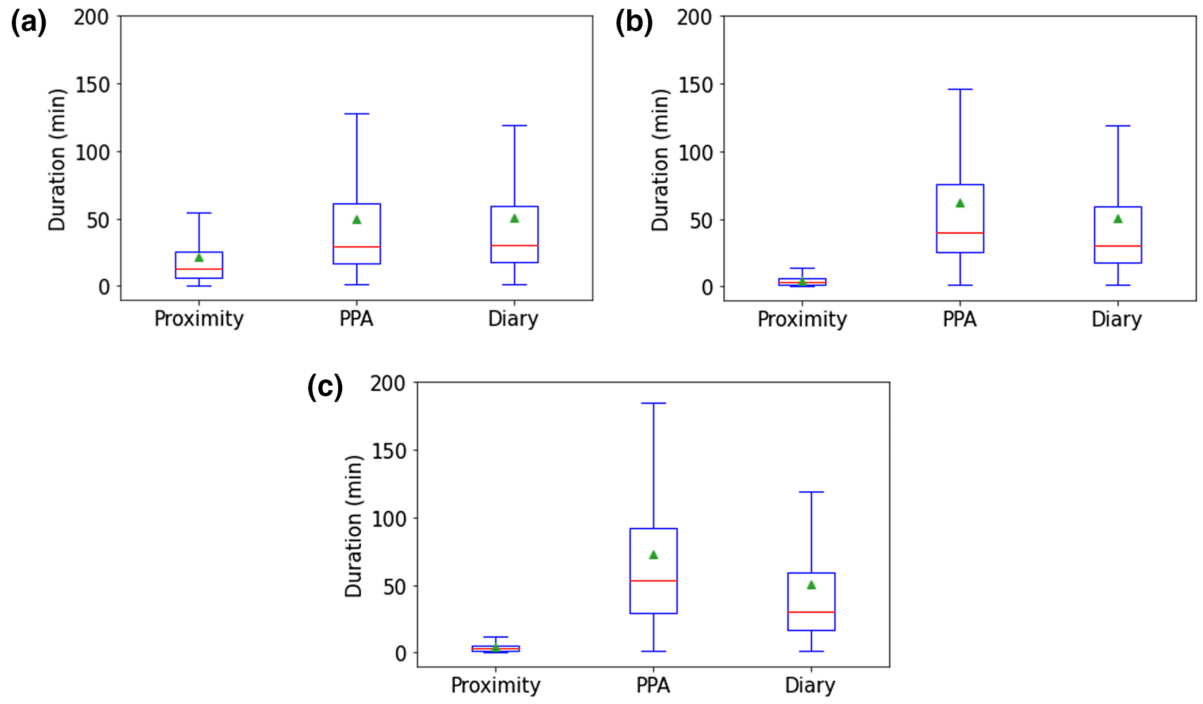



Figure 3.13: Box plots of the total duration of potential interactions for each pair of individuals using movement data of (a) 1 min, (b) 5 min, and (c) 10 min temporal resolution. Green triangles represent the mean value of each group of data.

proximity-based results almost become invalid. As shown in Table 3.4, for the 5 min sample (Figure 3.13b), $MAE \approx 22.91$ min and $RMSE \approx 36.5$ min for the PPA approach, while $MAE \approx 46.41$ min and $RMSE \approx 76.6$ min for the proximity-based approach. For the 10 min sample (Figure 3.13c), $MAE \approx 31.26$ min and $RMSE \approx 48.05$ min for the PPA approach, while $MAE \approx 46.79$ min and $RMSE \approx 76.83$ min for the proximity-based approach. Presumably, tracking data of coarser temporal resolutions are very likely to result in a bigger size of PPAs which make them more likely to intersect spatially, and eventually, this leads to an overestimation of the total duration of potential interactions. But for the proximity-based approach, increasing the sampling interval between consecutive tracking points can amplify the impact of asynchronous tracking issues, which lowers the chance of overlapping buffers for GPS fixes for different individuals in space and time.

A sensitivity analysis on the short time window parameter (i.e., $\tau$ for the PPA-based

Table 3.4: MAE and RMSE for the estimated total duration of interactions using the PPA-based approach or the proximity-based approach.

| Temporal resolution | Approach | MAE(min) | RMSE(min) |
|---|---|---|---|
| 1min | PPA-based | 17.28 | 29.61 |
| | Proximity-based | 32.84 | 58.3 |
| 5min | PPA-based | 22.91 | 36.5 |
| | Proximity-based | 46.41 | 76.6 |
| 10min | PPA-based | 31.26 | 48.05 |
| | Proximity-based | 46.79 | 76.83 |

approach and $\hat{\tau}$ for the proximity-based approach) is conducted to explore the impact of varying time windows when quantifying the duration of potential concurrent interactions. In this experiment, the time window parameter is varied from 1 min to a larger upper bound when judging if two PPAs or two spatial buffers intersect temporally. Allowing a very large time window may identify more delayed interactions rather than concurrent interactions. Therefore, for the 1 min sample, we only experiment on 1 min to 5 min time window. For the 5 min sample, we test time windows from 1 min to 25 min. For the 10 min sample, the time window parameter is varied from 1 min to 30 min. As illustrated in Figure 3.14, not surprisingly, as the time window parameter increases within a reasonable boundary, the PPA approach estimates the total duration less precisely while the performance of the proximity-based approach is improved. Having said that, the PPA approach still outperforms the proximity-based approach substantially according to the graphs of MAE and RMSE.

### 3.6.4 The impact of temporal scale on identifying potential interactions

The goal of this experiment is to provide further insights into the impact of temporal scale on identifying potential concurrent and delayed interactions between moving indi-

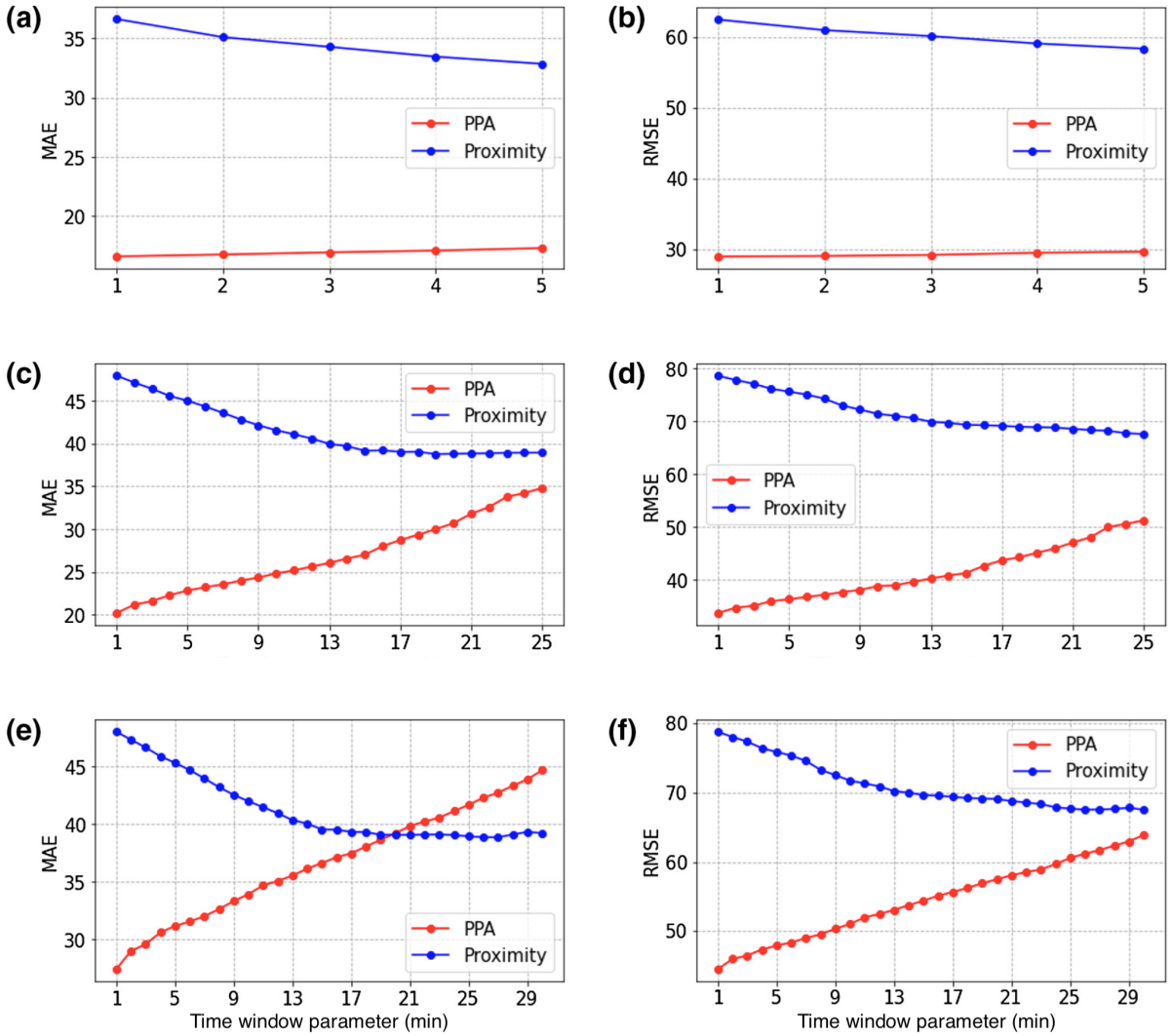



Figure 3.14: Sensitivity analysis of the time window parameter using (a)-(b) 1 min, (c)-(d) 5 min, and (e)-(f) 10 min samples.

viduals using ORTEGA and to compare these results with the traditional proximity-based approach. Following Section 3.6.3, the proximity-based approach is implemented by intersecting spatial buffers of a 100 meter distance threshold centered on synchronous GPS tracking points of two individuals. To relax the restriction of requiring synchronous fixes when identifying concurrent interactions, the parameter of a short time window $\hat{\tau}$ is set to 5 min when determining if two spatial buffers are intersected for the proximity-based

approach. Even though the PPA-based approach does not require a predefined distance buffer or a set time window, we allow the same 5-min time window (i.e., $\tau = \hat{\tau} = 5$ min) for intersection in this experiment to make it comparable with the proximity-based approach. In this case study, if the PPAs of the two individuals are intersected but the time lag is longer than 5 min, we consider this as a delayed interaction. The frequency of interactions among a group of people is then quantified by the count of pairs of individuals that have interacted with each other during the survey day for both the proximity-based and PPA-based approaches. Based on household composition information, we categorize the identified interactions into two groups: those occurring among individuals within the same household and those involving individuals from different households. Individuals within the same household tend to travel together and interact more, while individuals from different households might have occasional social interactions and various random contacts. This categorization allows for the incorporation of social relationship factors in the analysis of human interactions. The outcomes of these two approaches at different temporal scales are then compared as described below.

Figure 3.15 illustrates the impact of temporal scale on identifying concurrent interactions using the PPA-based approach versus the proximity-based approach between individuals of the same household and different households. In general, both approaches are sensitive to the temporal scale. The proximity-based approach can identify a greater number of contacts within household interactions when the temporal resolution is 1 min as opposed to coarser sampling rates beyond 1 min. Specifically, the total number of within-household interactions identified is 367 at 1-min interval, whereas this number drops to 361 at 5-min interval, and to 359 at 10-min interval and remains unchanged at 20-min and 30-min intervals. Likewise, the proximity-based approach can identify 73 outside household interactions at 1-min interval while substantially fewer outside household interactions are identified when the sampling interval increases (i.e., 6, 2, 2,

2 outside household interactions are identified at 5 min, 10 min, 20 min, 30 min interval, respectively). On the contrary, the PPA-based approach can identify substantially more outside-household concurrent interactions when coarse temporal scales are used. This is mainly because using larger sampling intervals results in bigger PPAs which are more likely to intersect with each other. However, the total number of within-household interactions remains stable as the sampling interval increases.

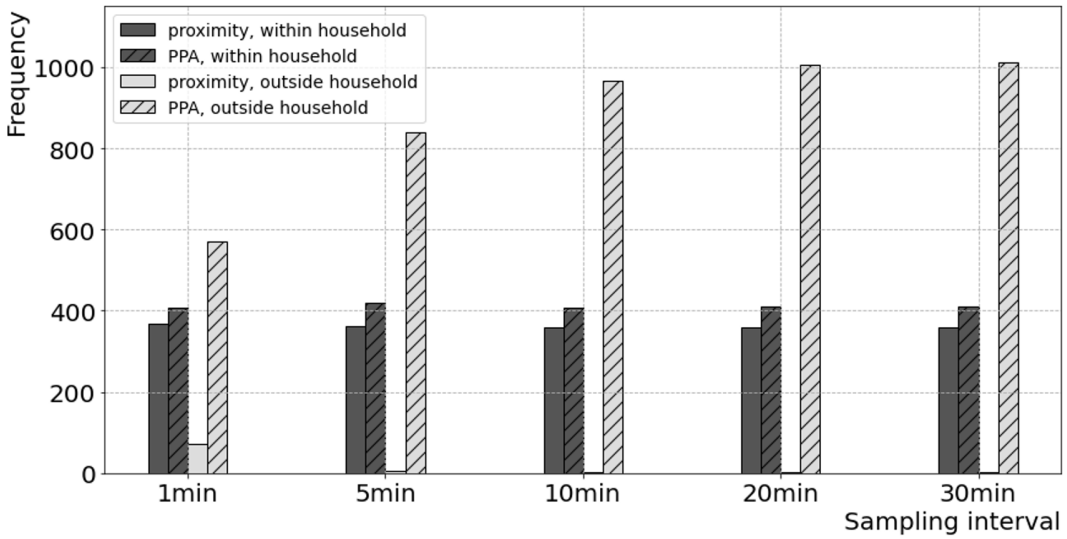



Figure 3.15: Illustration of the impact of temporal scale on identifying concurrent interactions (allowing 5 min time window) using the PPA-based approach versus the proximity-based approach (100m buffer) between individuals of the same household and different households.

People living in the same household are more likely to travel together during the day (e.g., parents escort children to school) [21], which increases the chance of physical contact with each other. Even when the sampling interval is increased, these people are still very likely to be tracked simultaneously. Hence, the PPA-based approach and the proximity-based can identify very closely the number of within-household interactions and the numbers remain stable along with the increasing sampling intervals. Unlike people living in the same household, people from different households are less likely to travel together and their physical interactions are often unintentional (e.g., two strangers

pass by each other on the street) and asynchronous (e.g., two persons visit the same grocery store at different times). The proximity-based approach is less effective in this situation because it requires a synchronous and predefined distance buffer threshold which is set by the user (such as 100m in this case) to determine if two moving entities come close to each other. However, the PPA-based approach does not have such constraint as it is based on the actual locations accessible to people, their time budget, and speed capacities. All of these parameters can be extracted from the given data without a need for a user-defined threshold. Therefore, we observe more outside household concurrent interactions identified by the PPA-based approach as opposed to the proximity-based approach.

As discussed in [50], the outcomes of the proximity-based approach are also sensitive to the buffer size. Specifically, they found that by applying a 500m distance buffer, the proximity-based approach can identify almost the same number of concurrent intrahousehold interactions but interhousehold interactions are still underestimated. Employing a distance buffer larger than 500m may not be reasonable as the potential concurrent interactions are defined as close contact between moving individuals in this experiment.

We next focus on the impact of temporal scale on identifying delayed interactions using the PPA-based approach. As shown in Figure 3.16, allowing a longer time lag can identify more delayed interactions within households as well as outside households. This is reasonable because allowing a longer time lag increases the chance of intersection between two PPAs. However, the increase in the amount of identified delayed interactions outside households is substantially higher than the interactions within households. This is mainly because people living in the same household are more likely to interact synchronously. Therefore, allowing a longer time lag does not help much with identifying delayed interactions within households. Nevertheless, people from different households are more likely to have asynchronous interactions, resulting in more delayed interactions

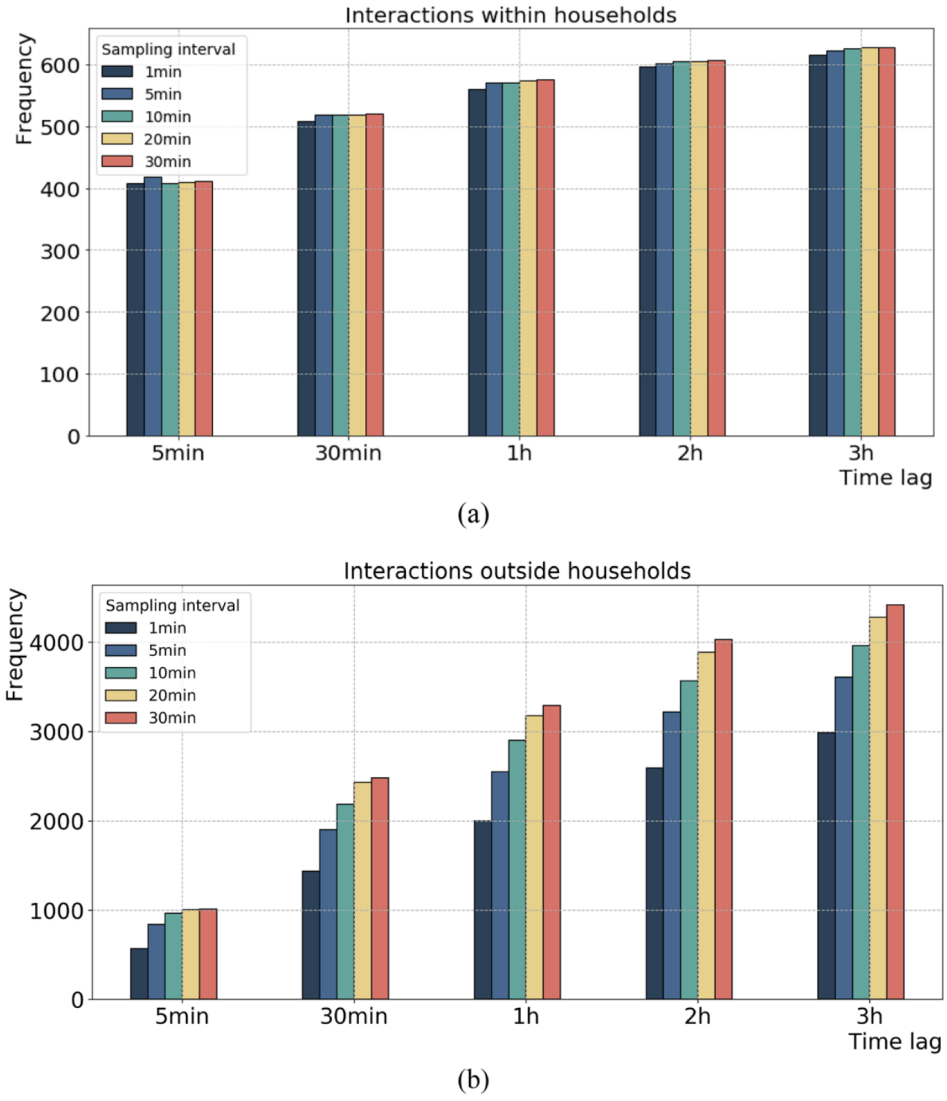



Figure 3.16: Illustration of the impact of temporal scale on identifying delayed interactions using PPA-based approach between individuals of (a) the same household and (b) different households.

when allowing longer time lags. In addition, using different sampling intervals of the data seems to have no significant impact on identifying delayed interactions between individuals of the same household, whereas more delayed interactions between individuals of different households can be identified if a coarser sampling rate of data is used. This is similar to our previous findings regarding concurrent interactions. Presumably, people

living in the same household are more likely to travel together and increasing sampling intervals results in larger PPAs but does not impact very much the chances of intersecting between PPAs. On the other hand, larger PPAs of individuals from different households as a result of a coarser time scale increase the probability of intersecting PPAs.

