TOWARD SMARTER CITIES: EXPLORING DATA-DRIVEN APPROACHES TO HUMAN-LAND INTERACTION BY ARTIFICIAL URBAN INTELLIGENCE

YAN LUO

PhD

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University
Department of Computing

Toward Smarter Cities: Exploring Data-Driven Approaches to Human-Land Interaction by Artificial Urban Intelligence

Yan LUO

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

July 2023

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_______________________ (Signed)

Yan LUO (Name of student)

Abstract

The rapid development of advanced technologies and the advent of the big data era have ushered in unprecedented opportunities for understanding and improving urban spaces. As urbanization continues to reshape our world, there is an increasing need to study human-land interactions and their impacts on urban environments. Artificial urban intelligence, a domain-specific application of artificial intelligence techniques for urban-related tasks, plays a crucial role in addressing this need. Emphasizing human-land interaction in urban applications is essential to developing smarter cities that are more sustainable, efficient, and adaptable. This thesis aims to contribute to the body of knowledge in urban environment comprehension, human mobility understanding, and location recommendation by investigating a series of challenges and limitations of existing methodologies, and proposing novel frameworks and techniques to overcome these obstacles.

In the first chapter, we provide an introduction to the background and scope of the research, highlighting the significance of human-land interaction in artificial urban intelligence applications. The second chapter reviews the current state of the art, examining the methods employed in urban environment comprehension and human mobility understanding.

In the third chapter, we propose a novel multi-graph framework called Region2Vec for urban region representation learning. The framework captures inter-region relations through human mobility, geographical contextual information via neighborhood data, and intra-region information using Point of Interest (POI) side information in knowledge graphs. Experiments on real-world datasets demonstrate the effectiveness of Region2Vec, consistently outperforming state-of-the-art baselines in various tasks and metrics.

The fourth chapter is divided into two parts: human mobility analysis and lo-

cation recommendation. We use tourist travel patterns as a case study and employ trip chains to model and discover fixed patterns. In the location recommendation section, we propose a novel Temporal Prompt-based and Geography-aware (TPG) framework, which excels in interval prediction on various real-world datasets.

In the final chapter, we provide a conclusion for the thesis, summarizing the key findings and contributions made to the field of artificial urban intelligence. The proposed techniques hold great potential for further development and application in the pursuit of smarter cities and more intelligent urban environments.

Publications arising from the thesis

1. **Luo, Y.**, Duan, H., Liu, Y., & Chung, F. (2023). "Timestamps as Prompts for a Geography-aware Next Location Recommendation." *In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23).* Association for Computing Machinery, New York, NY, USA, 1697–1706.

2. **Luo, Y.**, Liu, Y., Chung, F., Liu, Y., & Chen, C. (2024). "End-to-End Personalized Next Location Recommendation via Contrastive User Preference Modeling." *Under review (at TNNLS).*

3. **Luo, Y.**, Chung, F., Liu, H., & Chen, K. (2022). "Urban Region Profiling via A Multi-Graph Representation Learning Framework." *In Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM '22).* Association for Computing Machinery, New York, NY, USA, 4294–4298.

4. **Luo, Y.***, Jing, Z.*, Li, X., & Xu, X. (2022). "A Multi-Dimensional City Data Embedding Model for Improving Predictive Analytics and Urban Operations." *Industrial Management & Data Systems, 122*(10), 2199-2216.

5. Liu, Y., Li, H., Hu, C., Luo, S., **Luo, Y.**, & Chen, C. (2024). "Learning to Aggregate Multi-Scale Context for Instance Segmentation in Remote Sensing Images." *IEEE Transactions on Neural Networks and Learning Systems*, just accepted.

6. **Luo, Y.**, Leong, C., Chung, F., Li, W., Jiao, S., & Liu, G. (2023). "Geo-Tile2Vec: A Multi-Modal and Multi-Stage Embedding Framework for Urban Analytics." *ACM Transactions on Spatial Algorithms System, 9*(2), Article 10.

7. **Luo, Y.** (2022). "Characterizing Tourist Daily Trip Chains Using Mobile Phone Big Data." *arXiv:2205.14395.*

8. **Luo, Y.**, Xiang, L., Xu, Y., & Gui, Z. (2020). "Road Network Extraction from GPS Trajectories — A Tensor Voting Based Algorithm." *In Proceedings of the GISRUK 2020, London.*

Acknowledgements

I have journeyed a lengthy and arduous path to present this doctoral thesis to you. My twenty-year educational odyssey has been fraught with innumerable obstacles. Growing up in an impoverished community where education was often undervalued, I witnessed many children, especially girls, leaving school to work and help their families before completing junior high. The anticipated life trajectory for a girl in my community was to earn money, get married, have children, and become a housewife. Driven to challenge these expectations, I aspired to be the first in my family to pursue higher education at a university.

Without a role model to guide me, I relied on my own efforts to overcome the challenges of my upbringing. Not only did I face disrespect and bullying from children from more privileged families, but I also had to figure out my future path on my own. My parents never completed primary school, and no one in my family had ever attended university. My life's journey has been anything but smooth; when I started my undergraduate studies, my father passed away after squandering all our family savings, and my mother's health deteriorated. As a student, I had to bear the burden of supporting my family. In fact, my decision to pursue a Ph.D. was purely accidental. At the time, I had received a decent job offer, but I also had the good fortune to be awarded the Hong Kong Ph.D. Fellowship, which allowed me to support my family while continuing my academic pursuits. As a result, I decided to give it a try.

The journey to obtaining my Ph.D. was riddled with obstacles and unexpected events. In 2019, as I embarked on my Ph.D. studies after completing my bachelor's degree, Hong Kong was gripped by turmoil. Protesters vandalized our university, forcing students to leave the city. In 2020, as a Wuhan native, I experienced an extended lockdown for over half a year due to the pandemic. As the epicenter of the outbreak, I witnessed numerous life-altering separations during that year. My life mentor, who had a profound impact on me and encouraged me to embrace my uniqueness as a girl unwilling to settle for the status quo, was Ms. Wan, my junior high school homeroom teacher. Tragically, she passed away during the pandemic. I had never received much affection from my elders growing up, but a beloved aunt, who suffered from cancer, also lost her life because she could not receive chemotherapy during the outbreak. Following these events, I faced a department transfer and a significant shift in research direction.

However, throughout this journey, I have experienced significant personal growth and received considerable support and assistance. Firstly, I would like to express my heartfelt gratitude to my supervisor, Prof. Fu-lai Chung, for his tremendous support and guidance during my Ph.D. study. His exceptional patience, constant motivation, and vast knowledge have greatly assisted me over the past two and a half years and will have a lasting impact on my life. He taught me the importance of being an upright person and maintaining academic integrity. For a Ph.D. student, choosing the right supervisor is of utmost importance. If the supervisor is not a good fit, even a gold mine can be obscured by dust. I am a person with many ideas, always eager to explore various opportunities, and he always makes every effort to support me. He may not be the most knowledgeable in my specific area of research, but he is the person who can help me maximize my potential. Whenever I face difficulties, I am instantly filled with confidence, knowing that I have an unconditionally supportive supervisor. My every success is closely tied to his support. It is my great honor and fortune to have a supervisor like him. I will always remain indebted to my dear supervisor, and I hope that I can emulate him if I become a supervisor in the future. I am also grateful to Prof. Wenjie Li. When I initially shifted my focus to AI research, her insightful advice and critical judgment helped me tremendously in both research and life.

I sincerely thank my MIT supervisor, Prof. Kent Larson. I come from a family where my parents had no idea what MIT was. Before embarking on my doctoral journey, I never imagined that one day I would have the opportunity to study at

MIT. My experience at MIT has significantly broadened my horizons, as I met people from various backgrounds and witnessed the diverse possibilities in life. Kent is a co-founder of the Media Lab and a highly visionary person who proposed the concept of smart cities. His impact on me has been immense, and I can say that my experience at the Media Lab has reshaped my definition of innovation. I used to focus on technology itself with a narrow perspective, thinking that only technological advancements constituted true innovation. However, the demos from over thirty groups with diverse research directions at the Media Lab made me realize how impactful and awe-inspiring interdisciplinary works completed by scientists, engineers, artists, and politicians can be.

I would also like to express my gratitude to Prof. Kai Chen and Dr. Hao Liu at HKUST, as well as my colleagues from the HKUST RGC themed-based project team. When I had just changed my research direction, I saw that Prof. Chen was leading a project highly relevant to my area of study, so I proactively requested to join. He kindly accepted me, and the insightful discussions within the group laid a solid foundation for my subsequent research. Dr. Hao Liu once criticized my state of indecision when I was struggling to find a research direction. His frank advice, though hard to hear, proved beneficial and allowed me to quickly adjust my approach.

I would like to extend my gratitude to Prof. Yu Liu from Peking University, who welcomed me in 2019 when our school was vandalized, allowing me to study at Peking University for a period of time. Prof. Liu is a pioneer in the field of GeoAI in China, and this experience introduced me to the world of GeoAI, sparking a strong interest in the subject. The cultural atmosphere at Peking University was exceptional, and the days spent by the beautiful Weiming Lake remain vivid in my memory.

Last but not least, I would like to thank my mother for providing me with the opportunity to receive an education. I am grateful to my friends for believing in me when I lost faith in myself. I would like to express my appreciation to Prof. Xintao Liu and Prof. Wenzhong Shi, as well as the dean and the president, for their efforts in helping me transfer departments, allowing me the opportunity to pursue research in a field that truly interests me. I also wish to thank the RGC reviewers for awarding me the HKPFS that year, which provided me with the opportunity to pursue my Ph.D. Although the journey has been filled with challenges, the ultimate rewards have been worth it. The world's most astonishing, marvelous, and extraordinary sights are often hidden in remote and dangerous places, inaccessible to ordinary people. If I were to describe my future aspirations, I would simply say that I hope to maintain a youthful spirit after years of hard work and dedication. Even better, I aspire to make a meaningful social impact throughout my lifetime.

Table of contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The theory of human-land relationship, originating from the field of human geography, explores the connections between humans, their diverse social activities, and the geographic environment [118]. It suggests that humans constantly expand, transform, and utilize the geographical environment to meet their survival needs, while the environment significantly influences the regional characteristics and spatial differences of human activities. This relationship has been an intrinsic part of our existence, as the geographical environment consistently affects human survival and activities.

Since the 1960s, with the advent of the quantitative revolution, human geography research has delved deeper into micro-regional studies with practical implications, such as environmental protection, soil agriculture rotation, and desertification [10]. Nowadays, rapid urban development has led to cities becoming more modern and intelligent. As a result, the influence of humans on the urban environment and the impact of the urban environment on humans have become more prominent and faster-paced. For instance, in ancient times, due to limited transportation options, it was challenging for people to travel long distances, resulting in restricted human activity and a slower transition of land use (e.g., commercial areas, residential areas) determined by different types of human activities. Nowadays, with the convenience of airplanes and high-speed rail, the pace of urban change has accelerated significantly. For example, a family living in a rural area can easily purchase a shop in a provincial capital and start a business. If their business thrives, it may attract competitors, leading to the formation of a new commercial district. At the same time, extensive travel has led to a series of problems, such as climate change and pollution, which in turn affects human activity patterns, health, and so on.

The development of various advanced technologies has facilitated the generation of big data from heterogeneous sources in urban spaces, which in turn leads to a better understanding of city operations and offers an excellent opportunity to explore these interactions. In this thesis, the concept of **artificial urban intelligence** is explored, which refers to the methods that apply artificial intelligence technologies to urban domains. Simultaneously, it is necessary to redefine human-land relationships in urban spaces for the new era. Thus, the concept of **human-land interaction** is further explored, which refers to the mutual influence between humans and the urban environment in the age of big data. Compared to the original theory of human-land relationships, human-land interaction places a greater emphasis on quantifying various social phenomena, such as crime, mobility patterns, or energy usage, using big data-driven artificial urban intelligence technologies.

By delving deeper into human-land interactions, we can better utilize these insights to address various urban issues, including resource management, transportation systems, economic growth, quality of life, environmental sustainability, data-driven decision-making, and resilience. This understanding allows us to tackle urban challenges more efficiently, sustainably, and innovatively, ultimately enhancing the quality of life for city inhabitants and fostering long-term, sustainable growth. I argue that comprehending these interactions serves as the foundation for advancing toward smarter cities, which play a vital role in addressing the challenges of rapid urbanization and promoting a sustainable, prosperous future.

Based on the above observations, I propose the topic **exploring data-driven approaches to human-land interaction by artificial urban intelligence**. Our study focuses on the joint analysis of human-land interaction from urban environment comprehension and human mobility understanding in order to uncover hidden insights and facilitate intelligent decision-making. I explore three significant applications of artificial urban intelligence, as discussed in Section 1.2, that demonstrate the potential of our approach in addressing various urban challenges.

1.2 Research Framework and Scope

The framework of artificial urban intelligence includes several parts such as urban sensing, urban data management and analytics, and urban applications. It first acquires and stores data from various domains in urban spaces for urban sensing, such as human mobility, POI, and street view. Features are then extracted from multi-modal datasets. For different applications, data with correlations are jointly analyzed to acquire hidden insights and enable intelligent decision-making. The scope of this topic is focused on urban data analytics and urban applications. Specifically, I will study the requirements and issues related to human-land interaction for artificial urban intelligence applications.

This research topic focuses on two main areas: urban environment comprehension and human mobility understanding, which are interconnected through the concept of human-land interaction. Within urban environment comprehension, we investigate urban region representation learning, while within human mobility understanding, we examine human mobility analysis and next location recommendation. The relationship between mobility analysis and location recommendation is that basic and essential analysis is first required to understand patterns, which can then be used

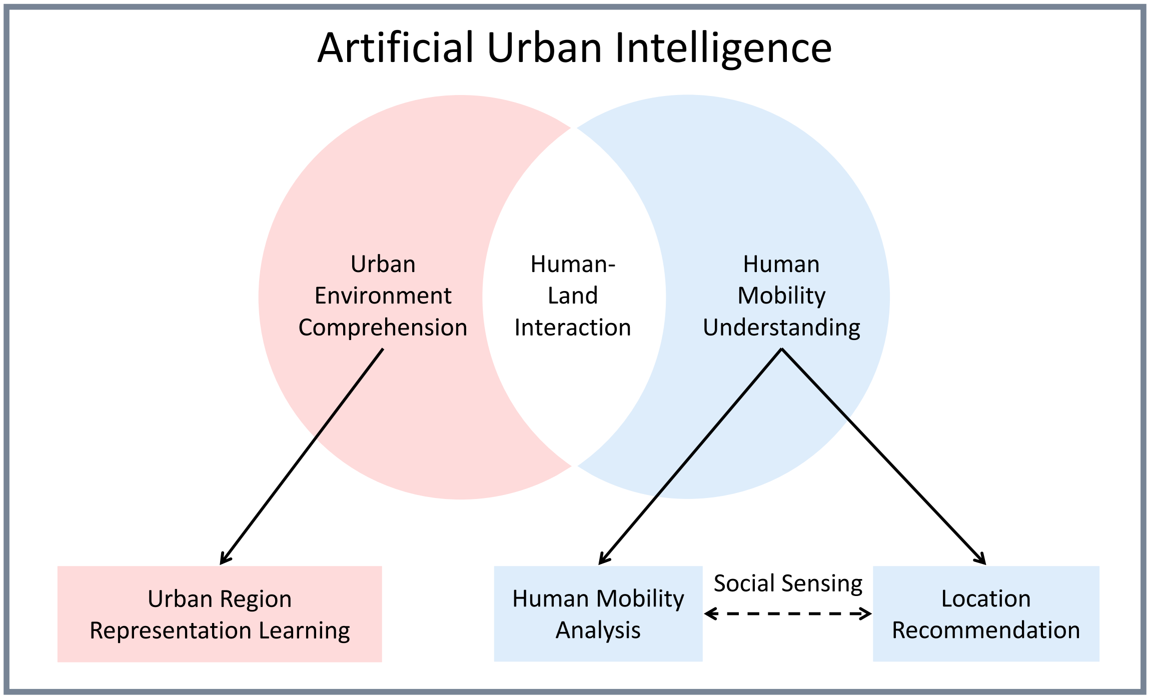

Figure 1.1: The logical structure of this thesis.

to make better predictions. This process is referred to as social sensing [70], which involves utilizing various geospatial big data sources, such as mobile phone locations, social media, taxi trajectories, and public transport card data, to extract human spatiotemporal behavior patterns and infer geographical features of human and socio-economic factors. By combining urban region representation learning, human mobility analysis, and next location recommendation, we aim to demonstrate how the integration of urban environment comprehension and human mobility understanding can generate hidden insights and enable intelligent decision-making in Artificial Urban Intelligence. The structure of this thesis can be seen in Figure 1.1.

Three unique challenges are identified in this research topic:

1. Understanding the characteristics of urban regions is essential for various downstream tasks, such as urban planning, business model development, and social welfare improvement. Conventional survey approaches, like the American

Community Survey (ACS) carried out by the U.S. Census Bureau at an annual expense of $250 million [26], can be both expensive and time-intensive when it comes to revealing the traits of neighborhoods. In the era of big data, we can try to learn urban region representations using a massive amount of unlabeled data in an unsupervised manner. However, various data sources, such as Points of Interest (POIs), street view images, and human trajectories, are collected to apply artificial urban intelligence. These datasets are naturally multi-modal and heterogeneous, making it difficult to effectively utilize them. Although existing studies have made great efforts to learn urban region representation from multi-modal urban data, there are still two limitations: (a) Most related methods focused merely on global-level inter-region relations while overlooking local-level geographical contextual signals and intra-region information; (b) Most previous work failed to develop an effective yet integrated fusion module which can deeply fuse multi-graph correlations. The challenge of effectively utilizing multi-modal urban big data for understanding urban regions and developing data-driven solutions for urban challenges requires further research and exploration.

2. Exploring human mobility patterns is crucial for numerous reasons, including enhancing urban planning and infrastructure development, optimizing traffic management to reduce congestion, improving emergency response and public safety, fostering economic growth and business development. There are limited quantitative works targeting the analysis of human activity patterns via big data, and the patterns of human mobility are still unclear. Establishing an effective model and extracting insights for mobility patterns is a worthwhile problem to investigate. Taking tourist travel patterns as an example, tourists tend to visit multiple destinations out of their variety-seeking motivations in

their trips. Thus, it is critical to discover travel patterns involving multi-destinations in tourism research. Existing relevant research most relied on survey data or focused on citizens due to the lack of large-scale, fine-grained tourism datasets. Several scholars have mentioned the notion of trip chains, but few works have been done towards quantitatively identifying the structures of trip chains.

3. Accurate human mobility prediction holds significant commercial value and can be applied to location-based services such as map services and local lifestyle services. However, predicting mobility involves complex spatiotemporal relationships. The famous geographic theory, Tobler's first law of geography[113], suggests that things are more related to nearby things than to distant things. Due to urban development, spatial heterogeneity[93] has emerged, complicating the matter. For example, if we take a point where a store is located as the center and draw a circle with a certain radius, the points on the circumference of the circle do not have the same properties. Points that intersect with the road network have higher accessibility than those that do not intersect with the road network. Capturing potential features from multi-relational spatial-temporal data and applying them to location recommendation to improve accuracy and user experience is a grand challenge.

In this thesis, we address these three major challenges in the realm of urban applications and big data, and make the following contributions:

1. We emphasize the importance of exploring human-land interaction in urban applications during the big data era. Understanding the complex interplay between humans and the urban environment is crucial for various urban planning, infrastructure development, and policy-making tasks. Investigating this

interaction is a necessary step towards building smarter and more sustainable cities.

2. We highlight the need to exploit geographical contextual information, intra-region information, and an effective fusion module when investigating urban-related problems. To address this need, we propose Region2Vec, a multi-graph representative learning framework for urban region profiling. This framework captures inter-region relations through human mobility data, geographical contextual information via neighborhood data, and intra-region information using Points of Interest (POI) side information in knowledge graphs. Additionally, the framework incorporates accessibility, vicinity, and functionality correlations among regions. The proposed Region2Vec outperforms state-of-the-art baselines in various tasks and metrics, demonstrating its potential as a tool for building general-purpose intelligent agents capable of handling diverse urban challenges.

3. We focus on tourist mobility patterns as a case study for understanding human-land interaction. We propose a model for quantitatively characterizing daily trip chains using mobile phone big data, enabling the discovery of underlying tourist travel patterns. This study not only uncovers complex daily travel trip chains from tourism big data, but also fills the gap in tourism literature on multi-destination trips by discovering significant and underlying patterns based on mobile datasets. This information can be invaluable for tourism industry stakeholders and urban planners in optimizing destination management and enhancing tourists' experiences.

4. We argue that explicitly modeling the timestamp of the location to be predicted is essential in real-world applications, particularly in the context of location

recommendation systems. To address this, we propose the Temporal Prompt-based and Geography-aware (TPG) framework. This innovative framework incorporates temporal information as a prompt for the recommendation system while using a shifted window mechanism to augment geographic data, thus avoiding the hard boundary problem when handling longitude and latitude of POIs with grids. Experiments on five real-world datasets (Gowalla, Brightkite, Foursquare-NYC, Foursquare-TKY, and Foursquare-SIN) show that TPG outperforms state-of-the-art counterparts under different settings and excels in interval prediction. Specifically, the model demonstrates its ability to predict a user's desired location at a given time, even when the most recent check-in data is masked, or predict a specific future check-in at a given timestamp, not just the next one. This advancement can significantly enhance location-based services and improve user experience.

1.3 Structure of Thesis

This thesis consists of six chapters:

- **Chapter 1** introduces the background and highlights the research framework and scope of studying human-land interaction for artificial urban intelligence applications;
- **Chapter 2** reviews the related work for methods in urban environment comprehension and human mobility understanding;
- **Chapter 3** presents the problem of urban region representation learning, which is a highly important topic in urban environment comprehension. It proposes a novel multi-graph framework to generate hidden insights for unsupervised learning in urban regions;

- **Chapter 4** is composed of two parts: human mobility analysis and location recommendation. In the human mobility analysis section, I use tourist travel patterns as an example and employ trip chains to model and discover three fixed patterns. In the location recommendation section, a Temporal Prompt-based and Geography-aware framework is proposed and demonstrates its effectiveness in real-world datasets;
- **Chapter 5** concludes the thesis.

Chapter 2

Literature Review

This chapter offers a comprehensive review of the two primary themes that constitute the foundation of this thesis: Urban Environment Comprehension and Human Mobility Understanding. Within each of these themes, we delve into the specific areas that are central to the focus of this thesis, such as Urban Region Representation Learning, Human Mobility Analysis, and Location Recommendation. By examining these areas in detail, we aim to shed light on the key concepts, methodologies, and challenges that underpin the research conducted throughout this thesis. This review serves as a basis for understanding the significance and potential applications of the work presented in subsequent chapters.

2.1 Urban Environment Comprehension

Urban Environment Comprehension is an essential aspect of understanding the complex dynamics within cities and urban spaces. The rapid growth of cities and increasing availability of geospatial data have provided researchers with new opportunities to explore the multifaceted nature of urban environments [6]. In recent years, numerous studies have emerged focusing on various aspects of urban environment comprehension, ranging from urban planning and design [69], land use classification [142], to transportation networks [149] and socioeconomic dynamics [22]. With the advent of big data, machine learning, and artificial intelligence, novel techniques such as deep learning [136] and graph-based models [9] have been employed to analyze and interpret urban data from multiple sources, such as satellite imagery, social media, mobile phone records, and points of interest (POIs) [31]. These advanced techniques have facilitated the extraction of valuable insights and patterns from vast amounts of data, enabling researchers to model urban environments with higher accuracy and granularity [119]. Consequently, urban environment comprehension has become a critical area of research with significant implications for sustainable urban development, policy-making, and urban management [36]. As this field continues to evolve, interdisciplinary approaches that integrate knowledge from geography, computer science, and urban studies will be essential for addressing the complex challenges facing urban environments in the 21st century [96].

Urban Analytics aims to explore the status of different urban regions and inter-regional connections by analyzing and computing heterogeneous information such as human mobility history and urban region attributes from physical environment. It serves for land use classification [134], POI recommendation [13], human mobility prediction [151] and other problems by providing solutions based on geographic information and regional functionality. Based on the nature of the utilized data, the existing Urban Analytics approaches can be classified as human activity data-based, physical environment data-based, and hybrid data-based.

The method based on human activity mainly uses the trajectories of human beings to mine the properties of different urban places and the relationships between them. Cesario et al. [12] explores the mobility patterns by discovering the regions that the trajectories frequently pass through. Comito [15] further adds the overall movement trends of different users to improve the accuracy of user's next location prediction. Due to the success of representation learning in natural language processing, DeepMove [151] applies Skip-Gram model [85] to movement data to learn representations of different locations. Similar to DeepMove, Yao et al. [134] uses the co-occurrence of origin-destination zones from taxi trajectories to learn region embeddings, while adding spatio-temporal characteristics of human mobility patterns. Moreover, Kim and Yoon [124] demonstrates the universality of such methods in characterizing the semantics of regions. Shimizu et al. [99] goes beyond the use of single-grained representation of locations in trajectories, and considers different levels of region size to learn region embeddings that contain multi-level information at the same time. Hu et al. [38] constructs taxi trajectories as road networks and captures the semantics in the road network topology of different regions by graph convolutional neural network so as to classify the regional functions. Recently, Lin et al. [68] proposed a transformer-based model to obtain location representation based on context-aware spatio-temporal information in trajectories and obtained state-of-the-art results in user's next location prediction task. Our proposed framework differs from these methods by considering not only mobility data but multimodal data.