## 3.7 Discussion

The experiments presented in this study suggest that the PPA-based approach is capable of differentiating different types of human interactions. In this section, we discuss some limitations of the work and ideas for further developments of the technique to improve its capabilities.

Even though we manually interpreted and classified several cases of human interactions as intentional or incidental for illustration in Section 3.6.2, the current algorithm is not fully equipped to automatically differentiate between intentional and incidental interactions as it does not consider context parameters such as street network configuration (e.g., [137]), travel mode, direction of movement, and social relationships between moving individuals [138]. Put more generally, it is necessary to extend the existing PPA-based approach with context-aware abilities or leverage network space-time prism or network PPA for further enriching and improvement of the identification of human interactions (i.e., encounters, concurrent and delayed interactions) and capture their nature. By doing so, for instance, it will be possible to differentiate close contact between individuals in indoor and outdoor settings by taking into account geographic and environmental variables. Additionally, by incorporating travel mode, interactions between individuals who traveled in different vehicles can also be distinguished from more chance interactions when individuals pass each other on the street or interact on foot in the risky context of disease transmission.

In addition, movement data granularity and temporal scale can be key factors influencing the performance and validity of the PPA-based approach for distinguishing different types of human interactions. Our results suggest that both the PPA-based approach and the proximity-based approach are sensitive to the temporal resolution of movement data, making interaction analysis more uncertain. Applying the proximity-based approach may result in more underestimation interactions if coarser data are used (as it may be hard to identify intersecting spatial buffers). In contrast, the PPA-based approach may lead to more overestimation results (as PPAs can be very large and easier to intersect with each other if coarser data are applied). If very fine tracking data are used (e.g., several seconds per GPS fix), there may not be much difference between the PPA-based approach and the proximity-based approach as the PPAs and the spatial buffers may cover similar accessible areas. In addition, using data with various sampling intervals seems to have no significant impact on identifying delayed interactions between individuals of the same household, whereas more delayed interactions between individuals of different households can be identified with a coarser temporal scale. The outcomes suggest that movement data of fine-grained temporal granularity such as one minute per fix is required for a more reliable interaction analysis.

## 3.8 Conclusions

This study presented a classification framework and a time-geographic (or PPA-based) approach and its open-source Python package called ORTEGA to identify various types of human interactions using GPS tracking data. The proposed framework classifies human interactions into encounter, concurrent interaction, and delayed interaction based on the temporal aspect (lag and duration) of how the movement paths of individuals overlap or intersect. By implementing PPA intersection methods, it is possible to iden-

tify and distinguish different types of interactions. This research also systematically evaluated the proposed approach for measuring the duration of interaction and discussed the impact of temporal scale on human interaction analysis. The results using a large sample of individuals showed that the PPA-based approach can estimate the duration of potential interactions more accurately, while the proximity-based approach underestimates potential interactions, which may result in more intermittent interactions with shorter duration. A sensitivity analysis of the time window parameters showed that as the time window parameter increases within a reasonable boundary, the PPA-based approach estimates the total duration less precisely.

In terms of substantial findings regarding the impact of temporal scale on human interaction analysis, we found that the concurrent interactions between individuals identified by the proximity-based approach can substantially be underestimated at coarser sampling intervals. In contrast, the PPA-based approach can identify substantially more concurrent interactions outside households when a coarser temporal scale is used. The use of movement data with different sampling intervals seems to have no significant impact on identifying delayed interactions between individuals of the same household, whereas more delayed interactions between individuals of different households can be identified with a coarser temporal scale. In addition, using a coarser temporal resolution of movement data may lead to an overestimation of the duration of interaction using the PPA-based approach, while the proximity-based results can become invalid.

# Chapter 4

# Spatiotemporal Dynamics of Racial-ethnic Diversity and Segregation: Insights from Mobile Phone Data

**Abstract** Racial diversity and segregation persist as pressing issues in US cities. Understanding these issues is crucial for promoting social equity and justice. Existing research has predominantly focused on residential racial diversity while overlooking racial diversity and segregation in individuals' daily activities and place visits, due in part to data limitations. This study introduces a dynamic measure of racial diversity and dominance at the finest spatial scale, specifically at the Points of Interest (POI) level and various temporal contexts. Leveraging one month of privacy-enhanced mobile phone location data in Los Angeles County, California, this study explores the racial diversity and spatial segregation simultaneously in POI visits in LA County. Empirical results indicate a significantly higher level of racial diversity in POI visits during weekdays compared to

weekends. Our findings confirm that individuals' daily mobility in urban areas enhances social integration among various racial-ethnic groups. The outcomes also suggest significant spatial clusters of isolated and integrated areas regarding racial mix in people's everyday activities. Notably, the Latino population tends to stay in their own communities and experience a higher level of segregation in their daily activities. The overall diversity level of daily visits reaches its peak at noon or in the afternoon, as opposed to the morning or evening, indicating a higher level of social mixing among different groups during lunchtime and afternoon activities. Moreover, visitors from outside the Greater LA region in California contribute positively to the observed racial diversity in visits to POIs on weekends in LA County.

## 4.1 Introduction

The challenge of racial diversity within communities and the persistence of residential segregation remain critical concerns as urbanization and suburbanization continue to expand in the United States [139]. Racial segregation, especially affecting marginalized communities or disadvantaged groups, is closely linked to adverse consequences in education, housing, employment, poverty, public safety, environmental justice, and access to healthcare resources [140, 141, 142, 143, 144]. Gaining insights into racial diversity and segregation is imperative for city planners to promote social integration via creating opportunities and local environments that foster more intergroup contact and interaction among individuals from diverse racial and ethnic backgrounds [141, 145]. Promoting social integration is essential because it fosters positive interactions with socially heterogeneous individuals, contributing to personal integration into a community and enhancing social cohesion more broadly, which, in turn, leads to increased solidarity, trust, pro-social behavior, well-being, and reduced violence and crime in society [5, 6].

The prevailing body of research has traditionally centered on residential racial diversity and segregation, primarily drawing from static census data [146, 139, 147]. While these census-based studies have provided valuable insights into residential segregation and its implications for racial stratification, they tend to overlook the multifaceted dimensions of racial diversity and segregation within individuals' daily lives and their interactions with various places beyond their residences. Some scholars argued that segregation is a dynamic phenomenon that evolves across both spatial and temporal dimensions, suggesting the need to extend the study of segregation beyond the residential context [148, 149, 120, 6, 150, 151]. Nevertheless, many of these efforts faced constraints, often characterized by either restricted sample sizes, especially when employing conventional travel diaries to document individuals' daily activities [148, 120], or issues related to data representativeness and demographic bias when utilizing georeferenced social media posts [6, 151].

The advent of Location-Aware Technologies (LATs) has ushered in an era of abundant movement data, encompassing diverse temporal granularities and collected through various means. These data sources include GPS data recorders, call detail records (CDR) from mobile phones, radio-frequency identification (RFID) tags, Wi-Fi and Bluetooth sensors, and georeferenced social media applications [55]. The proliferation of movement data provides a wealth of information characterized by both volume and variety, particularly in urban settings. Among these sources, GPS records of mobile phone data have gained increasing popularity in segregation studies in recent years [152, 150]. This is due to its extensive spatial and temporal coverage, large sample size, and low cost of data collection. As a result, mobile phone data provides an unprecedented opportunity for understanding human behavior and conducting research on racial diversity and segregation across finer spatial and temporal scales beyond residential perspectives.

This research introduces a novel approach that leverages massive mobile phone lo-

cation data to examine dynamic racial diversity and segregation in individuals' visits to places at the finest spatial scale. Specifically, we explore shared visits (i.e., co-location) of individuals at the Points of Interest (POIs) level, and in various temporal contexts (e.g., weekdays versus weekends, certain periods of the day such as morning, noon, afternoon, and evening). A POI is a particular place or location that individuals might find appealing or useful for different activities. Examples of POIs include restaurants, hotels, gas stations, hospitals, public transportation stops, shopping centers, recreation areas, and more. Leveraging one month of anonymized, fine-resolution mobile phone location data in Los Angeles County, California, this study explores the racial diversity and spatial segregation in POI visits in LA County. Our research intends to answer the following questions: do people from different social groups living in racially segregated residential spaces also present segregated patterns in their daily visits to different types of places, and how does that vary across space and different temporal contexts? The contributions of this study are twofold. Firstly, this is the first attempt to explore dynamic racial diversity using massive mobile phone location data at the most granular spatial resolution currently available, specifically at the level of POI visits, which directly represents the racial composition of individuals' daily activity locations. Secondly, our research unveils the spatial patterns of segregation in POI visits across diverse social groups and temporal contexts. The findings can have significant implications for urban planners and policymakers to design targeted solutions and policies in order to promote social equity, social integration, and equal access to public amenities and opportunities in urban spaces [153, 154].

## 4.2 Background

Racial and ethnic segregation has been studied extensively (e.g., see a review in [155]). There has been a debate about conceptualizing segregation [156]. For instance, [156] suggested five dimensions of segregation: evenness, exposure-isolation, concentration, centralization, and clustering. Some researchers, though, argued that these can be simplified into evenness-clustering and exposure-isolation [157], or evenness-concentration and clustering-exposure [158]. Another perspective combines aspects into two broader categories: separateness (which includes evenness, isolation, and clustering) and location (which includes concentration and centralization)[159]. [156] proposed that segregation is a multifaceted phenomenon and should be assessed using diverse indices that emphasize various dimensions. Of the dimensions mentioned above, *evenness* and *exposure* are the most commonly discussed in segregation studies.

To measure the evenness aspect of segregation, the most widely used measurement might be the dissimilarity index $D$ [160]. The dissimilarity index is commonly defined as one-half of the sum of the absolute difference between the two proportions of the two population groups across all predefined discrete spatial units within the study area, as below.

$$D = \frac{1}{2} \sum_{\alpha} |\frac{p_{\alpha,i}}{T_i} - \frac{p_{\alpha,j}}{T_j}| \tag{4.1}$$

where $\alpha$ is the index of spatial unit, $i$ and $j$ represent two population groups, $T_i$ and $T_j$ denote the respective total population of the two groups in the whole study area, $p_{\alpha,i}$ and $p_{\alpha,j}$ are the population of groups $i$ and $j$ respectively in spatial unit $\alpha$. The dissimilarity index falls within the range of 0 to 1, where a value of 0 indicates perfect integration, implying that both groups are evenly distributed across all areas. This global index has limitations as it can only provide a summary of the overall segregation within the

entire study area, such as a city or region, and it primarily focuses on residential areas while neglecting the places individuals visit during the day. To overcome these limitations, some extended the index by incorporating spatial relationships between population groups (e.g., distance or adjacency relationships between spatial units) [161, 157, 162]. In addition, various spatial autocorrelation statistics have been used for quantifying segregation [163]. Besides the similarity-based measures, there has been a growing focus on entropy-based methods [164, 146, 139], which can assess the level of racial diversity in residential areas and are adaptable to situations involving more than two racial groups. These entropy-based indices typically provide values that span from 0 (indicating no diversity) to 1 (maximum diversity).

Another major dimension of segregation, exposure-isolation, is commonly measured using the exposure and isolation indices [165, 166, 167, 168, 156]. These indices measure the extent to which members of one group (e.g., a specific racial or ethnic group) are exposed to members of another group in a given area or neighborhood. On the other hand, when evaluating an individual's experience of segregation, the exposure dimension is often regarded as the most effective indicator [169, 148, 170, 171, 120]. Several researchers claimed that an individual's experience of segregation is shaped by their daily activities across multiple locations and interactions with various individuals. For example, [169] devised socio-spatial isolation indices to capture an individual's degree of isolation from others, by considering several specific types of people's daily activities. [148] extended the original exposure index to accommodate all individuals' daily activity locations beyond residences, regardless of the type of socio-geographical spaces based on the concept of activity space [172]. [120] developed an individual-level spatiotemporal proximity index accounting for individuals' daily activities in space and time observed from travel diaries. However, in practice, capturing all intentional or significant physical interactions between individuals is exceptionally challenging due to a scarcity of appropriate data.

Another advancement in racial-ethnic segregation studies involves replacing global measures with analyses that classify neighborhoods into distinct groups representing various levels of racial mixing [173, 146, 139]. [146] suggested that racial diversity and segregation are related to each other and should be understood jointly. To achieve this goal, they expanded the conventional entropy-based method to categorize neighborhoods according to both their compositional diversity and racial dominance. Specifically, they employed the scaled entropy index to calculate the racial diversity of aggregated geographic areas, such as census tracts, using racial and ethnic composition from traditional census data (See formulations in Section 4.4.2). The scaled entropy index provides a multi-group perspective that transcends the "majority-minority divide", enabling equal consideration to all social groups. Following this, predefined threshold criteria can be applied to differentiate diversity levels, and these can be subsequently subdivided by the dominant racial group. The neighborhood classification approach turns out to be effective in capturing the processes of neighborhood change over time in terms of racial-ethnic configuration [173, 146, 139, 174]. However, to the best of our knowledge, this idea has not been fully incorporated into the understanding of racial diversity and segregation beyond the residential perspective. In addition, few studies have directly examined racial diversity at the POI level, which can be considered the finest spatial resolution for assessing social interactions. In this study, we bridge this gap by developing a dynamic place-based racial diversity measure based on the scaled entropy index to assess racial diversity and segregation patterns simultaneously at the POI level, taking into account the spatiotemporal dynamics of individuals' activities and visits to various types of places.

## 4.3 Mobile Phone Location Data

The mobile location records used in this study are obtained from Spectus[1], a location data intelligence company, that provides anonymized, privacy-enhanced, and high-resolution mobile location pings of millions of users in the United States. Spectus collects mobility data through its software development kit that is integrated into various mobile applications and services in smartphones (e.g., maps, navigation, weather, and geo-specific retail), all while maintaining strict privacy adherence[2]. The data collection relies on the voluntary sharing of location by device owners through a process compliant with the General Data Protection Regulation and the California Consumer Privacy Act compliant frameworks. Spectus mandates application partners to disclose their relationships with Spectus in their privacy policies. The mobile phone location data used in this study encompasses the daily movement data of 283,556 anonymized individuals (more than one billion location records) during March 2019, comprising 10 weekend days and 21 weekdays. Each entry in the mobility data includes an anonymized person identifier, location coordinates (in longitude and latitude), and timestamp. The month of March was selected for our research because it does not include any special holiday seasons, making it a suitable representation of typical human activity patterns before the COVID-19 pandemic. The study area is Los Angeles County, the most populous county in the United States with an estimated population of 10,105,518 according to the 2018 American Community Survey [175]. LA is also renowned for its cultural diversity, which makes it suitable for this study.

[1] `https://spectus.ai/`

[2] `https://spectus.ai/privacy/privacy-policy/`

## 4.4 Methodology

### 4.4.1 Data processing

As shown in Figure 4.1, in order to prepare the data for the main racial diversity analysis, several essential data-processing steps are conducted using the Spectus curated mobile phone location data. These steps include: (1) approximating the home location of each individual based on mobile phone pings, (2) inferring individual race-ethnic type based on their home census data, and (3) identifying their POI visits based on mobile phone pings and POI data. Among these, steps (1) and (3) are already automated and incorporated in Spectus data.

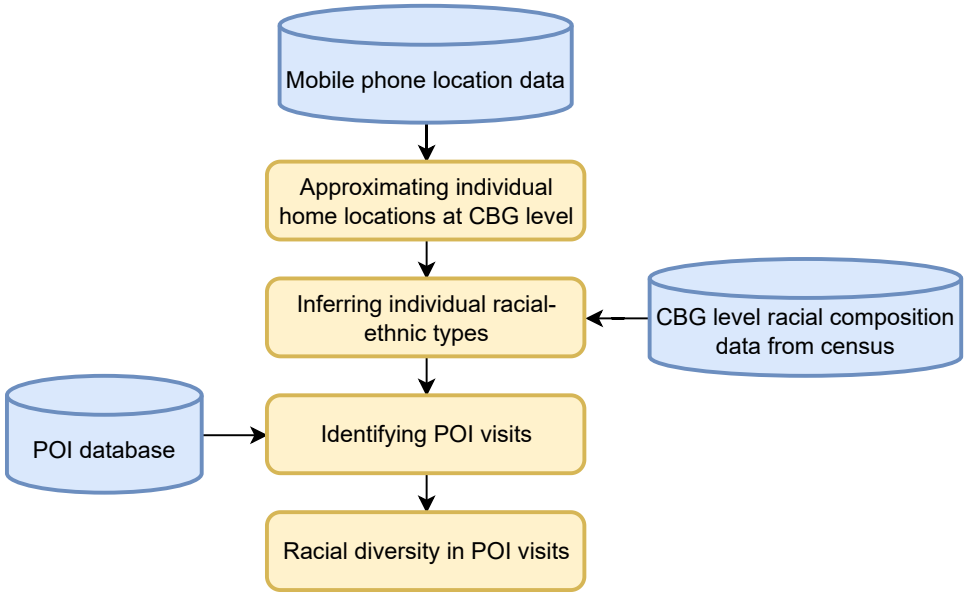



Figure 4.1: Workflow of data processing and analysis for racial diversity in POI visits.

#### Approximating individual home locations

In order to preserve privacy, Spectus provides the approximate home location at the Census Block Group (CBG) level for each individual using their most recurring location

between 10 p.m. and 6 a.m. every week[3]. According to the U.S. Census Bureau, CBGs generally have a population of 600 to 3,000 individuals with similar sociodemographic characteristics, ensuring data privacy. Specifically, Spectus utilizes raw mobile phone pings to generate three variables: the duration of stay in a particular location within the past month, the average number of hours spent daily in that location, and the time of day visited that location. These variables are used to calculate a score, which indicates the likelihood of a location being considered as home. Subsequently, the location with the highest score is determined as the individual's home location. To ensure privacy protection, the specific home locations are obfuscated by mapping the latitude and longitude coordinates of each home location to the corresponding CBG.

**Inferring individual racial-ethnic types**

Since mobile phone pings do not capture racial-ethnic characteristics of the individuals, we infer individual racial-ethnic categories based on home CBG's census data. The census is from the 2018 American Community Survey [175]. In this research, we consider six racial-ethnic groups including Hispanic or Latino/a ("Latino" herein), Black or African American ("Black" herein), American Indian or Alaska Native ("American Indian" herein), Asian American or Pacific Islander ("Asian" herein), White, and "Other" (encompasses all other racial-ethnic groups that do not belong to the above five groups, such as multiracial and multiethnic groups). For simplicity, note that in this study, these six racial-ethnic groups are mutually exclusive. For example, someone who identifies as White is not counted as Latino (i.e., the person is non-Latino White). We use the racial-ethnic composition of the CBG where individuals' estimated homes are located as a proxy for their racial-ethnic category. Instead of assigning a single distinct race-ethnic

[3]See the details in the Spectus documentation: `https://docs.spectus.ai/Getting%20Started/User_Guides/Data_Assets/Device_Recurring_Areas_and_Sensitive_Locations/`

category to each individual, we compute each individual's probabilities of belonging to each of the six racial-ethnic groups according to their home CBG's racial-ethnic compositions from the census. For example, for an individual with an identified home located in a CBG with 60% White, 30% Latino, 6% Black, and 4% Asian, the probability of this individual belonging to White is 0.6, Latino 0.3, Black 0.06, Asian 0.04, American Indian 0, and Other 0.

### Identifying POI visits

Spectus offers POI visit data that is generated through a spatial join between stops and nearby POIs [176]. The POI dataset is also curated by Spectus in compliance with its Sensitive Points of Interest Policy[4]. Each POI represents a point with an attribute indicating its geographic boundary when available; otherwise, a spatial buffer is assigned based on the POI type. For instance, airports and malls have a larger radius (500 meters and 300 meters, respectively), while others such as bus stops have a smaller radius (5 meters). For a more accurate estimation of POI visits, the geographic boundaries or spatial buffers of the POI, the open hours of the POI, and the minimum and maximum dwell time are all taken into account in this procedure[5]. Each POI visit entry encompasses a unique and anonymized person identifier, POI identifier, POI type according to Standard Industrial Classification (SIC) codes[6], and start and end time of the visit. The POI data contain over 160 distinct categories according to the SIC code, such as eating places, grocery stores, physical fitness facilities, gasoline service stations, and commercial banks. To streamline these categories, we also employ a higher-level classification system consisting of ten major business divisions according to the SIC system, which encompasses: 1) Agriculture, forestry, and fishing, 2) Mining, 3) Construction, 4) Man-

[4] `https://spectus.ai/privacy/spoi-policy/`
[5] See the details in the Spectus documentation: `https://docs.spectus.ai/`
[6] `https://www.osha.gov/data/sic-manual`

ufacturing, 5) Transportation and public utilities, 6) Wholesale trade, 7) Retail trade, 8) Finance, insurance, real estate, 9) Services, and 10) Public administration. The POI visit data are subsequently utilized to calculate the racial diversity of visitors at various types of places.

### 4.4.2 Racial diversity computation

#### Residential racial diversity index

A widely used measure for assessing racial-ethnic diversity is the scaled entropy index, which is formally defined as follows: [177, 178, 146].

$$E_\alpha = \frac{\sum_{j=1}^{n} p_{\alpha,j} ln(1/p_{\alpha,j})}{ln(n)} \tag{4.2}$$

where $E_\alpha$ is the scaled entropy index of the spatial unit $\alpha$ (e.g., census tract, CBG, grid cell, etc), $p_{\alpha,j}$ denotes a studied racial group $j$'s proportion of the total population observed at the spatial unit $\alpha$, and $n$ is the total number of racial groups. In this study, $n$ equals six racial-ethnic groups as introduced in Section 4.4.1. This index is widely used to assess residential racial diversity based on racial composition from census data. The scaled entropy index falls within the range of 0 to 1. A value of 0 signifies no diversity, indicating the presence of only one racial group at the spatial unit $\alpha$. Conversely, a value of 1 indicates maximum diversity, suggesting that all racial groups are equally represented.

We utilize Equation 4.2 to calculate residential racial diversity spatially at the CBG level using the census data. In essence, the racial composition at the CBG level, as obtained from the census data, corresponds to $p_{\alpha,j}$ and can be used for calculating the scaled entropy index for each CBG.

### Dynamic place-based racial diversity index

We extend the scaled entropy index to assess racial diversity at the POI level, which can be considered the finest spatial resolution for capturing social interactions. This extension accounts for the spatiotemporal dynamics of individuals' daily activities and their visits to these locations. The dynamic place-based scaled entropy at the POI level can be formalized as follows.

$$E(l,t) = \frac{\sum_{j=1}^{n} p_j(l,t) ln(1/p_j(l,t))}{ln(n)} \tag{4.3}$$

where $E(l,t)$ is the scaled entropy index of the POI $l$ at time $t$, $p_j(l,t)$ denotes racial group $j$'s proportion of the total individuals who visited POI $l$ at time $t$. The interpretation of $E(l,t)$ values is the same as $E_\alpha$.

For each POI, the racial group composition is quantified based on the inferred race categories of individuals, which is determined according to the racial configuration of their home CBGs (see Section 4.4.1). To ensure a sufficient number of POI visit records, we calculate racial diversity for each POI across four distinct periods throughout the day: morning (7:00 am - 10:59 am), noon (11:00 am - 1:59 pm), afternoon (2:00 pm - 5:59 pm), and evening (6:00 pm - 11:59 pm). It is worth noting that the proposed dynamic scaled entropy at the POI level is flexible enough to be calculated at any temporal resolution (e.g., minute-by-minute, hourly) as long as a sufficient number of POI visits are available at the given temporal aggregation. The variable $t$ is used to represent these four periods in Equation 4.3. The premise is that people of different racial groups might tend to or prefer visiting certain POIs at different times during the day. For example, Black, Latino, or White might prefer certain and different POIs for socializing after work hours, while they tend to visit other POIs more during the day. Similarly, certain groups might avoid certain POIs at night time due to concerns for safety or perceived risks. Furthermore,

we distinguished between weekdays and weekends, as they exhibit significant differences in human daily behaviors [179].

**Accounting for sampling bias**

To deal with the potential sampling bias issues in mobile phone location data, we apply a post-stratification technique. Post-stratification is a well-known sampling tool, commonly employed to investigate the influence of sampling biases in mobile phone location data or geo-tagged social media data on a wide range of subsequent data analyses [180]. Specifically, each individual's POI visit count is weighted by an expansion factor [181] which is the ratio of the population of their home CBG to the number of individuals observed in the same CBG from the mobile phone data. Subsequently, racial group $j$'s proportion of the population who visited POI $l$ during time $t$ can be calculated as follows.

$$p_j(l,t) = \frac{\sum_i w_i \rho(i \in G_j)}{\sum_i \sum_{j'=1}^{n} w_i \rho(i \in G_{j'})} \tag{4.4}$$

where $w_i$ denotes the expansion factor of person $i$ who visited POI $l$ at time $t$, $\rho(i \in G_j)$ denotes the probability of person $i$ belonging to racial group $G_j$ (the probability is inferred based on their home CBG's racial configuration, see Section 4.4.1). To avoid the potential bias caused by small samples, the POIs with fewer than ten visit counts over the entire study timeline are excluded from this measure.

### 4.4.3 Racial diversity and dominance classification

Following [146]'s approach, we further adopt a classification framework to differentiate different types of racial diversity for POI visits. This approach incorporates a multiplicity of racial groups, moving beyond the focus on single-group numerical dominance or even pairs of groups [146, 139]. It is done by first distinguishing three levels of diversity

including *low*, *moderate*, and *high diversity*, and subsequently subdividing by a dominant racial group. Specifically, a low diversity POI is one with scaled entropy values less than or equal to 0.3707 (this is the maximum entropy when one of the six racial groups constitutes 85 percent of the population) or when one group makes up more than 80 percent of the total population. A POI with high diversity is characterized by scaled entropy values greater than or equal to 0.7414, and no single group constituting more than 45 percent of the total population (this ensures that the third and fourth-ranked groups maintain significant representation). Other POIs that do not fall into either the low or high diversity categories are classified as POIs with moderate diversity.

Subsequently, the three levels of diversity are further categorized based on the predominant racial group. In essence, there could be 13 unique types of racial diversity and dominance for the POIs in this study: Latino-dominated low diversity, Latino-dominated moderate diversity, Asian-dominated low diversity, Asian-dominated moderate diversity, Black-dominated low diversity, Black-dominated moderate diversity, White-dominated low diversity, White-dominated moderate diversity, American Indian-dominated low diversity, American Indian-dominated moderate diversity, Other-dominated low diversity, Other-dominated moderate diversity, and High diversity. Note that the High diversity category is not further classified by a dominant racial group because, by definition, a CBG with High diversity has no dominant racial or ethnic group. The reader may refer to [146] for details of the classification criteria.

## 4.5 Results

### 4.5.1 Residential racial diversity

Figure 4.2 (all the maps in this study were generated using Esri ArcGIS Pro v3.1.2 software) illustrates the spatial distribution and percentage composition of census block groups of LA County classified by residential racial diversity and dominance using the 2018 American Community Survey. In general, the percentage composition aligns with the racial-ethnic configuration of LA County from the census data [175], where Latinos made up 48.64%, Whites 25.92%, Asians 14.58%, Blacks 7.76%, American Indians 0.19%, and Other races 2.91%. Furthermore, these dominant residential racial groups are spatially segregated. For example, many White-dominated CBGs (i.e. CBGs with White low or moderate diversity) are concentrated along the coast and in suburban and remote mountain areas. Latino-dominated CBGs are away from the coast and more prevalent in downtown LA and its surrounding areas, as well as suburban areas to the west of Burbank city and several neighborhoods in the eastern part of LA County. Black-dominated CBGs cluster primarily in South Central Los Angeles. Asian-dominated CBGs are predominantly located in neighborhoods such as Chinatown, Koreatown, Torrance City, San Gabriel Valley, and Rowland Heights.

### 4.5.2 Racial diversity in POI visits

Next, we evaluate the dynamic racial diversity at the POI level by considering the spatiotemporal dynamics of individuals' activities and visits to various types of places. To ensure sufficient observations, the POI visits for the entire month of March 2019 were aggregated and only differentiated by weekdays and weekends. The weekday dataset consists of 34,205 unique POIs and the weekend dataset has 31,382 POIs. The difference

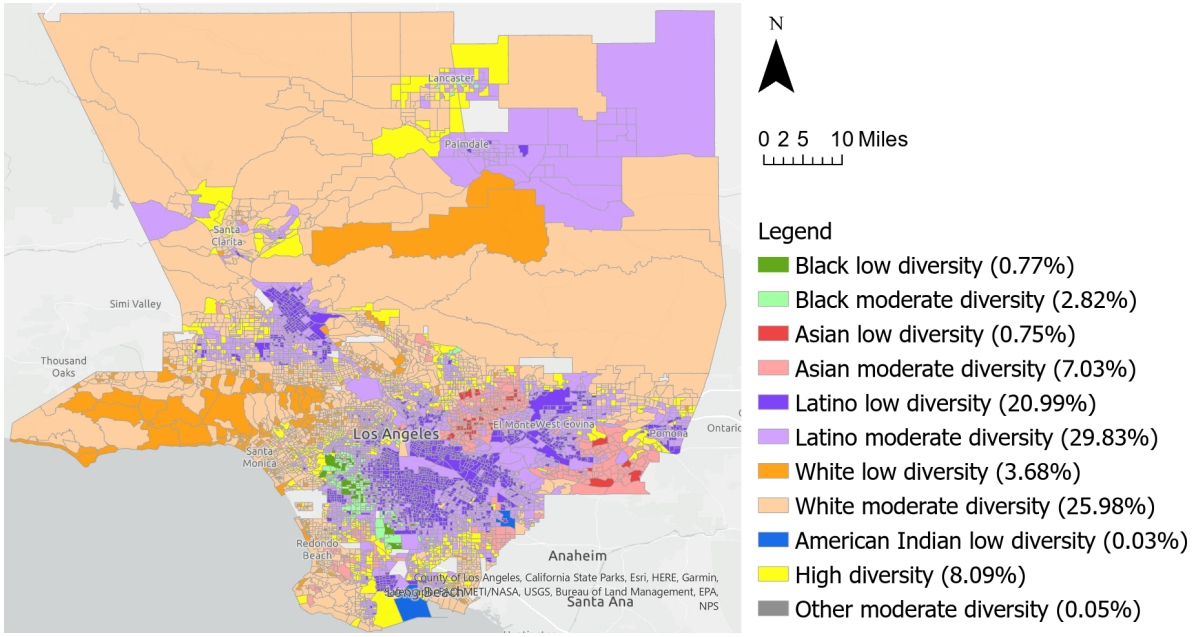



Figure 4.2: Los Angeles County's census block groups classified by residential racial diversity and dominance based on racial-ethnic composition from the 2018 American Community Survey.

in the number of POIs between weekdays and weekends is reasonable as certain businesses may be closed partially or completely during the weekend (e.g., banks, schools, government offices, some restaurants, and small local businesses). After aggregating the one-month POI visits, the weekday data contain 2,963,269 POI visit records from 242,108 distinct individuals, while the weekend data comprise 1,617,700 records from 226,710 distinct individuals. The racial diversity and dominance of POI visits are then computed based on these two datasets separately. The unique individuals in our data are not only from LA County but from across the entire state of California. In the next section, we discuss the disparities when considering different populations, such as those residing within the Greater LA region as opposed to the visitors from the entire California.

Figure 4.3 shows the spatial distribution of racial diversity and dominance types of POI visits on weekdays and weekends. Table 4.1 summarizes the percentage composition of POI racial diversity types during weekdays and weekends. On both weekdays

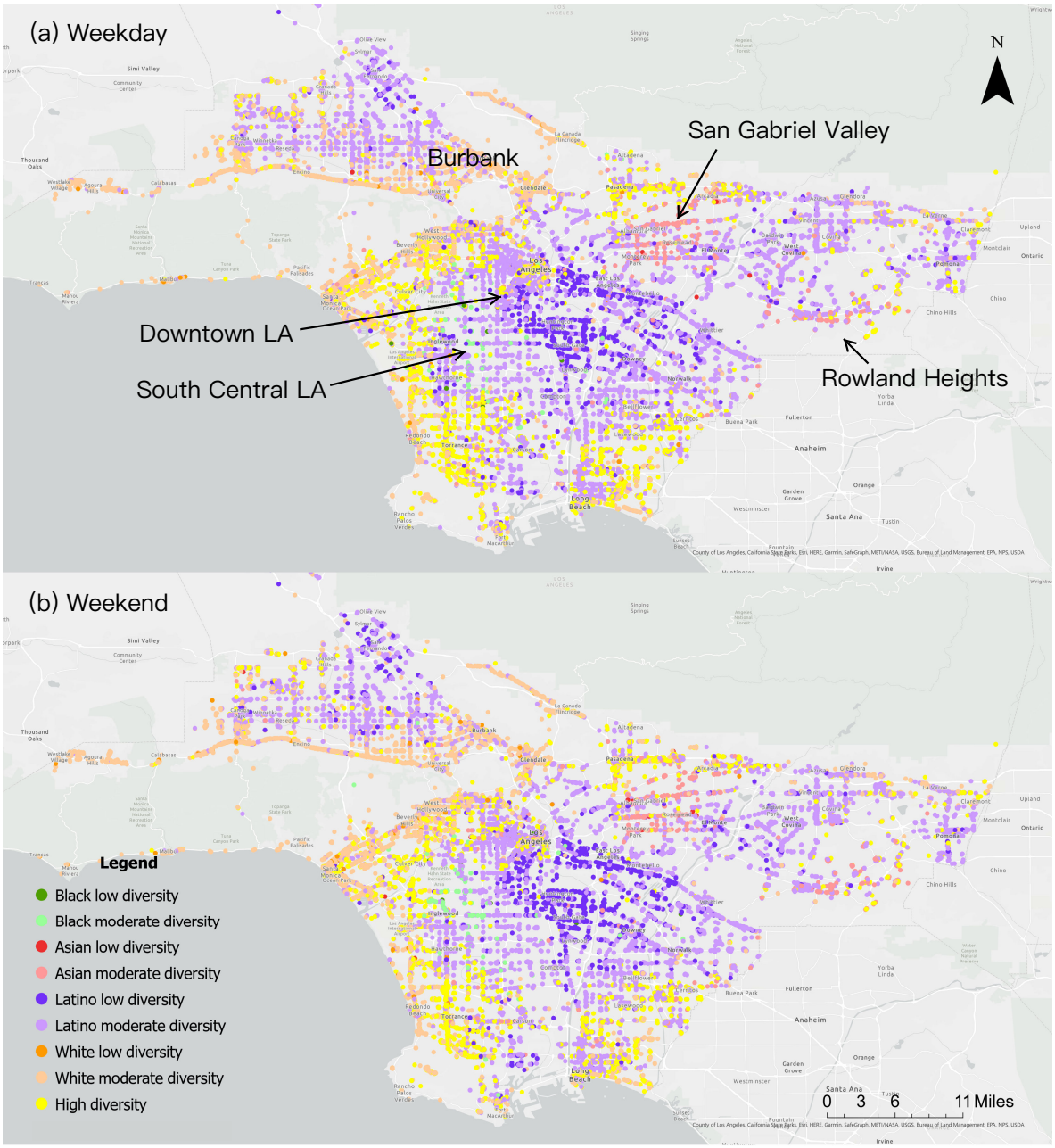



Figure 4.3: Spatial distribution of racial diversity and dominance categories of POI visits on (a) weekdays and (b) weekends in Los Angeles County based on visitors' home CBGs and mobile phone location data collected in March 2019. The map cropped out the northern suburban and less densely populated areas with fewer POIs, to better display the distribution of POIs in the main LA County areas.

and weekends, the majority of POI visits display Latino-dominated moderate diversity, followed by White-dominated moderate diversity, High diversity, Latino-dominated low

Table 4.1: Percentages of POI racial diversity and dominance types on weekdays and weekends.

| POI racial diversity type | Weekday | Weekend |
|---|---|---|
| Latino moderate diversity | 47.11% | 45.11% |
| White moderate diversity | 23.62% | 24.83% |
| High diversity | 20.43% | 19.25% |
| Latino low diversity | 5.28% | 6.55% |
| Asian moderate diversity | 2.68% | 2.88% |
| Black moderate diversity | 0.68% | 0.89% |
| White low diversity | 0.15% | 0.39% |
| Asian low diversity | 0.03% | 0.05% |
| Black low diversity | 0.02% | 0.04% |
| Total number of POIs | 34,205(100%) | 31,382(100%) |

diversity, and Asian-dominated moderate diversity. Other POI categories account for less than 1% of the total. When compared to the percentages of racial diversity types in residential CBGs, as depicted in Figure 4.2, Latino-dominated moderate diversity and White-dominated moderate diversity continue to be the top two types. The average scaled entropy value is 0.66 on weekdays but slightly drops to 0.64 during the weekend. A Mann-Whitney U test indicates a significant difference between the two values ($p$-value$<$0.01). This implies a significantly higher level of racial diversity (i.e., lower level of segregation) in POI visits on weekdays compared to weekends in LA County. This finding is consistent with existing literature [149]. One plausible explanation is that on weekends, individuals may prefer to socialize with their family and members of their own social group, but on weekdays they might be exposed more to other groups based on their work locations and the location and nature of their weekday activities.

The overall geographic distribution of various POI racial diversity types closely mirrors the residential racial diversity of the CBGs, as depicted in Figure 4.2. Presumably, individuals of the same racial-ethnic background tend to visit similar places. Black-dominated POIs cluster primarily in South Central Los Angeles. Latino-dominated

POIs are distributed widely across LA County. White-dominated POIs are predominantly located in coastal areas and along major freeways. Asian-dominated POIs are concentrated in several Asian-populated neighborhoods, such as San Gabriel Valley and Rowland Heights. Comparing Figures 4.2 and 4.3, it is worth noting that the High diversity POIs display a noticeably wider geographical distribution compared to the High diversity CBGs of LA residents. Notably, the High diversity category constitutes 8% of the total CBGs, while there are nearly 20% High diversity POIs. These observations suggest that individuals' daily mobility in urban areas enhances social integration among various racial-ethnic groups, particularly at these POIs characterized by a diverse mixing of races and ethnicities.