Other common physical environment data include POI, street view images, satellite images, etc. As POI is one of the most intuitive data to describe the nature of physical locations, existing work [131, 139, 108, 141] has shown the importance and effectiveness of POI in characterizing regions. Zhang et al. [140] uses information reconstruction and graph learning to learn a low-dimensional representation of regions that preserves correlations between temporal, spatial, and tagged text units. Huang et al. [40] aggregates POIs embeddings into embeddings of corresponding regions based on the spatial co-occurrence patterns of POIs and the semantic information of the hierarchical categories of POIs. Crivellari and Resch [17] evaluates the consistency of region embedding methods based on mobility patterns as well as on the distribution of local POIs in characterizing region function. As for location-related image information, which is difficult to be described, Gebru et al. [27] for the first time identifies national socioeconomic trends through large-scale street view image data, illustrating that geographically relevant image information can be used to characterize the nature of regions through Convolutional Neural Network (CNN). Law and Neira [62] uses Convolutional AutoEncoder to reconstruct street views as well as street map images through unsupervised reconstruction to predict urban characteristics such as street-level enclosures and street network density. Instead of unsupervised reconstructing images, Tile2Vec [46] uses triplet loss to explicitly learn a CNN encoding model that maintains the geographic proximity of satellite images. Unlike the above works that focus on a single city, some studies [34, 130, 50] propose methods for learning location representations that are transferable across cities based on mentioned region features.

However, as geotagged data becomes increasingly rich in modalities and volume, relying on a single modality of geographical data is no longer sufficient to adequately represent the properties of regions. Consequently, several methods [145, 24, 143] have been developed to consider both intra-regional attributes and inter-regional movement relationships. For instance, Jenkins et al. [48] incorporate satellite image information to end-to-end integrate multimodal information for obtaining region representations. Urban2Vec [123] learns region embedding by simultaneously maintaining the proximity of vectors for geographically close street images and POI comments. Building upon Urban2Vec, M3G [39] further incorporates edge sampling to include inter-region relations. Nonetheless, most existing approaches primarily focus on inter-region correlations while neglecting region-wise inherent features. Moreover, ensuring the robustness and comprehensiveness of urban region representation is essential for the performance of the region embedding framework. Mere concatenation of graphs proves insufficient for extracting features from multi-modalities, resulting in suboptimal performance in downstream tasks [23]. This underlines the pressing need for a multimodal framework with a comprehensive fusion module that can effectively combine all modalities to learn urban representation.

2.2 Human Mobility Understanding

Human mobility understanding has become an increasingly important research area due to the rapid development of location-based services and the widespread use of smartphones and GPS-enabled devices. This field investigates the patterns and dynamics of human movement, aiming to uncover the underlying mechanisms and characteristics of mobility. Researchers have utilized various data sources, such as mobile phone records, social media check-ins, and taxi GPS trajectories, to study human mobility patterns and behavior [30, 150, 149]. A popular method for analyzing human mobility is trajectory data mining, which includes techniques like clustering, classification, and sequential pattern mining to explore and model the spatial-temporal properties of mobility data [149].

Recent studies have focused on understanding the relationships between human mobility and the built environment, as well as social and economic factors [22, 69]. For instance, Fras-Martnez et al. [22] investigated the correlation between mobility patterns and the socio-economic environment. Moreover, research has been conducted to develop methods for predicting human movement, such as destination prediction and travel demand forecasting, leveraging machine learning and deep learning techniques [133, 119]. Additionally, studies have explored the potential applications of human mobility understanding in various domains, including urban planning, public health, and transportation management. Despite the significant advancements in this field, challenges remain in modeling the complex and dynamic nature of human mobility, as well as addressing the issues of data sparsity, noise, and privacy concerns.

2.2.1 Human Mobility Analysis

Human mobility analysis has been an area of great interest for researchers, as understanding human movement patterns can provide valuable insights into urban planning, traffic management, and public health. Several studies have focused on the mobility of citizens, such as examining the relationship between commuting patterns and urban structure [101], investigating the impact of social networks on human movement [52], and exploring the influence of urban environments on individuals' daily activity spaces [138]. However, there is a lack of research focusing specifically on tourist mobility. In this thesis, we narrow our focus to the study of tourist mobility patterns, aiming to fill this gap in the literature and provide a deeper understanding of the factors that drive tourist movement within urban environments.

A trip involves multiple activities across hierarchical stages of travel experiences [47]. That is, a trip is not a simple origin-destination mechanism; rather, it entails multiple destinations [72]. To explicate patterns associated with these multi-trips, tourism scholars have proposed various approaches, from spatial configuration [72, 64] to activity-based perspectives [125, 126]. These relevant studies argue that it is fallacious to assume travelers go to a single place (or destination) after leaving home. Individuals show multi-destination patterns in which they make numerous activity decisions that influence behavior in an interactive fashion. The pattern of multi-destination trips is associated with the notion of a rational behavior wherein individuals are likely to minimize time and cost associated with travel, an effect that potentially increases accrual of benefits and fulfills the desire for variety in destinations.

In this sense, the current study suggests the concept of trip chain to elucidate multi-destination trip patterns, as have been widely discussed in the literature on human mobility and transportation (see [94]). The notion of trip chain varies depending on the various contexts in which it has been applied [60, 83]. From the viewpoint of transportation, [94] summarized two commonly used definitions: (1) "A sequence of trip segments begins at the home activity and ends when the individual returns home"; (2) "A sequence of trip segments between a pair of activities:

home, work, or school" (page 58). The trip chain approach enables researchers to identify similarity/regularity of public transit patterns [75], design activity schedule linking primary (e.g., home) and secondary (e.g., work) activities [94], and compare different usages of transportation facilities according to socio-demographic features [80, 147]. According to human mobility, trip chain can be regarded as a mobility network in essence whereby the activity (or a place visited) can be denoted as a node and the spatial pattern/links (or flow) can be indicated as edges. In this sense, another similar concept mobility motif, which refers to highly-repeated multidimensional subsequences in a complex mobility network structure, is also proposed to clarify the mobility patterns [28, 37, 78, 87, 146]. For instance, [97] described a kind of individual daily movement network as a mobility motif if it occurred more than 0.5% in the datasets. Indeed, while different names were used (e.g., "trip chain" and "mobility motif"), these consistently represent methods to interpret individuals' travel patterns by distinguishing between locations. Principally, "trip chain" refers to complex relationships between a set of activities and the interdependence of temporal (e.g., timing, duration, length, and sequence of trips) and spatial (e.g., location) characteristics associated with human mobility. Thus, the trip chain model in tourism can be defined as a sequential pattern of trip activities (or places visited for travel activities) made by travelers on a day-to-day basis [29, 82].

Tourism scholars have applied the notion of trip chain as a means of comprehending travel movement behaviors. [72], in a pioneering study, conceptually and primarily suggested a trip-chaining pattern as a type of spatial model of pleasure vacation trips, illustrating visitations involving numerous focal activities. Along with their own work, several tourism scholars (e.g., [105]) tried to demonstrate the trip chain patterns in their own terms but on consistent ideas. For example, [61] presented a chaining loop as part of tourist movement patterns describing a certain pattern of visiting multiple destinations. Likewise, [64] proposed conceptual linear path models of tourist behavior in intra-destinations consisting of point-to-point patterns (Type I), circular patterns (Type II), and complex patterns (Type III). The previous tourism literature, in theory, has discussed the notion of trip chains and its application to deconstruct travel movement patterns.

Importantly, however, most studies contain the limitation with lack of quantitative verifications of the models. This may be attributable to challenges faced in accessing tourism big data that provides comprehensive insights into the phenomenon and can be used to calibrate the models. The tourism studies typically have collected data on travel behaviors using surveys. Such approaches typically require a substantial financial outlay and expenditure of effort. They also contain the potential for response errors such as cognitive bias of respondents [100]. With the evolution of information technology, a number of tourism researchers have adopted social media data such as Flicker, Twitter, and Weibo [106, 89, 57]. Nevertheless, analysis of social media content suffers from sparseness of data, making it difficult to discern the comprehensive travel patterns. Thus, our study applies a trip-chaining method to tourist mobile phone data, enabling tourism researchers to overwhelm the shortcomings from traditional data and to uncover hidden patterns, which ultimately discover underlying spatial behaviors of tourists.

Travel mobility is closely related to travel distance that reflects individual efforts being consumed to reach his/her goals (e.g., arrival to a place; [30]). A broad theory of the principle of least effort (PLE) proposed by George Kingsley Zipf (1949) supports this argument. The model claims that people tend to choose the method requiring the least effort to finish tasks. As an example, Zipf discovered a certain speech pattern in which people tend to use short words for their daily communication. That is, the distribution between word frequency used by speakers and hearers and word rank is largely skewed and demonstrated by a mathematical formula, now called Zipf's law [76]. Applying PLE to travel mobility, the task encompasses movement;

travelers would likely to seek out the "optimal way" to minimize the total movement required. It can be thus argued that travel distance is related to different types of trip chain models. Other than understanding of human mobility, the PLE has been widely applied to explain a variety of human behaviors including information-seeking behavior [5, 90], human mobility [11], pedestrian mobility [56], and street networks [79].

2.2.2 Location Recommendation

Location recommendation, i.e., POI recommendation, which draws considerable attention recent years due to great business value, can be viewed as a special sub-task of sequential recommendation with spatial information [74]. Regarding the use of spatio-temporal information in next location recommendation, many previous works only use spatio-temporal intervals between two successive visits in a recurrent layer. For example, DeepMove [21] combines an attention layer for learning long-term sequential regularity. LSTPM [107] proposes a geo-dilated RNN that aggregates locations visited recently, but only for shot-term preference. Inspired by sequential item recommendation [55], GeoSAN [66] uses self-attention model in next location recommendation within the trajectory. STAN [74] adopts a spatial-temporal attention network that aggregates all relevant check-ins in trajectories. GETNext [132] proposes a novel graph enhanced Transformer model by exploiting the extensive collaborative signals.

However, these models suffer from limitations in geographic and temporal information modeling. In this thesis, we take the challenge of the hard boundary problem in grid mapping, and propose the shifted window mechanism. We also discard the implicit way to fuse the temporal information in our proposed model. Temporal prompt is proposed for explicitly modeling timestamp of locations to be predicted.

Chapter 3

Urban Environment Comprehension

3.1 Background and Motivation

To sustainably develop urban areas, effective frameworks for urban analytics and modeling are compelling needed [18, 35, 128]. The significance of studying the representations of urban regions should be highlighted due to two reasons: (1) Sometimes the downstream tasks are ambiguous; (2) A general-purpose urban intelligence framework can be achieved if we incorporate several data sources while the framework is designed as task-agnostic. Such an architecture is capable of handling various applications. It can provide us with a better understanding about the patterns of urban spaces, which will produce insights for urban planning, and make cities more livable and sustainable.

Recently, the advent of information and communication technologies and sensing technologies leads to the proliferation of urban datasets. This allows researchers to explore and investigate the characteristics of urban spaces via data-driven approaches. Taking human mobility (e.g., vehicle trajectories, human movement data) as an example, correlations among regions can be well-mined from such activities. Remote regions may form a community since the concept 'daily life circle' exists.

State-of-the-art results are achieved in [120, 135] based on this intuition. However, these models only consider human mobility. Although they successfully explored region correlations, the inherent flaw of a single modality is that it has less information than multi-modalities. The most important benefit of using multi-modal data is that the information can be adopted cooperatively to achieve better performance. In many cases, the similarities between objects may be manifested differently by different modalities.

Considering that not all types of urban data can be easily accessed in all urban areas, except from mobility flow, we only further take POI data into account, which is the most common type of urban data. Several existing studies [14, 23] also tried to use both POI data and human mobility to characterize region features. However, most of them focus on inter-region correlations. Despite promising results achieved, region-wise inherent features are largely overlooked in modeling urban regions. We notice that POI data has many attributes (e.g., category, subclass). At the same time, recent research in graph embedding tends to take a graph as the input and leverage the auxiliary information to facilitate the embedding [53, 54]. Thus, we treat POI attributes as side information and construct a knowledge graph for POIs to discover intra-region properties. Besides, we also incorporate geographical neighborhood into the framework as the geographical contextual signals since adjacent regions naturally show direct correlations according to the First Law of Geography [112]. By incorporating multi-modalities including global-level mobility flow, local-level geographical neighborhood, and region-wise POI side information, the urban spaces are comprehensively depicted from "man-land-dynamic-static", which is the four classical dimensions in Geography. Figure 3.1 shows an example of our idea. For our chosen region 1, besides its neighborhood region 3, it is also correlated to region 2 which has similar accessibility patterns. Region 4 will also affect region 1 since their POI functionality is similar. Thus, regions 2, 3, 4 are considered more significant when we analyze region 1, while irrelevant regions 5 are less important. In addition, to ensure the performance of the region embedding framework, the robustness and comprehensiveness of urban region representation is extremely significant. The key point here is how to effectively fuse all graphs. Simply taking the concatenation of them [23] is insufficient to extract features from multi-graphs and multi-modalities, which leads to suboptimal performance in downstream tasks. A comprehensive fusion module is urgently needed for learning urban representation.

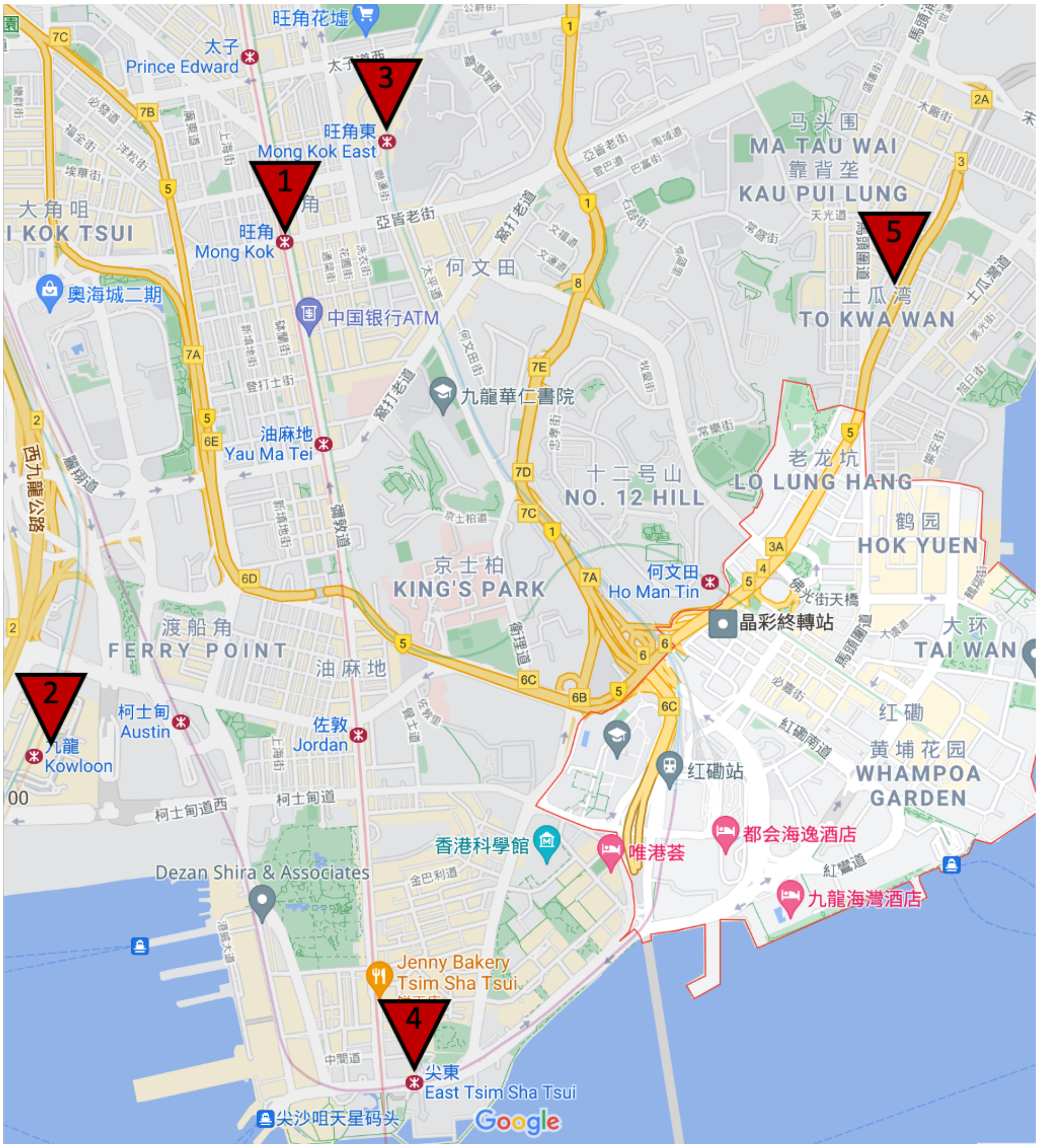

Figure 3.1: An example of different types of correlations among regions.

To address these challenges mentioned above, the goal of this work is to propose a multi-graph and multi-modal representation learning framework, namely Region2Vec, to investigate urban region profiling problem. In Region2Vec, inter-region relations, geographical contextual signals, as well as intra-region information are all captured via leveraging most common urban datasets including human mobility, geographical neighborhood, and POIs. Through mobility pattern similarity analysis, topology analysis, and constructing knowledge graph, we can then encode accessibility, vicinity, functionality correlations among regions by constructing graphs. To better propagate information for every single modality, the graph attention network is employed. At last, to promote the cooperation of different graph representations, a multi-graph fusion module with some designed learning objects is proposed to model the underlying correlations among graphs in a joint manner. Overall, the contributions of our work are mainly four-fold:

- We emphasize the importance of exploiting geographical contextual information and intra-region information, as well as an effective fusion module, when investigating urban related problems.
- A novel region embedding framework for urban region profiling is proposed. The final urban region representations preserve global-level inter-region correlations, local-level geographical contextual signals, and inherent region-wise attributes, via exploiting common urban data such as human mobility, geographical neighborhood, and POI side information. Graph attention networks are employed to propagate information within each modality.
- A multi-graph fusion module is also proposed to integrate multiple graphs. It is capable of fusing multi-modal urban data into comprehensive latent representations, with the collaboration of the global encoder and accessibility/ vicinity/ functionality correlation decoder.
- We conduct experiments on real-world datasets to demonstrate the effectiveness of Region2Vec. The results show that proposed Region2Vec has the great ability to learn the comprehensive representation from multi-graphs and multi-modalities for urban regions. Especially, Region2Vec consistently outperforms

state-of-the-art baselines by at least 7.80% in different tasks and various metrics. Through such a task-agnostic representation learning architecture, our model makes a step towards building general-purpose intelligence agents capable of handling various applications.

3.2 Preliminaries and Problem Statement

Definition 1 (Human Mobility). Human mobility can be defined as a set of trips conducted by citizens in urban spaces. A trip in human mobility datasets starts from an origin point (O) and ends by a destination point (D). Thus, a trip can also be named an OD. If we link the O/D of a trip with the urban regions which they belong to, then we can denote human mobility dataset M as:

$$M = \left\{\overrightarrow{m_0}, \overrightarrow{m_1}, \ldots, \overrightarrow{m_{|M|}}\right\}, \vec{m} = (r_o, r_d) \tag{3.1}$$

where $\vec{m}$ is a trip that can be represented as a two dimensional vector. r_o is the origin region and r_d is the destination region.

Definition 2 (Geographic Neighborhood). Geographic neighborhood of a region is described based on the spatial adjacency. In this study, we use 8 neighborhoods to define the geographic neighborhood. That is to say, if two regions have pixels connected, then they are adjacent, including 4 neighborhoods and diagonal neighborhoods. It is worth noting that the number of geographic neighborhoods of different regions maybe different due to irregular shapes of urban regions. Examples are given in Figure 3.2. Yellow areas are the chosen areas. Their geographic neighborhoods are numbered, while diagonal neighborhoods are all numbered as 1.

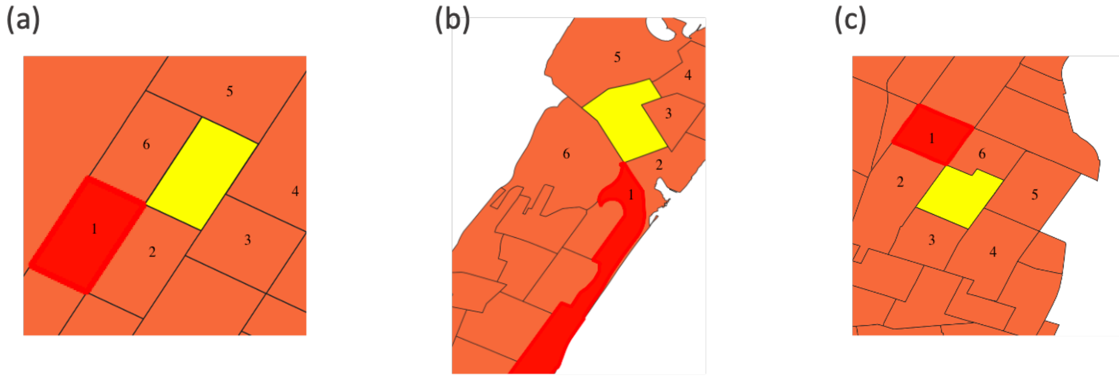

Figure 3.2: Three examples of geographic neighborhood.

Through traversing all the urban regions, for each region, we can get a vector of variable length of dimension. Supposed there are N urban regions in total. The geographic neighborhood dataset can be denoted as:

$$N = \overrightarrow{n_0}, \overrightarrow{n_1}, \ldots, \overrightarrow{n_N}, \vec{n} = \left(r_1, r_2, \ldots, r_{|\vec{n}|}\right) \tag{3.2}$$

where $\overrightarrow{n}$ is a geographic neighborhood vector for an urban region; r is a neighborhood for the urban region.

Definition 3 (POI Side Information). POI side information refers to different attributes of POI. Since POIs are the direct representations of urban functions, features from POI side information can be regarded as meta-knowledge, which reflect region functional attributes. Side information can help to establish correlations among those POIs, then to model relations among urban regions. Firstly, we map POIs to the located region. Then, POI side information dataset can be denoted as follows:

$$S = \overrightarrow{s_0}, \overrightarrow{s_1}, \ldots, \overrightarrow{s_N}, \vec{s} = \left(s_1, s_2, \ldots, s_{|\vec{s}|}\right) \tag{3.3}$$

where $\overrightarrow{s}$ is a POI side information vector for a urban region. s is a kind of POI attribute.

In this study, we utilize side information described in 3.1.

Table 3.1: Field description of POI dataset

Attribute field	Description
PLACEID	The unique identifier for each POI.
SOURCE	Agency that defined the POI.
FACILITY_T	Categories of POIs.
FACI_DOM	Subclasses of POIs.
SEGMENTID	POI is assigned the closest roadbed SEGMENTID.
PRI_ADD	POI has PRI_ADD field if the POI is related to any address point.
BIN	Point is assigned a Building Identification Number (BIN) if it falls within a building.
SOS	Indicates which side of the street the POI is on.
SAFTYPE	Point is assigned a SAFTYPE if it is a part of a Complex.
COMPLEXID	Point is assigned a COMPLEXID if it is a part of a Complex.

Problem Statement (Urban Region Embedding). Given three sets of vectors M,N,S, this research aims to learn a distributed and low dimensional embedding v_i for each urban region r_i. The embedding set can be denoted as:

$$V = \overrightarrow{v_0}, \overrightarrow{v_1}, \ldots, \overrightarrow{v_N}, v_i \in \mathbb{R}^d \tag{3.4}$$

where d is the uniform dimension for every urban region r_i. The embedding set V should preserve information of human mobility, geographic neighborhood, and POI side information.

3.3 A Multi-Graph Representation Learning Framework for Urban Region Profiling

In this section, we mainly introduce the proposed multi-graph and multi-modal representation learning framework, namely Region2Vec, for urban region embedding. Firstly, we roughly present an overview of the framework. Then we elaborate on three main modules for in our framework.

3.3.1 Framework Overview

Figure 3.3 shows the pipeline of our proposed multi-graph and multi-modal representation learning framework. Different modalities of urban data, including human mobility, geographic neighborhood, and POI side information can be encoded using multiple graphs. First, a correlation modeling module is introduced to construct multi-graphs based on multi-modal data. Then, a graph attention network [117] is used to aggregate and update information in each graph. After that, we propose a multi-graph fusion module, to deeply integrate multi-graph information. In this way, the final embedding incorporates non-Euclidean correlations among regions based on human mobility, geographic neighborhood, and POI side information.

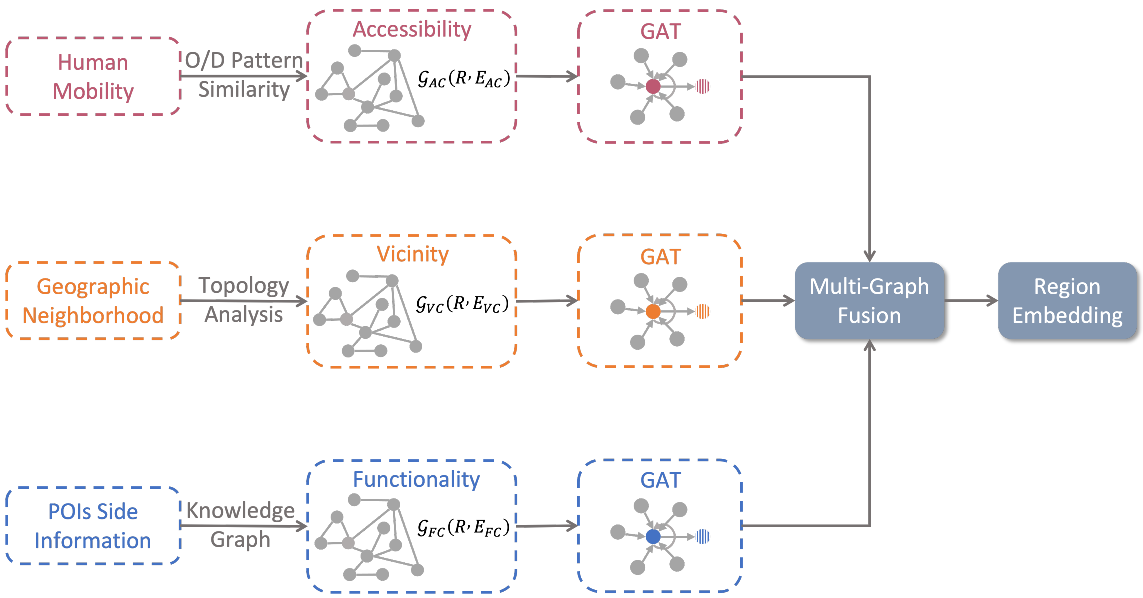

Figure 3.3: Overall architecture of the proposed multi-graph representation learning framework Region2Vec.