**Spatial clusters of racial diversity**

Subsequently, we apply Local Indicators of Spatial Association (LISA) [182] analysis (specifically Local Moran's I) on the scaled entropy index to assess spatial segregation in terms of the racial diversity level in POI visits. Local Moran's I is a widely used tool for identifying spatial concentrations of high values or low values, and spatial outliers. In the context of our study, the outcomes of Local Moran's I indicate four types of significant spatial association: high-high clusters, representing POIs with high racial diversity spatially clustered with neighboring POIs having high racial diversity; low-low clusters, indicating POIs with low racial diversity spatially clustered with POIs with low racial diversity; high-low clusters, indicating POIs with high racial diversity surrounded primarily by POIs with low racial diversity; low-high clusters, denoting POIs with low racial diversity surrounded primarily by POIs with high racial diversity. In particular, we are interested in low-low and high-high clusters, as they may indicate the presence of spatial isolation, segregation, and integration areas in terms of racial mixing.

Figure 4.4 depicts the spatial distribution, counts and percentages of high-high and

low-low clusters ($p$-value<0.05) of racial diversity of POI visits during weekdays and weekends, as analyzed using Local Moran's I through Esri ArcGIS Pro v3.1.2. The decrease in the percentage of high-high clusters from weekdays to weekends further reinforces our previous analysis that the average scaled entropy is lower during weekends. The high-high clusters (shown in red) indicate places where higher levels of social integration or social mixing occurred, whereas the low-low (shown in blue) clusters may suggest more isolated areas where only one or very few ethnic-racial groups are predominantly present. The high-high clusters tend to be close to coastal regions, suburban areas, and major freeways, while the low-low clusters predominantly concentrate in inland areas. In general, there is no significant variation in the spatial segregation of high and low racial diversity observed between weekdays and weekends.

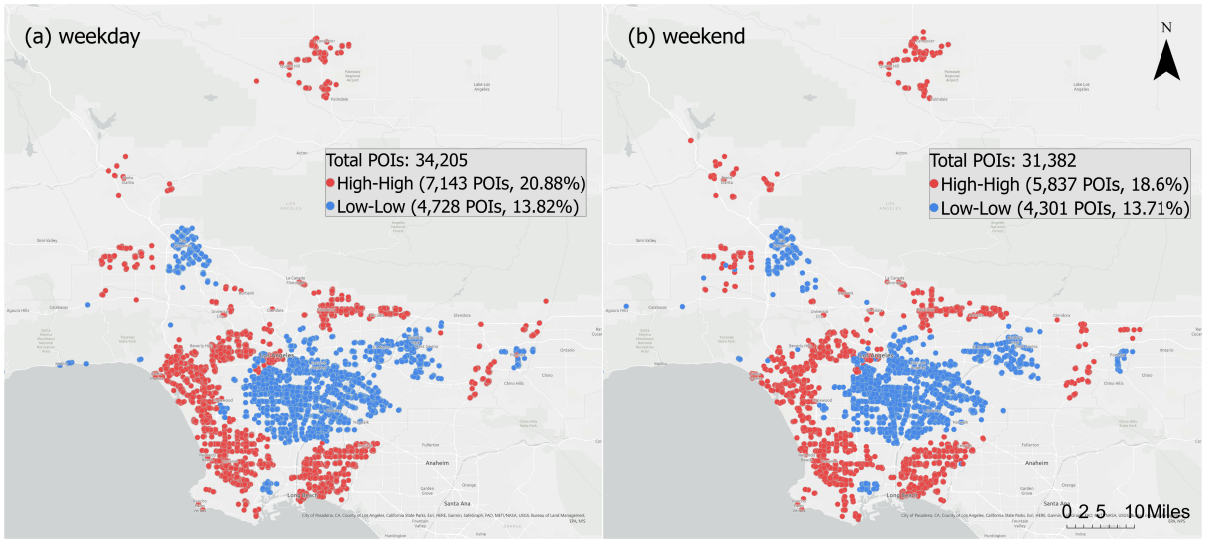



Figure 4.4: Significant high-high clusters (in red) and low-low clusters (in blue) of racial diversity in POI visits on (a) weekdays and (b) weekends using Local Moran's I ($p$-value<0.05).

These high-high clusters and low-low clusters are subsequently associated with their respective racial diversity and dominance categories (see Figure 4.3). Table 4.2 summarizes the POI count and percentage of each POI racial diversity type in high-high

and low-low clusters. As expected, more than half of the POIs in high-high clusters are intrinsically High diversity types in terms of racial mixing. The rest of the POIs at least exhibit a moderate level of racial diversity. Within the 4,728 POIs classified into low-low clusters, two major Latino-dominant types emerge: 3,065 (64.83%) POIs exhibit Latino moderate diversity, and 1,552 (32.83%) POIs display Latino low diversity. In other words, approximately one-third of the POIs in low-low clusters are primarily or, at times, almost exclusively visited only by the Latino population. This suggests that the Latino population tends to stay in their own communities and experience a higher level of segregation in their daily activities. Similar to the observation on weekdays, on weekends, more than half of the total number of POIs within high-high clusters exhibit High diversity, followed by Latino moderate diversity, White moderate diversity, Asian moderate diversity, and Black moderate diversity. In terms of low-low clusters on the weekend sample, Latino moderate diversity and Latino low diversity remain the two predominant types.

Table 4.2: POI count and percentage of each POI racial diversity type in high-high and low-low clusters.

| POI racial diversity type | Weekday | | Weekend | |
|---|---|---|---|---|
| | High-high | Low-low | High-high | Low-low |
| High diversity | 4,480 (62.72%) | 0 | 3,471 (59.47%) | 0 |
| Latino moderate diversity | 1,620 (22.68%) | 3,065 (64.83%) | 1,340 (22.96%) | 2,479(57.64%) |
| White moderate diversity | 946 (13.24%) | - | 906 (15.52%) | - |
| Asian moderate diversity | 77 (1.08%) | - | 106 (1.82%) | - |
| Black moderate diversity | 20 (0.28%) | - | 14 (0.24%) | - |
| Latino low diversity | 0 | 1,552 (32.83%) | 0 | 1,736(40.36%) |
| Total number of POIs | 7,143 (100%) | 4,728 (100%) | 5,837 (100%) | 4,301 (100%) |

Note: The dash indicates below 1%. Other types in low-low clusters less than 1% are also omitted in this table.

We then associate these spatial clusters with their corresponding POI business categories to better contextualize the observed spatial segregation patterns in racial diversity. As described in Section 4.4.1, each POI is categorized with one of the ten major business

divisions. Table 4.3 summarizes the percentage composition of each POI business category in high-high clusters and low-low clusters, with the total number of POIs on weekdays (34,205 POIs) and on weekends (31,382 POIs) as the denominators, respectively. The retail trade POIs have the highest count and percentage among all high-high or low-low clusters on weekdays and weekends. This observation is consistent with the fact that the retail trade category stands out as the dominant one, constituting 39.63% and 41.24% of the total POIs on weekdays and weekends, respectively. The retail trade division encompasses establishments involved in the sale of merchandise for personal or household consumption, including businesses like grocery stores, eating and drinking places, home furniture, furnishings, and equipment stores, among others. Overall, we observe a lower count of POIs for each POI business type in high-high or low-low clusters during weekends compared to weekdays. This could be attributed to a reduced number of businesses operating during the weekends. Furthermore, the percentage changes in low-low clusters between weekdays and weekends are comparatively smaller than the percentage changes in high-high clusters. This observation may indicate that low-low clusters, corresponding to racially isolated areas, tend to exhibit greater stability between weekday and weekend activities, whereas high-high clusters, representative of racially diverse areas, can demonstrate a higher spatial variation in their activities between weekdays and weekends.

**Time of day variation in racial diversity in POI visits**

Three types of businesses (according to SIC codes) frequently visited in people's daily lives which may provide opportunities for social interaction are selected to provide additional context for understanding the time-of-day variation in racial diversity and social mixing. The box plots in Figure 4.5 indicate the time of day variation of the scaled entropy index of eating places[7] (in total 8,734 POIs), grocery stores (1,704 POIs), and

[7]See the definition of eating places: `https://www.osha.gov/sic-manual/5812`

Table 4.3: POI count and percentage of each POI business category in high-high and low-low clusters.

| POI business category | Weekday | | Weekend | |
|---|---|---|---|---|
| | High-high | Low-low | High-high | Low-low |
| Retail trade | 2,933(8.58%) | 1,924(5.63%) | 2,523(8.04%) | 1,822(5.81%) |
| Finance, insurance and real estate | 1,178(3.44%) | 861(2.52%) | 912(2.91%) | 777(2.48%) |
| Manufacturing | 990(2.89%) | 467(1.37%) | 842(2.68%) | 419(1.34%) |
| Transportation and public utilities | 844(2.47%) | 598(1.75%) | 600(1.91%) | 545(1.74%) |
| Services | 697(2.04%) | 537(1.57%) | 567(1.81%) | 465(1.48%) |
| Agriculture, forestry and fishing | 493(1.44%) | 331(0.97%) | 389(1.24%) | 262(0.83%) |
| Public administration | 7(0.02%) | 10(0.03%) | 3(0.01%) | 10(0.03%) |
| Total number of POIs | 34,205(100%) | 34,205(100%) | 31,382(100%) | 31,382(100%) |

Note: POI business categories less than 0.01% are omitted in this table, such as Mining and Construction.

physical fitness facilities (798 POIs) in LA County on weekdays in March 2019. These POIs are grouped according to the residential racial diversity type of the CBG in which the POI is situated. Several groups with less than ten POIs are excluded from this figure. The box plots to the left of the vertical dashed line illustrate the distribution of residential racial diversity. This can be considered a default diversity level for each POI without considering people's daily activities. The box plots labeled 'Morning', 'Noon', 'Afternoon', and 'Evening' illustrate the distribution of the scaled entropy index for the same groups as in the 'Default' group. The Mann-Whitney U test results confirm that all changes in morning, noon, afternoon, and evening compared to the default entropy values are statistically significant ($p$-value$<0.05$).

At the start of the day, we can observe a substantial rise in the racial diversity level of the three selected types of POIs located in low diversity CBGs, including those dominated by Black, Asian, Latino, and White populations. For eating places, on average the Black-dominated low diversity group increases from 0.33 (default value) to 0.59 (daily average of the four values of morning, noon, afternoon, and evening), Asian-dominated low diversity increases from 0.33 to 0.62, Latino-dominated low diversity increases from

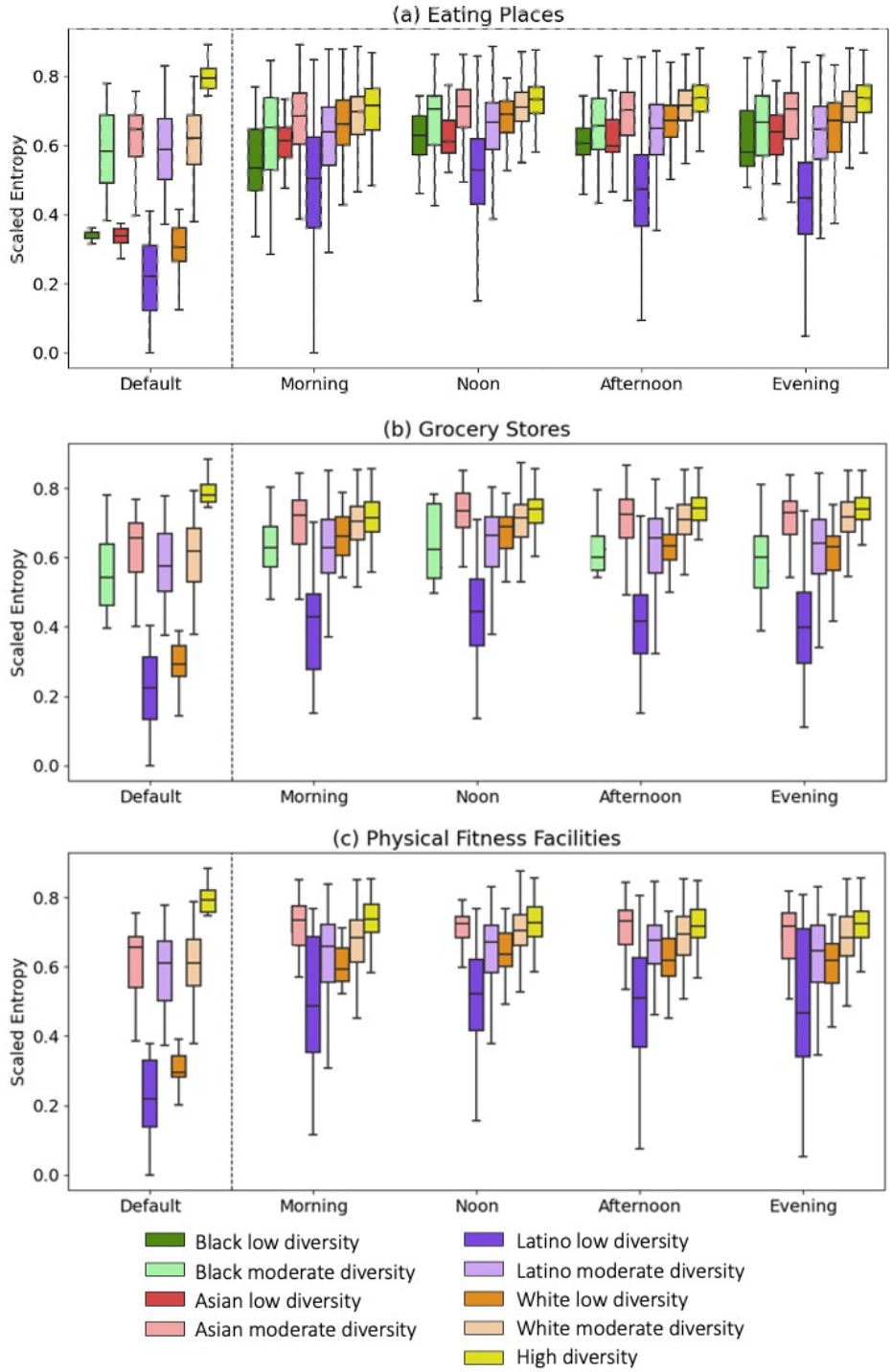



Figure 4.5: Time of day variation in the scaled entropy index for visits to (a) eating places, (b) grocery stores, and (c) physical fitness facilities in LA County on weekdays in March 2019, grouped by POI's residential racial diversity and dominance type.

0.21 to 0.48, and White-dominated low diversity increases from 0.29 to 0.66. This suggests that despite these eating places being situated in areas with low diversity, daily mobility and activities of individuals contribute to more racial mixing and social integration. Note that the rise of the scaled entropy index in the Latino-dominated low diversity group is the most modest among the four groups. Presumably, Latino-dominated low diversity areas are more preferred by the Latino population. The increase in the White-dominated low diversity group is the biggest which may suggest that restaurants operating in White-dominated low diversity areas attract the most racially diverse population.

The increase in racial diversity levels in visits to grocery stores is relatively smaller than the increase observed in restaurants. This is logical since individuals usually tend to visit grocery stores near their residences. Specifically, within the low diversity groups of POIs dominated by Latino and White populations, we observe the following changes: the Latino-dominated low diversity group increases from 0.22 to 0.42, and the White-dominated low diversity group increases from 0.3 to 0.63. Similar to the observations in restaurants, the increase in racial diversity level in visits to grocery stores of the Latino-dominated low diversity group is the smallest among all POI groups. Presumably, grocery stores near Latino neighborhoods are more attractive to the Latino population. Grocery stores in the White-dominated low diversity group also display the biggest increase in racial diversity levels. These grocery stores in White-dominated low diversity areas might not have the same racial specificity as those in Latino-dominated low diversity areas, such as Mexican-specific grocery stores.

In terms of physical fitness facilities, we observe an increase from 0.22 to 0.5 in the Latino-dominated low diversity group, and the White-dominated low diversity group from 0.3 to 0.61. The rise in the racial diversity level in visits to physical fitness facilities in the Latino-dominated low diversity group is nearly the same as the increase observed in restaurants and surpasses the increase in grocery stores. This makes sense as physical

fitness facilities typically have no racial specificity, and individuals may choose gyms that are not necessarily close to their homes but can be situated near their workplaces, for instance. In addition, the values of scaled entropy in the Latino-dominated low diversity group exhibit the widest distribution (i.e., high variance) in comparison to all other POI groups. This may suggest that some gyms (e.g., especially POIs within the first quartile of the box plot) located in Latino-dominated neighborhoods attract mostly Latino customers.

The scaled entropy index peaks at noon or afternoon compared to morning and evening as depicted in Figure 4.5. Specifically, for all eating places, the overall scaled entropy reaches 0.66 during noon hours and 0.65 during afternoon hours, both of which surpass the entropy values during mornings (0.63) and evenings (0.64). Similarly, for all grocery stores, the overall scaled entropy is 0.65 during noons and 0.644 during afternoons, both of which exceed the values noted during mornings (0.63) and evenings (0.642). In the case of all physical fitness facilities, the overall scaled entropy stands at 0.67 for both noons and afternoons, which is higher than the values seen during mornings (0.66) and evenings (0.66). This pattern is likely because noontime and afternoon are typically the busiest times of the day when people frequently visit restaurants, grocery stores, and gyms, leading to increased social interactions among individuals from diverse racial groups.

Furthermore, among the four moderate diversity groups, the entropy value, in general, also increases from its default value derived from residential racial diversity. Nevertheless, these increases are relatively smaller compared to the four low diversity groups. A noticeable declination is observed in the High diversity group: for eating places, the scaled entropy decreases from 0.8 to 0.72; for grocery stores, it drops from 0.79 to 0.73; and for physical fitness facilities, it decreases from 0.79 to 0.72. This outcome is unsurprising, given that the initial entropy value for the High diversity group is already exceptionally

high, and it remained notably high across all four periods.

### 4.5.3 Comparison between the Greater LA and the entire California

Previous analyses of racial diversity in POI visits focused on individuals from across the entire state of California, including those residing in LA County, who visited POIs in LA County. It's reasonable to assume that individuals who live close to LA County may have a more significant impact on the dynamics of racial diversity in POI visits in LA County compared to individuals residing elsewhere in California. Building upon this insight, this section explores the disparities in racial diversity when examining populations within the Greater Los Angeles region as opposed to visitors from the entire California. We examine the Greater LA region rather than solely LA County because it is common in California, especially among residents in the Greater LA region, to commute from neighboring counties to LA County for work or other purposes. The Greater Los Angeles region encompasses Ventura County in the west, extending to San Bernardino County and Riverside County in the east, with Los Angeles County at its center, Kern County in the north, and Orange County to the south. By focusing on individuals whose home locations are situated in the Greater LA region, the new weekday dataset comprises 2,907,007 POI visit records (34,172 unique POIs) from 226,170 unique individuals, while the weekend dataset consists of 1,569,466 POI visit records (31,315 unique POIs) from 210,377 unique individuals.