3.3.2 Correlation Modeling Module

Correlations among urban regions can be described in different aspects. From the aspect of human mobility, a trip has an origin region and a destination region, which can form a correlation. For many trips, origin regions/destination regions can also be related to other origin regions/destination regions based on mobility patterns in terms of accessibility. As for geographic neighborhood, vicinity in space can be revealed.

As the representation of urban functions, POI side information reflects functionality correlations. Similar regions in terms of accessibility, vicinity, and functionality will show high correlations due to proximity in non-Euclidean space. In our study, we construct three types of region correlations based on human mobility, geographic neighborhood, and POI side information.

Accessibility Correlation Modeling Based on Human Mobility.

Human mobility directly reveals the inter-region interaction movement between people and urban spaces. It is found that if trips have the same O/D regions, then the different D/O regions of trips are similar [135]. That is to say, through the O/D pattern similarity, important underlying accessibility correlation can be modeled and captured based on human mobility. Suppose we have a human mobility dataset M, the similarity value between region r_i and region r_j is computed as:

$$s_{r_j}^{r_i} = |(r_i, r_j) \in M| \tag{3.5}$$

where $(r_i$,$r_j)$ form a trip in M, and $|.|$ calculates the length of set. Then, the O/D pattern similarity and accessibility correlations among regions can be defined as:

$$p_o\left(r \mid r_i\right) = \frac{s_{r_i}^{r}}{\sum_r s_{r_i}^{r}}, p_d\left(r \mid r_i\right) = \frac{s_r^{r_i}}{\sum_r s_r^{r_i}} \tag{3.6}$$

$$\begin{aligned} AC_o^{ij} &= simi\left(p_o\left(r \mid r_i\right), p_o\left(r \mid r_j\right)\right), \\ AC_d^{ij} &= simi\left(p_d\left(r \mid r_i\right), p_d\left(r \mid r_j\right)\right) \end{aligned} \tag{3.7}$$

where simi(.) is the function for calculating the cosine similarity; AC_o^{ij} is the accessibility correlation between two O regions; AC_d^{ij} denotes the accessibility correlation between two D regions.

Vicinity Correlation Modeling Based on Geographic Neighborhood.

Spatial vicinity is a kind of important correlations due to the Frist Law of Geography [112]. Adjacent regions in space naturally are more similar. According to formula 4.5, through topology analysis, the geographic neighborhood dataset N contains geographic neighborhood vector $\overrightarrow{n}$ for each urban region r. These vectors actually represent vicinity correlations among regions. The vicinity correlations are described as:

$$VC^{ij} = simi\left(\overrightarrow{n_i}, \overrightarrow{n_j}\right) \tag{3.8}$$

where VC^{ij} is the vicinity correlation between region r_i and r_j.

Functionality Correlation Modeling Based on POI Side Information.

The POI side information of a region reveals the functionality, and also reflect intra-region features. To include more accurate, diverse, and explainable information, it is necessary to go beyond POI itself and take POI attributes (i.e., POI side information) into account. We here choose to use the knowledge graph of POI side information to construct the functionality correlation model. Generally speaking, knowledge graph is a relational network obtained by connecting different kinds of information. For a typical KG G, it expresses data as a directed graph $G = \{E, R, T\}$, where E, R and T denote the sets of entities, relations and facts respectively. Each triple $(h, r, t) \in T$ indicates a relation $r \in R$ between head entity $h \in E$ and tail entity $t \in E$ exists. We can then conduct knowledge graph embedding, which is an effective way to parameterize entities and relations as vectors, while preserving the graph structure.

When we construct KG for POI side information, as illustrated in Figure 3.4, there are two methods for defining nodes of knowledge graph: (a) only including POIs and side information (i.e., entities); (b) including regions, POIs, and side information.

Different types of relations among nodes are regarded as different types of edges. Here, we list only 3 types of relations in (a) and 4 relations in (b) for toy instance. Some simple connectivity can be easily found such as $p_1 - R_2 - e_3$ in (a) and $r_1 - R_1 - p_1 - R_3 - e_3$ in (b). Two triples (r_1,R_1,p_1) (p_1,R_3,e_3) can be extracted from the case in (b). However, if further we treat the edges as reversible, we can capture the long-range connectivity in knowledge graph, such as: $p_1 - R_2 - e_3 - (-R_2) - p_3 - R_3 - e_5$ in (a) and $r_1 - R_1 - p_1 - R_3 - e_3 - (-R_3) - p_3 - R_4 - e_5$ in (b). The case in (b) indicates that in addition to e_3 and p_3, entity e_5 will also have some impacts on region r_1. Considering that functionality region embeddings base on POI side information can be obtained neither from direct knowledge graph embedding for regions or averaging knowledge graph embedding for POIs which are located in regions, we will have 6 types of approaches for getting functionality region embeddings as shown in Table 1.

In this work, TransD [49] is employed on knowledge graph embedding for getting functionality region embeddings $\overrightarrow{s}$. The functionality correlations are described as:

$$FC^{ij} = simi\left(\overrightarrow{s_l}, \overrightarrow{s_J}\right) \tag{3.9}$$

where FC^{ij} is the functionality correlation between region r_i and r_j.

KG type	Method	Edges	Embeddings
KG01	(a)	not reversible	averaging
KG02	(a)	reversible	averaging
KG03	(b)	not reversible	direct
KG04	(b)	not reversible	averaging
KG05	(b)	reversible	direct
KG06	(b)	reversible	averaging

Table 3.2: 6 types of approaches for getting functionality region embeddings.

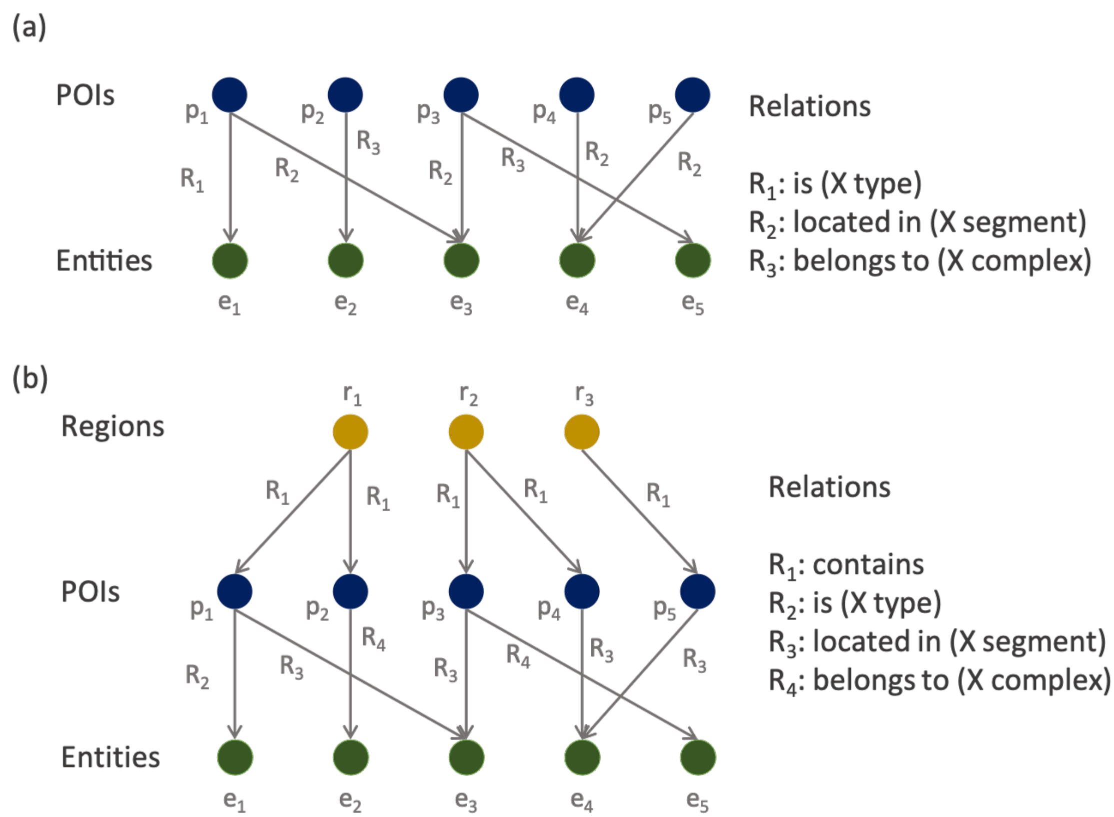

Figure 3.4: Two methods of constructing knowledge graph. Here, we list only 3 relations in (a) and 4 relations in (b) for toy instance.

3.3.3 Graph Attention Network Module

We construct graphs for accessibility correlation AC, vicinity correlation VC, and functionality correlation FC, respectively. Each graph can be denoted as $\mathcal{G}(R, E)$, where $R = \{r_i\}_{i=1}^{n}$ represents n regions and $E = \{E_i\}_{i=1}^{n}$ indicates the edges connected with the set of n nearest neighbors of each node. Therefore, we have $\mathcal{G}_{\mathcal{AC}}$, $\mathcal{G}_{\mathcal{VC}}$, and $\mathcal{G}_{\mathcal{FC}}$ based on different kinds of correlations. Then, the graph attention network [17] is applied to integrate and update node representations of each graph. The attention mechanism on the graph-structured data can automatically learn weights of information from neighbors of a node during the propagation. The output node representation of graph attention network (GAT) module of each graph can be denoted as E_{AC}, E_{VC}, and E_{FC}, respectively. The output node representation contains information of its neighbors.

3.3.4 Multi-Graph Fusion Module

After the GAT module integrates and updates different types of region information for each graph, the multi-graph fusion module is needed for fusing all types of information into the final region embeddings. Region correlations from different graphs are highly related. Taking human mobility and POI side information as an example, similar OD pairs in the morning peak and evening peak generally represent commuting between residence districts and business districts. As for geographical neighborhood and human mobility/POI side information, we can use the First Law of Geography to explain it — ”Everything is related to everything else, but near things are more related to each other”. Such relationships among these three graphs give us an intuition that incorporating information from multi-graphs will not only improve the performance but also enhance the learning process for each graph. A multi-graph fusion module can endow the Region2Vec with the capability to incorporate spatial semantics from region-wise side information, local-level geographical adjacent relations, and global-level mobility pattern. As shown in Figure 3.5, we employ an encode-decode architecture for enabling the effective integration among multiple graphs. Then we design a loss function as our overall learning object.

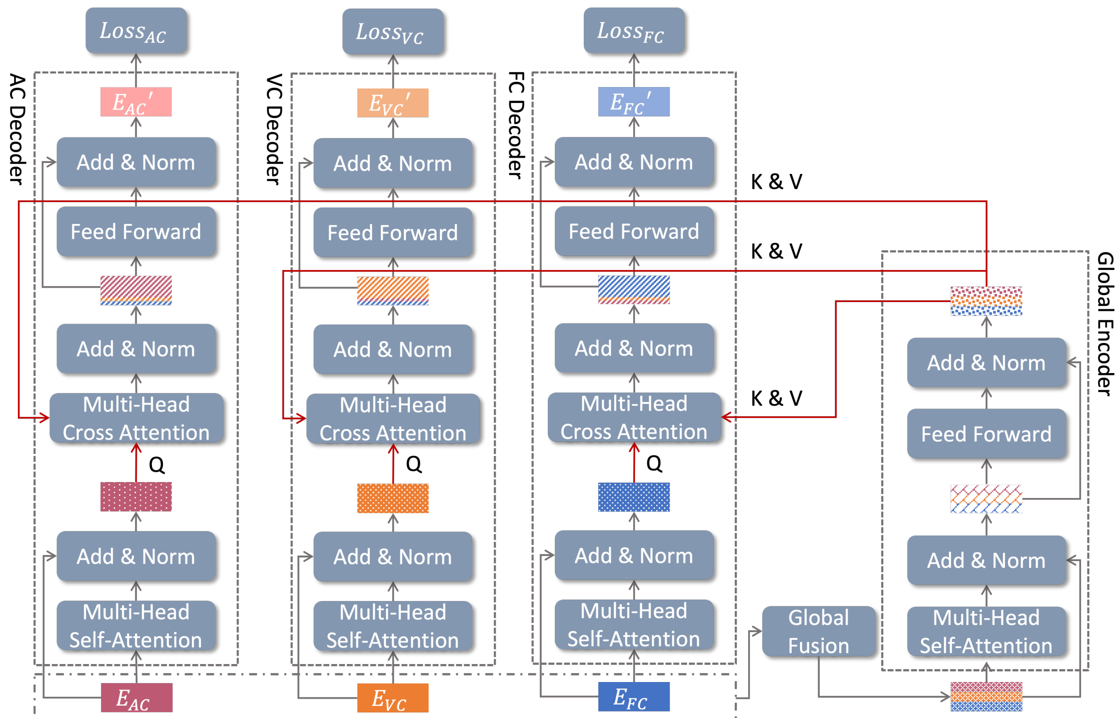

Figure 3.5: The architecture of multi-graph fusion module.

Global Encoder.

Multi-graph representations E_{AC}, E_{VC}, and E_{FC} are first concatenated and fed into the global fusion layer, which generates comprehensive region embeddings by a single layer MLP. For each graph feature E_m, the fusion process can be described as:

$$E = \sum \sigma \left(E_m W + b \right) E_m. \tag{3.10}$$

Following the transformer architecture [115], in the global encoder layer, we use two sub-layers a multi-head self-attention mechanism for enabling further integration of information, and a fully connected feed-forward network for deeply feature extraction. Residual connection and layer normalization are employed around each of the two sub-layers.

AC/VC/FC Decoder.

In addition to the two sub-layers mentioned in the encoder layer, the AC/FC/VC decoder applies a multi-head cross-modal attention, which treating the output of the encoder layer as the key and value, and using the attention result of themselves as the query. For the result $E_{m'}$ via the self-attention mechanism for each graph representation E_m, the query matrix $Q \in \mathbb{R}^{n \times d}$, key matrix $K \in \mathbb{R}^{n \times d}$ and value matrix $V \in \mathbb{R}^{n \times d}$ can be defined as:

$$Q = EW_Q, K = E_{m'}W_K, V = E_{m'}W_V. \tag{3.11}$$

Residual connection and layer normalization are also employed around each of the three sub-layers.

Loss Function Designation.

Through the multi-graph fusion module, features of each graph are updated. Various learning tasks are then designed based on these updated graph representations.

- **Accessibility Correlation Reconstruction**

We aim to reconstruct accessibility correlation by maximizing the probability of O/D occurrence. We expect that the possibility of predicting the O/D region given the D/O region based on the region representations will be the highest. Then the accessibility correlation reconstruction loss between region r_i and region r_j can be computed as:

$$\begin{aligned}
\hat{p}_o\left(r_j \mid r_i\right) &= \frac{\exp\left(E_o^{i^T} E_d^j\right)}{\sum_j \exp\left(E_o^{i^T} E_d^j\right)}, \\
\hat{p}_d\left(r_j \mid r_i\right) &= \frac{\exp\left(E_d^{i^T} E_o^j\right)}{\sum_j \exp\left(E_d^{i^T} E_o^j\right)}
\end{aligned} \tag{3.12}$$

$$Loss_{AC} = \sum_{(r_i,r_j)\in\mathcal{M}} -log\hat{p}_o\left(r_j|r_i\right) - log\hat{p}_d\left(r_j|r_i\right) \tag{3.13}$$

where E_o^i means the accessibility region representation (i.e., E_{AC}) when i-th region acts as O region.

- **Vicinity Correlation Reconstruction**

We design the vicinity correlation reconstruction loss to make the final region representation preserve the information from geographical neighborhood. The vicinity correlation reconstruction loss can be computed as:

$$Loss_{VC} = \sum_{i,j}\left(A^{ij} - E_{VC}^i E_{VC}^j\right)^2 \tag{3.14}$$

where A^{ij} is the vicinity correlation between region r_i and region r_j.

- **Functionality Correlation Reconstruction**

We design the functionality correlation reconstruction loss to make the final region representation preserve the information from POI side information. The functionality correlation reconstruction loss can be computed as:

$$Loss_{FC} = \sum_{i,j}\left(B^{ij} - E_{FC}^i E_{FC}^j\right)^2 \tag{3.15}$$

where B^{ij} is the functionality correlation between region r_i and region r_j.

- **Overall Learning Object**

Then, the final loss function can be represented as:

$$\mathcal{L} = Loss_{AC} + Loss_{VC} + Loss_{FC}. \tag{3.16}$$

Table 3.3: Details of NYC datasets for urban environment comprehension.

Dataset	Details
Census blocks	180 block boundaries split by streets in Manhattan.
Taxi trips	$\sim$ 10 million taxi trip records accumulated in one month in Manhattan.
POI data	$\sim$ 6 thousand POIs (including 9 types of attributes) in Manhattan.
Crime data	$\sim$ 30 thousand crime records during one year in Manhattan.
Check-in data	$\sim$ 80 thousand check-in records during one year in Manhattan.
District division	Manhattan is divided into 12 districts based on the land usage.

3.4 Experiments

In this section, we conduct several experiments on real-world datasets to evaluate the performance of Region2Vec. The objective of our experiments is to answer the following questions:

RQ1: How well does Region2Vec perform in various downstream urban analytics tasks?

RQ2: How do different modules of Region2Vec contribute to the model performance?

RQ3: For POI side information, which kind of knowledge graph should we choose?

3.4.1 Study Area and Datasets

We choose Manhattan borough in New York City as our study area. In this study, the area is divided into 180 regions. From NYC Open Data,[*] we collect real-world datasets such as census block shapefile, taxi trips, POI data, crime data, and check-in data. We also find the district division by the community boards from [8]. The details of these datasets can be found in Table 3.3.

3.4.2 Downstream Tasks for Evaluation

Region Clustering Visualization.

Regions may fall into the same category if their land-use type is similar. To verify whether our obtained region embeddings effectively fuse multi-graphs to contain the

[*] http://opendata.cityofnewyork.us/

information of land use, we cluster region embeddings using K-means and visualize the results to intuitively interpret them. As shown in Figure 3.6 (a), the district division data from community boards [8] which divide Manhattan into 12 components are used as ground truth. Thus, we partition the study area into 12 clusters. For the clustering result, regions with the same land use type should be in the same group.

Region Clustering Evaluation.

We use the following two metrics to further quantitively evaluate region clustering results of the proposed embedding method and baselines:

Normalized Mutual Information (NMI): It describes the purity of region clustering results. NMI is defined as:

$$NMI = \frac{I(X,Y)}{[H(X)+H(Y)]/2} \tag{3.17}$$

where X is the set of prediction (i.e., clusters) and Y is the set of labels (i.e., ground truth). $I(X,Y)$ denotes the mutual information between elements in X and Y. $H(X)$ and $H(Y)$ denote the entropy of clusters and ground truth respectively.

Adjusted Rand Index (ARI): It is the corrected-for-chance version of the Rand Index (RI). ARI is defined as follows:

$$\begin{aligned} ARI &= \frac{RI - E(RI)}{\max(RI) - E(RI)}, \\ RI &= \frac{TP+TN}{TP+FP+TN+FN} \end{aligned} \tag{3.18}$$

where $E(RI)$ denotes the expectation of RI. $TP/TN/FP/FN$ denotes true positive/ true negative/ false positive/ false positive/ false negative, respectively. The value of ARI is between -1 and 1. A value closes to 0 means random labeling, while a value close to 1 means perfect match.

Crime Prediction.

Lasso regression model [110] is employed in this task to predict the number of crime events. The independent variables are region embeddings, while the dependent variable is the number of crime events. In this part, we use three metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2), to measure the performance of different approaches. The first two metrics are to quantitate errors of prediction, while the last one is to estimate the goodness of fit of models. We calculate these three metrics by K-Fold cross-validation, where K is set as 5. The value of the L1 normalization weight of Lasso regression is chosen by grid searching.

Popularity Prediction.

Check-in volume is regarded as the popularity. The process in popularity prediction is the same as crime prediction.

3.4.3 Performance Comparison (RQ1)

To demonstrate the performance of our proposed Region2Ve, we have selected four classical models from the traditional graph embedding methods, as well as three of the latest models based on deep learning methods published in top-tier conferences, to serve as baselines.

I. Graph Embedding Baselines.

GAE: We apply Graph Auto-Encoder (GAE) proposed in [58] on multi-graphs and make the GAE model of each graph share the middle layer to learn region embeddings.

DeepWalk: We apply the DeepWalk model proposed in [91] on multi-graphs and concatenate the embeddings of each graph to get region embeddings.

LINE: We apply the LINE model proposed in [109] on multi-graphs and concatenate the embeddings of each graph to get region embeddings.

Node2Vec: We apply node2vec proposed in [32] on multi-graphs and concatenate the embeddings of each graph to get region embeddings.

II. State-of-the-Art Methods.

ZE-Mob: ZE-Mob proposed in [135] learns region embeddings by considering the co-currency relation of regions in human mobility trips.

MV-PN: MV-PN proposed in [23] learns region embeddings with a multi-view POI network within the region.

MV-Embedding: MV-Embedding proposed in [144] learns region embeddings based on both human mobility and inherent region properties.

In the experiments, the embedding size of ZE-Mob is set 96 as recommended by the authors. Thus, the embedding sizes of baselines, our model and variants are all set as 96. To enhance the GAT performance, 8-head attention mechanism is employed in each GAT layer. For the multi-graph fusion module, we set the head number h of multi-head self-attention as 4.

Region Clustering Visualization.

In Figure 3.6, the same color marks regions in the same cluster. We can see from the comparison between (c)-(e) and (b) in Figure 3.6 that multi-graph and multi-modal methods are much better than uni-graph and uni-modal method. In addition, our proposed Region2Vec has more ideal clustering than another state-of-the-art method MV-Embedding, in terms of consistence with real boundaries of ground truth.

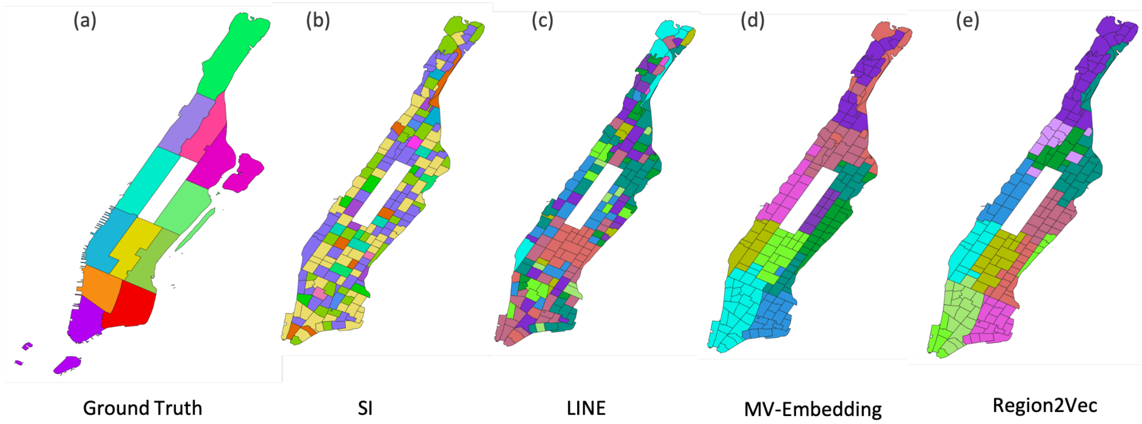

Figure 3.6: Region clustering results for some methods in Manhattan borough. (a) Ground Truth; (b) SI; (c) LINE; (d) MV-Embedding; (e) Region2Vec.

Region Clustering Evaluation

Figure 3.7 shows the NMI and ARI of region clustering results on all approaches. We can find that: (1) Our proposed Region2Vec outperforms all baseline approaches. Compared with state-of-the-art methods, it has 16.64% increase in performance in NMI and 12.66% increase in performance in ARI. (2) Methods for a simple combination of multi-graphs, such as GAE, LINE, and Node2Vec, obviously cannot make full use of multi-graph information.

Crime Prediction

Results are presented in Figure 3.8. We can observe that our proposed Region2Vec achieves the best performance compared with all approaches. Region2Vec has 9.95%, 8.82%, and 13.89% improvement in MAE, RMSE and R2, respectively.

Popularity Prediction

Results are presented in Figure 3.9. Compared with state-of-the-art methods, Region2Vec has 10.34%, 7.80% and 12.88% improvement in MAE, RMSE and R2, respectively.

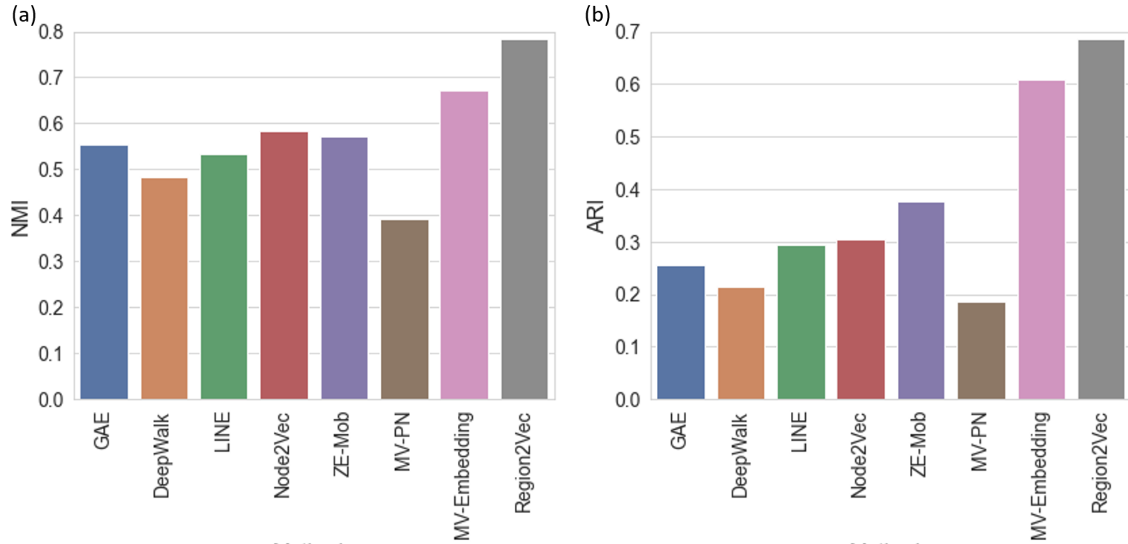

Figure 3.7: Results of region clustering using baselines and state-of-the-art methods.

3.4.4 Ablation Study (RQ2)

To better understand the effect of each module on multi-graph multi-task training with Region2Vec, we perform following ablation experiments:

I. Ablation Study for Correlation Modeling Module

HM: Region2Vec applied merely on human mobility (HM).

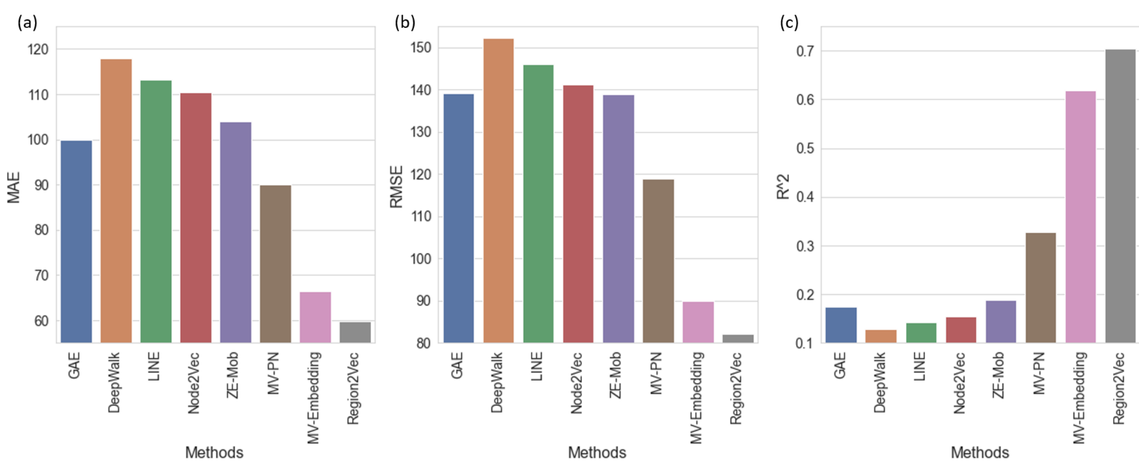

Figure 3.8: Results of crime prediction using baselines and state-of-the-art methods.

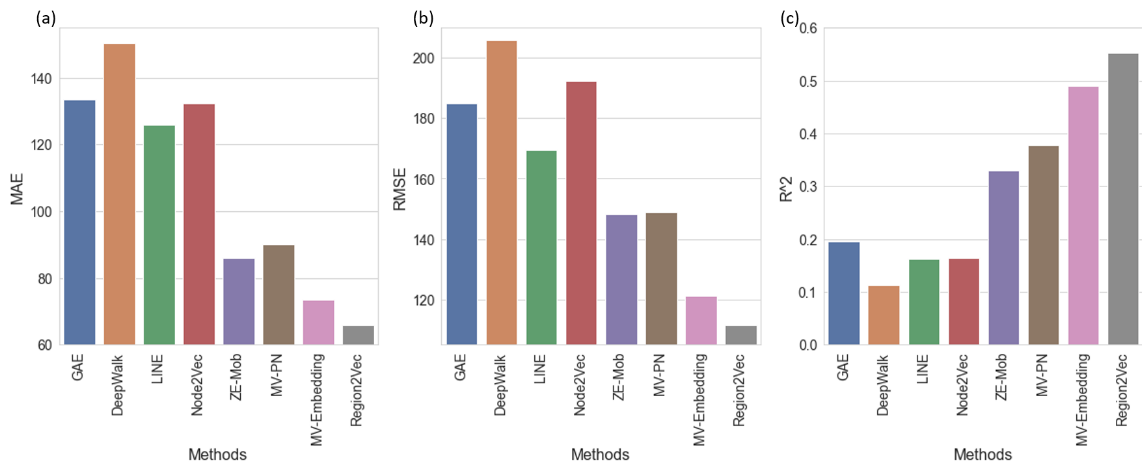

Figure 3.9: Results of popularity prediction using baselines and state-of-the-art methods.

GN: Region2Vec applied merely on geographic neighborhood (GN).

SI: Region2Vec applied merely on side information (SI).

HM+GN: Region2Vec applied merely on HM and GN.

HM+SI: Region2Vec applied merely on HM and SI.

II. Ablation Study for GAT Module

R2V-g: Region2Vec without GAT module to propagate and update information for each graph.

III. Ablation Study for Multi-Graph Fusion Module

R2V-f: Region2Vec without multi-graph fusion module to deeply integrate all graphs. Every graph is assigned equal weights when obtaining the final region embedding result.

R2V-m: We use the fusion module in MV-embedding while disabling multi-graph fusion module for Region2Vec.

Region Clustering Visualization.

We can see from the comparison between (c)-(e) and (b) in Figure 3.6 that multi-graph methods are much better than the uni-graph methods.

Region Clustering Evaluation.

We can see from Figure 3.10 that: (1) Every module in Region2Vec is necessary. (2) The methods for bi-graphs (i.e., HM+GN/HM+SI) generally have better performance than the methods for uni-graph (i.e., HM/GN/SI). Among uni-graph methods, HM marginally outperforms the other two methods, which indicates the importance of accessibility correlation in region clustering tasks. (3) If we just simply combine the information from multi-graphs (i.e., R2V-g, R2V-f, and R2V-m), it will cause at least 8.73% and 9.78% reduction of performance in NMI and ARI, respectively.

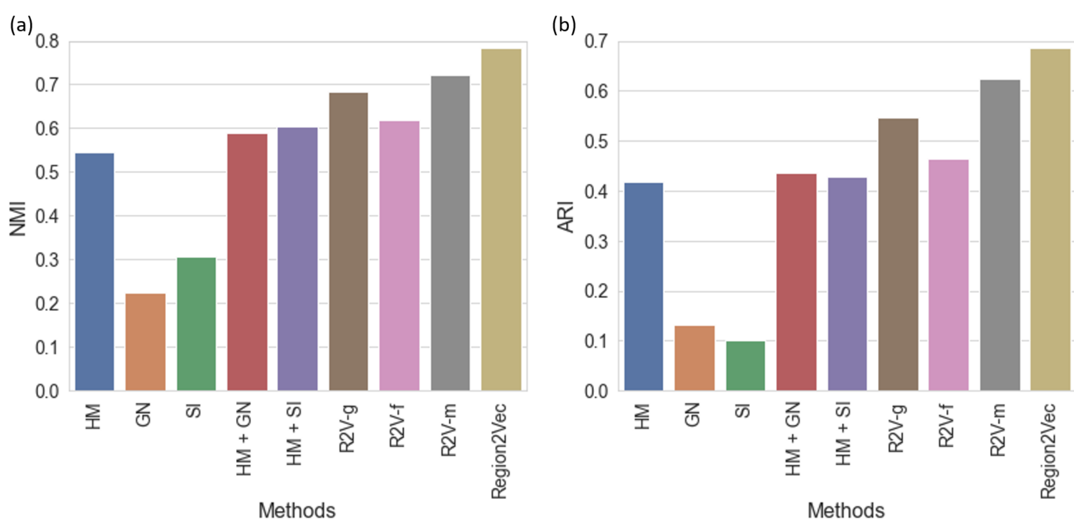

Figure 3.10: Results of region clustering evaluation using variants of Region2Vec.

Crime Prediction.

Results are presented in Figure 3.11. Region2Vec has at least 7.78%, 6.49%, and 12.8% improvement in MAE, RMSE and R2, respectively.

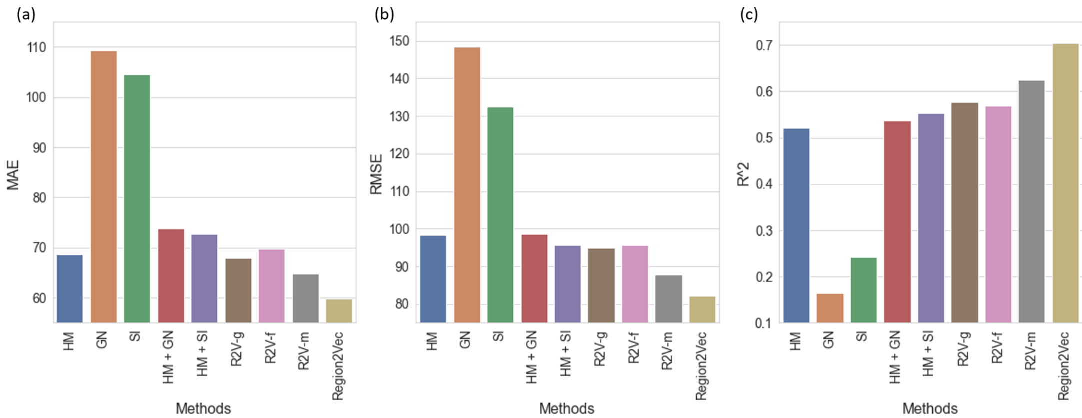

Figure 3.11: Results of crime prediction using variants of Region2Vec.

Popularity Prediction.

Results are presented in Figure 3.12. Region2Vec has at least 5.64%, 6.11% and 8.45% improvement in MAE, RMSE and R2, respectively. The necessity of each module is soundly verified.

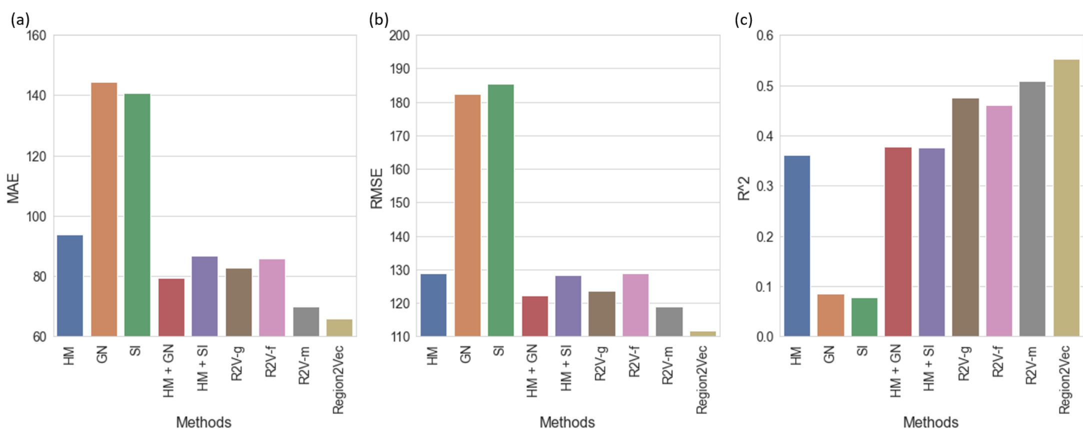

Figure 3.12: Results of popularity prediction using variants of Region2Vec.

3.4.5 Knowledge Graph Selection (RQ3)

We here use two downstream tasks — region clustering evaluation and crime prediction — to select the most appropriate knowledge graph structure for Region2Vec. As shown in Figure 3.13 and Figure 3.14, KG02 generally achieves the best performance in 6 KGs. KG02 corresponds to constructing knowledge graph by only considering POIs and POI side information. The edges in KG02 are reversible. The region knowledge representation is obtained from taking the average of POI knowledge representation.

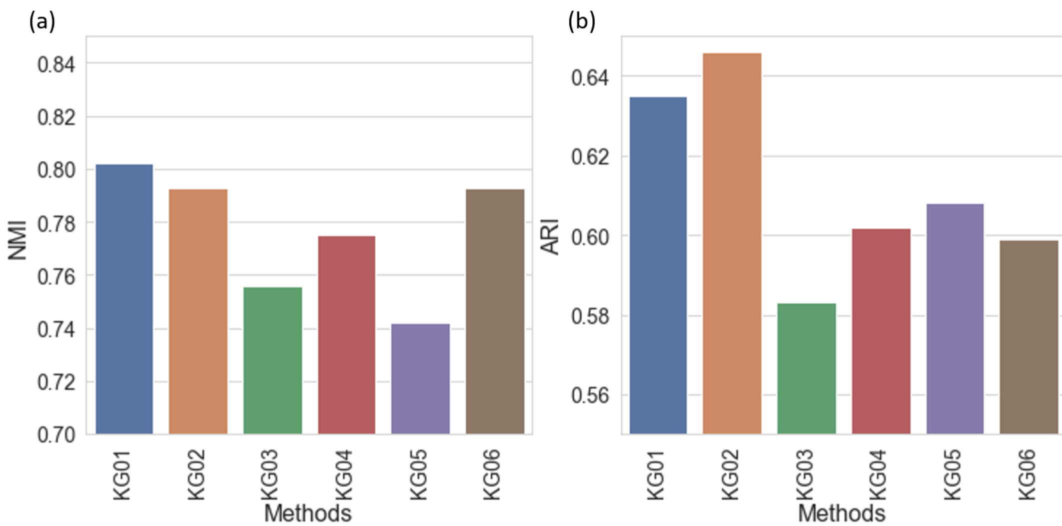

Figure 3.13: Metric values of region clustering evaluation for 6 KGs.

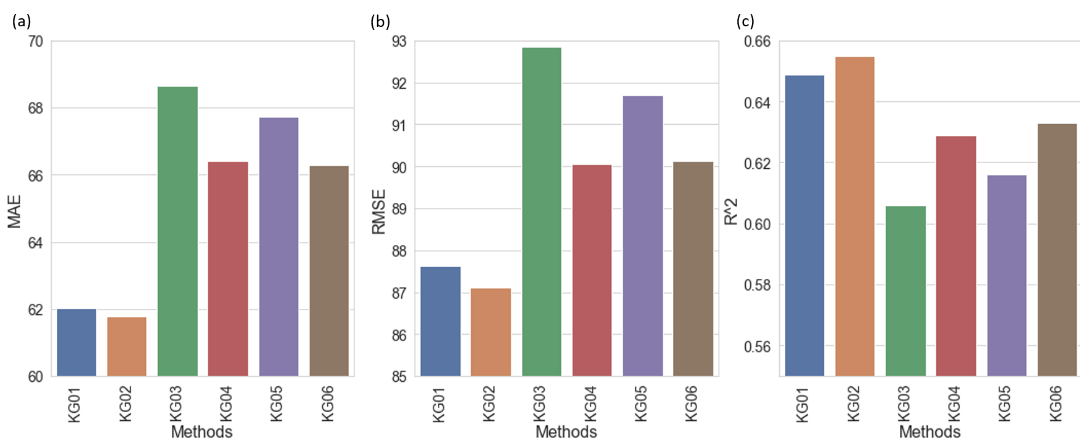

Figure 3.14: Metric values of crime prediction for 6 KGs.

3.5 Conclusions

In this study, we investigate urban region profiling problem. Region2Vec, a multi-graph and multi-modal representation learning framework, is proposed to learn an embedding space for urban regions. Through urban region profiling, the task-agnostic framework is capable of handling various applications. In particular, multiple modalities of urban data, including human mobility, geographic neighborhood, and POI side information are encoded using multi-graphs for representing inter-region relations, geographical contextual information, and intra-region information, respectively. Accessibility, vicinity, and functionality correlations among regions are then constructed based on multi-graphs. After the GAT module is employed to aggregate and update information in each graph, we use a multi-graph fusion module to jointly learn comprehensive representations. Experiments on real-world datasets demonstrate the effectiveness of Region2Vec in capturing the latent representation from multi-graphs and multi-modalities for urban regions. We apply both quantitative and qualitative methods to evaluate the proposed model. Results show that Region2Vec outperforms all state-of-the-art baselines in four downstream tasks. Our future work includes making our framework more task-oriented and focusing on the interpretability of our model.

Chapter 4

Human Mobility Understanding

4.1 Human Mobility Analysis

4.1.1 Introduction and Motivation

With the rapid advancement of information and location-aware technologies, large-scale movement datasets are now accessible, greatly enhancing empirical research on human travel behavior. Numerous studies have seized this opportunity to investigate human travel patterns, particularly from the perspective of trip chains [19, 30, 51, 147]. A trip chain is typically defined as a series of trip segments connecting two significant activity locations [81]. Analyzing trip chains allows researchers to uncover how people organize their activities across space and time, as well as the topological relationships among different activity locations. A deeper understanding of trip chains provides valuable insights for transportation planners and policymakers, benefiting land-use planning, improving urban accessibility, and even forecasting and controlling the global spread of epidemics [4, 42, 71, 80, 82].

The majority of existing research on trip chaining behavior has focused on the activities of city dwellers. These studies have discovered that people's travel patterns exhibit a limited number of predictable and characteristic trip chains that adhere to simple rules [102]. In other words, a few trip chains are sufficient to capture the primary characteristics of populations [97]. Moreover, citizens' travel decisions are influenced by the principle of least effort, with individuals of diverse backgrounds and socioeconomic characteristics aiming to minimize their effort across various aspects of human life [33, 41, 153]. However, limited effort has been devoted to examining the trip chaining behavior of tourists, primarily due to a lack of large-scale, fine-grained datasets capturing tourist movement patterns. Although some models of tourist travel patterns have been proposed [61], their validation often relies on small-scale datasets such as travel diaries [84]. Research quantifying tourist trip chains using fine spatiotemporal resolution movement datasets remains scarce. In today's context, analyzing tourist travel patterns is essential for helping tourists plan their trips more efficiently. If tourism-related businesses can develop marketing strategies catering to tourists based on travel patterns, they will experience accelerated growth. A deep understanding of tourists' movement patterns derived from big data can ultimately benefit both tourists and the tourism industry [16, 43, 84, 114] and have direct applications in tourism management activities, such as tour product development, attraction planning, and accommodation development. In light of these considerations, this research focuses on mobile phone big data of South Korean tourists and aims to uncover their frequent travel patterns.

We acknowledge that limited quantitative measurements have been proposed for modeling tourists' trip chains. Accordingly, this study aims to address the following research questions: (1) What are the major topological characteristics of tourists' daily trip chains? (2) Do visitors organize their travels similarly across different days during their stay in a city? (3) Is the principle of least effort reflected in their travel behavior?

To answer these questions, we analyze a large-scale mobile phone dataset collected in two South Korean cities (Jeonju and Gangneung), which captures the location footprints of international travelers who visited these cities during a one-year period. This study seeks to provide answers to the aforementioned questions using a novel model capable of capturing and reproducing the spatiotemporal structures and regularities in tourist trip chains. Specifically, we extract trip chains from raw mobile phone trajectories and further examine the most popular ones to identify typical tourist trip chains. To address the first research question, we analyze daily trip chains by categorizing them into "intra-city" chains and "hybrid" chains. Our goal is to comprehend how tourists organize their travels during their time within a city or on the first and last days of their visits to a city. For the second research question, we strive to identify inherent patterns concerning individual tourist mobility and assess the predictability of tourist movement. Regarding the third question, we employ two metrics — average degree and average travel distance of trip chains — to evaluate tourist travel behavior in relation to the principle of least effort.

4.1.2 Study Area

South Korea is a country with a well-developed travel and tourism industry. Jeonju and Gangneung, as shown in Figure 4.1, which are two popular cities to international travelers in South Korea, are selected as areas of study. Jeonju, the capital city of Jeollabuk-do Province, is an important tourist center famous for traditional Korean food, historic buildings, sports activities, and festivals. It has an area of 206 km2 and a population of 0.65 million (as of 2017). Gangneung sits on the east coast of South Korea. The city has many tourist attractions, such as Jeongdongjin, a very popular area for watching the sun rise, and Gyeongpo Beach. It's also the city that hosted all the ice events for the 2018 Winter Olympics. As a city in Gangwon-do Province, it has an area of 1040 km2 and a population of 0.21 million (as of 2019).

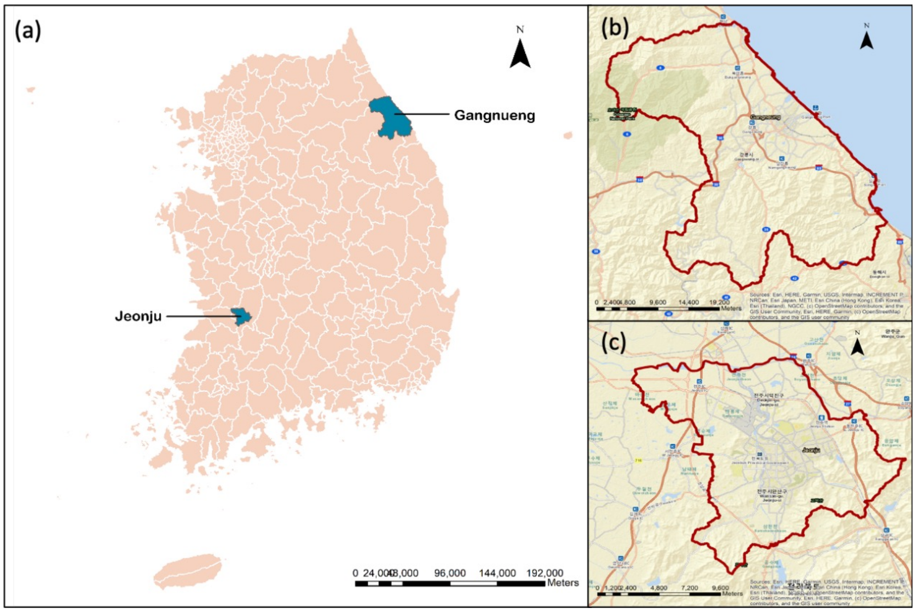

Figure 4.1: Study areas: (a) the whole South Korean, (b) Gangnueng, and (c) Jeonju.

This research uses a mobile roaming dataset collected by a major telecom company in South Korea. This anonymized dataset tracks the location footprints of 18,625 and 33,219 tourists who visited Jeonju and Gangneung respectively between August 1st, 2017 to July 31st, 2018. Since the timespan of dataset covers the 2018 Winter Olympics and mobility patterns of tourists could be different during this special event, we filter out this part of the data (from Jan 20th, 2018 to Feb 26th, 2018). Thus, the number of tourists in Gangneung changes to 15,095. Note that in this dataset, as long as an individual has visited Jeonju or Gangneung, the sequence of locations that he or she stayed when travelling in any other city of South Korea were also documented. This reveals additional information on when an individual entered/left Jeonju or Gangneung, which enables precise quantification of his/her trip chain on the first or last day of visit to a city. Table 4.1 shows an example of an individual's phone records. Each row in the table represents one stay activity and the time periods in between indicate trips among locations. For example, the first two rows in Table 1 indicate that the user stayed at two different locations during

[07:16:00 - 12:33:00] and [12:46:00 – 12:52:00] respectively, and a trip was possibly conducted by the user in between (i.e., [12:33:00 - 12:46:00]).

Table 4.1: Example of an individual's mobile phone records

User ID	Date	Starting Time	Ending Time	Longitude	Latitude
28***	2017-08-25	07:16:00	12:33:00	127.***	35.***
28***	2017-08-25	12:46:00	12:52:00	127.***	36.***
28***	2017-08-25	13:08:00	13:24:00	127.***	36.***
...	...	...	...	...	...
28***	2017-09-06	15:01:00	15:14:00	126.***	35.***
28***	2017-09-06	15:43:00	16:07:00	126.***	35.***
28***	2017-09-06	16:41:00	17:00:00	126.***	35.***

The locations of users were positioned at the level of cellphone tower and their densities in space define the spatial resolution of the dataset. The numbers of cellphone towers in Jeonju and Gangneung are 782 and 704, respectively. To better understand their spatial arrangement in the two cities, we calculate the statistics of the distance from each cellphone tower to its nearest peer. The average distance is 250 meters in Jeonju and 420 meters in Gangneung. Overall, the dataset provides a fine-grained view of tourist mobility in time and space, which allows for reliable extraction of tourist trip chains.

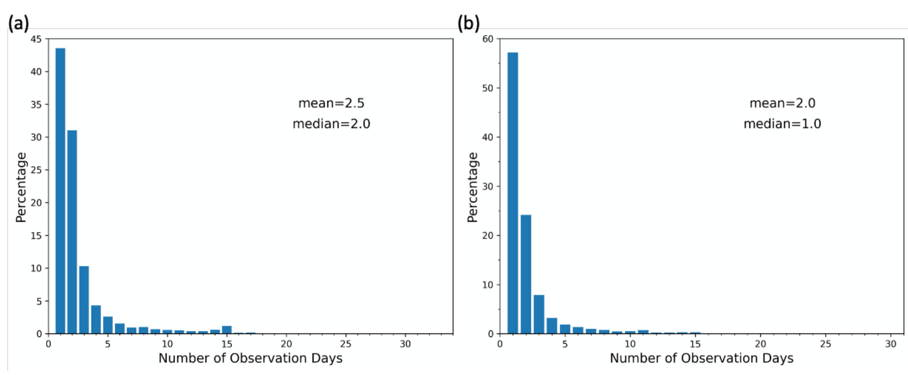

Figure 4.2: The number of observation days of users in two cities.

As shown in Figure 4.2, the number of observation days of users (i.e., the number of days with records for a user) show similar distribution patterns in two cities. Most of the users stayed for only a few days. The median values of observation days are 2.0 for Jeonju and 1.0 for Gangneung, and the mean values are 2.5 for Jeonju and 2.0 for Gangneung. The maximum values of observation days are 34 days for Jeonju and 30 days for Gangneung, respectively. Note that for some travellers, they would stay in a particular city for a few days and left, and then came back for another visit. In this study, we do not perform trip chain analysis for individuals with gap days. Such individuals account for 7.22% and 12.66% of visitors in Jeonju and Gangneung, respectively.