Table 4.4 summarizes the descriptive statistics and Mann-Whitney U test results for the scaled entropy index of the Greater LA sample and the entire California sample. This comparison is made for weekdays and weekends across four distinct periods. We also break down the analysis by more than 160 POI categories according to SIC code instead

of using all POIs in the significant tests; however, no significant differences were found for any POI category. As indicated in the table, the racial diversity of POI visits tends to be lower during weekends compared to weekdays. Specifically, weekend mornings exhibit the lowest diversity level, followed by weekend evenings, weekend noons, and weekend afternoons. On weekdays, the lowest diversity level is observed during the evenings, followed by mornings, afternoons, and noons. The overall scaled entropy index suggests a decrease of 0.001 to 0.003 when only considering the Greater LA sample. The scaled entropy index for all periods on weekends, as observed in the Greater LA sample, exhibits a significant decrease compared to the entire California sample. However, no significant differences are observed during weekdays. These findings suggest that individuals from regions outside the Greater LA areas in California contribute positively to racial diversity observed in POI visits or social integration in LA County during weekends. Presumably, individuals visiting from outside Greater LA during weekends are often tourists, and they are likely to come from more diverse racial and ethnic backgrounds.

Table 4.4: Descriptive statistics and Mann-Whitney U test results for the scaled entropy index of the Greater LA sample and the entire California sample.

| Period | The Greater LA | | | California | | | Difference |
|---|---|---|---|---|---|---|---|
| | POI count | Mean | SD | POI count | Mean | SD | |
| Weekday morning | 29,502 | 0.626 | 0.15 | 29,584 | 0.627 | 0.149 | -0.001 |
| Weekday noon | 31,831 | 0.645 | 0.136 | 31,891 | 0.647 | 0.135 | -0.002 |
| Weekday afternoon | 32,600 | 0.642 | 0.139 | 32,655 | 0.643 | 0.138 | -0.001 |
| Weekday evening | 27,220 | 0.625 | 0.152 | 27,275 | 0.626 | 0.152 | -0.001 |
| Weekend morning | 23,888 | 0.603 | 0.165 | 24,051 | 0.606 | 0.164 | -0.003* |
| Weekend noon | 28,287 | 0.626 | 0.152 | 28,393 | 0.628 | 0.151 | -0.002** |
| Weekend afternoon | 28,330 | 0.629 | 0.149 | 28,416 | 0.631 | 0.148 | -0.002* |
| Weekend evening | 22,178 | 0.614 | 0.159 | 22,288 | 0.616 | 0.158 | -0.002* |

Note: *$p$ <0.1; **$p$ <0.05.

## 4.6 Discussion and Future Directions

This study introduces a novel approach for assessing dynamic racial diversity in Point of Interest (POI) visits, building upon the traditional residential-based racial diversity framework. Leveraging one-month privacy-enhanced mobile phone location data collected in Los Angeles County, this study explores the racial diversity and spatial segregation in POI visits in LA County.

Empirical results from LA County show a significantly higher level of racial diversity in POI visits during weekdays compared to weekends. The outcomes suggest a notable spatial segregation in the racial diversity and dominance in POI visits within LA County. The overall geographic distribution of various POI racial diversity types closely mirrors the patterns of residential racial diversity. This suggests that individuals of the same racial-ethnic background tend to visit similar places. Notably, POIs classified under the High diversity category exhibit a more extensive geographic presence when compared to High diversity Census Block Groups (CBGs). This observation suggests that individuals' daily mobility in urban areas enhances social integration among various racial-ethnic groups, particularly at these POIs characterized by a diverse mixing of races and ethnicities. While the analysis reported here is based on visit counts, we also considered the duration of individual activity spent at each POI. However, no significant differences were observed.

A LISA analysis suggests significant high-high and low-low clusters in POI's racial diversity. The outcomes reveal that the high-high clusters tend to be close to coastal regions, suburban areas, and major freeways, while the low-low clusters predominantly concentrate in inland areas. Approximately one-third of the POIs in low-low clusters are primarily or, at times, almost exclusively visited only by the Latino population. This suggests that the Latino population tends to stay in their own communities and

experience a higher level of segregation in their daily activities. By associating spatial clusters with unique POI types, we find that low-low clusters, corresponding to racially isolated areas, tend to exhibit greater stability between weekday and weekend activities, whereas high-high clusters, representative of racially diverse areas, can demonstrate a higher spatial variation in their activities between weekdays and weekends. Furthermore, POIs located in areas with low diversity experience a substantial increase in diversity levels during the day, especially around noon and afternoon. This observation aligns with the reality that noontime and afternoon are typically the busiest times when people often visit restaurants and various stores, leading to increased social interactions among diverse racial groups. Lastly, a comparison between visitors from the Greater LA region and the entire state of California identifies a significant decrease in racial diversity in POI visits on weekends when focusing solely on the Greater LA sample. This implies that individuals from outside the Greater LA region in California have a positive impact on the racial diversity observed in visits to POIs on weekends in LA County.

Several limitations need further investigation. A limitation of the mobile phone data is that it may not accurately represent the actual population, and also we cannot discern the purpose of visits or the types of social interactions. For example, the visits do not represent whether they involve strangers co-locating randomly in these places without socially interacting or friends catching up. Hence, the racial diversity metrics computed in our study should be regarded as a proxy for all meaningful social interactions and as an upper limit for experienced racial diversity. Second, our current focus is primarily on investigating the dynamics of racial diversity within urban areas. Nevertheless, it is important to acknowledge that racial and ethnic relations can differ significantly across various socio-geographical spaces [183]. Future studies may explore the disparities in racial diversity among various urban settings in different locations and incorporate people's socioeconomic characteristics and built environment to provide further insights into

the segregation patterns in terms of racial diversity in people's daily destination choice [184].

# Chapter 5

# Conclusions and Future Work

In this chapter, I articulate the main contributions of this research and respond to the three proposed research questions of this dissertation. This is followed by a discussion of the limitations of this dissertation and plans for future research endeavors.

## 5.1 Research Contributions

The main contribution of this dissertation is the advancement of computational methods for quantifying and identifying spatiotemporal patterns of human daily activities and interactions, using both conventional travel diaries and massive movement data of individuals. This dissertation also contributes significant new insights into understanding human activity and interaction patterns in various socio-geographical spaces, such as the heterogeneity in daily activity-travel patterns between two occupational groups including telecommuters and commuters, the differences between intra-household and inter-household interactions identified from massive continuous and long-term movement data, and social mixing and segregation among multiple social groups in a metropolitan area. These insights are further summarized in my response to each research question in

Section 5.2. Below, I outline each specific contribution made in this dissertation.

**A novel joint pattern recognition method for daily activity-travel pattern analysis.** Characterizing multi-faceted human daily behaviors is also known as the challenge of the "curse of dimensionality" in travel behavior studies [185]. To tackle this challenge, Chapter 2 proposes a new joint pattern recognition method that combines network motif-based analysis with activity sequence analysis. The combination of building a taxonomy of daily mobility patterns using the concept of motif with the taxonomy of minute-by-minute daily time allocation patterns derived from sequence analysis enables the study of interdependency in space and time of multiple facets of behavior that are fundamental in building activity-based models without imposing a priori behavioral assumptions. This method can be used directly in simplifying modeling and simulation of daily activity schedules by decreasing the number of possible combinations of the many variables used in activity microsimulation. The proposed method can also be used to compare daily patterns across diverse cultures and under different policies [186]. For example, exploring the destination choice and activity duration in different motif and time allocation patterns can be correlated with land use types. This enables us to infer the shifts in destination choice and time allocation changes under different scenarios of land use, supporting the investigation of land use dynamics in urban areas.

**A new time-geographic approach to trace interactions between moving individuals using movement data.** Recent advancements have utilized time geography or PPA intersections to identify potential concurrent or delayed interactions between moving individuals based on their movement data [47, 48, 49, 50]. However, current methodologies fall short in their ability to quantify the duration of dynamic interactions and/or the frequency of contact between moving individuals, crucial factors for distinguishing different types of human interactions. To fill this gap, Chapter 3 extends

existing time-geographic-based interaction analytics to incorporate the quantification of the duration or frequency of interaction between moving individuals using movement data. The proposed method extracts subsequences of continuous interaction segments to estimate the duration of interaction, considering the temporal order or sequential order of the intersecting PPAs.

**A theoretical contribution to conceptualizing taxonomy of human interactions using movement data.** While current approaches have advanced the capability to identify and trace patterns of interaction in movement tracking data, the definitions and classifications of interaction between moving individuals remain isolated. Existing classifications for interaction analysis fall short of fully capturing the spatiotemporal properties of human interaction or reflecting the duration and nature of interactions [50, 112, 172], critical information for quantifying and contextualizing human interaction. To address this challenge, Chapter 3 introduces a novel classification framework for dynamic human interaction patterns observed from movement data. In this proposed framework, interaction between moving individuals is categorized into three types based on both the time it occurs and its duration: encounter (brief duration), concurrent interaction (long duration), and delayed interaction. An empirical analysis of interactions among 986 people using their movement tracking data demonstrates the efficacy of this classification framework in capturing these three types of human interactions.

**An open-source Python package ORTEGA v1.0 for time-geographic-based interaction analysis using movement data and contextual variables.** While there is an exciting growth in research in computational methods for interaction analysis using movement data, there remain challenges regarding reproducibility and replicability of the existing approaches. The existing movement interaction analysis tools are often designed for the dyadic interaction of animals (i.e., a pair of moving animals) [187].

They are less accessible or tested for broader use in applications for human interaction of multiple individuals. To fill this gap, this dissertation presents ORTEGA v1.0, an Object-oRiented TimE-Geographic Analytical approach, as an open-source Python package for analyzing potential interactions between a pair or multiple moving entities based on the observation of their movement. The ORTEGA v1.0 package contributes two significant capabilities: 1) the functions to identify potential interactions (e.g., brief encounters, concurrent interactions, delayed interactions) from movement data of two or more entities using a time-geographic-based approach; and 2) the capacity to compute attributes of potential interaction events including start time, end time, interaction duration, and difference in movement parameters such as speed and moving direction, and also contextualize the identified potential interaction events. To the best of my knowledge, the present ORTEGA is the first open-source Python package that offers a time-geographic-based interaction analytical tool.

**A dynamic place-based measure for the diversity of interaction at activity locations between diverse social groups.** A widely used measure for assessing diversity such as racial-ethnic diversity in Chapter 4 is the scaled entropy index, which is commonly used to study residential segregation [177, 178, 146]. I extend the scaled entropy index to assess place-based racial diversity at the POI level, accounting for the spatiotemporal dynamics of individuals' daily activities and their visits to these locations. The methodology can be applied at the finest spatial scale, specifically at the POI level, and within various temporal contexts such as time of the day, day of the week, weekday versus weekend, etc. While the case study in this dissertation focused on racial-ethnic diversity, this approach can be extended to comprehend dynamic interactions and the underlying socioeconomic activities among moving individuals across diverse social contexts, such as various age groups, different income levels, and comparisons between

new immigrants and local residents, among others.

## 5.2 Discussions

This dissertation delves into important questions to understand heterogeneity in people's daily activity-travel patterns and interactions across different socio-geographical spaces. To conclude how I achieve this goal, I revisit my three initial research questions below by summarizing key findings and discussing their implications for each research question.

**Research Question 1***: Are there significant differences in daily activity-travel patterns between different occupational groups, such as telecommuters and commuters?*

In response to RQ 1, first, approximately 20% of the telecommuters stay at home all day during a workday, while only 8% of the commuters do the same. For those who at least make one trip on workdays, around 2% of telecommuters present motif 1 pattern characterized by a single location-based loop trips pattern (mostly are taking a walk around home), while very few commuters (less than 1%) use this motif pattern. In addition, telecommuters have by far greater variety in their motifs with more complex tour formation and a distinct time of day allocation to activities and travel from the commuters. Commuters show a substantial proportion in routine morning and afternoon peaks of arriving at and departing from work. Telecommuters do not perform work tasks only from home. Instead, during the day they travel to a variety of locations to visit customers and/or use their spatiotemporal flexibility to perform work tasks anywhere they want. In contrast, more than 80% of commuters perform work at their workplace.

Recall that in Chapter 1, I discussed that travel behavior researchers are interested in understanding the group decision-making process regarding how people allocate time to themselves and others in activity and travel. For instance, [101] studied the intra-

household interactions of dual-earning household members and found a very strong association between telecommuting frequency and time allocation. With the proposed approach, I explored these associations without imposing the strong assumptions required in regression-like statistical models. In Chapter 2, I conducted an intra-household analysis to understand the combinations of telecommuters and commuters within dual-earner and three-earner households, as well as how a household member's daily activities and travel can influence other household members' activities and travel on the same day. The results show that two commuters is the most popular combination in dual-earner households; likewise, three commuters is the dominant combination for three-earner households. The second most popular combinations for dual-earner and three-earner households are combinations that have one telecommuter plus one commuter for dual-earner households and one telecommuter plus two commuters. I found that 54.67% of telecommuters are women while 44.68% of commuters are women in dual-earner households, which demonstrates that a higher percentage of women are more likely to be telecommuters in dual-earner households. The combination of daily time-of-day patterns of activities and travel for the one telecommute plus one commuter households shows that given a household member works as a commuter, the other household member working as a telecommuter has much more diversity in daily time-of-day patterns and takes on more household responsibilities such as escorting kids to school, grocery shopping, and so forth. This also reflects the substitution characteristics in van Wissen's classification of intra-household interactions [29].

**Research Question 2***: What can long-term movement data reveal about dynamic interactions between individuals?*

In this study, interactions identified through movement observations suggest potential contact, though such contact does not necessarily imply physical or social interaction.

However, it may present opportunities for potential social interaction. In response to RQ 2, I developed a new time geographic approach extending the existing Object-oRiented TimE-Geographic Analytical approach ("ORTEGA" herein), initially introduced in [50], to analyze potential interactions between pairs of moving individuals based on the long-term observation of their movement. The novelty of the proposed approach lies in its incorporation of the capability to measure the duration and frequency of interactions. This is achieved by considering the spatiotemporal continuity properties of intersecting PPAs between moving individuals along their movement paths. By considering the duration, distinctions can be made between various types of interactions, such as brief encounters on the street and long-duration interactions when individuals travel together or share the same space. The proposed method was contrasted with the traditional proximity-based approach (which relies on the intersection of buffers of two individuals along their movement paths) in the experiment, aiming to compare its efficacy in identifying and quantifying different types of interactions.

Specifically, I utilized the travel diaries of the same participants to validate the intra-household interactions identified from GPS tracking data because individuals might travel together if they report common trips in their travel diaries. The findings illustrate that ORTEGA can provide a reasonably accurate estimation of the duration of intra-household interactions. However, the proximity-based approach underestimates potential interactions, which may result in more intermittent interactions with shorter durations. In terms of the impact of the temporal scale of movement data on interaction analysis, I found that the proposed ORTEGA and the proximity-based can identify very closely the number of intra-household interactions and the numbers remain stable along with the increasing sampling intervals. ORTEGA can also identify substantially more inter-household interactions while the the proximity-based approach almost failed to capture any. Presumably, people living in the same household are more likely to travel together

during the day [21], which increases the chance of physical contact with each other. Even when the sampling interval is increased, these people are still very likely to be tracked simultaneously when they engage in the same activity or travel together. In contrast, people from different households are less likely to travel together and their physical interactions are often unintentional (e.g., two strangers pass by each other on the street) and asynchronous (e.g., two persons visit the same grocery store at different times). The proximity-based approach is less effective in this situation because it requires synchronous tracking and a predefined distance buffer to determine if two moving entities come close to each other. However, ORTEGA does not have such constraints as it is based on the actual locations accessible to people, their time budget, and speed capacities.

As mentioned in Chapter 1, conventional travel surveys typically face limitations in terms of sample size and observation period due to the high costs associated with data collection. The ability of ORTEGA to capture the characteristics of various interpersonal interactions from movement data, especially inter-household interactions, presents a promising avenue to supplement travel surveys. This opens the door for future investigations into the decision-making process when individuals allocate time to activities and travel in the company of household members and individuals from different households.

**Research Question 3***: To what degree can the analysis of individual movement data contribute to our understanding of collective behaviors between communities, such as social mixing and segregation between diverse groups?*

In exploring RQ 1, I have examined the spatiotemporal variations in the daily activity-travel patterns of different occupational groups, namely telecommuters and commuters. In addressing RQ 2, I elucidated the distinct characteristics of intra-household and inter-household interactions that occurred along movement paths or activity spaces of moving individuals. The investigation of RQ 3 centered on understanding human daily activities

and interactions in a broader socio-geographical space, particularly social mixing and segregation among multiple social groups in a metropolitan area. This is also relevant to the discussion of the representativeness of movement data and the generalizability of movement research [188, 189, 190, 191]. Below, I summarize the key findings in response to RQ 3 and engage in a discussion on the representativeness and generalizability of movement research.

In response to RQ 3, Chapter 4 developed a dynamic place-based measure of the diversity in human interactions among different social groups at the POI level and various temporal contexts. Leveraging one month of privacy-enhanced mobile phone location data of 283,556 anonymized individuals collected in Los Angeles County along with the POI data, the conducted case study explores the racial diversity and spatial segregation simultaneously in people's daily interactions at various types of places. Empirical results indicate a significantly higher level of racial diversity in POI visits during weekdays compared to weekends. The findings confirm that individuals' daily mobility in urban areas enhances social integration among various racial-ethnic groups. The outcomes also suggest significant spatial clusters of isolated and integrated areas regarding racial mixing in people's everyday activities. For instance, the Latino population tends to stay in their own communities and experience a higher level of segregation in their daily activities. Places with higher levels of social integration or social mixing tend to be close to coastal regions (e.g., tourist areas) and along major freeways (e.g., commercial areas).

Movement data, such as individual traces collected from mobile phones, represent smaller subsets of the user population [190]. Consequently, this limitation may further restrict the generalizability of conclusions drawn from the analysis of such data [188, 189, 190, 191]. In addition, new location-aware technologies are generating abundant data, but these data points often lack detailed attribution [11]. This poses the challenge of identifying who these moving individuals are (e.g., sociodemographic char-

acteristics), and what actual activities they are participating in (i.e., activity type or trip purpose). Ultimately, it constrains the ability to comprehend the underlying socioeconomic activities or environments through the analysis of human movement and interaction [192, 193]. This dissertation has, in part, tackled these challenges and alleviated bias in mobile phone data through three approaches: (1) Utilizing census data (i.e., American Community Survey) at the Census Block Group (CBG) level to infer individual demographic characteristics, such as race-ethnic type, based on the census of their residential CBG. (2) A post-stratification technique to address potential sampling bias issues in mobile phone location data. Specifically, each individual's POI visit count was weighted by an expansion factor which is the ratio of the population of their home CBG to the number of individuals observed in the same CBG from the mobile phone data. (3) Inferring individual activity types according to visits to POI, and the POI business categories and open hours. The outcomes as I summarized above demonstrate the capability of utilizing passive and massive mobile phone traces to capture intricate human interactions, particularly in understanding how different social groups engage with one another on a large metropolitan scale.

## 5.3 Limitations and Future Works

In this section, I discuss several limitations inherent in this dissertation and propose potential avenues for future research that could improve this dissertation.