4.1.3 Characterizing Tourist Daily Trip Chains Using Mobile Phone Big Data

To understand and quantitatively model tourists' mobility patterns, we designed a pipeline to analyze the trip chains of tourists. This pipeline contains four steps: (1) extracting meaningful activity locations (”anchor points”) from tourists' cellphone trajectories; (2) constructing daily trip chains of tourists and examining their key characteristics; (3) quantifying the day-to-day transitions of individual trip chains; (4) exploring decisive factors that shape tourists' trip chains.

Deriving Individual Activity Anchor Points from Cellphone Trajectories

Identifying meaningful locations of travelers is an essential step for trip chain analysis. However, the cellphone tower locations documented in the dataset do not always reflect the actual locations of users for several main reasons: (1) the signal of mobile phones could switch between adjacent cellphone towers, producing the so-called ”ping-pong” effect [44, 95]; (2) the signal transmitted and received by cellphone towers will be compromised during propagation, leading to the inaccuracy of cellphone tower positioning [45]. Therefore, we argue that the combination of neighboring cellphone towers could better represent a location or place that is meaningful to a traveler.

Activity anchor points have been used in previous studies to describe a person's major activity locations [2, 20, 98]. In this study, we define an anchor point (AP) as a set of cellphone towers that are close to each other and where an individual has stayed over a certain period of time.

Given a cellphone trajectory $T = (v_1, t_1^s, t_1^e), (v_2, t_2^s, t_2^e), \ldots, (v_n, t_n^s, t_n^e)$ that documents a user's location footprints. Here $v_i = (\mathrm{lng}_i, \mathrm{lat}_i)$ symbolizes the cellphone tower location of the i^{th} record; t_i^s and t_i^e represent the starting and ending time of i^{th} record. The anchor point extraction works as follows. First, we calculate the total amount of time the individual stayed at each cellphone tower, and sort all the cellphone tower in descending order based on total stay duration. We start from the cellphone tower with the largest stay duration and group other cellphone towers within a roaming distance (Δd) of the selected cellphone tower into a cluster. Then from left cellphone towers that have not been assigned to any cluster, we conduct the same process on the one with longest stay duration. Iterating above steps until all the cellphone towers in trace T are tackled with, the trajectory is processed into a sequence at anchor point level. Since the average nearest distance between cellphone towers in both Jeonju and Gangneung are below 500m, we set Δd as 500m for both cities when performing the anchor point extraction.

In this way, an individual's trajectory can be represented as a sequence of APs. Figure 4.3 displays an example of how APs are extracted from the cellphone trajectory. The five vertical line segments mean five rows of records. A to E represent five cellphone towers in the trajectory T. R_1 to R_3 denote the extracted activity anchor points. Given the individual's cellphone trace $T = (A, t_A^s, t_A^e), (B, t_B^s, t_B^e), \ldots, (E, t_E^s, t_E^e)$, the travel patterns can be denoted as a sequence of cellphone tower locations tra-

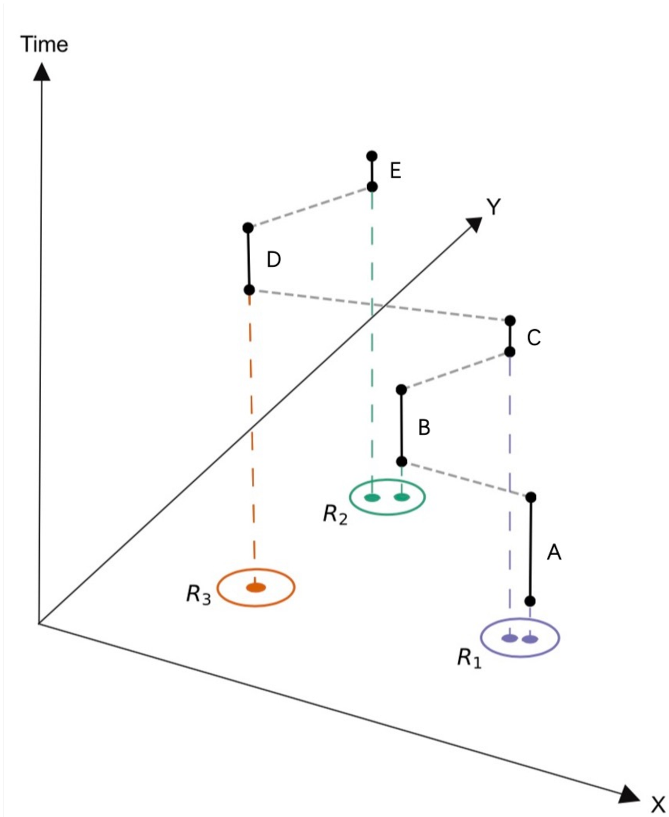

Figure 4.3: An example of extracting anchor points from a trajectory.

versed: A → B → C → D → E. In order to implement the above workflow, the cellphone tower with the longest total duration (A) is selected to create the buffer. The distance between cellphone tower C and A is less than 500m, so they are grouped as AP R_1. Then, the next cellphone tower with the highest amount of time (B) is selected and grouped with E to form R_2. The left cellphone tower D forms AP R_3 itself. At this point, the individual's travel patterns can be redefined as a sequence of APs: $R_1 \rightarrow R_2 \rightarrow R_1 \rightarrow R_3 \rightarrow R_2$. And the trace is processed to $T' = (R_1, t^s_{R_1}, t^e_{R_1}), (R_2, t^s_{R_2}, t^e_{R_2}), \ldots, (R_2, t^s_{R_2}, t^e_{R_2})$, where $R_j = (\mathrm{lng}_j, \mathrm{lat}_j)$ denotes the location and ID of the j^{th} AP. So far, we have converted the original observations at the cellphone tower level to a sequence at the AP level.

Constructing Tourist Daily Trip Chains

To further depict tourist daily travel patterns, we construct trip chains with different topological structures, which reveal how tourists organize their daily travels. Unlike previous human mobility research focusing on residents [3, 19, 51, 97], our research objects in this study are tourists. Residents' movements are usually within a given city; while for those tourists who visited several cities in one travel, they have records outside the given city. Thus, when we construct daily trip chains for a cross-city traveler, we have two kinds of trip chains: (1) hybrid" trip chains, which are composed of all records of an individual including within-city and out-of-city observations, and (2) "intra-city" trip chains, which refer to trip chains constructed with only intra-city records. For the "hybrid" trip chains, since the purpose of keeping the out-of-city records is to explore the travel pattern features of the day when cross-city travelers come or leave a given city, we only need to count the relevant records before arriving or after leaving the city as one aggregated AP, respectively.

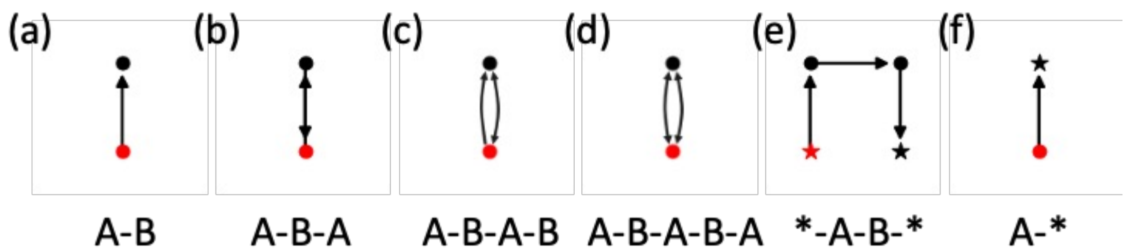

Figure 4.4: Examples of daily trip chain networks.

In Figure 4.4, (a) denotes an individual with a daily trip chain from anchor A to anchor B (A-B). (b) denotes an individual with a daily trip chain from anchor A to anchor B then to anchor A (A-B-A). (c) denotes an individual with a daily trip chain A-B-A-B. (d) denotes an individual with a daily trip chain A-B-A-B-A. (e) denotes an individual with a daily trip chain -A-B-, who is out of the city at the beginning and end and visited two "within city" anchors in the process. (f) denotes an individual with a daily trip chain A-*, who stays in a "within city" anchor A at the beginning, then leaves the city to an "out of city" anchor. When we plot the structure of trip chains, we use an edge with a one-way arrow to represent an individual moving from one anchor to another. We use an edge with a two-way arrow to represent a round trip between two APs. The red symbol represents the start anchor. A point in a round shape denotes anchor points within the city. A point in a star shape denotes anchor points outside the city. We use "*" to symbolize "out of city" anchors, and capital letters to symbolize "within city" anchors. Figure 4.4 (a) - (d) show typical "intra-city" trip chains, while (e) and (f) show typical "hybrid" trip chains.

Day-to-Day Transition of Tourist Trip Chains

To understand how an individual changes their travel patterns on the base of consecutive days, we need to analyze the transition between continuous daily trip chains. As we have mentioned in Figure 4.2, the median values of observation days are 2.0 for both cities. Thus, it is meaningful to explore the transition patterns of tourists' daily mobility between two consecutive days. In this part, we focus specifically on whether an individual will change their type of travel chain on the base of a consecutive sequence. By doing so, inherent patterns in tourist individual mobility will be explored, which may provide new insights into tourist mobility.

Note that days coming or leaving the given city are not comparable to days staying in the city, so we only focus on "intra-city" trip chains in this section. In this part, only individuals with at least two consecutive days of "intra-city" trips are analyzed. Suppose a tourist has records for Q days of "intra-city" trips, we can then extract $Q - 1$ pairs of consecutive days. For each pair of consecutive days, we define the trip chain of the former day as the original chain, and the trip chain of the latter day as the transferred chain. A transition matrix will then be constructed to count the frequency of different combinations of trip chain transitions. Row indexes of the transition matrix are different types of original chains, and column indexes are different types of transferred chains. We first investigate the frequency value of each element in the transition matrix, then we turn frequency into probability by summing each row. By doing so, each element value in the matrix presents the likelihood that people convert one type of original chain to different types of transferred chains. For instance, suppose a person has records in a given city for W continuous days (W is not less than 2). They stayed in one place (chain type: A) for the whole day on the first day and conducted a round trip with two nodes A, B (chain type: A-B-A) on the second day. Then the frequency of the matrix element corresponding to the original chain A to the transferred chain A-B-A should add one. Then we count the frequency for the transition of the other $W-2$ pairs of consecutive days for this individual. After all tourists in the given city are performed using the above procedures, the probability value of one original chain type changing to other transferred chain types can be calculated.

The Principle of Least Effort in Trip Chaining Behavior

We argue that the principle of least effort also acts as a driving force in trip chaining behavior. To demonstrate that it is a decisive factor in tourist travel behavior, we need to adopt some statistical metrics. Distance is a key metric in human mobility [88], and average degree is an important metric in network-like topological structure. Thus, in this work, we use two indicators to represent two aspects of trip chaining behavior: (1) average degree K, and (2) average distance D. These two indicators are defined and calculated as follows:

$$K = \frac{E}{N-1}, \tag{4.1}$$

$$D = \frac{1}{E}\sum_{i=1}^{N-1} d_{i,i+1}, \tag{4.2}$$

where E is the edge number of each trip chain, and N represents the number of anchor points (which we denote as N afterwards). i and $i+1$ are a pair of consecutive anchor points in a chain, and $d_{i,i+1}$ represents the Euclidean distance between anchor points i and $i+1$.

The purpose of investigating average degree K is to explore whether there is a connection between travel efficiency and people's preference of choosing daily trip chains. K can be regarded as a proxy of travel efficiency. Through the network structure of trip chains, we can clearly know whether people prefer to visit different locations in a single round tour before returning to the starting location, or if they prefer to return to their starting location before visiting another location. Obviously, the most effective way to conduct an itinerary with M anchor points is a round trip with M segments, in which case, K is equal to 1. If a person moves multiple times between anchor points, then K is larger than 1. The higher the value of K, the less efficient the individual travel is.

As for average distance D, it makes sense that people usually think ahead about the next day's itinerary, including how many places they want to go and the distances between these places in space. We calculate average distance for each sample to combine the information of anchor point number N and Euclidean distance between two consecutive anchor points in the trip chain. To some extent, the average travel distance of a trip chain can be considered as an integration of multiple factors related to the principle of least effort when people choose trip chains, such as travel cost, spatial range, and mobility regularity. Average distance can reflect the cost people

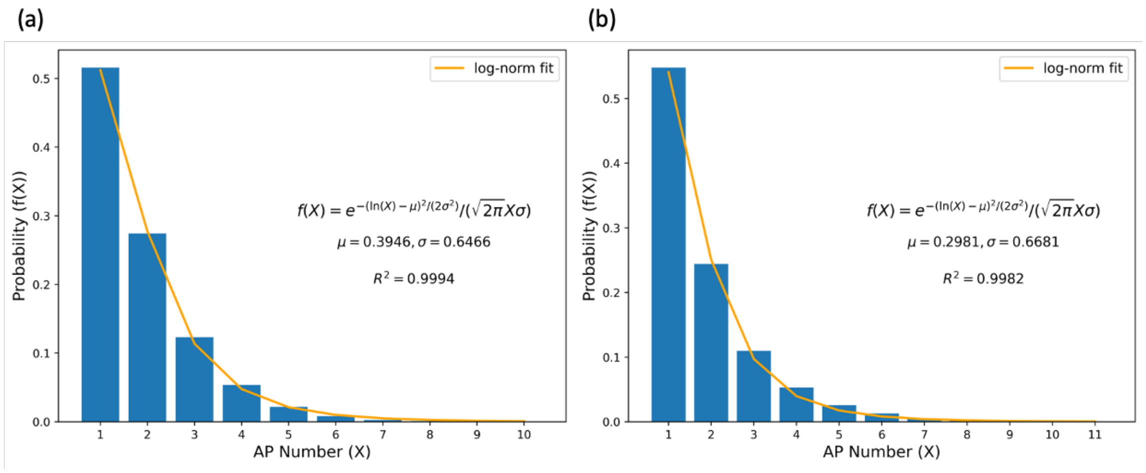

Figure 4.5: Distributions of AP number for (a) Jeonju dataset, (b) Gangneung dataset.

pay for travel in terms of time and space. From this point of view, the travel distance metric can offer another standpoint towards understanding tourist travel behaviors.

4.1.4 Results and Discussions

Distribution Patterns of Anchor Points

In this section, we report the distribution of the number of activity anchor points extracted from tourists' cellphone trajectories in Jeonju and Gangneung. We then explore the spatial patterns of these anchor points to gain insights into the tourists' spatial preferences in the two cities.

Numerical Distribution of Anchor Points. We first investigate the number of daily visited AP on individual basis. Through the numerical distribution of AP, we can have an insight into one aspect of tourisms' daily travel preference.

The distribution of AP as well as the fitting of the probability of travelers who visited a given AP number is presented in Figure 4.5. Tourists in Jeonju and Gangneung visited up to 10 and 11 APs in a daily basis respectively. The values of blue bars are the observed probabilities of different number of nodes. The yellow lines denote the best-fitted distributions — log-normal. The fitting was conducted via the least squares fit. The corresponding function and parameters are shown in each figure. The probabilities P of different AP number are: $P_{\text{Jeonju}}(X = 1) = 0.52$, $P_{\text{Jeonju}}(X = 2) = 0.27$, $P_{\text{Jeonju}}(X = 3) = 0.12$; $P_{\text{Gangneung}}(X = 1) = 0.55$, $P_{\text{Gangneung}}(X = 2) = 0.24$, $P_{\text{Gangneung}}(X = 3) = 0.11$. In addition, it's seems that daily human mobility patterns follow a universal law. The number of daily visited APs can be approximated with a log-normal fit:

$$f(X) = \frac{e^{-(\ln X - \mu)^2/2\sigma^2}}{X\sigma\sqrt{2\pi}} \tag{4.3}$$

with the parameters $\mu_{\text{Jeonju}} = 0.3946$, $\sigma_{\text{Jeonju}} = 0.6466$, $\mu_{\text{Gangneung}} = 0.2981$, $\sigma_{\text{Gangneung}} = 0.6681$.

The parameters are calculated with 95% confidence bounds, and R-squares of fittings are both above 0.99. The two datasets demonstrate great internal heterogeneity through log-normal fitting, which indicates travelers are diverse in terms of the number of places visited in a day. However, the similar values of μ and σ in two datasets reveal that distributions extracted from Jeonju and Gangneung dataset show similar patterns. In other words, the variance in travelers' spatial behavior are comparable between the two cities. The average AP number $\hat{X} \approx 2$ (1.83 for Jeonju and 1.84 for Gangneung) is small; hence, most people visit only a few places. In fact, 99.60% / 99.24% of the population visit less than seven APs on a daily basis in Jeonju / Gangneung.

The tail of the distributions shows that although most people visit less than four APs, a small fraction of tourists visit quite a lot APs within a day in some cities.

Spatial Patterns of Tourist Activities. In order to explore the hot spots of our study areas through meaningful activity anchor points, we use kernel density method to plot heatmap of anchors. A smaller radius of kernel density can make the local patterns obvious, while a larger radius can generate a smoother global surface. Note that the radiuses for kernel density need to be larger than the average distance of cellphone towers (250 m in Jeonju and 420 m in Gangneung). We also need to take the area of cities into account (206.22 km^2 for Jeonju and 1040 km^2 for Gangneung). Since our goal is to identify hotspots by kernel density, the search radiuses for Jeonju and Gangneung are set as 1000 m and 3000 m respectively.

The preliminary spatial distribution results are shown in Figure 4.6. In (a) (c), popular areas are derived through visited times of AP. The deeper the blue, the lower the value of kernel density; the deeper the red, the higher the value of kernel density. In (b) (d), the blue marks correspond to locations of spots in Jeonju and Gangneung respectively. The red star area in (d) represents Winter Olympic venues. Area 1, 2, 3 in (a) has the same locations with Area 1, 2, 3 in (b); Area 1, 2, 3, 4 in (c) has the same locations with Area 1, 2, 3, 4 in (d). As shown in Figure 4.6, even though some areas of spots are not detected, the distributions of popular areas of two cities derived by kernel density generally match with the ground truth provided by TripAdvisor. Note that the kernel density maps are based on one year's mobile phone data, while maps of TripAdvisor are based on Points of Interests. Thus, this section can be seen as a validation of our deriving activity anchor point method. Moreover, results of kernel density also imply which areas are popular in 2017-2018.

Significant Trip Chain Types

In order to discover the most popular ways for tourists in South Korea to organize their daily travels, we construct significant trip chains and visualize them. To keep it consistent, for "hybrid" trip chains and "intra-city" trip chains, we regard the top 13 trip chains among each of them as significant chain types in this research, since they all account for more than 1% of the total trip chains in the respective analysis.

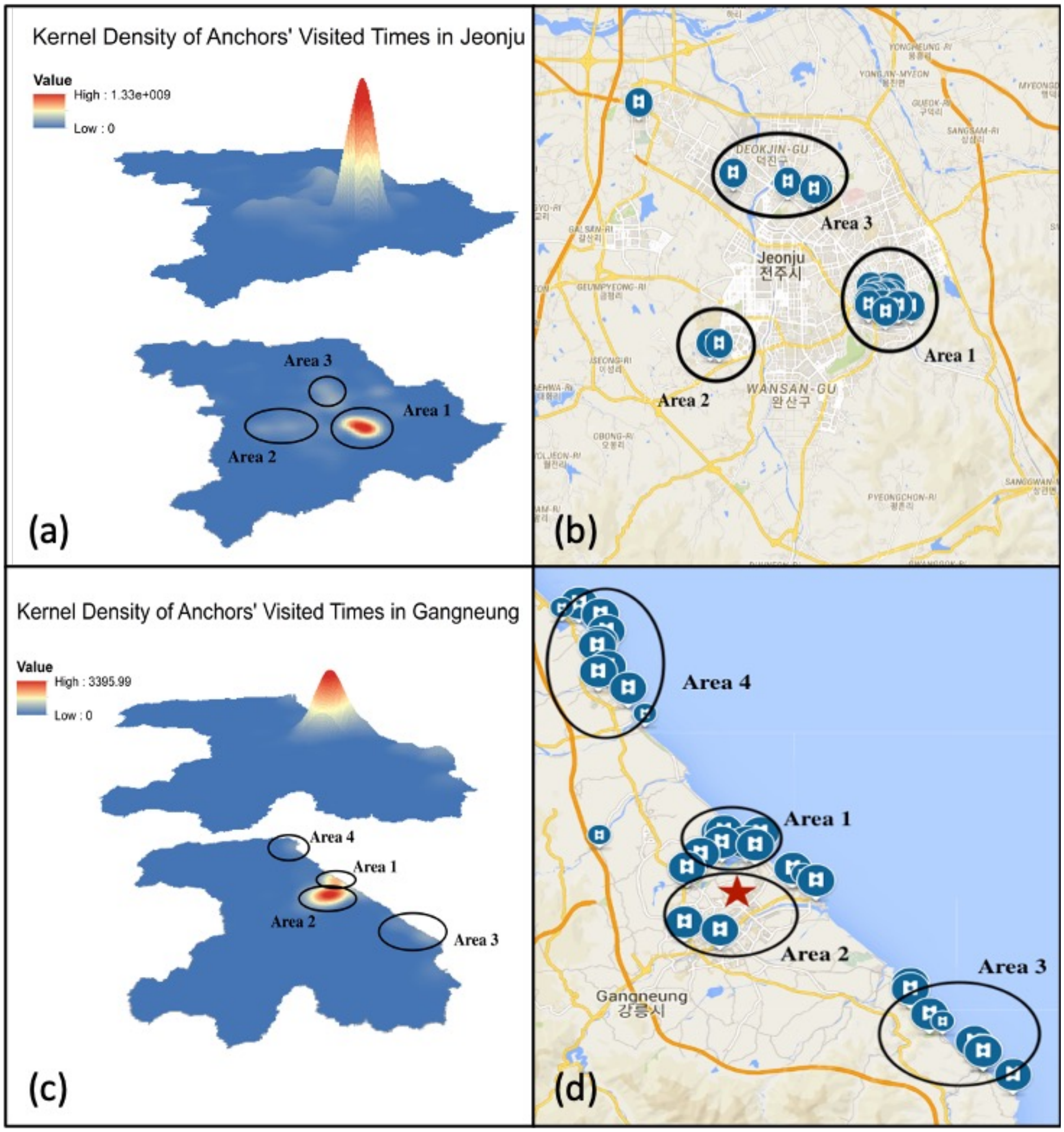

Figure 4.6: Popular areas derived from datasets (a) Jeonju (c) Gangneung, and hot spots marked by TripAdvisor (b) spots in Jeonju (d) Gangneung.

Significant "Hybrid" Trip Chains. We argue that if we extend trip chains to "hybrid" perspective, more patterns about days leaving or coming to the given city can be observed.

Figure 7 shows the top 13 "hybrid" chains, which include both individual trip chains which are all in the city and individual trip chains which are sometimes out of the city. This would reveal the behavioral diversity of travelers on an average day. To summarize, up to 76.41% and 75.99% of the measured "hybrid" trip chain types can be described with only 13 different daily trip chains in Jeonju and Gangneung respectively. In general, the trip chains can be grouped into four main categories based on the start nodes and end nodes:

- (C1) staying in the city: the start point and end point of the trip chain are both in round shape, which means the user were in the city at the beginning and the end. Or the motif just has one single point, which means the user spent all day in one anchor point;

- (C2) passing by the city: the start point and end point of the trip chain are both in star shape, which means the user were out of the city at the beginning and the end;

- (C3) coming to the city: the start point of the trip chain is in star shape while the end point of the trip chain is in round shape, which means the user were in the city at the beginning and out of the city at the end;

- (C4) leaving the city: the start point of the trip chain is in round shape while the end point of the trip chain is in star shape, which means the user were out of the city at the beginning and in the city at the end.

Figure 4.7 shows the significant "hybrid" trip chain types. There are 1866 different chain types in Jeonju and 1655 in Gangneung. The different colors of bar indicate

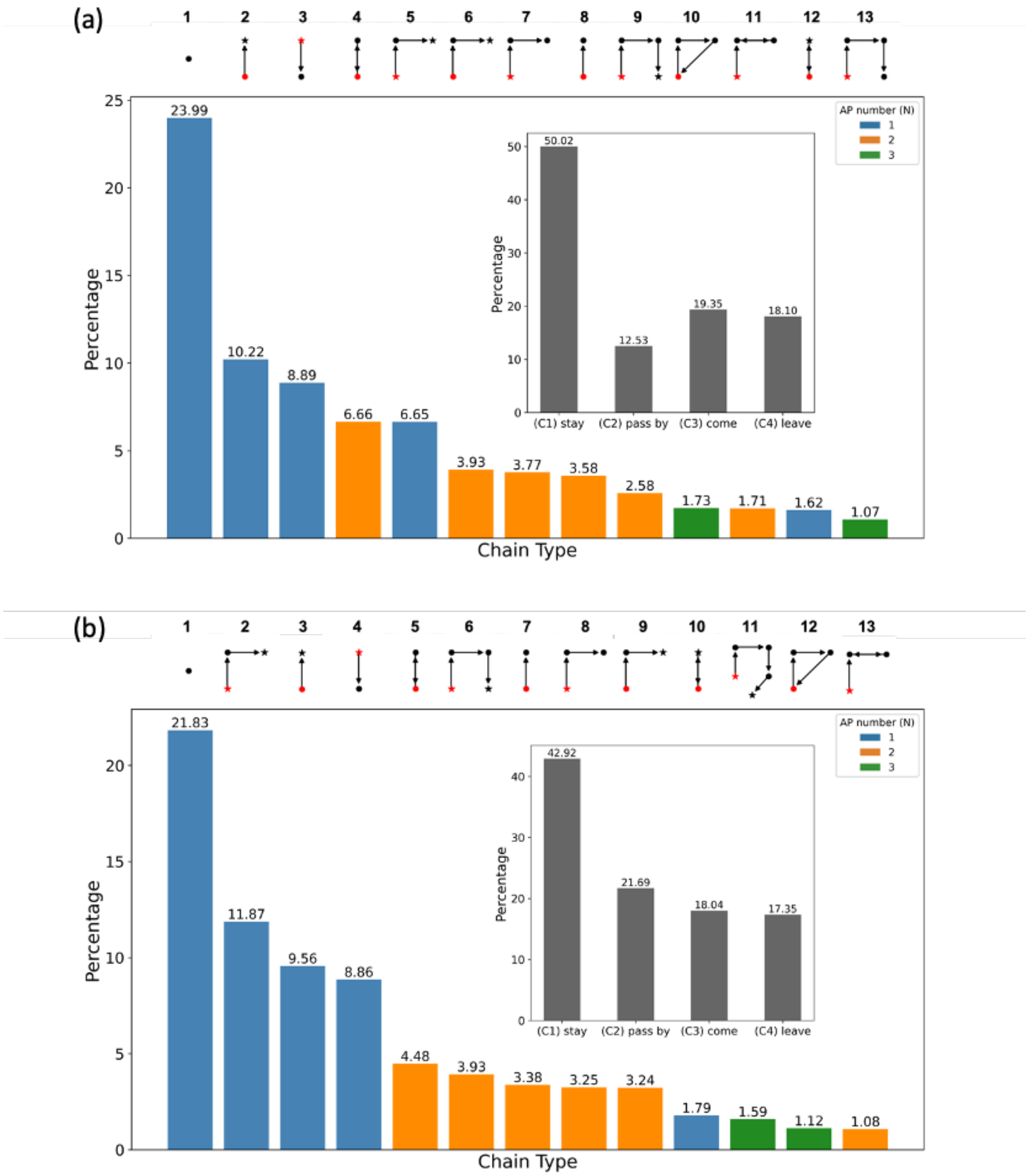

Figure 4.7: Significant “hybrid” trip chain types in (a) Jeonju, (b) Gangneung.

the number of “within city” AP in a chain. The topological structures and their ID (1-13) are shown at the top. In addition, we divide tourists into four categories: (C1) staying in the city, (C2) passing by the city, (C3) coming to the city, (C4) leaving the city. The small bar graphs with grey bars show the proportion of them. We can see from Figure 4.7 that, about half of individual daily trip chains belong to the category of staying in the two cities; the amount of individual daily trip chains of coming to the city and leaving the city is almost the same in these two study areas; the amount of individual daily trip chains of passing by Gangneung is larger than that of Jeonju, accounting for 21.69% and 12.53% respectivelyrespectively. Chain type 3 in Jeonju and chain type 4 in Gangneung indicate that for most tourists who come to the given cities, they prefer to go directly to a place (e.g., hotel) and stay there after the first day. Chain type 5 in Jeonju and chain type 2 in Gangneung indicate that most tourists who pass by the given cities only visit one spot.

Significant “Intra-City” Trip Chains. Figure 4.7 suggests that the category (C1): staying in the city accounts for about half of the samples. In this section, we focus on this category and analyze significant “intra-city” trip chains, to discover the most popular patterns for tourists at the intra-city level.

The distribution of intra-city trip chain type samples is displayed in Figure 4.8. There are 743 different chain types in Jeonju and 467 in Gangneung. The different colors of bar indicate the number of “within city” AP in a chain. The topological structures and their ID (1-13) are shown at the top. 89.55% and 90.89% of the measured intra-city chain types can be identified with 13 different daily networks in Jeonju and Gangneung respectively. The first three kinds of chain types (ID 1, 2, 3) account for a quite large part (75.47% for Jeonju and 68.13% for Gangneung) of all chain types. Due to the phone positioning mechanism of the roaming dataset, if a person keeps moving in a period, then no record will be documented during

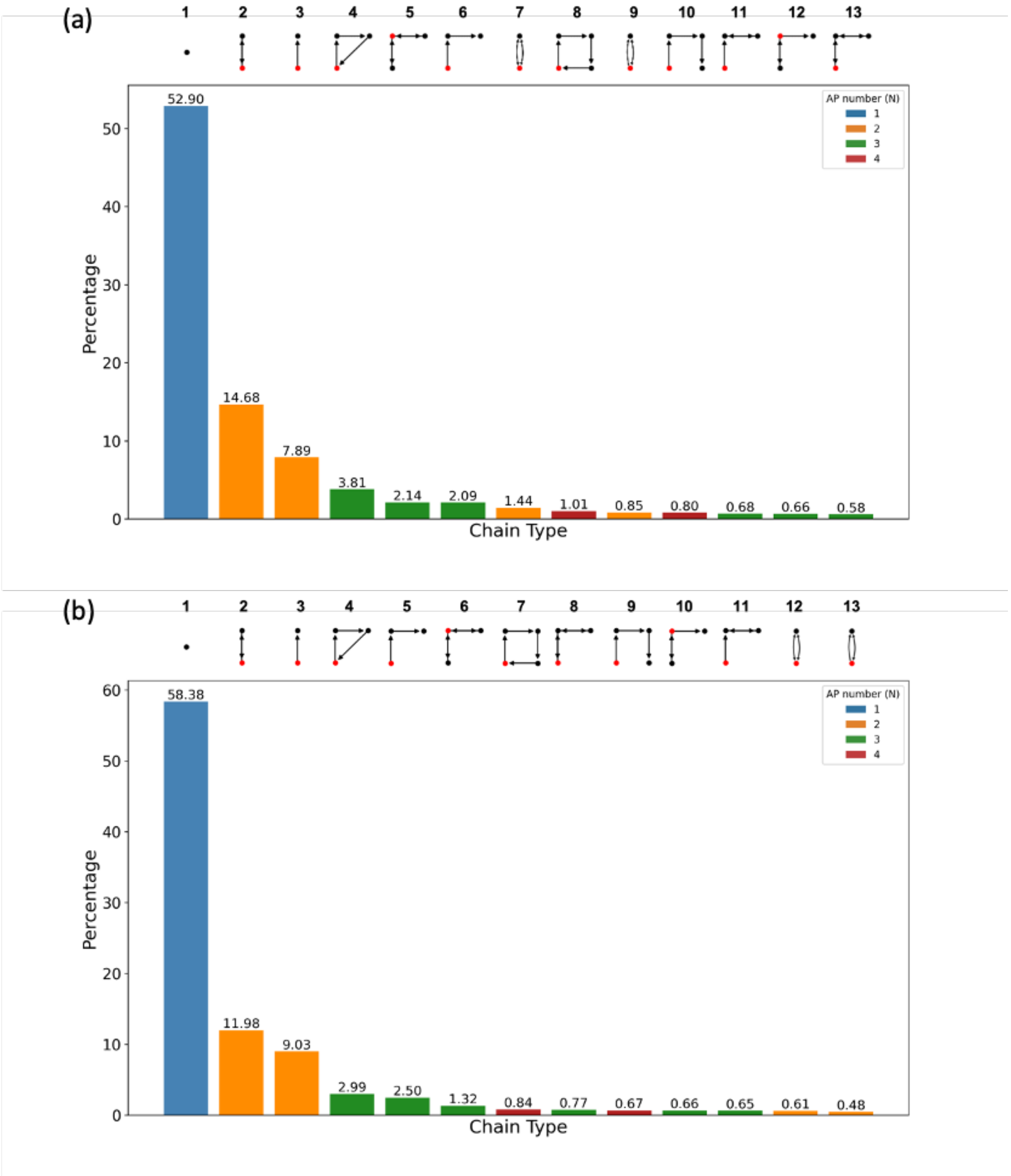

Figure 4.8: Significant "intra-city" trip chain types in (a) Jeonju, (b) Gangneung.

this period. Thus, the first three kinds of chain types correspond to several travel behaviors such as staying at a hotel all day, hanging out near the hotel, and visiting very limited places within a day. It implies that tourists tend to conduct quite simple daily tours in intra-city tourism. In addition, the similarity of results in the two cities also demonstrates the intrinsic properties of tourist mobility. These findings may provide some insights for tourism planning. For example, most tourists in these two cities may prefer integrated attractions rather than decentralized and monotonous ones.

Day-to-Day Transition of Tourist Trip Chains

Figure 4.9 and 4.10 show the transition matrix for the two cities. Every row represents a kind of trip chain that individuals conduct on the first day, every column represents a kind of trip chain that individuals conduct on the second day. For trip chains that are not significant, we group them into "others". We can see that the transition daily mobility patterns of consecutive sequence-based chains in the two cities are quite similar. The values, which equal the probabilities of corresponding two kinds of trip chains in two consecutive days, show how more or less likely an original travel chain transfers to another under the condition that the individual has mobility on two continuous days. The darker the colors, the higher the values. By locating grids with dark color, it can be found that almost all kinds of original chains have a rather high probability to transfer to either the first two transferred chains, or "others", in both Jeonju and Gangneung.

The emerging patterns of transitions could be interpreted in two aspects. First, Jeonju and Gangneung are cities with not many tourist attractions since the median values of observation days of tourists are 2 for both cities. For some tourists who just pass by these cities, they prefer to stay in a hotel to get fully rested, or go to one spot for short sightseeing or tasting local food. For other tourists who want to

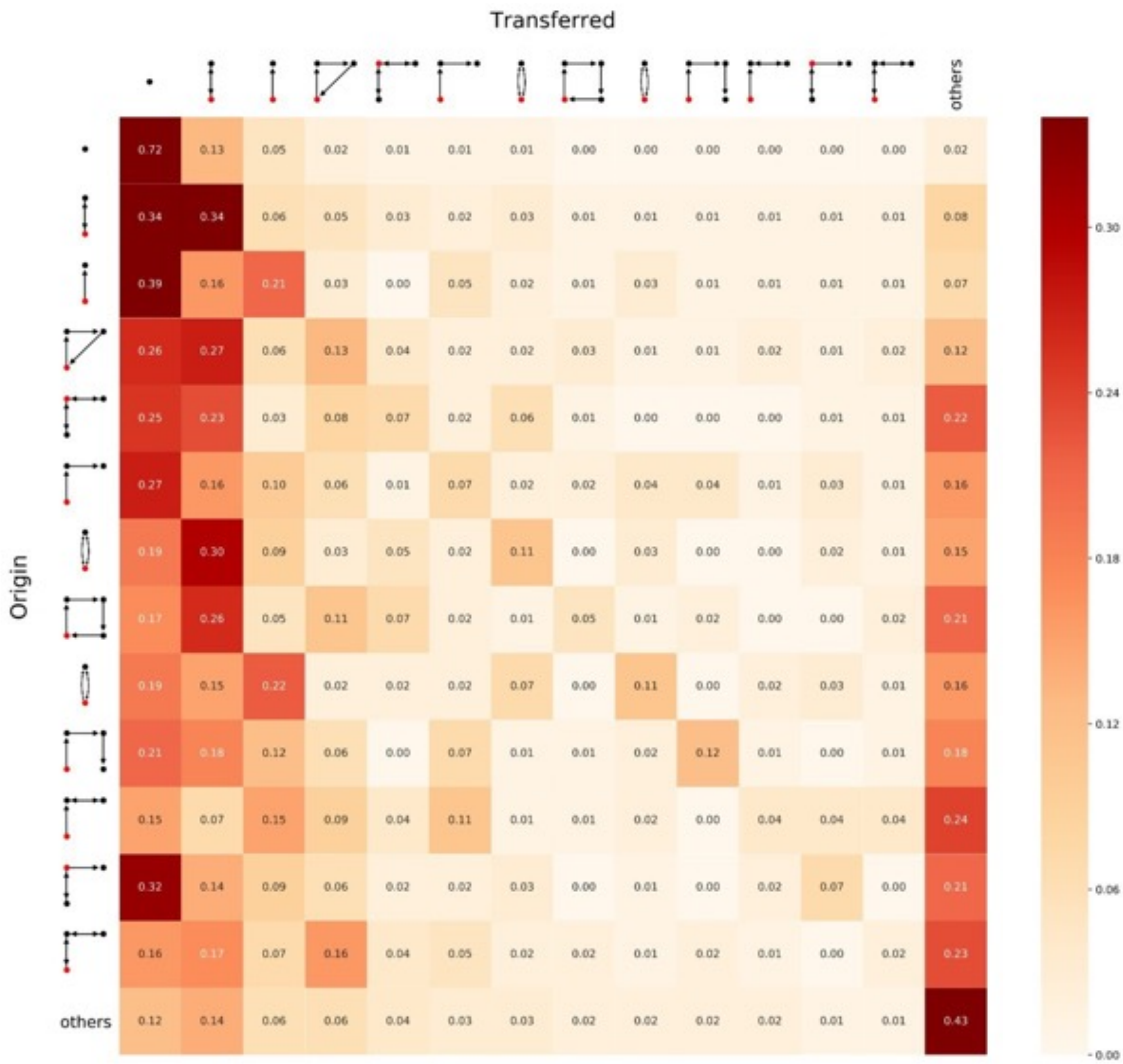

Figure 4.9: Transition matrix for Jeonju dataset.

explore these cities, they may also like to take one day to get fully rested the next day after they have a whole-day sightseeing. Second, the high probability of transferring to “other” may be related to the long tail principle. The type “others” consists of 730 chain types in Jeonju and 454 chain types in Gangneung; when all the situations are accumulated, the probability of significant original chain types transferring to “others” will become relatively high.

When we aggregate the chains by AP number N, we can clearly see from Table

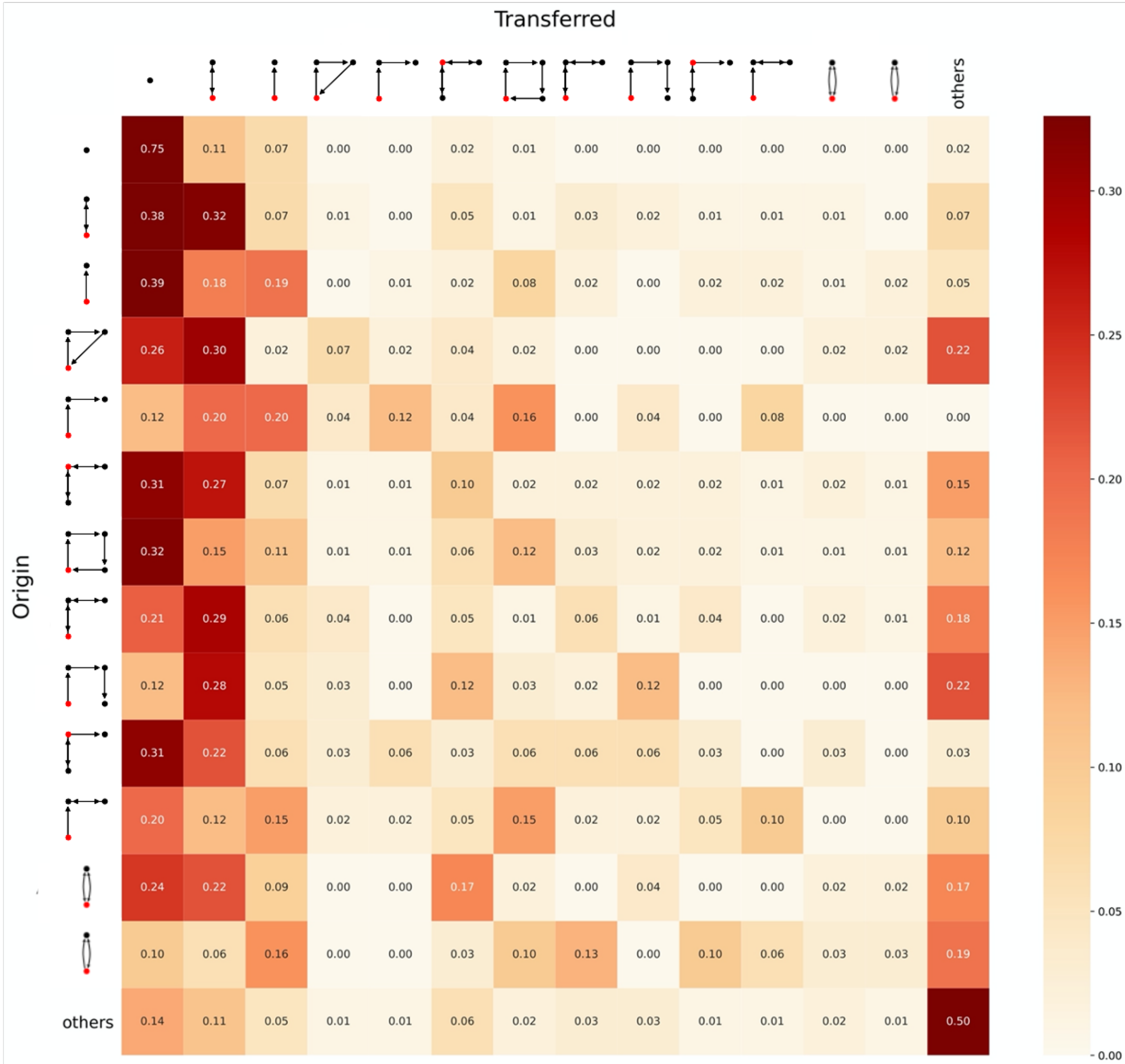

Figure 4.10: Transition matrix for Gangneung dataset.

Table 4.2: The probability of transition for chains with different AP numbers in Jeonju.

Transferred \ Origin	N = 1	N = 2	N = 3	N = 4
N = 1	0.72	0.19	0.06	0.01
N = 2	0.33	0.44	0.12	0.02
N = 3	0.25	0.32	0.22	0.04
N = 4	0.19	0.33	0.20	0.10

Table 4.3: The probability of transition for chains with different AP numbers in Gangneung.

Transferred \ Origin	N = 1	N = 2	N = 3	N = 4
N = 1	0.75	0.18	0.04	0.01
N = 2	0.37	0.41	0.14	0.02
N = 3	0.27	0.34	0.23	0.02
N = 4	0.18	0.27	0.31	0.05

4.2 and Table 4.3 that for the top 13 significant chains in each city, there is a high probability for them transferring to chains with one or two APs. The results further demonstrate our above conclusions.

The Principle of Least Effort in Trip Chaining Behavior

As mentioned in section 4.1.3, we hypothesize that travel efficiency is the substantial factor in tourist travel behavior. We checked the proxy of travel efficiency, chain degree K, in order to validate our assumption.

We group trip chains by different node number. Since the sample size is very small when node $\geqslant 7$ in both Jeonju (117 individual daily trip chains) and Gangneung (73 individual daily trip chains), we treat samples with node $\geqslant 7$ as one group in each of the two cities. Figure 4.11 (a) and (b) generally show a negative correlation between the maximum value of degree and AP number N, whereas a slight positive correlation between the minimum value of degree and N is demonstrated. The distribution of degree tends to be more concentrated to the median when the N increases. Moreover, for chain groups with N = 2 and N = 3, the medians of degree are 1 in two cities; for chain groups with more N, the medians of degree are similar and just over 1. Specially, we can see that for groups with N = 2 and N = 3 in two cities, the frequency distributions of degree are similar but rather uneven, which can be attributed to the limitations of N. When N is small, there are not many possible topological structures of corresponding trip chains.

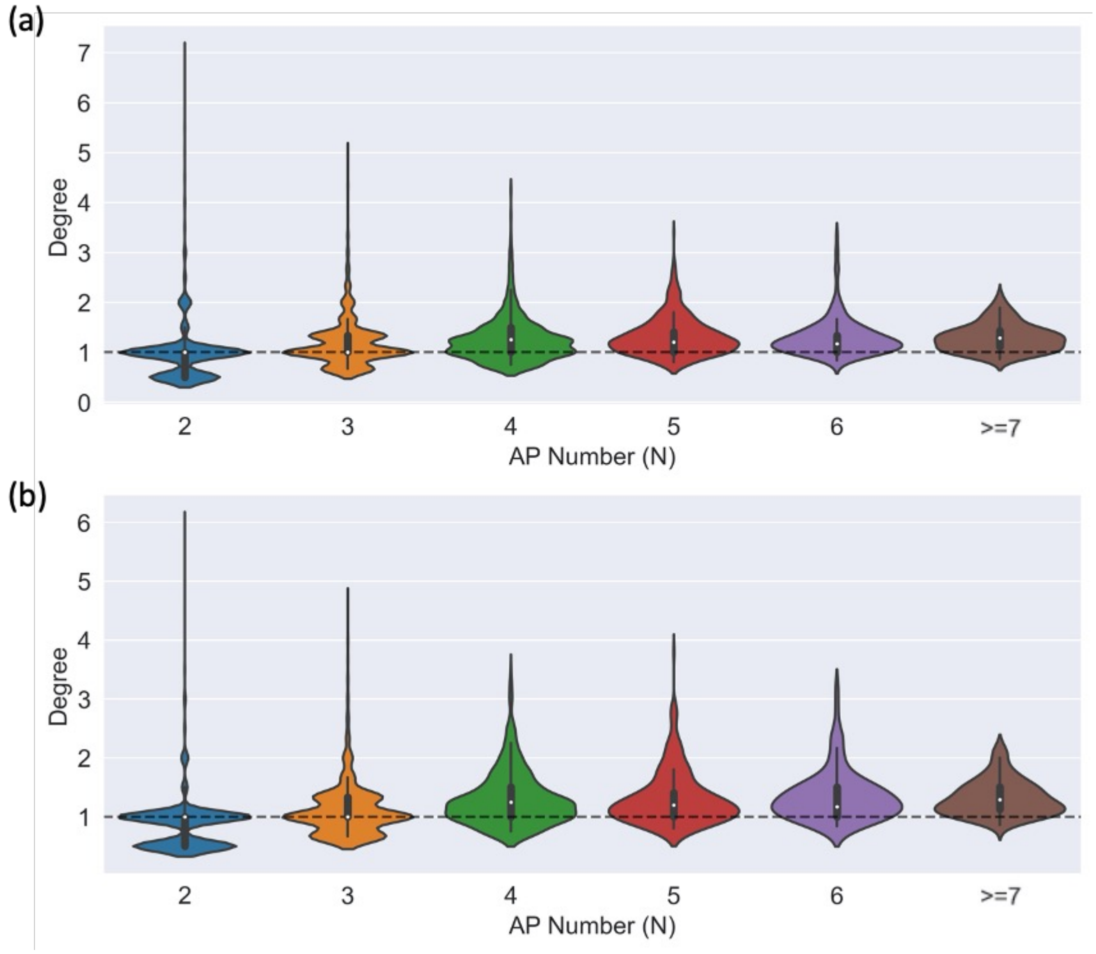

Figure 4.11: Violin plots for degree of trip chains by node number in (a) Jeonju, (b) Gangneung. The horizontal dash lines correspond to degree value 1.

The results above can be interpreted from two perspectives. First, for groups with fewer APs, the heterogeneity is more significant due to the more discrete distribution of chain degree. Thus, groups with fewer APs may consist of a larger variety of tourists. On the contrary, groups with more APs appear to be more homogeneous due to the lower extreme high degree value and higher extreme low degree value. This indicates even samples with extreme high degree value are more likely to conduct daily travel with lower degree trip chains. Second, the median values of degree do not significantly increase as the N increases, which implies for groups with more

APs, most of them have almost the same travel efficiency as groups with fewer APs. Combining these two perspectives, we discover the principle of least effort and its impacts on tourist travel behavior. The effect is more obvious when AP number N becomes larger. This illustrates even though visitors plan their itinerary with special proclivities, such as unpopular or distant spots, they still tend to choose the most convenient trip chains to achieve the highest travel efficiency.

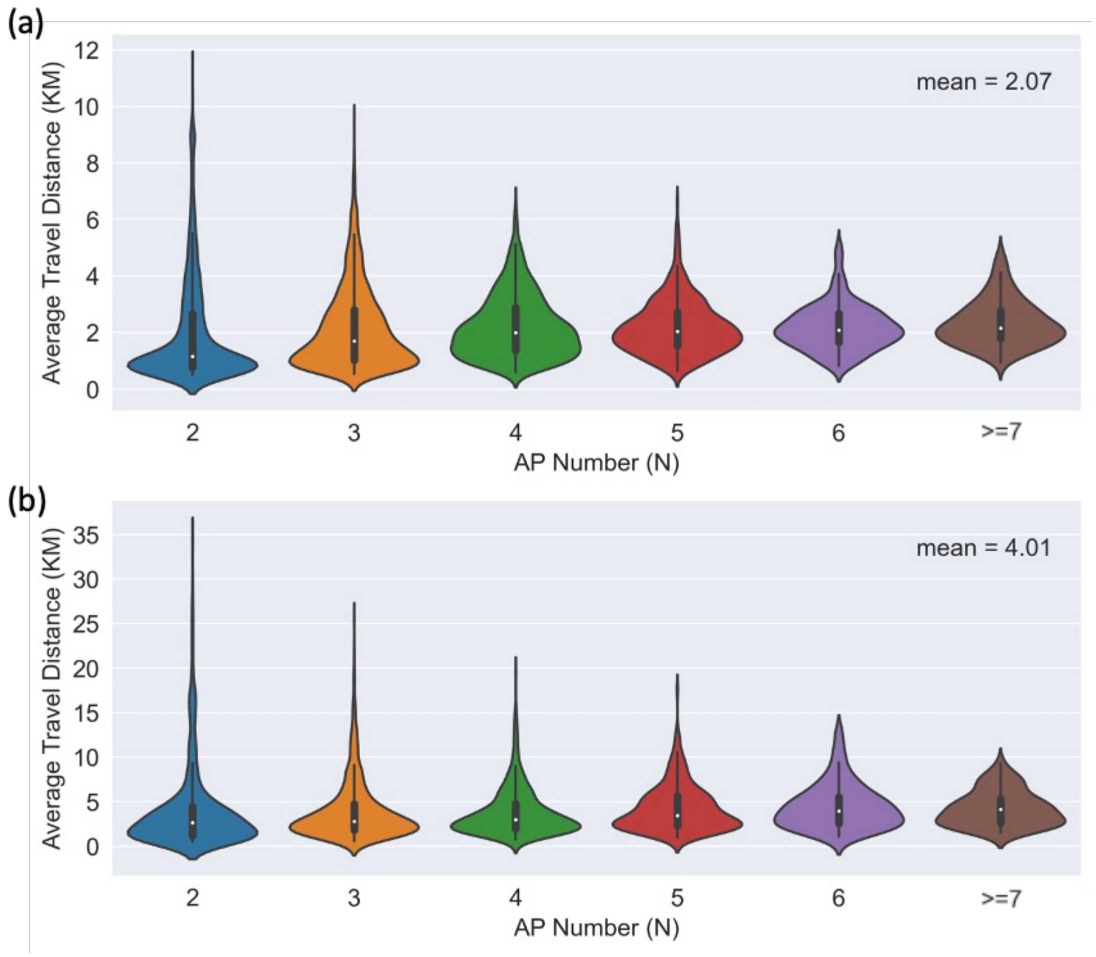

Figure 4.12: Violin plots for average travel distance of trip chains in (a) Jeonju, (b) Gangneung.