**Ground truth and validation for interaction analysis.** Deciphering the purpose of visits or the nature of social interactions can be exceptionally challenging when relying on movement data [193], such as the GPS tracking data utilized in Chapter 3, or mobile phone traces as discussed in Chapter 4. For example, the POI visits do not represent whether they involve strangers co-locating randomly in these places without

socially interacting or friends catching up. Hence, the interaction analysis in this dissertation should be regarded as an upper limit for all possible human interactions. To tackle this challenge, future research could leverage multidimensional contextual data such as regional household travel surveys to validate the analytical outcomes derived from these massive movement data [194, 195]. A regional household travel survey gathers detailed self-reported information including locations visited and transportation modes used for each trip. One example is the validation of intra-household interactions (derived from ORTEGA) in Chapter 3 using CHTS survey data. However, it is challenging to validate inter-household interactions, such as between strangers, using household travel surveys. In addition to its role as a validation dataset, it may be more reliable to generate inclusive conclusions based on comparing the spatial and temporal patterns resulting from multidimensional contextual data and massive movement data. Travel surveys can also be integrated into models for imputing trip purposes, thereby addressing missing values within the extensive movement data such as mobile phone traces [196]. In a word, multidimensional contextual data sets can be used not only for validation purposes but also as valuable complements in refining models and enriching analyses based on large-scale movement data.

**Limitations of time geography assumptions for interaction analysis.** Even though using the time-geography framework was confirmed to be a more robust way to identify interactions between moving individuals, several limitations remain open problems. First, the assumption that potential concurrent interactions between two individuals can occur when their PPAs intersect or overlap spatially and temporally might be problematic. Human movement is confined to many factors such as the physical environment (e.g., transportation network, road network, transit routes), individual capabilities, etc. It is important to incorporate the physical environment, individual characteristics,

and individual perception of accessibility when measuring the accessible locations for moving individuals [197]. For example, a possible extension of the current interaction analysis techniques may be to consider using network-based potential path areas to more accurately depict the accessible area to moving individuals [198, 199]. Second, the current approaches are limited to deterministic design which means the algorithms can only identify whether or not a potential interaction is possible between mobile individuals. Future efforts should focus on developing probabilistic approaches that can not only identify potential interactions but also present the probability of interactions during the movement across the potential path areas [115, 116, 117, 118]. For instance, Context-Aware Movement Analysis (CAMA) might be a promising approach to this goal [200, 201, 202, 203]. This idea originates from the movement ecology community. The goal is to integrate auxiliary contextual data in the modeling of animal movement to understand their responses to environmental conditions or disruptions [201]. For example, by incorporating movement parameters and contextual variables such as step length, displacement from contact, and percentage of forest cover, [108] inferred behavioral changes of white-tailed deer before, during, and after contact with mock hunters. However, the idea has not been fully incorporated into human movement analysis literature.

**Sampling bias in mobile phone location data.** This dissertation uses mobile phone traces to explore racial diversity and spatial segregation in people's daily visits to various places (see Chapter 4). Massive and passive movement data such as cell phone traces typically provide relatively precise locations in terms of activities and movement paths during travel. However, a drawback of mobile phone data lies in its potential sampling bias, as it may not accurately represent the entire population [189]. Given smartphones are available nowadays to the majority of the population, the new passive data collection is biased towards the younger group as they are more likely to be familiar with new tech-

nologies. To address these limitations, future analyses could explore the integration of both survey data and digital trace data [204]. This approach may create synthetic populations that offer a more accurate representation of the overall population, acknowledging and mitigating potential biases associated with relying solely on mobile phone data. An alternative method involves scrutinizing and quantifying sampling bias in mobile phone data through a comparison with the demographic configuration outlined in the census population. Furthermore, future research should take into account the inclusiveness issue of this type of big data [205]. Existing big data collections often exclude individuals from economically disadvantaged backgrounds, social minority groups, or marginalized populations (e.g., homeless individuals without cell phones) [190]. This exclusion may have negative implications or biases for economic opportunities, social mobility, and various inferences derived from the analysis of the big data. In some cases, it could even lead to reinforcing and exacerbating existing problems [190]. Future research leveraging big data such as mobile phone traces should consider empathy and equity issues during data collection as well as when making inferences [205].

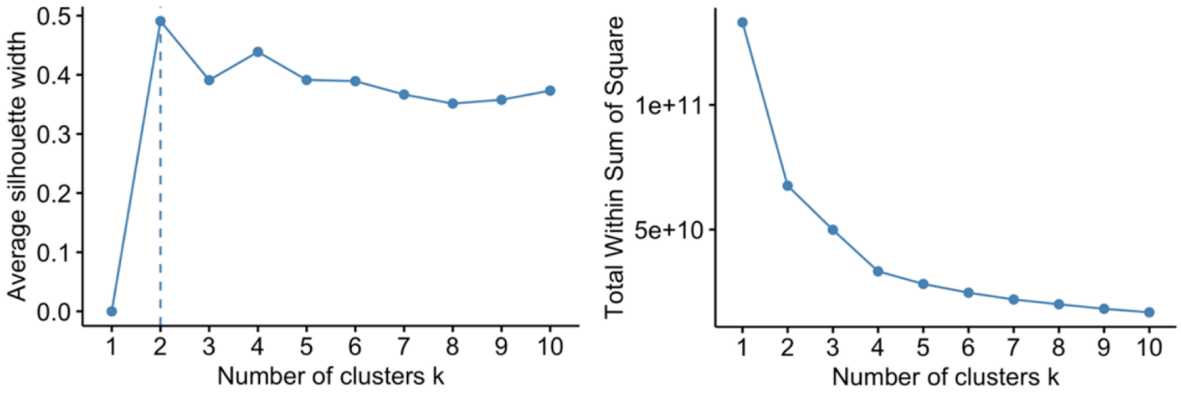
**Uncertainty in movement data.** Global navigation satellite systems, such as the Global Positioning System (GPS), have supported ubiquitous sensors for collecting the movement of objects in geographical space, typically recorded as trajectories. Despite the unprecedented accuracy in global positioning provided by satellite navigation, GPS trajectories are susceptible to positional uncertainty or errors [206, 207]. Two main types of errors inherent in movement data are measurement error and interpolation error [206]. Measurement error arises from the inherent limitation in determining the actual position of an object by the measurement system. Interpolation error refers to the limitation in representing the actual motion between two sampled points, influenced by the sampling rate. Both types of errors are present in trajectories recorded with GPS, and they prop-

agate through further computational analyses [208, 209]. In my specific case, the impact of uncertainty in the original GPS positions can propagate through movement parameter computation (e.g., speed, movement direction, time budget), a fundamental step for constructing PPAs, and subsequently, causing inaccurate inferences by sophisticated interaction analysis presented in Chapter 3 (e.g., duration analysis, differentiating between brief duration encounters and long duration concurrent interactions). These errors may also affect the analysis of racial diversity in Chapter 4, introducing uncertainties in the estimation of individuals' activity locations. Consequently, this may lead to problematic inferences about interactions between social groups. In my future research, I aim to evaluate the influence of positional uncertainty in human movement data on interaction analysis. Specifically, one possible approach is to explore the existence of spatial autocorrelation in GPS measurement errors within the datasets.

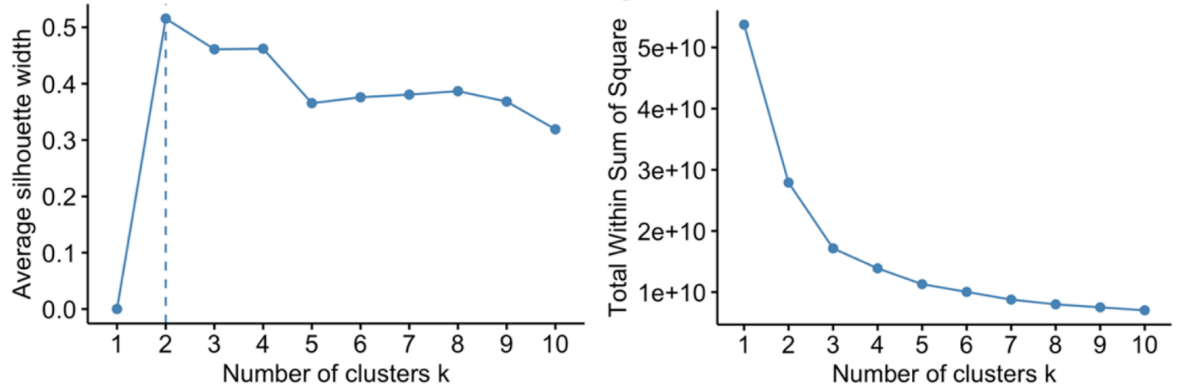
# Appendix A

## A.1 Determining the optimal number of clusters for motif groups

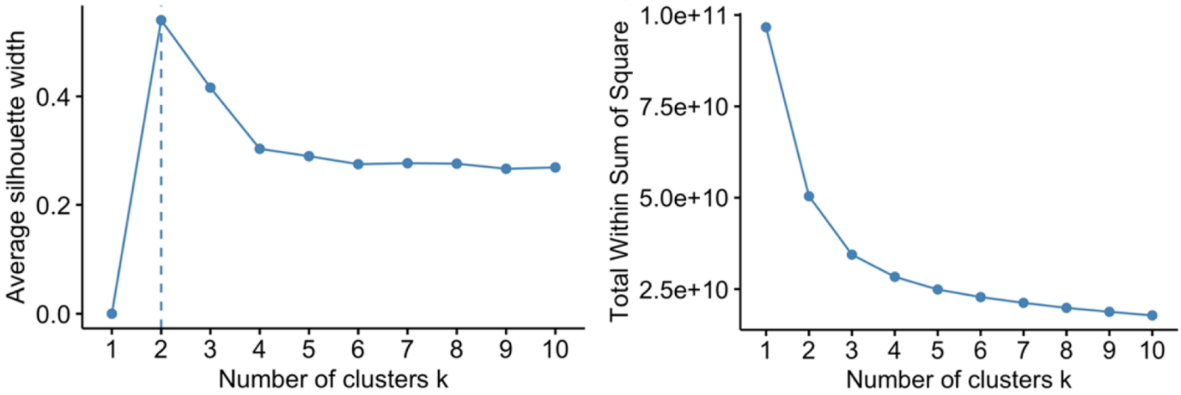
The silhouette graphs suggest the best solution concerning dissimilarity among clusters is the 2-cluster solution with the highest average silhouette coefficient for all three motif groups. However, selecting 4 clusters yields very small WSS and may not be worth selecting a solution with more than 4 clusters. These two indicators show that 2 or 4 clusters would be an optimal solution. Selection then between 2 and 4 clusters is left to the researcher using behavioral interpretation of the patterns produced by the 2-cluster and 4-cluster solutions. By comparing the time allocation patterns produced by the 2-cluster and 4-cluster solutions, we finally chose 4 clusters as the optimal solution because it could identify more distinct activity sequence patterns, for example, the Mostly Out of Home Day and Long Work from Home Day patterns. In contrast, the 2-cluster solution cannot (see Figures A.1 and A.2).




(a) Group T.III

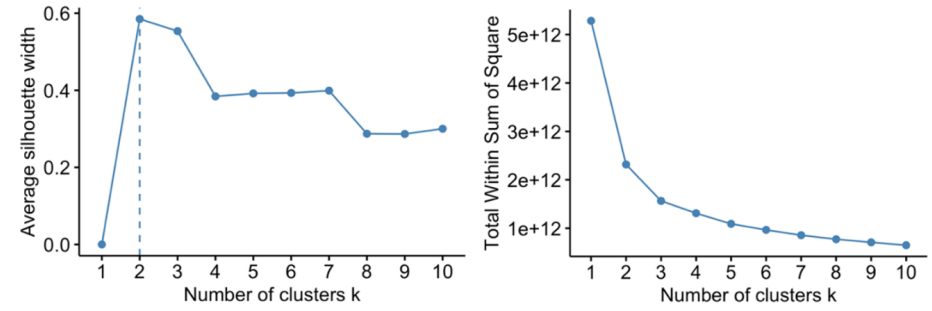



(b) Group T.IV

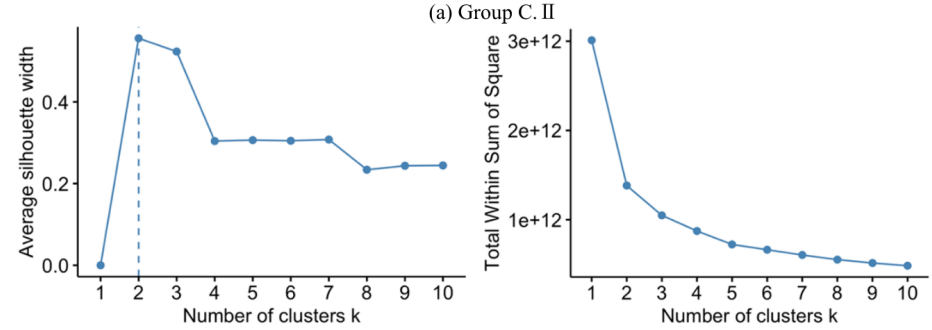



(c) Group T.V

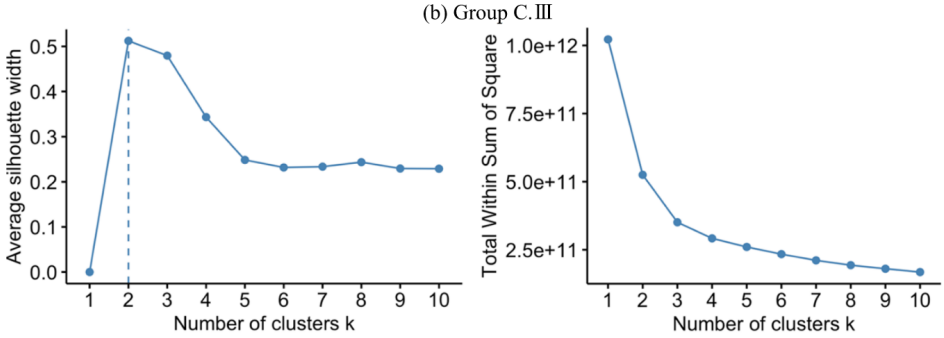
Figure A.1: Silhouette and WSS elbow methods for motifs Groups T.III, T.IV, and T.V.



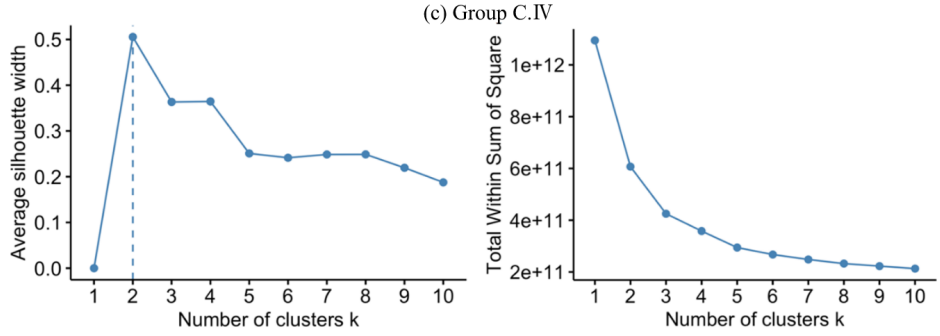

(a) Group C. II




(b) Group C. III




(c) Group C.IV




(d) Group C. V

Figure A.2: Silhouette and WSS elbow methods for motifs Groups C.II, C.III, C.IV, and C.V.The two indicators show that 2, 4, or 5 clusters would be an optimal solution. By comparing the time allocation patterns produced by the 2-cluster, 4-cluster, and 5-cluster solutions, we finally adopt the 4-cluster solution for Groups C.II and C.V, and the 5-cluster solution for Groups C.III and C.IV.

# Bibliography


[1] A. H. Maslow, *A theory of human motivation.*, *Psychological Review* **50** (1943), no. 4 370–396. Place: US Publisher: American Psychological Association.

[2] J. S. House, K. R. Landis, and D. Umberson, *Social relationships and health*, *Science* **241** (1988), no. 4865 540–545. Publisher: American Association for the Advancement of Science.

[3] J. S. Coleman, *Social capital in the creation of human capital*, *American journal of sociology* **94** (1988) S95–S120. Publisher: University of Chicago Press.

[4] R. Forrest and A. Kearns, *Social Cohesion, Social Capital and the Neighbourhood*, *Urban Studies* **38** (Nov., 2001) 2125–2143.

[5] R. D. Putnam, *Bowling alone: The collapse and revival of American community.* Simon and schuster, 2000.

[6] N. E. Phillips, B. L. Levy, R. J. Sampson, M. L. Small, and R. Q. Wang, *The Social Integration of American Cities: Network Measures of Connectedness Based on Everyday Mobility Across Neighborhoods*, *Sociological Methods & Research* **50** (Aug., 2021) 1110–1149.

[7] R. Kitamura, *An evaluation of activity-based travel analysis*, *Transportation* **15** (1988), no. 1-2 9–34. ISBN: 0049-4488 Publisher: Springer.

[8] K. W. Axhausen and T. Gärling, *Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems*, *Transport reviews* **12** (1992), no. 4 323–341. ISBN: 0144-1647 Publisher: Taylor & Francis.

[9] C. R. Bhat and F. S. Koppelman, *Activity-based modeling of travel demand*, in *Handbook of transportation Science*, pp. 35–61. Springer, 1999.

[10] R. Joo, M.-P. Etienne, N. Bez, and S. Mahévas, *Metrics for describing dyadic movement: a review*, *Movement Ecology* **6** (Dec., 2018) 26.

[11] H. J. Miller, S. Dodge, J. Miller, and G. Bohrer, *Towards an integrated science of movement: converging research on animal movement ecology and human mobility*

*science*, *International Journal of Geographical Information Science* **33** (May, 2019) 855–876.

[12] S. Dodge, *A Data Science Framework for Movement*, *Geographical Analysis* **53** (Jan., 2021) 92–112.

[13] U. Demšar, J. A. Long, F. Benitez-Paez, V. Brum Bastos, S. Marion, G. Martin, S. Sekulić, K. Smolak, B. Zein, and K. Siła-Nowicka, *Establishing the integrated science of movement: bringing together concepts and methods from animal and human movement analysis*, *International Journal of Geographical Information Science* (Feb., 2021) 1–36.

[14] T. Hägerstrand, *What about people in regional science?*, in *Papers of the Regional Science Association*, vol. 24, pp. 1–12, 1970.

[15] B. Lenntorp, *Paths in space-time environments: a time-geographic sudy of movement possibilities of individuals*, *Lund Studies in Geography Number 44, Royal University of Lund, Sweden* (1976).

[16] L. Burns, *Transportation, Temporal and Spatial Components of Accessibility.*, *Lexington, MA: Lexington Books* (1979).

[17] H. J. Miller, *A Measurement Theory for Time Geography*, *Geographical Analysis* **37** (Jan., 2005) 17–45.

[18] S.-L. Shaw and H. Yu, *A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space*, *Journal of Transport Geography* **17** (Mar., 2009) 141–149.

[19] T. F. Golob and M. G. McNally, *A model of activity participation and travel interactions between household heads*, *Transportation Research Part B: Methodological* **31** (June, 1997) 177–194.

[20] P. Jones, *Developments in dynamic and activity-based approaches to travel analysis*, *Oxford Studies in Transport* (1990).