It is worth mentioning that the area of Jeonju and Gangneung is 206.22 km^2 and 1040 km^2 respectively. Thus, it makes sense that the mean value of average travel distances of Gangneung (4.01KM) are longer than that of Jeonju (2.07KM). However, such a slight difference implies even though tourist's average distances are relevant to city's area, the activity space of tourists is rather limited. As can be seen from Figure 4.12, for each city, the maximum value of average travel distance drops as the N increases while the minimum value of average travel distance increases as the N increases. When N gets higher, the distribution of average distance for each group becomes more concentrated. This result also demonstrates the heterogeneity for groups with fewer APs and homogeneity for groups with more APs. Another finding from Figure 4.12 is that the median values of average travel distance have very slight growth when the N becomes higher for two cities. This indicates although some tourists prefer to have trajectories with multiple stops, they will also make their trips as short as possible, which demonstrates the least effort principle in tourist travel behavior from the perspective of travel distance.

4.1.5 Conclusions and Implications

Tourism studies suggest that the assumption that travel decisions involving a single destination is misleading. Travelers tend to visit multiple destinations/attractions to meeting their variety-seeking motivations and maximizing benefits from their trips. Hence, the identification of travel patterns involving a range of sites visited has been a critical issue in tourism research [111]. Along with the transportation literature, this work proposes tourist daily trip chains to discover underlying travel movement patterns. This model has important theoretical implications for the tourism knowledge base. In fact, there have been several tourism scholars who have mentioned the notion of trip chains, but the attempt to quantitatively identify the structures of trip chains is paucity (e.g., [61]; [104]). This challenge is somehow attributable to restrictions on accessing proper data enabling researchers to detect individuals' movement in a comprehensive and detailed manner. Most previous studies relied primarily on survey data. This approach, however, contains several limitations (e.g., substantial cost and effort, the potential for response errors; cf. [100]).

Taking advantage of mobile sensor data providing fine-grained spatiotemporal resolution of travel behaviors, this study discovered 13 key trip chains that account for approximately 76% of "hybrid" and 90% of "intra-city" chain types. The benefits of mobile technology facilitating the collection of digital footprints—within as well as outside a focal destination—makes it possible to explore two types of trip chains including "hybrid" (containing inter-city and intra-city components) and "intra-city" (including only intra-city patterns). The results revealed different formations of trip chains including semantic places and directional movements between two types of trip chains, which demonstrate the complex daily travel network from tourism big data. Specifically, travelers who show "hybrid" trip chains are likely to prefer to go directly to the hotel after arriving at the focal city and take a rest the first day, and most tourists who pass by the given cities only visit one spot. Based on the result of over 90% measured "intra-city" trip chains, it can be concluded that tourists tend to prefer integrated attractions rather than decentralized and monotonous ones. These findings fill the theoretical gap in tourism literature on multi-destination trips by discovering significant and underlying patterns based upon a full travel trajectory.

Furthermore, this research applies the principle of least efforts (PLE) proposed in evolutionary biology and information systems into travel mobility [11]. Indeed, the broad theory of PLE serves to explain empirical tourist mobility patterns and regularities exhibited by international travelers. More specifically, this research suggests two indicators—trip chain degree and average travel distance—that quantify individual travel efficiency and reveal the presence of PLE. This framework helps to reveal that travelers make their trips as short as possible and that this affects their structures of trip chains. Tourism researchers have discussed gravity theory (e.g., [86]) and distance decay ([63]) as means of understanding travel movement behaviors.

Importantly, this work suggests an additional theory to interpret human behaviors (achieving tasks in PLE): travel distance (efforts spent in PLE). It also contributes to methodological approaches to examining these phenomena. Introducing such an analytical framework offers a path toward greater understanding of tourist mobility patterns. It also provides valuable input for many applications (e.g., personalized location-based services for tourism; smart city and smart tourism; and sustainable city planning). In addition, as demonstrated by our analysis results, a constantly growing number of mobile phone data sources contribute a great deal to geographic data mining and knowledge discovery in the age of instant access.

This work also suggests important managerial implications. The findings discovering underlying trip chains should be beneficial for travel organizers in developing new products. Based on flow-based destination planning, the structure of trip chains considering directions and sequences of travel movement can become fundamental knowledge in transportation and crowd management as well as the development of travel packages and routes. Recently, DMOs are likely to cooperate with big data firms (e.g., telecommunication companies). This collaboration and cooperation can generate innovative opportunity to access real-time information of travel behaviors, and to collect "big data." Approach to analyzing the trip chain this study suggests should guide for DMOs to not only how to analyze mobile big data, but to better understand travel spatial behaviors, which should be essential to accomplish smart tourism destination.

However, there still exist a few limitations we wish to point out. First, while these documented locations of stop points have higher accuracy than that of years ago, they still contain errors that might mislead the understanding of underlying mobility patterns. Second, our work just examined two cities as study areas. In the future, more cities—especially different types of cities —can be added into our research to achieve generalized results. Another promising direction for future work is that we can infer spatiotemporal information regarding tourist itineraries by mapping the trip chains to the spatial context. In addition, for metropolises with tourists of high average observation days, it is suggested to dig deep into the transition patterns in continuous days.

One important point to note is that while the focus of this research is on tourists, the analytical methods can also be generalized to urban residents. Due to the lack of suitable datasets, I have not yet conducted research on urban residents. However, principles such as the minimum effort observed in tourists' behaviors and their preferences are expected to provide valuable insights for mobility research among urban residents in the future.

4.2 Location Recommendation

4.2.1 Introduction and Motivation

Location recommendation is a product of necessity in the era of big data. On the one hand, the prevalence of location-based social networks (LBSNs), like Foursquare and Gowalla, has led to a tremendous amount of user check-in data. On the other hand, while a typical city has thousands of POIs, most users visit very limited POIs in and out of his/her hometown [103, 73]. By exploiting user check-in data, location recommendation can provide people with their most interesting locations out of numerous POIs. Location recommendation can not only satisfy a user's travel needs but also has great commercial value for applications such as precision advertising.

A large number of efforts have been devoted to improving the quality of location recommendation. Some studies adopt the most popular solution used by recommender systems, matrix factorization [148, 67]. However, they do not consider inherent spatial-temporal sequential behaviors and hence achieve suboptimal performance. Traditional statistics-based Markov chain models [7, 25] have been widely

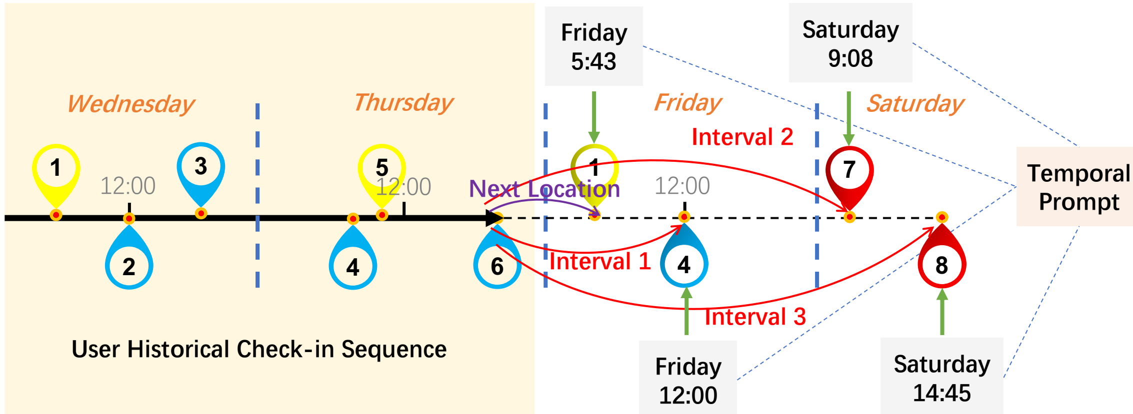

Figure 4.13: An illustration of how TPG performs the next location recommendation and interval predictions.

adopted to solve this sequential recommendation problem, but they have limitations, such as only considering the influence from the last check-in activity. With the increase of data volume and the development of deep learning technology, recurrent neural network (RNN) based methods [152, 21] are employed to consider long historical information with much stronger representation ability than Markov chain-based models. To tackle the sparsity issue in matrix factorization-based models, some graph-based methods [127, 122] are proposed and can achieve great performances. Recently, the attention mechanism has been proposed and can achieve impressive performance to model long-range temporal and spatial dependencies in human trajectories for location recommendation [66, 129]. As for inputs of location recommender systems, except simple check-in sequences, scholars also try many information sources, such as social relationships and geography information [129].

Nevertheless, there are two main challenges not yet addressed among these methods. The first one is that most context-aware state-of-the-art methods implicitly incorporate the temporal information associated with check-ins. They fused time interval information or timestamps of check-ins together with check-in history. This results in inflexible and inaccurate POI prediction outputs when the model needs to

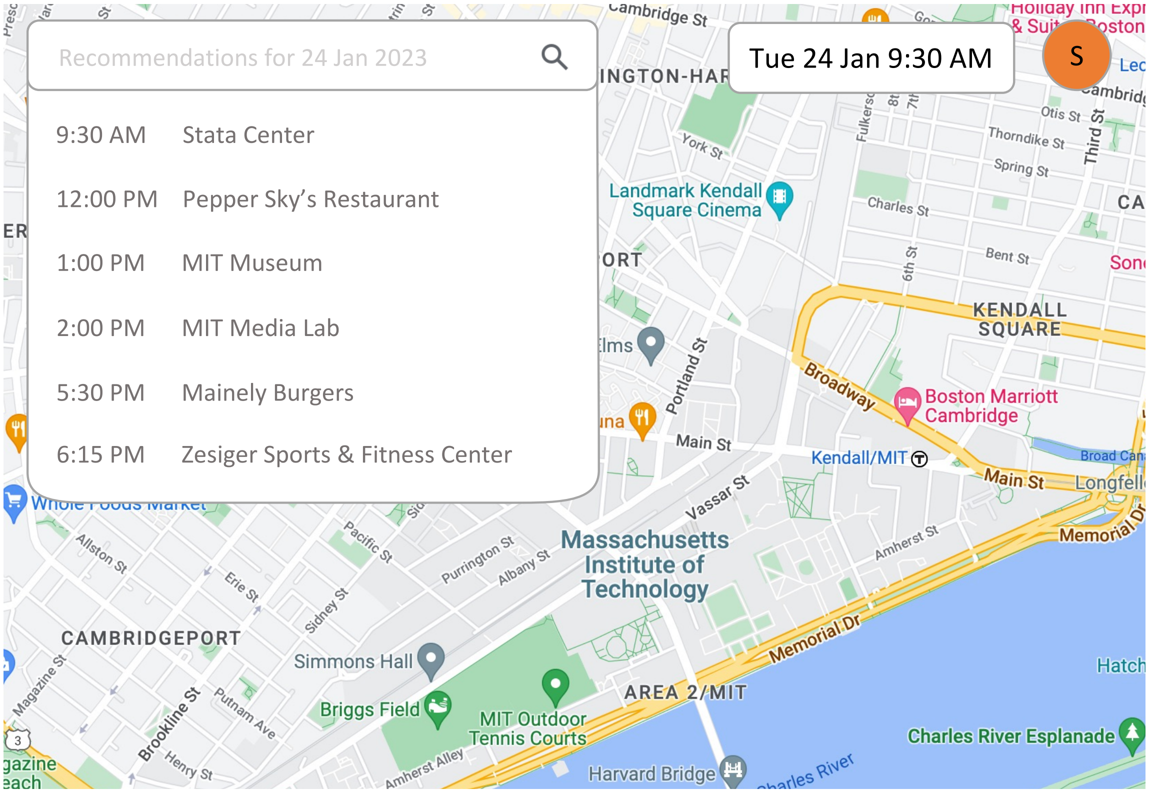

Figure 4.14: One possible use case of TPG is in map services.

predict not only the next location but also further locations. For example, STAN [74] builds spatial-temporal matrices for all check-ins within the trajectory slice. CARA [77] leverages both sequences of check-ins and contextual information associated with the sequences. If we want to predict the 101-st check-in and the 102-nd check-in based on the first 100 historical check-ins, the inputs (*i.e.*the first 100 historical check-ins and their associated contextual information) for both the 101-st and the 102-nd check-in prediction are the same. In other words, these models do not consider different timestamps of the locations to be predicted. However, the fact is that people will tend to visit different locations at different times. For example, a user usually goes to his/her work in the morning and returns home in the evening. Thus, it is important for location recommender systems to have the ability to generate specific prediction(s) based on certain timestamp(s). It is worth noting that methods such as STAN and CARA of course can predict the 102-nd check-in by involving the 101-st predicted check-in into the first 100 historical check-ins. However, in this way this type of models all need to re-train themselves, which will require a large computational overhead.

The second main challenge is that the geographic information is very important in location recommendation, since it is a spatial-temporal problem. State-of-the-art methods do not make effective use of geographic information. MobTCast [129] directly feeds the latitude and longitude of POIs into an encoder. However, since check-in data is extremely sparse [1], processing geographic information in this way makes it difficult to capture the physical proximity and dependency between locations. GeoSAN [66] further proposes to use hierarchical grids to model the spatial clustering phenomenon in human mobility. However, it suffers from the hard boundary problem, meaning that the POIs near the grid boundary are manually separated.

To tackle these issues, we propose a Temporal Prompt-based and Geography-aware (TPG) framework for location recommendation. We make TPG a Transformer-based framework, because Transformer [116] is originally designed for sequential data with uncertain length, and can differentiate the informativeness of different check-ins and aggregate all check-ins in the trajectory simultaneously for prediction. Firstly, a geography-aware encoder is designed to capture the geographic correlations among POIs. To avoid the aforementioned hard boundary problem, we propose a shifted window mechanism in the geography-aware encoder. It can bridge the proximity gaps between two adjacent grids and connect adjacent grids by aggregation. Subsequently, the information of user, POI, time, and geography from historical check-in sequences are incorporated by a history encoder, which is designed to learn a comprehensive representation of user travel preference. Afterwards, by using a timestamp as a prompt and regarding it as the query, a temporal prompt-based decoder is utilized to predict the future location(s). In this way, TPG explicitly incorporates the timestamp of the location to be predicted, separating historical check-in sequences and timestamp information. Thus, TPG is very flexible with respect to multiple scenarios. It can not only perform next location recommendations, but also handle interval predictions by using temporal prompts based on those future locations. It should be noted that there are two equivalent scenarios for interval prediction: (a) predicting some further check-ins (*e.g.*the 102-nd, the 103-rd) based on a fixed length of history trajectory slice (*e.g.*the first 100 check-ins), (b) predicting a future location (*e.g.*the 100-th) while the most recent check-in behavioral data being masked (*e.g.*using the first 95 or 96 check-ins). Figure 4.13 is a simple example demonstrating how TPG performs the next location recommendation and interval prediction. The next location recommendation is denoted by the purple line, and interval predictions by using temporal prompts is denoted by red lines. Different colored markers denote different categories of POIs. As shown in Figure 4.13, given the user historical check-in sequence is POI 1-6 from Wednesday to Thursday, the model can know the next four locations the user will visit are POI 1 at 5:43 Friday, POI 4 at 12:00 Friday, POI 7 at 9:08 Saturday, and POI 8 at 14:45 Saturday. Predicting POI 1 at 5:43 Friday is the task of next location recommendation. By making use of temporal prompts, TPG can also predict the location that a user wants to go at a certain time (*i.e.*interval prediction). For example, the model can predict POI 4 at 12:00 Friday (interval 1), POI 7 at 9:08 Saturday (interval 2), and POI 8 at 14:45 Saturday (interval 3), only based on historical check-in sequence POI 1-6. Figure 4.14 is one possible use case for a real-world application. The left column is a recommendation list generated by TPG for multiple timestamps. The upper right orange circle displays the user information. The bar to the left of the orange circle is the current timestamp.

To summarize, the contributions of this work can be listed as follows:

- We argue that the explicitly modeling timestamp of the location to be predicted is essential in real-world applications. A novel and effective Transformer-based framework named TPG is proposed. Temporal information is regarded as a prompt for our recommendation system.

- To effectively utilize geographic information, we propose a geography-aware encoder with a shifted window mechanism devised to avoid the hard boundary problem when treating longitude and latitude of POIs with grids.

- Experimental results on five real-world datasets, namely, Gowalla, Brightkite, Foursquare-NYC, Foursquare-TKY, and Foursquare-SIN, show that our model outperforms the state-of-the-art counterparts under different settings. We also demonstrate that TPG's interval prediction perform much better than baselines.

4.2.2 Timestamps as Prompts for Geography-Aware Location Recommendation

In this section, more details about the proposed TPG framework are elaborated. We first give the problem statement and provide an overview of the framework. Then, we elaborate on the three main modules of TPG, *i.e.*geography-aware encoder, history encoder, and temporal prompt-based decoder.

Overview

Each check-in $c_i^u = (u, t_i, p_i)$ is a user, POI, time tuple, which denotes a behavior that a user u visits POI p_i at time t_i. Each POI p_i has its own geographic coordinates (x_i, y_i). Each user u has a sequence of historical check-ins $C_{1\rightarrow n}^u = \{c_i^u\}_{i=1}^n$. Given the historical check-in sequences of users, the goal of next location recommendation is to predict the next POI $\rho_{t_{n+1}}$ that a certain user u will visit at a certain time t_{n+1}.

The overall architecture of our TPG framework is described in Figure 4.15. Detailed explanations of notations in Figure 4.15 are described in Section 5.3. Based on

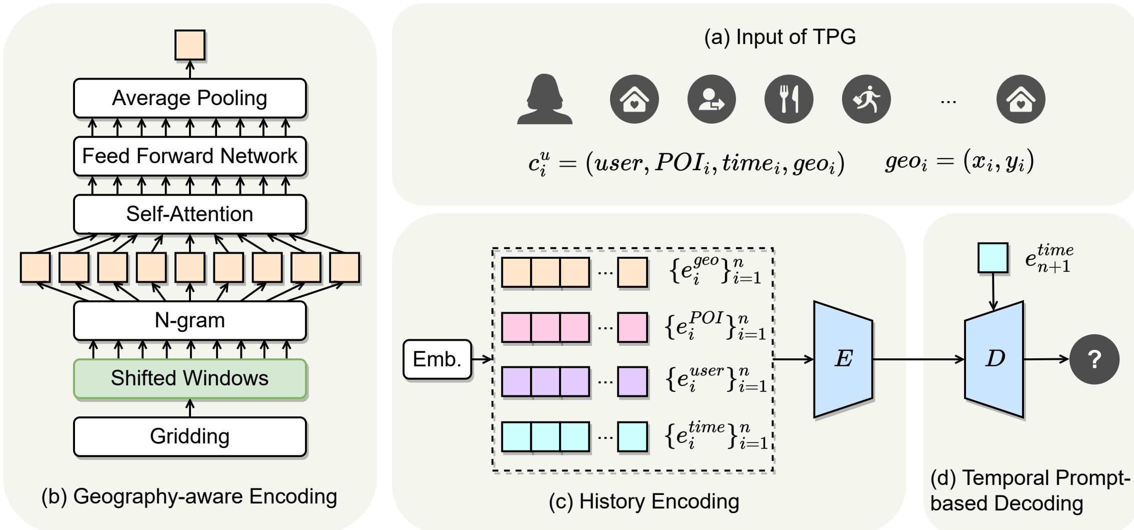

Figure 4.15: The overall architecture of the proposed TPG.

the Transformer's encoder-decoder structure, TPG can be divided into three parts, *i.e.*geography-aware encoder, history encoder, and temporal prompt-based decoder. For each check-in, the geographic coordinate of POI can be fed into the geography-aware encoder to get geographical representation e_i^{geo}. The historical check-in sequences including POI, user, and time information are then fed into the multi-modal embedding module to generate hidden representations $\{e_i^{POI}\}_{i=1}^n$, $\{e_i^{user}\}_{i=1}^n$, and $\{e_i^{time}\}_{i=1}^n$. Together with $\{e_i^{geo}\}_{i=1}^n$ from the geography-aware encoder, these representations are processed by a history encoder to generate user travel preference representation. Using temporal information of t_{n+1} as prompt, the temporal prompt query and user travel preference memory are then forwarded to the decoder, which is capable of generating more accurate predictions for the next locations.

Geography-aware Encoder

Sparsity issue is a key challenge in recommendation problem. In particular, as the check-in data gives implicit feedback of visiting behavior [67], the data sparsity problem is even worse for POIs compared other item recommendation such as movies or goods, of which users usually only express their opinion with ratings. Thus, when it comes to encoding geographic information of POIs, directly feeding the coordinates of POIs into the learning model makes it difficult for the model to capture geographic correlations. GeoSAN [66] embeds the exact position of locations by mapping latitude and longitude into hierarchical grids using tile map system*as exemplified in Figure 4.16. The given example is about mapping a location into grids at level 16-18, whose quadkeys are annotated. The tile map system is a hierarchical multi-resolution pyramid model. The world map is obtained by projecting the entire world into a flat plane by Mercator† The scale of the plane starts with 512×512 pixels. It grows by a factor of 2 with the increase of levels. For better retrieval and display, the plane is further divided into grids of 256×256 pixels each. From the low level bottom to the high level top of the tile pyramid, the resolution becomes lower and lower, but the geographic range is unchanged via sub-gridding one grid into four grids of the same size. Since the partition of grids is like quadtree, each grid can be identified with a unique quadtree key (quadkey for short). Quadkeys consist of the characters from the set {"0", "1", "2", "3"}. The length of it equals the level of grid.

It indeed can alleviate the sparsity problem to some extent. However, for grids at the same level, the boundary of grids may damage the physical spatial proximity of two POIs around the boundary, which is a violation of Tobler's First Law of Geography [113]. In other words, lacking connections across adjacent grids limits modeling power of geography-aware encoder. To introduce cross-grid connections, we propose a shifted window mechanism in our geography-aware encoder. As illustrated in Figure 4.17, for each grid, we move the shifted window along the X and Y direction (and both) by a certain step, which is part of the length of the grid size. In this way, we will get nine grids for each grid, *i.e.*itself and eight augmented neighbor grids.

*https://www.maptiler.com/google-maps-coordinates-tile-bounds-projection

†https://www.britannica.com/science/Mercator-projection

Now the remaining task of geography-aware encoder is transforming quadkeys of these nine grids into a continuous latent embedding with rich information. For each quadkey, it is actually a character sequence. Each character in the sequence denotes the index of the grid partition at a certain level. If we want to make effective use of hierarchical spatial information of grids, an intuitive and straightforward approach is to conduct self-attention between these characters. However, since the cardinality of the character set is very small (*i.e.*only 4), treating a quadkey at character-level cannot achieve the goal of fully encoding the geographic correlations between POIs. Therefore, we consider dividing the character sequence by n-gram, and converting it into a sequence at n-gram-level. In this way, the vocabulary size of the sequence increases from 4 to 4n. For example, if a quadkey is "013201233", the result of using four-gram is 0132-1320-3201-2012-0123-1233. We then use a stacked self-attention network and a point-wise feed forward network for capturing dependencies among these n-grams. After that, for each grid among these nine grids, average pooling can be utilized to aggregate the sequence of n-gram representations. We then obtain geographic embedding e_i^{geo} for the given location POI_i by average pooling on aggregated representations of these nine grids. The shifted window mechanism is helpful to reduce possible bias caused by the arbitrary partition of grids via including information from augmented neighbor grids.