[21] J. H. Lee and K. G. Goulias, *Companionship and time investment in social fields at different life cycle stages: Implications for activity and travel modeling and simulation*, *Research in Transportation Economics* **68** (Aug., 2018) 18–28.

[22] K. G. Goulias, E. C. McBride, and R. Su, *Life cycle stages, daily contacts, and activity-travel time allocation for the benefit of self and others*, in *Mobility and Travel Behaviour Across the Life Course* (J. Scheiner and H. Rau, eds.). Edward Elgar Publishing, Dec., 2020.

[23] S. Srinivasan and C. R. Bhat, *Modeling household interactions in daily in-home and out-of-home maintenance activity participation*, *Transportation* **32** (Sept., 2005) 523–544.

[24] K. K. Srinivasan and S. R. Athuru, *Analysis of within-household effects and between-household differences in maintenance activity allocation*, *Transportation* **32** (Sept., 2005) 495–521.

[25] M. Bradley and P. Vovsha, *A model for joint choice of daily activity pattern types of household members*, *Transportation* **32** (Sept., 2005) 545–571.

[26] P. A. Samuelson, *Social indifference curves*, *The quarterly journal of economics* **70** (1956), no. 1 1–22. Publisher: MIT Press.

[27] T. A. Townsend, *The effects of household characteristics on the multi-day time allocations and travel/activity patterns of households and their members*. Unpublished Ph.D. Dissertation. Evanston, IL: Northwestern University, 1987.

[28] F. Koppelman and T. Townsend, *Task allocation among household members: theory and analysis*, in *Fifth International Conference on Travel Behavior, Aix-en-Provence, France*, 1987.

[29] L. J. van Wissen, *A model of household interactions in activity patterns*, *In: Presented at the International Conference on Dynamic Travel Behavior Analysis, Kyoto.* (1989).

[30] J. C. Tronto, *Women and caring: What can feminists learn about morality from caring?[1989]*, in *Justice and care*, pp. 101–116. Routledge, 1995.

[31] S. Srinivasan and C. R. Bhat, *An exploratory analysis of joint-activity participation characteristics using the American time use survey*, *Transportation* **35** (May, 2008) 301–327.

[32] K. G. Goulias and T.-G. Kim, *An analysis of activity type classification and issues related to the with whom and for whom questions of an activity diary*, *Berkeley: University of California Transportation Center* (2005).

[33] K. M. N. Habib and J.-A. Carrasco, *Investigating the Role of Social Networks in Start Time and Duration of Activities: Trivariate Simultaneous Econometric Model*, *Transportation Research Record: Journal of the Transportation Research Board* **2230** (Jan., 2011) 1–8.

[34] S. Milgram, *The familiar stranger: An aspect of urban anonymity*, 1977. Pages: 51–53 Publication Title: The individual in a social world: Essays and experiments.

[35] P. Van Den Berg, A. Kemperman, and H. Timmermans, *Social Interaction Location Choice: A Latent Class Modeling Approach*, *Annals of the Association of American Geographers* **104** (Sept., 2014) 959–972.

[36] J. Zhou, Y. Yang, H. Ma, and Y. Li, *"Familiar strangers" in the big data era: An exploratory study of Beijing metro encounters*, *Cities* **97** (Feb., 2020) 102495.

[37] E. R. Dugundji, A. Páez, T. A. Arentze, J. L. Walker, J. A. Carrasco, F. Marchal, and H. Nakanishi, *Transportation and social interactions*, *Transportation Research Part A: Policy and Practice* **45** (May, 2011) 239–247.

[38] M. A. Hogg, *Chapter 6 Social Identity Theory*, in *Contemporary Social Psychological Theories* (P. J. Burke, ed.), pp. 111–136. Stanford University Press, Redwood City, 2006.

[39] N. Agarwal, H. Liu, S. Murthy, A. Sen, and X. Wang, *A social identity approach to identify familiar strangers in a social network*, in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 3, pp. 2–9, 2009. Issue: 1.

[40] S. Y. A. Tsui and A. S. Shalaby, *Enhanced system for link and mode identification for personal travel surveys based on global positioning systems*, *Transportation Research Record* **1972** (2006), no. 1 38–45. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[41] P. Widhalm, P. Nitsche, and N. Brändie, *Transport mode detection with realistic smartphone sensor data*, in *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*, pp. 573–576, IEEE, 2012.

[42] J. Wolf, R. Guensler, and W. Bachman, *Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data*, *Transportation Research Record* **1768** (2001), no. 1 125–134. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[43] P. R. Stopher, Q. Jiang, C. FitzGerald, and others, *Processing GPS data from travel surveys*, *2nd international colloqium on the behavioural foundations of integrated land-use and transportation models: frameworks, models and applications, Toronto* (2005).

[44] P. Stopher, N. Swann, and C. FitzGerald, *Using an odometer and a GPS panel to evaluate travel behaviour changes*, *TRB Transp. Appl* (2007).

[45] H. Li, R. Guensler, and J. Ogle, *Analysis of morning commute route choice patterns using global positioning system–based vehicle activity data*, *Transportation Research Record* **1926** (2005), no. 1 162–170. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[46] J. Hood, E. Sall, and B. Charlton, *A GPS-based bicycle route choice model for San Francisco, California*, *Transportation letters* **3** (2011), no. 1 63–75. Publisher: Taylor & Francis.

[47] J. A. Downs, D. Lamb, G. Hyzer, R. Loraamm, Z. J. Smith, and B. M. O'Neal, *Quantifying spatio-temporal interactions of animals using probabilistic space–time prisms*, *Applied Geography* **55** (Dec., 2014) 1–8.

[48] J. A. Long, S. L. Webb, T. A. Nelson, and K. L. Gee, *Mapping areas of spatial-temporal overlap from wildlife tracking data*, *Movement ecology* **3** (2015), no. 1 1–14. Publisher: Springer.

[49] B. A. Hoover, J. A. Miller, and J. Long, *Mapping areas of asynchronous-temporal interaction in animal-telemetry data*, *Transactions in GIS* **24** (June, 2020) 573–586.

[50] S. Dodge, R. Su, J. Johnson, A. Simcharoen, K. Goulias, J. L. Smith, and S. C. Ahearn, *ORTEGA: An object-oriented time-geographic analytical approach to trace space-time contact patterns in movement data*, *Computers, Environment and Urban Systems* **88** (July, 2021) 101630.

[51] B. Kuijpers and W. Othman, *Modeling uncertainty of moving objects on road networks via space–time prisms*, *International Journal of Geographical Information Science* **23** (Sept., 2009) 1095–1117.

[52] B. Kuijpers, R. Grimson, and W. Othman, *An analytic solution to the alibi query in the space–time prisms model for moving object data*, *International Journal of Geographical Information Science* **25** (Mar., 2011) 293–322.

[53] J. A. Long, T. A. Nelson, S. L. Webb, and K. L. Gee, *A critical examination of indices of dynamic interaction for wildlife telemetry studies*, *Journal of Animal Ecology* **83** (2014), no. 5 1216–1233. Publisher: Wiley Online Library.

[54] K. Cukier and V. Mayer-Schoenberger, *The rise of big data: How it's changing the way we think about the world*, *Foreign Aff.* **92** (2013) 28. Publisher: HeinOnline.

[55] M. Batty, *Smart Cities, Big Data*, *Environment and Planning B: Planning and Design* **39** (Apr., 2012) 191–193.

[56] P. Widhalm, Y. Yang, M. Ulm, S. Athavale, and M. C. González, *Discovering urban activity patterns in cell phone data*, *Transportation* **42** (July, 2015) 597–623.

[57] S. Dodge, R. Weibel, S. C. Ahearn, M. Buchin, and J. A. Miller, *Analysis of movement data*, *International Journal of Geographical Information Science* **30** (May, 2016) 825–834.

[58] S. K. Boell, D. Cecez-Kecmanovic, and J. Campbell, *Telework paradoxes and practices: the importance of the nature of work*, *New Technology, Work and Employment* **31** (July, 2016) 114–131. Publisher: John Wiley & Sons, Ltd.

[59] P. Mokhtarian, *If telecommunication is such a good substitute for travel, why does congestion continue to get worse?*, *Transportation Letters* **1** (Jan., 2009) 1–17. Publisher: Taylor & Francis.

[60] L. Taskin and V. Devos, *Paradoxes from the Individualization of Human Resource Management: The Case of Telework*, *Journal of Business Ethics* **62** (Nov., 2005) 13–24.

[61] U. S. O. of Personnel Management, *Status of Telework in the Federal Government: Report to Congress Fiscal year 2016*, *OPM.GOV and Telework.gov* (2017).

[62] U. S. O. of Personnel Management, *Telework Enhancement Act of 2010, Public Law 111-292*, *Telework.gov* (2017).

[63] A. Sekar, E. Williams, and R. Chen, *Changes in Time Use and Their Effect on Energy Consumption in the United States*, *Joule* **2** (Mar., 2018) 521–536.

[64] G. W. Analytics, *Telecommuting Statistics*, *https://globalworkplaceanalytics.com/telecommuting-statistics* (2020).

[65] U. S. B. of Labor Statistics, *American Time Use Survey Summary*, *https://www.bls.gov/news.release/atus.nr0.htm* (2020).

[66] G. Tal, *Reduced Overestimation in Forecasting Telecommuting as a Travel Demand Management Policy*, *Transportation Research Record* **2082** (Jan., 2008) 8–16. Publisher: SAGE Publications Inc.

[67] B. Vilhelmson and E. Thulin, *Who and where are the flexible workers? Exploring the current diffusion of telework in Sweden*, *New Technology, Work and Employment* **31** (Mar., 2016) 77–96. Publisher: John Wiley & Sons, Ltd.

[68] J. I. Gimenez-Nadal, J. A. Molina, and J. Velilla, *Telework, the Timing of Work, and Instantaneous Well-Being: Evidence from Time Use Data*, IZA Discussion Papers 11271, Institute of Labor Economics (IZA), Bonn, 2018.

[69] J. C. Messenger and L. Gschwind, *Three generations of Telework: New ICTs and the (R)evolution from Home Office to Virtual Office*, *New Technology, Work and Employment* **31** (Nov., 2016) 195–208.

[70] Pliskin Nava, *The telecommuting paradox*, *Information Technology & People* **10** (Jan., 1997) 164–172. Publisher: MCB UP Ltd.

[71] C. Weinbaum, B. Triezenberg, E. Meza, and D. Luckey, *Understanding Government Telework: An Examination of Research Literature and Practices from Government Agencies*. RAND Corporation, 2018.

[72] B. A. Lautsch, E. E. Kossek, and S. C. Eaton, *Supervisory approaches and paradoxes in managing telecommuting implementation*, *Human Relations* **62** (May, 2009) 795–827. Publisher: SAGE Publications Ltd.

[73] V. Illegems and A. Verbeke, *Telework: What Does it Mean for Management?*, *Long Range Planning* **37** (Aug., 2004) 319–334.

[74] J. M. Nilles, F. R. Carlson, P. Gray, and G. Hanneman, *Telecommuting-an alternative to urban transportation congestion*, *IEEE Transactions on Systems, Man, and Cybernetics* (1976), no. 2 77–84. Publisher: IEEE.

[75] J. M. Nilles, *Traffic reduction by telecommuting: A status review and selected bibliography*, *Transportation Research Part A: General* **22** (July, 1988) 301–317.

[76] R. Kitamura, J. Nilles, P. Conroy, and D. M. Fleming, *Telecommuting as a Transportation Planning Measure:Initial Results of California Pilot Project*, *University of California Transportation Center* (1990).

[77] P. J. Valk and R. Hellot, *Telecommuting in Southern California in 1995*, *Transportation Research Record* **1606** (Jan., 1997) 86–94. Publisher: SAGE Publications Inc.

[78] P. L. Mokhtarian, *Defining Telecommuting*, *UC Davis: Institute of Transportation Studies* (1991).

[79] I. Salomon and M. Salomon, *Telecommuting: The employee's perspective*, *Technological Forecasting and Social Change* **25** (Feb., 1984) 15–28.

[80] R. M. Pendyala, K. G. Goulias, and R. Kitamura, *Impact of telecommuting on spatial and temporal patterns of household travel*, *Transportation* **18** (Dec., 1991) 383–409.

[81] R. S. Gajendran and D. A. Harrison, *The good, the bad, and the unknown about telecommuting: Meta-analysis of psychological mediators and individual consequences.*, *Journal of Applied Psychology* **92** (Nov., 2007) 1524–1541.

[82] P. L. Mokhtarian, G. O. Collantes, and C. Gertz, *Telecommuting, Residential Location, and Commute-Distance Traveled: Evidence from State of California Employees*, *Environment and Planning A: Economy and Space* **36** (Oct., 2004) 1877–1897. Publisher: SAGE Publications Ltd.

[83] P. L. Mokhtarian, S. L. Handy, and I. Salomon, *Methodological issues in the estimation of the travel, energy, and air quality impacts of telecommuting*, *Transportation Research Part A: Policy and Practice* **29** (July, 1995) 283–302.

[84] Kitou Erasmia and Horvath Arpad, *Transportation Choices and Air Pollution Effects of Telework*, *Journal of Infrastructure Systems* **12** (June, 2006) 121–134. Publisher: American Society of Civil Engineers.

[85] P. Zhu, L. Wang, Y. Jiang, and J. Zhou, *Metropolitan size and the impacts of telecommuting on personal travel*, *Transportation* **45** (Mar., 2018) 385–414.

[86] H. Asgari and X. Jin, *Impacts of Telecommuting on Nonmandatory Activity Participation: Role of Endogeneity*, *Transportation Research Record* **2666** (2017), no. 1 47–57. _eprint: https://doi.org/10.3141/2666-06.

[87] C. M. Schneider, C. Rudloff, D. Bauer, and M. C. González, *Daily travel behavior: lessons from a week-long survey for the extraction of human mobility motifs related information*, in *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing - UrbComp '13*, (Chicago, Illinois), p. 1, ACM Press, 2013.

[88] J. Cao, Q. Li, W. Tu, and F. Wang, *Characterizing preferred motif choices and distance impacts*, *PLOS ONE* **14** (Apr., 2019) e0215242.

[89] M. G. McNally, *The four step model*, *Handbook of transport modelling* **1** (2000) 35–41. Publisher: Pergamon Oxford.

[90] S. Jiang, J. Ferreira, and M. C. González, *Clustering daily patterns of human activities in the city*, *Data Mining and Knowledge Discovery* **25** (2012), no. 3 478–510. Publisher: Springer.

[91] J. Drucker and A. J. Khattak, *Propensity to Work from Home: Modeling Results from the 1995 Nationwide Personal Transportation Survey*, *Transportation Research Record* **1706** (Jan., 2000) 108–117. Publisher: SAGE Publications Inc.

[92] X. Jin and J. Wu, *Propensity to Telecommute: Exploring the National Household Travel Survey*, *Transportation Research Record* **2231** (Jan., 2011) 110–119. Publisher: SAGE Publications Inc.

[93] A. Gabadinho, G. Ritschard, N. S. Müller, and M. Studer, *Analyzing and Visualizing State Sequences in R with TraMineR*, *Journal of Statistical Software* **40** (2011), no. 4.

[94] E. C. McBride, A. W. Davis, and K. G. Goulias, *Fragmentation in Daily Schedule of Activities using Activity Sequences*, *Transportation Research Record: Journal of the Transportation Research Board* **2673** (Apr., 2019) 844–854.

[95] E. C. McBride, A. W. Davis, and K. G. Goulias, *Sequence analysis of place-travel fragmentation in California*, in *Mapping the Travel Behavior Genome*, pp. 371–398. Elsevier, 2020.

[96] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*, vol. 344. John Wiley & Sons, 2009.

[97] C.-H. Joh, T. A. Arentze, and H. J. P. Timmermans, *A Position-Sensitive Sequence-Alignment Method Illustrated for Space–Time Activity-Diary Data*, *Environment and Planning A: Economy and Space* **33** (Feb., 2001) 313–338.

[98] H. J. Miller and J. Han, *Geographic data mining and knowledge discovery*. CRC press, 2009.

[99] M. Dijst and V. Vidakovic, *Travel time ratio: the key factor of spatial reach*, *Transportation* **27** (2000), no. 2 179–199.

[100] M.-P. Kwan, *Gender differences in space-time constraints*, *Area* **32** (June, 2000) 145–156.

[101] R. Paleti and I. Vukovic, *Telecommuting and Its Impact on Activity–Time Use Patterns of Dual-Earner Households*, *Transportation Research Record* **2658** (Jan., 2017) 17–25. Publisher: SAGE Publications Inc.

[102] P. J. Jackson and J. Van der Wielen, *Teleworking: international perspectives: from telecommuting to the virtual organisation*. Psychology Press, 1998.

[103] M. T. Institute, *University of Maryland COVID-19 Impact Analysis Platform, https://data.covid.umd.edu, accessed on NOV 9, University of Maryland, College Park, USA*, .

[104] Y. D. Pouri and C. R. Bhat, *On Modeling Choice and Frequency of Home-Based Telecommuting*, *Transportation Research Record* **1858** (Jan., 2003) 55–60. Publisher: SAGE Publications Inc.

[105] H. Asgari, X. Jin, and A. Mohseni, *Choice, Frequency, and Engagement: Framework for Telecommuting Behavior Analysis and Modeling*, *Transportation Research Record* **2413** (Jan., 2014) 101–109. Publisher: SAGE Publications Inc.

[106] H. Asgari and X. Jin, *Toward a comprehensive telecommuting analysis framework: Setting the conceptual outline*, *Transportation Research Record* **2496** (2015), no. 1 1–9. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[107] J. A. Miller, *Towards a better understanding of dynamic interaction metrics for wildlife: a null model approach*, *Transactions in GIS* **19** (2015), no. 3 342–361. Publisher: Wiley Online Library.

[108] J. A. Long, S. L. Webb, S. M. Harju, and K. L. Gee, *Analyzing Contacts and Behavior from High Frequency Tracking Data Using the wildlifeDI R Package*, *Geographical Analysis* (Sept., 2021) gean.12303.

[109] A. D. Marra, H. Becker, K. W. Axhausen, and F. Corman, *Developing a passive GPS tracking system to study long-term travel behavior*, *Transportation Research Part C: Emerging Technologies* **104** (July, 2019) 348–368.

[110] R. A. Kleinman and C. Merkel, *Digital contact tracing for COVID-19*, *Cmaj* **192** (2020), no. 24 E653–E656. Publisher: Can Med Assoc.

[111] G. Cencetti, G. Santin, A. Longa, E. Pigani, A. Barrat, C. Cattuto, S. Lehmann, M. Salathe, and B. Lepri, *Digital proximity tracing on empirical contact networks for pandemic control*, *Nature communications* **12** (2021), no. 1 1–12. Publisher: Nature Publishing Group.

[112] C. P. Doncaster, *Non-parametric estimates of interaction from radio-tracking data*, *Journal of theoretical biology* **143** (1990), no. 4 431–443. Publisher: Elsevier.

[113] H. J. Miller, *Modelling accessibility using space-time prism concepts within geographical information systems*, *International journal of geographical information systems* **5** (Jan., 1991) 287–301.

[114] R. Su, S. Dodge, and K. G. Goulias, *Understanding the impact of temporal scale on human movement analytics*, *Journal of Geographical Systems* (Feb., 2022).

[115] S. Winter, *Towards a probabilistic time geography*, in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, (Seattle Washington), pp. 528–531, ACM, Nov., 2009.

[116] S. Winter and Z.-C. Yin, *Directed movements in probabilistic time geography*, *International Journal of Geographical Information Science* **24** (Aug., 2010) 1349–1365.

[117] J. A. Downs, *Time-geographic density estimation for moving point objects*, in *Geographic Information Science: 6th International Conference, GIScience 2010, Zurich, Switzerland, September 14-17, 2010. Proceedings 6*, pp. 16–26, Springer, 2010.

[118] Y. Song and H. J. Miller, *Simulating visit probability distributions within planar space-time prisms*, *International Journal of Geographical Information Science* **28** (Jan., 2014) 104–125.

[119] D. Sui and M. Goodchild, *The convergence of GIS and social media: challenges for GIScience*, *International Journal of Geographical Information Science* **25** (Nov., 2011) 1737–1748.

[120] Y. M. Park and M.-P. Kwan, *Beyond residential segregation: A spatiotemporal approach to examining multi-contextual segregation*, *Computers, Environment and Urban Systems* **71** (Sept., 2018) 98–108.

[121] N. Shoval and M. Isaacson, *Sequence Alignment as a Method for Human Activity Analysis in Space and Time*, *Annals of the Association of American Geographers* **97** (June, 2007) 282–297.

[122] M. Batty, J. DeSyllas, and E. Duxbury, *The discrete dynamics of small-scale spatial events: agent-based models of mobility in carnivals and street parades*, *International Journal of Geographical Information Science* **17** (2003), no. 7 673–697. Publisher: Taylor & Francis.

[123] C. for Disease Control and Prevention, *COVID-19 Employer Information for Office Buildings*, *https://www.cdc.gov/coronavirus/2019-ncov/community/office-buildings.html (Accessed 8 Feb 2021)* (2020).

[124] N. Oliver, B. Lepri, H. Sterly, R. Lambiotte, S. Deletaille, M. De Nadai, E. Letouzé, A. A. Salah, R. Benjamins, C. Cattuto, V. Colizza, N. de Cordes, S. P. Fraiberger, T. Koebe, S. Lehmann, J. Murillo, A. Pentland, P. N. Pham, F. Pivetta, J. Saramäki, S. V. Scarpino, M. Tizzoni, S. Verhulst, and P. Vinck, *Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle*, *Science Advances* **6** (June, 2020) eabc0764.

[125] M. Génois and A. Barrat, *Can co-location be used as a proxy for face-to-face contacts?*, *EPJ Data Science* **7** (Dec., 2018) 11.

[126] E. Noi, A. Rudolph, and S. Dodge, *Assessing COVID-induced changes in spatiotemporal structure of mobility in the United States in 2020: a multi-source analytical framework*, *International Journal of Geographical Information Science* **36** (Mar., 2022) 585–616.

[127] R. Su and K. Goulias, *Evolution of the Chinese spring festival travel network during the COVID-19 early outbreak*, *Transportation Letters* (Mar., 2021) 1–9.

[128] Y. Kang, S. Gao, Y. Liang, M. Li, J. Rao, and J. Kruse, *Multiscale dynamic human mobility flow dataset in the US during the COVID-19 epidemic*, *Scientific data* **7** (2020), no. 1 390. Publisher: Nature Publishing Group UK London.

[129] M. U. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. Du Plessis, N. R. Faria, R. Li, W. P. Hanage, and others, *The effect of human mobility and control measures on the COVID-19 epidemic in China*, *Science* **368** (2020), no. 6490 493–497. Publisher: American Association for the Advancement of Science.

[130] J. Long and T. Nelson, *Home range and habitat analysis using dynamic time geography: Home Range and Dynamic Time Geography*, *The Journal of Wildlife Management* **79** (Apr., 2015) 481–490.

[131] I. Gijbels, A. Pope, and M. P. Wand, *Understanding exponential smoothing via kernel regression*, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61** (Feb., 1999) 39–50.

[132] J. D. Hunter, *Matplotlib: A 2D graphics environment*, *Computing in Science & Engineering* **9** (2007), no. 3 90–95. Publisher: IEEE COMPUTER SOC.

[133] C. R. Harris, K. J. Millman, S. J. v. d. Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. v. Kerkwijk, M. Brett, A. Haldane, J. F. d. Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, *Array programming with NumPy*, *Nature* **585** (Sept., 2020) 357–362. Publisher: Springer Science and Business Media LLC.

[134] S. Gillies, C. van der Wel, J. Van den Bossche, M. W. Taves, J. Arnott, B. C. Ward, and others, *Shapely*, Dec., 2022.

[135] NuStats, *2010-2012 California Household Travel Survey Final Report*, tech. rep., California Department of Transportation, 2013.

[136] G. Boeing, *OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks*, *Computers, Environment and Urban Systems* **65** (Sept., 2017) 126–139.

[137] G. Boeing, *Urban spatial order: street network orientation, configuration, and entropy*, *Applied Network Science* **4** (Dec., 2019) 67.

[138] K. Goulias and D. Janelle, *GPS Tracking and Time Geography: Applications for Activity Modeling and Microsimulation, Position Papers*. FHWA-sponsored Peer Exchange and CSISS Specialist Meeting, 2005.

[139] R. Wright, M. Ellis, S. R. Holloway, and S. Wong, *Patterns of Racial Diversity and Segregation in the United States: 1990–2010*, *The Professional Geographer* **66** (Apr., 2014) 173–182.

[140] D. S. Massey, G. A. Condran, and N. A. Denton, *The effect of residential segregation on black social and economic well-being*, *Social Forces* **66** (1987), no. 1 29–56. Publisher: The University of North Carolina Press.

[141] S. A. Perlin, D. Wong, and K. Sexton, *Residential Proximity to Industrial Sources of Air Pollution: Interrelationships among Race, Poverty, and Age*, *Journal of the Air & Waste Management Association* **51** (Mar., 2001) 406–421.

[142] S. C. Grady, *Racial disparities in low birthweight and the contribution of residential segregation: a multilevel analysis*, *Social science & medicine* **63** (2006), no. 12 3013–3029. Publisher: Elsevier.

[143] M. Hall and J. Stringfield, *Undocumented migration and the residential segregation of Mexicans in new destinations*, *Social science research* **47** (2014) 61–78. Publisher: Elsevier.

[144] N. Brazil, *Environmental inequality in the neighborhood networks of urban mobility in US cities*, *Proceedings of the National Academy of Sciences* **119** (Apr., 2022) e2117776119.

[145] M. Jones and A. R. Pebley, *Redefining Neighborhoods Using Common Destinations: Social Characteristics of Activity Spaces and Home Census Tracts Compared*, *Demography* **51** (June, 2014) 727–752.

[146] S. R. Holloway, R. Wright, and M. Ellis, *The Racially Fragmented City? Neighborhood Racial Segregation and Diversity Jointly Considered*, *The Professional Geographer* **64** (Feb., 2012) 63–82.

[147] D. T. Lichter, D. Parisi, and M. C. Taquino, *Toward a New Macro-Segregation? Decomposing Segregation within and between Metropolitan Cities and Suburbs*, *American Sociological Review* **80** (Aug., 2015) 843–873.

[148] D. W. S. Wong and S.-L. Shaw, *Measuring segregation: an activity space approach*, *Journal of Geographical Systems* **13** (June, 2011) 127–145.

[149] S. Silm and R. Ahas, *The temporal variation of ethnic segregation in a city: Evidence from a mobile phone use dataset*, *Social Science Research* **47** (Sept., 2014) 30–43.

[150] B. Moya-Gómez, M. Stepniak, J. C. García-Palomares, E. Frías-Martínez, and J. Gutiérrez, *Exploring night and day socio-spatial segregation based on mobile phone data: The case of Medellin (Colombia)*, *Computers, Environment and Urban Systems* **89** (Sept., 2021) 101675.

[151] M. Wu, X. Liu, Y. Qin, and Q. Huang, *Revealing racial-ethnic segregation with individual experienced segregation indices based on social media data: A case study in Los Angeles-Long Beach-Anaheim*, *Computers, Environment and Urban Systems* **104** (Sept., 2023) 102008.

[152] O. Järv, A. Masso, S. Silm, and R. Ahas, *The Link Between Ethnic Segregation and Socio-Economic Status: An Activity Space Approach*, *Tijdschrift voor Economische en Sociale Geografie* **112** (July, 2021) 319–335.

[153] G. Orfield and C. Lee, *Why segregation matters: Poverty and educational inequality.*, *Civil Rights Project at Harvard University (The)* (2005). Publisher: ERIC.

[154] R. Johnston, M. Poulsen, and J. Forrest, *Segregation matters, measurement matters*, in *Social-Spatial Segregation*, pp. 13–44. Policy Press, 2014.

[155] J. Yao, D. W. Wong, N. Bailey, and J. Minton, *Spatial Segregation Measures: A Methodological Review*, *Tijdschrift voor Economische en Sociale Geografie* **110** (July, 2019) 235–250.

[156] D. S. Massey and N. A. Denton, *The dimensions of residential segregation*, *Social forces* **67** (1988), no. 2 281–315. Publisher: The University of North Carolina Press.

[157] S. F. Reardon and D. O'Sullivan, *3. Measures of Spatial Segregation*, *Sociological Methodology* **34** (Dec., 2004) 121–162.

[158] L. A. Brown and S.-Y. Chung, *Spatial segregation, segregation indices and the geographical perspective*, *Population, Space and Place* **12** (Mar., 2006) 125–143.

[159] R. Johnston, M. Poulsen, and J. Forrest, *Ethnic and racial segregation in US metropolitan areas, 1980-2000: The dimensions of segregation revisited*, *Urban Affairs Review* **42** (2007), no. 4 479–504. Publisher: Sage Publications Sage CA: Thousand Oaks, CA.

[160] O. D. Duncan and B. Duncan, *A Methodological Analysis of Segregation Indexes*, *American Sociological Review* **20** (Apr., 1955) 210.

[161] M. J. White, *The measurement of spatial segregation*, *American journal of sociology* **88** (1983), no. 5 1008–1018. Publisher: University of Chicago Press.

[162] D. W. Wong, *Formulating a general spatial segregation measure*, *The Professional Geographer* **57** (2005), no. 2 285–294. Publisher: Taylor & Francis.

[163] D. O'Sullivan and D. W. Wong, *A surface-based approach to measuring spatial segregation*, *Geographical analysis* **39** (2007), no. 2 147–168. Publisher: Wiley Online Library.

[164] H. Theil, *Statistical decomposition analysis*, *Amsterdam: North Holland* (1972).

[165] W. Bell, *A Probability Model for the Measurement for Ecological Segregation*, *Social Forces* **32** (1954) 357–364. Publisher: HeinOnline.

[166] S. Lieberson, *Measuring population diversity*, *American Sociological Review* (1969) 850–862. Publisher: JSTOR.

[167] S. Lieberson and D. K. Carter, *A Model for Inferring the Voluntary and involuntary causes of residential segregation*, *Demography* **19** (Nov., 1982) 511–526.

[168] S. Lieberson and D. K. Carter, *Temporal Changes and Urban Differences in Residential Segregation: A Reconsideration*, *American Journal of Sociology* **88** (Sept., 1982) 296–310.

[169] I. Schnell and B. Yoav, *The Sociospatial Isolation of Agents in Everyday Life Spaces as an Aspect of Segregation*, *Annals of the Association of American Geographers* **91** (Dec., 2001) 622–636.

[170] S. Farber, A. Páez, and C. Morency, *Activity spaces and the measurement of clustering and exposure: A case study of linguistic groups in Montreal*, *Environment and Planning A* **44** (2012), no. 2 315–332. Publisher: SAGE Publications Sage UK: London, England.

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

[172] R. G. Golledge, *Spatial behavior: A geographic perspective*. Guilford Press, New York, 1997.

[173] R. Johnston, M. Poulsen, and J. Forrest, *The Geography of Ethnic Residential Segregation: A Comparative Study of Five Countries*, *Annals of the Association of American Geographers* **97** (Dec., 2007) 713–738.

[174] W. A. V. Clark, E. Anderson, J. Östh, and B. Malmberg, *A Multiscalar Analysis of Neighborhood Composition in Los Angeles, 2000–2010: A Location-Based Approach to Segregation and Diversity*, *Annals of the Association of American Geographers* **105** (Nov., 2015) 1260–1284.

[175] United States Census Bureau, *American Community Survey 2018*, 2018.

[176] L. Xiang, M. Gao, and T. Wu, *Extracting Stops from Noisy Trajectories: A Sequence Oriented Clustering Approach*, *ISPRS International Journal of Geo-Information* **5** (Mar., 2016) 29.

[177] M. J. White, *Segregation and Diversity Measures in Population Distribution*, *Population Index* **52** (1986), no. 2 198.

[178] C. R. Farrell and B. A. Lee, *Racial diversity and change in metropolitan neighborhoods*, *Social Science Research* **40** (July, 2011) 1108–1123.

[179] R. Su, E. C. McBride, and K. G. Goulias, *Unveiling daily activity pattern differences between telecommuters and commuters using human mobility motifs and sequence analysis*, *Transportation Research Part A: Policy and Practice* **147** (May, 2021) 106–132.

[180] M. J. Salganik, *Bit by bit: Social research in the digital age*. Princeton University Press, 2019.

[181] T. Yabe, B. G. B. Bueno, X. Dong, A. Pentland, and E. Moro, *Behavioral changes during the COVID-19 pandemic decreased income diversity of urban encounters*, *Nature Communications* **14** (Apr., 2023) 2310.

[182] L. Anselin, *Local Indicators of Spatial Association-LISA*, *Geographical Analysis* **27** (1995), no. 2 93–115.

[183] M. Ellis, R. Wright, and V. Parks, *Work Together, Live Apart? Geographies of Racial and Ethnic Segregation at Home and at Work*, *Annals of the Association of American Geographers* **94** (Sept., 2004) 620–637.

[184] R. Su and K. Goulias, *Untangling the relationships among residential environment, destination choice, and daily walk accessibility*, *Journal of Transport Geography* **109** (May, 2023) 103595.

[185] E. Cherchi and C. A. Guevara, *A Monte Carlo experiment to analyze the curse of dimensionality in estimating random coefficients models with a full variance–covariance matrix*, *Transportation Research Part B: Methodological* **46** (Feb., 2012) 321–332.

[186] V. M. Garikapati, R. M. Pendyala, E. A. Morris, P. L. Mokhtarian, and N. McDonald, *Activity patterns, time use, and travel of millennials: a generation in transition?*, *Transport Reviews* **36** (2016), no. 5 558–584. Publisher: Taylor & Francis.

[187] R. Joo, M. E. Boone, T. A. Clay, S. C. Patrick, S. Clusella-Trullas, and M. Basille, *Navigating through the r packages for movement*, *Journal of Animal Ecology* **89** (Jan., 2020) 248–267.

[188] G. Ranjan, H. Zang, Z.-L. Zhang, and J. Bolot, *Are call detail records biased for sampling human mobility?*, *ACM SIGMOBILE Mobile Computing and Communications Review* **16** (Dec., 2012) 33–44.

[189] M. F. Goodchild, *The quality of big (geo)data*, *Dialogues in Human Geography* **3** (Nov., 2013) 280–284.

[190] J. Lerman, *Big Data and Its Exclusions*, *SSRN Electronic Journal* (2013).

[191] Z. Zhao, S.-L. Shaw, Y. Xu, F. Lu, J. Chen, and L. Yin, *Understanding the bias of call detail records in human mobility research*, *International Journal of Geographical Information Science* **30** (Sept., 2016) 1738–1762.

[192] N. Eagle and A. (Sandy) Pentland, *Reality mining: sensing complex social systems*, *Personal and Ubiquitous Computing* **10** (May, 2006) 255–268.

[193] Y. Liu, X. Liu, S. Gao, L. Gong, C. Kang, Y. Zhi, G. Chi, and L. Shi, *Social Sensing: A New Approach to Understanding Our Socioeconomic Environments*, *Annals of the Association of American Geographers* **105** (May, 2015) 512–530.

[194] W. Bohte and K. Maat, *Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands*, *Transportation Research Part C: Emerging Technologies* **17** (2009), no. 3 285–297. Publisher: Elsevier.

[195] L. Gong, X. Liu, L. Wu, and Y. Liu, *Inferring trip purposes and uncovering travel patterns from taxi trajectory data*, *Cartography and Geographic Information Science* **43** (Mar., 2016) 103–114.

[196] L. Shen and P. R. Stopher, *A process for trip purpose imputation from Global Positioning System data*, *Transportation Research Part C: Emerging Technologies* **36** (Nov., 2013) 261–267.

[197] F. J. Pot, B. van Wee, and T. Tillema, *Perceived accessibility: What it is and why it differs from calculated accessibility measures based on spatial data*, *Journal of Transport Geography* **94** (June, 2021) 103090.

[198] T. Neutens, N. Van de Weghe, F. Witlox, and P. De Maeyer, *A three-dimensional network-based space–time prism*, *Journal of Geographical Systems* **10** (Mar., 2008) 89–107.

[199] H. J. Miller, *Time Geography and Space-Time Prism*, in *International Encyclopedia of Geography* (D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu, and R. A. Marston, eds.), pp. 1–19. John Wiley & Sons, Ltd, Oxford, UK, Mar., 2017.

[200] P. Laube, *Computational Movement Analysis*. SpringerBriefs in Computer Science. Springer International Publishing, Cham, 2014.

[201] V. Brum-Bastos, M. Łoś, J. A. Long, T. Nelson, and U. Demšar, *Context-aware movement analysis in ecology: a systematic review*, *International Journal of Geographical Information Science* (Aug., 2021) 1–23.

[202] S. Dodge, *Chapter 11: Context-dependent movement analysis*, in *Handbook of Spatial Analysis in the Social Sciences*, pp. 187 – 207. Edward Elgar Publishing, Cheltenham, UK, 2022.

[203] S. Dodge and T. A. Nelson, *A framework for modern time geography: emphasizing diverse constraints on accessibility*, *Journal of Geographical Systems* (Feb., 2023).

[204] S. Stier, J. Breuer, P. Siegers, and K. Thorson, *Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field*, *Social Science Computer Review* **38** (Oct., 2020) 503–516.

[205] T. A. Nelson, M. F. Goodchild, and D. J. Wright, *Accelerating ethics, empathy, and equity in geographic information science*, *Proceedings of the National Academy of Sciences* **119** (May, 2022) e2119967119.

[206] M. Schneider, *Uncertainty management for spatial datain databases: Fuzzy spatial data types*, in *Advances in Spatial Databases: 6th International Symposium, SSD'99 Hong Kong, China, July 20—23, 1999 Proceedings 6*, pp. 330–351, Springer, 1999.

[207] P. Ranacher, R. Brunauer, W. Trutschnig, S. Van Der Spek, and S. Reich, *Why GPS makes distances bigger than they are*, *International Journal of Geographical Information Science* **30** (Feb., 2016) 316–333.

[208] G. B. Heuvelink, *Error propagation in environmental modelling with GIS*. CRC press, 1998.

[209] A. S. Furtado, L. O. C. Alvares, N. Pelekis, Y. Theodoridis, and V. Bogorny, *Unveiling movement uncertainty for robust trajectory similarity analysis*, *International Journal of Geographical Information Science* **32** (Jan., 2018) 140–168.