History Encoder

Each check-in is a tuple consisting of user, time and POI information. To tackle with such discrete and heterogeneous data, we need a multi-modal embedding module to transfer check-in data into interpretable information for TPG. Specifically, for encoding time information, timestamps are firstly discretized into $24 \times 7 = 168$ types learnable vectors. For encoding user and POI information, the type number of learnable vectors equals to the unique number of users and POIs in datasets. All

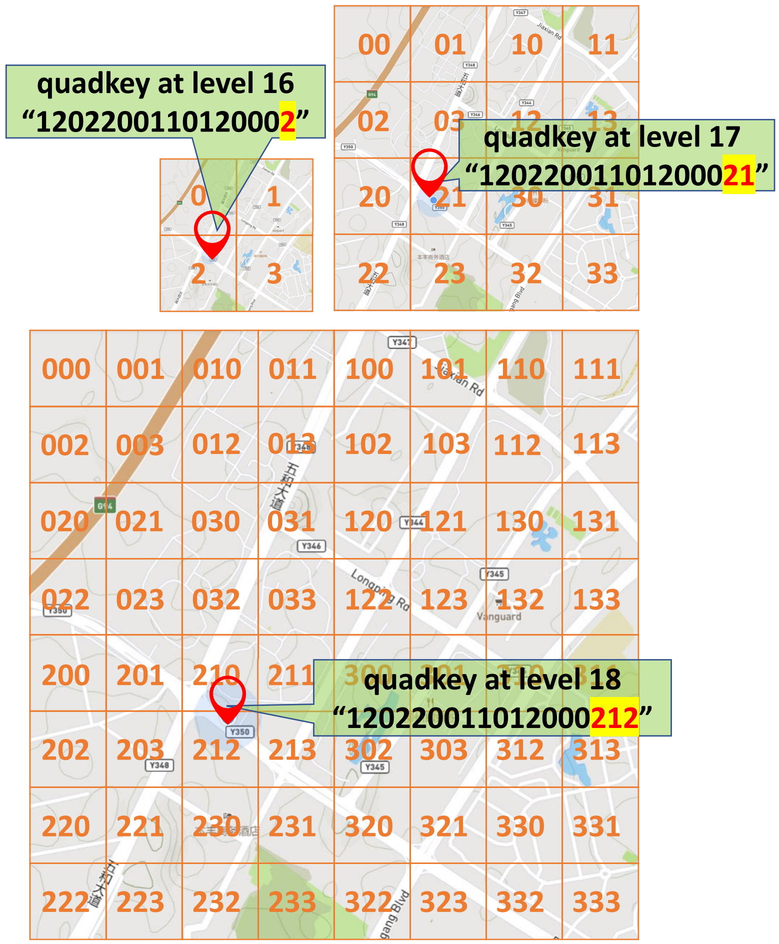

Figure 4.16: An illustration of hierarchical gridding based on the tile map system.

these vectors are then linearly projected into d-dimensional embeddings $e_i^{time} \in \mathbb{R}^d$, $e_i^{user} \in \mathbb{R}^d$, and $e_i^{POI} \in \mathbb{R}^d$. In this way, for the user u, the historical check-in sequence $\{c_i^u\}_{i=1}^n$ can be further denoted as $(\{e_i^{POI}\}_{i=1}^n, \{e_i^{user}\}_{i=1}^n, \{e_i^{time}\}_{i=1}^n, \{e_i^{geo}\}_{i=1}^n)$. Note that since check-in data requires a certain order of precedence, learnable positional embedding is also added into inputs for history encoder.

Compared with previous RNN-based methods, Transformer architecture [116] can not only avoid recurrence, allowing parallel computing to reduce training time, but also migrate performance degradation problem with regard to long-term dependencies in RNNs. To better capture long range spatial-temporal dependencies in users' historical check-in sequences, we stack Transformer encoder layers [116] for constructing the history encoder. Each Transformer encoder layer involves a multi-head self-attention module and a point-wise feed-forward network. We also keep the residual connection and layer normalization employed in Transformer encoder layers. Dividing the attention mechanism into multiple heads to form multiple sub-spaces allows the model to focus on different aspects of information. For each attention head, self-attention result for a check-in c_i can be computed as

$$ATTENTION(e_i^c) = w_z \sum_{j=1}^{N_v} \frac{\exp(w_q e_i^c \times w_k e_j^c)}{\sum_{m=1}^{N_v} \exp(w_q e_i^c \times w_k e_m^c)} w_v e_j^c + e_i^c \tag{4.4}$$

where e_i^c is the input check-in embedding for c_i, e_j^c is the embedding for contextual check-in in the sequence, and $w_{\{q,k,v,z\}}$ denotes linear transform weights for the query, key, value, and output matrices. The self-attention mechanism aggregates the global context information into each check-in features. After multi-head self-attention results are obtained by concatenating every self-attention result, the encoder is able to jointly attend to information from different representation sub-spaces at different positions. Then, the feed-forward network contains two linear transformations with a ReLU activation in between, which can be denoted as

$$FFN(e_{i'}^c) = max(0, e_{i'}^c W_1 + b_1) W_2 + b_2 \tag{4.5}$$

where $e_{i'}^c$ is the input embedding after multi-head self-attention, W_1 and W_2 denote linear transform weights, and b_1 and b_2 are linear transform offsets.

Temporal Prompt-based Decoder

The normal way for existing methods to generate the predictions of next location is based on historical check-ins and the associated contextual information. The model will not explicitly indicate the output with respect to a special temporal information. However, in real world applications, it is important to consider the exact time for next

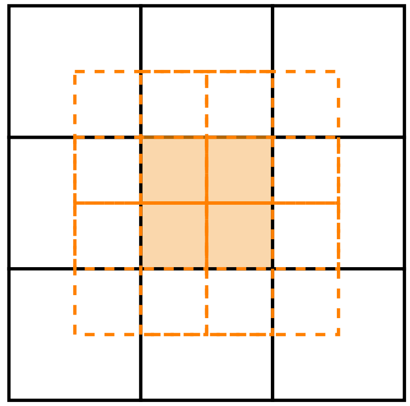

Figure 4.17: An illustration of shifted windows for the grid marked by orange (rolling step as 0.5).

location prediction. Location recommendation is usually employed in map services such as Google Map and location based services such as Foursquare. We can regard the time of user opening the app or clicking the query box in the app as the timestamp of next location. Human mobility has the periodicity [137]. People will revisit POIs of the same category around the same hour of different days. While at different hour of a day, people tend to visit different types of POIs. For example, it is the simple fact that if a user clicking the query box in the map app at noon, he/she is probably looking for a restaurant. While in the morning, there is a big chance that he/she wants to search for the route to his/her company. Therefore, if the model does not know the timestamp of next location, it cannot produce results of next locations with high confidence. Thus, an intuitive idea is that we can directly tell the model about the timestamp of next location. A simple method for it is to incorporate this timestamp into inputs. However, since a check-in is a tuple of user, time, and POI, directly adding this single timestamp into inputs is not appropriate. To this end, we propose a temporal prompt-based decoder, using timestamps as prompts and queries for the decoder. There are several advantages of utilizing temporal prompts: (1) It separates historical check-in sequences and temporal information of locations to be predicted, making the model more flexible for generating any future check-in. (2) By explicitly modeling temporal information, the prediction is greatly correlated to the given timestamp. As mentioned above, people's travel choices are strongly related to timestamps, so this design will most likely improve model performances.

In greater detail, the preference decoder takes queries (*i.e.*time representation e_{n+1}^{time}) and encoder memory (*i.e.*user travel preference representation e^C) as inputs. We construct the temporal prompt-based decoder by stacking Transformer decoder layers[116], each of which consists of a multi-head self-attention sub-layer, an encoder-decoder attention sub-layer, and a feed-forward network sub-layer. User travel preferences and timestamp information are deeply fused in each encoder-decoder attention sub-layer. Each attention head of the encoder-decoder attention sub-layer can be represented by

$$ATT(e^C, e_{n+1}^{time}) = w_z \sum_{j=1}^{N_v} \frac{\exp(w_q e^C \times w_k e_{n+1}^{time})}{\sum_{m=1}^{N_v} \exp(w_q e^C \times w_k e_{n+1,m}^{time})} w_v e_{n+1}^{time} + e^C \tag{4.6}$$

where notations are consistent with Eq. 4.4.

After adding up and feeding the results into the feed-forward networks for further projection, the output embedding decodes the fused check-in features and has the same length as the query embedding. The output here is actually the embedding of predicted next location.

As for the training scheme, we adopt the negative log likelihood with sampled Softmax as the recommendation loss for each user u. The recommendation loss can be depicted as:

$$\mathcal{L}_{rec}(\tilde{y}) = -\log \frac{\exp(\tilde{y}y^+)}{\exp(\tilde{y}y^+) + \sum_{y^- \in \mathcal{Y}^-} \exp(\tilde{y}y^-)} \tag{4.7}$$

where $\tilde{y}$, y^+, and y^- indicate the inferred location embedding, the ground truth of location which user u visits at time t_{n+1}, and the randomly sampled negative data

Table 4.4: Location recommendation dataset statistics.

	Gowalla	Brightkite	NYC	TKY	SIN
#users	31,708	5,247	1,010	6,771	367
#locations	131,329	48,181	5,135	14,590	3,104
#check-ins	2,963,373	1,699,579	140,229	871,200	136,847

which user does not visit at t_{n+1}, respectively.

4.2.3 Experiments and Evaluations

In this section, we report the extensive experiments conducted to evaluate the performance and show the superior performance of our method. The results also demonstrate the effectiveness and utility of TPG. Further experiments also validate the rationality of each component of TPG.

Experimental Settings

Datasets. We use five publicly available real-world Location-Based Social Network datasets to evaluate our method: Gowalla[‡], Brightkite[§], NYC, TKY[¶], and SIN[‖]. Gowalla and Brightkite contain worldwide data while the NYC, TKY, and SIN are extracted from Foursquare global dataset, which only focuses on a single city/region. We adopted LibCity's [121] pre-processing pipeline. Table 4.4 gives a rough sketch of the statistics of the five datasets.

Evaluation Metrics. We adopt two widely-used metrics of ranking evaluation: Recall and normalized discounted cumulative gain (NDCG), to evaluate recommendation performance. Recall@k counts the rate of true positive samples in all positive

[‡] `https://snap.stanford.edu/data/loc-gowalla.html`

[§] `http://snap.stanford.edu/data/loc-brightkite.html`

[¶] `https://drive.google.com/file/d/0BwrgZ-IdrTotZ0U0ZER2ejI3VVk/view?usp=sharing&resourcekey=0-rlHp_JcRyFAxN7v5OAGldw`

[‖] `https://www.ntu.edu.sg/home/gaocong/data/poidata.zip`

Table 4.5: Overall comparison with eight baselines for location recommendation.

	Gowalla				Brightkite				NYC				TKY				SIN			
	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10
HSTLSTM	44.6	30.9	54.4	35.1	46.8	37.3	50.4	41.3	25.7	19.7	31.1	21.4	29.1	22.1	35.7	24.3	14.9	11.2	21.0	13.4
DeepMove	49.9	35.9	59.1	39.0	50.3	39.1	58.7	43.2	29.8	21.6	36.5	23.8	33.3	25.1	39.8	27.2	16.8	11.8	24.7	13.4
LSTPM	42.2	30.4	53.2	33.0	43.8	34.9	52.4	38.5	22.8	16.1	30.7	18.6	37.2	28.2	45.4	30.9	13.1	9.2	20.9	12.0
TMCA	44.3	32.5	55.4	35.5	45.5	36.4	55.9	41.6	24.6	18.3	33.1	20.2	39.9	27.9	41.7	29.5	15.5	11.3	22.3	14.0
CARA	50.2	36.6	60.0	40.8	51.6	40.2	54.7	41.2	28.0	20.2	37.5	24.0	31.8	24.3	37.2	28.0	15.04	11.8	20.8	14.0
MobTCast	54.3	37.9	65.5	46.6	52.5	43.6	59.5	46.4	31.3	21.3	41.3	28.1	59.7	48.4	65.4	52.6	17.0	12.4	24.6	15.1
STAN	58.7	41.6	70.3	46.6	57.2	45.4	69.8	47.3	32.1	22.3	45.9	27.3	61.2	50.1	71.3	54.4	18.0	11.3	29.5	15.7
GeoSAN	56.2	41.4	69.9	45.9	55.8	42.3	67.2	46.0	30.2	20.8	44.5	25.3	58.8	48.4	69.1	51.8	17.4	11.8	30.0	15.9
TPG	**63.2**	**45.5**	**74.6**	**50.1**	**62.4**	**47.9**	**74.4**	**51.8**	**37.3**	**26.8**	**52.8**	**31.8**	**65.4**	**53.1**	**76.5**	**56.7**	**19.4**	**14.3**	**34.3**	**19.1**
Improv.	7.7	9.4	6.1	7.5	9.1	5.5	6.6	9.5	16.2	20.2	15.0	13.2	6.9	6.0	7.3	4.2	7.8	15.3	14.3	20.1

samples, which in our case means the rate of the label in the top-k probability samples, NDCG rewards method that ranks positive items in the first few positions of the top-k ranking list. We report k=5 and k=10 in our experiments.

Baselines. To show the effectiveness of our proposed methods, we compare our proposed TPG with several baselines. Here, I have selected five models based on the LSTM method and three based on the Transformer method, which are widely recognized for this task.

- HSTLSTM [59]: a LSTM based method which introduces spatio-temporal transfer factors and uses an encoder-decoder structure for prediction.
- DeepMove [21]: an attentional recurrent network which capture the complicated sequential transitions and the multi-level periodicity.
- LSTPM [107]: a long- and short-term preference modeling framework which consists of a nonlocal network for long-term preference modeling and a geo-dilated RNN for short-term preference learning.
- CARA [77]: a novel contextual attention recurrent architecture that leverages both sequences of feedback and contextual information associated with the sequences to capture the users' dynamic preferences.

- TMCA [65]: a novel temporal and multi-level context attention LSTM-based encoder-decoder framework which is able to adaptively select relevant check-in activities and contextual factors for next POI preference prediction

- GeoSAN [66]: a geography-aware sequential recommender based on the self-attention network that uses hierarchical gridding of GPS locations for spatial discretization and uses self-attention layers.

- STAN [74]: a spatial-temporal attention network that explicitly aggregates all relevant check-ins in trajectories, not only just successive ones.

- MobTCast [129]: a Transformer-based context-aware network combined with a location prediction branch as an auxiliary task. It captures temporal, semantic, social and geographical contexts.

Implementation Details. For the check-in sequence of each user, we take the last check-in record on a previously unvisited location as ground truth in evaluation, and check-in sequence before that for training. The maximum sequence length is set to 100.

Different from GeoSAN which directly uses ground truth for negative sampling in both train and evaluation setting and may cause label leakage, we consider a practical scenario where each user's next physical position is unknown, and the negative samples have to be drawn from the vicinity of the immediately preceding check-in location. To be more specific, 100 of the 2000 nearest locations from user's current GPS coordinates are chosen randomly as negative samples. Recall and NDCG can then be computed based on the ranking of these 101 locations.

We run all the experiment on NVIDIA V100 GPUs. For our TPG model, we set the dimension of location and region embeddings to 50 respectively, and time embedding to 100. The step size of the shifted windows is set to a quarter of the side

Table 4.6: Interval Prediction Performances.

	Gowalla				Brightkite				NYC				TKY				SIN			
	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10
STAN	58.7	41.6	70.3	46.6	57.2	45.4	69.8	47.3	32.1	22.3	45.9	27.3	61.2	50.1	71.3	54.4	18.0	11.3	29.5	15.7
int. 1	54.2	36.5	66.5	42.2	52.1	40.6	66.2	43.3	30.0	20.3	42.5	24.5	57.4	46.9	67.7	52.1	16.8	10.3	27.5	14.4.
int. 2	48.2	31.3	62.2	37.0	50.4	40.9	65.2	42.0	24.6	14.5	36.5	18.3	52.5	43.8	64.4	49.4	13.5	8.3	25.5	13.4
int. 3	42.4	28.1	58.4	33.3	47.0	37.5	63.3	40.3	20.3	10.9	32.1	15.6	47.9	40.2	62.5	45.9	12.1	8.5	23.6	12.1
GeoSAN	56.2	41.4	69.9	45.9	55.8	42.3	67.2	46.0	30.2	20.8	44.5	25.3	58.8	48.4	69.1	51.8	17.4	11.8	30.0	15.9
int. 1	54.3	39.5	67.8	43.9	52.1	40.6	66.2	43.3	29.0	18.6	43.4	23.2	56.3	46.3	67.7	49.3	16.4	**12.0**	29.5	14.4
int. 2	53.2	39.6	65.3	41.2	50.4	40.9	65.2	42.0	28.7	18.3	42.6	22.9	54.3	44.8	65.4	47.6	15.2	10.4	28.6	13.4
int. 3	50.1	35.4	62.1	38.5	47.0	37.5	63.3	40.3	25.7	16.9	39.6	21.0	55.3	43.2	65.2	46.2	14.2	9.4	26.4	13.8
TPG	63.2	45.5	74.6	50.1	62.4	47.9	74.4	51.8	37.3	26.8	52.8	31.8	65.4	53.1	76.5	56.7	19.4	14.3	34.3	19.1
int. 1	**63.7**	44.3	73.7	49.1	60.4	46.2	73.0	50.3	**38.0**	26.7	**53.6**	**32.1**	64.5	52.9	75.6	56.4	**19.9**	13.3	33.1	18.1
int. 2	60.5	**45.9**	72.8	48.6	59.9	45.6	72.5	49.9	**37.5**	26.5	**53.2**	30.1	64.1	52.2	75.1	55.8	19.1	12.6	32.2	18.7
int. 3	59.7	**46.5**	72.0	47.1	59.8	45.3	72.3	49.8	37.0	25.3	52.0	29.7	64.9	**53.1**	75.1	**56.7**	**19.6**	**14.3**	31.3	17.6

length of the grid. We train our model using the Adam optimizer with a learning rate of 0.001 and set the dropout ratio to 0.5. The number of training epochs is set to 50 for all four datasets. For baselines except STAN, we follow their implementation and best settings which they claim in their papers. STAN builds matrices for all historical check-ins of each user, which results in extremely time consuming and memory consuming. Running the original version of STAN caused an out-of-memory (OOM) error on our server with 768GB memory. To test the performance of STAN, we choose to select a part of users to train model at a time and test performance on these users. For NYC and SIN datasets, we use all users. For other datasets, we select the first 2000 users to test model performance due to the large number of users on these datasets.

Overall Performance Comparison

We compare the performance of our proposed TPG with baselines mentioned above. Table 4.5 reports the performance of TPG and eight baselines in terms of Recall@k and NDCG@k on five real world datasets. The "Improv." column refers to the improvement rate of TPG compared to the second best model. Based on the results, we observe that:

(1) Our proposed TPG significantly outperforms all the baseline methods on all

datasets w.r.t. both NDCG@k and Recall@k and all values of k. Our proposed method achieves up to 20.2% and 16.2% improvements over the best-performing baseline in terms of NDCG@5 and Recall@5. It demonstrates the effectiveness and superiority of our proposed TPG, which makes effective usage of geographic information and temporal signal of the next location.

(2) Compared with RNN-based approaches, pure attention-based methods such as MobTCast, STAN, GeoSAN, and our proposed TPG clearly achieve better performances. It is reasonable since attention mechanism can capture global contextual information in spatial-temporal check-in sequences, while RNN-based methods suffer from the risk of forgetting past long-range information. Among RNN-based models, DeepMove and CARA generally have relatively better performances than others, which attributes to their consideration of spatial-temporal modelling, and short-term and long-term periodicity modeling. These designations make up for the inherent defects of RNN to a certain extent. Compared with attention-based state-of-the-art model MobTCast, STAN, and GeoSAN, the substantial improvement achieved by TPG demonstrates the importance of explicitly using temporal signal of the next location and shifted window mechanism for geo-gridding. Although STAN generally performs better than MobTCast and GeoSAN, it costs extremely large memory overhead and calculation time overhead due to matrix operations for all historical check-ins. Our method TPG is far superior to other methods in terms of computing time, memory, and accuracy.

(3) The density of check-in records for different datasets can be represented by #check-ins / (#users $\times$ #locations). The sparsities are 0.001, 0.007, 0.027, 0.009, and 0.120 for Gowalla, Brightkite, NYC, TKY, and SIN, respectively. This can explain why the improvement brought by TPG is larger in NYC and SIN than in TKY. Besides, it is obvious that TPG has a strong ability to handle sparse data like Gowalla.

Table 4.7: Performances of ablation studies for TPG.

	Gowalla				Brightkite				NYC				TKY				SIN			
	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10
TPG	63.2	45.5	74.6	50.1	62.4	47.9	74.4	51.8	37.3	26.8	52.8	31.8	65.4	53.1	76.5	56.7	19.4	14.3	34.3	19.1
I. − TP	60.4	44.0	73.0	48.4	61.1	47.2	73.6	50.4	34.3	23.7	47.1	28.2	62.2	50.5	72.9	54.0	18.3	12.2	29.7	15.9
II. − TE	61.7	44.0	73.7	49.0	57.3	43.7	69.0	47.5	36.9	25.0	50.8	29.4	65.4	53.3	76.4	56.9	21.3	15.9	32.2	18.4
III. − SW	58.77	43.6	72.6	48.0	61.5	47.4	73.7	51.3	36.8	25.6	52.3	31.7	64.0	51.3	75.1	55.0	19.9	13.3	30.5	16.7
IV. − GE	60.3	42.6	72.1	47.6	54.7	42.2	65.1	45.6	34.5	25.2	50.3	28.50	56.2	43.3	68.1	47.4	19.7	14.3	33.7	17.1
V. + UE	62.1	45.1	74.0	48.5	57.9	44.1	70.1	48.0	36.3	25.4	51.1	32.0	53.2	40.1	65.9	44.3	18.5	13.1	26.7	15.7

Interval Prediction Performances

By introducing temporal prompts, TPG is able to make interval predictions with accurate timestamps of locations to be predicted. Relevant results are given in Table 4.6. We here mask one ("int. 1" in Table 4.6), two ("int. 2" in Table 4.6), and three ("int. 3" in Table 4.6) most recent check-in(s) of users to test TPG and two baselines STAN and GeoSAN 's performances on all datasets. The detailed setting is using first 96, 97, 98 check-ins to predict the 100-th check-in in each user's trajectory.

For STAN and GeoSAN, we observe that compared with using all check-in data, the more the latest check-in(s) is/are masked, the more the performances drop marginally. It is reasonable since some previous studies [92] have demonstrated that a user's recent behavior has a great impact on the user's next behavior. However, the situation is totally different for TPG's interval prediction. The performances are sometimes even better than the next location recommendation. These impressive results are very strong arguments for the benefit of explicit temporal information modeling by using timestamps as prompts.

Ablation Study

We conduct an extensive ablation study on TPG, to dive into the effectiveness of each module in our proposed model. It should be emphasized that our base model (denoted as TPG in the first row of Table 4.7) does not have user embedding. We consider the following variants of our model for ablation:

- Remove TP (Temporal-based Prompt): We use locations as query of the decoder, instead of using temporal-based prompts.

- Remove TE (Time Embedding): We remove time embedding of check-in sequences, only using POI embedding and geography embedding as inputs of history encoder.

- Remove SW (Shifted Window Mechanism): We remove the shifted window mechanism in the geography-aware encoder.

- Remove GE (Geography Encoder): We remove the geography encoder and only use POI embedding and time embedding as inputs of history encoder.

- Add UE (User Embedding): We add user embedding into the history encoder by concatenating it with POI embedding, geography embedding, and time embedding.

The performance comparisons are shown in Table 4.7. From the comparisons, we have several findings:

(1) The overall performance of the model drops without temporal-based prompts, especially on Foursquare datasets with local regions. This phenomenon indicates the significance of explicitly incorporating temporal signal of the next location. Furthermore, results of "Remove TE" is generally slightly better than "Remove TP" in Gowalla Foursquare datasets, while still worse than TPG. It indicates that if time information is not used appropriately, it brings additional noise to the model and degrades the performances. It also proves that the temporal-based prompt strategy proposed in this work makes effective use of time information.

(2) The results of "Remove SW" is worse than original TPG in most cases. It demonstrates that our proposed shifted window mechanism is an efficient augmentation method for spatial information. It can bridge the semantic gap of adjacent grids

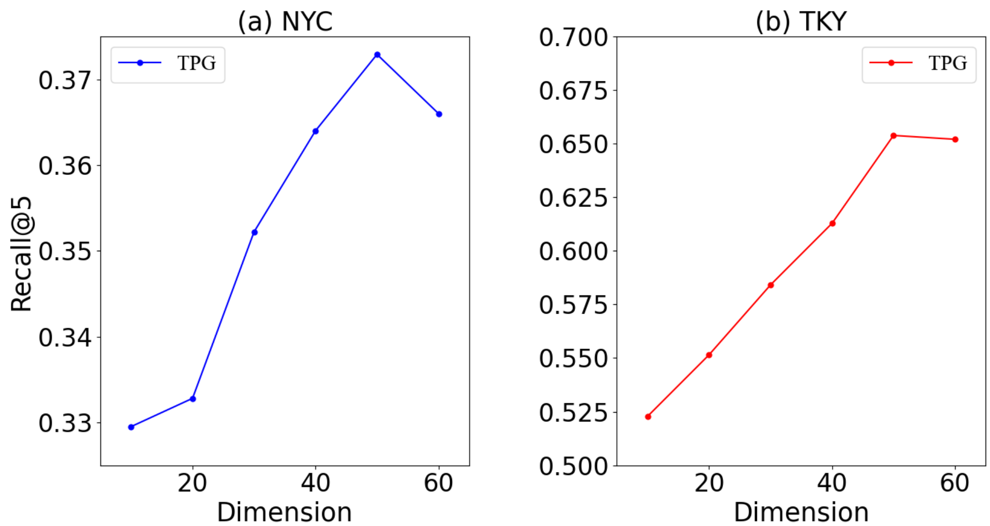

Figure 4.18: The impact of geography embedding dimension for model performances.

in terms of spatial distribution, and improves the accuracy and stability of the model. Performances of "Remove GE" is even worse than "Remove SW". It is reasonable since geographic information is of great significance in location-based applications.

(3) "Add UE" generally leads to great performance reduction compared with TPG. This is because adding user embedding to inputs of history encoder may contribute to the misaligned between the vector space of check-in sequences and the vector space of locations. This suggests future work to appropriately use user information to enhance the performance.

Parameter Sensitivity Analysis

Due to the emphasis on the innovation of the geography-aware Encoding encoder in this study, we chose to conduct sensitivity experiments on two parameters, namely, geography embedding dimension and step size of the shifted windows, within this module. We first investigate model sensitivity with regard to geography embedding dimension. We vary the dimension used in the geography-aware encoder from 10 to 60 with a step of 10. The experimental results on two datasets NYC and TKY are reported in Figure 4.18. We can come to a conclusion that a small dimension for geography embedding will make the performance very poor. This is because small dimensions are difficult to describe the complex geographic relationships between POIs, which will cause great information loss. The model performance reaches the peak when the geography dimension is 50. When the dimension increases to 60, the performance decreases a bit. This may be explained that the size of the semantic space formed by the geography-aware encoder is certain. When the embedding dimension is too high, the information is unsaturated, and noise may be introduced instead.

We further investigate model sensitivity with regard to the step size of the shifted windows. We vary the step size used in the shifted window mechanism from 0.25 to 1 with a step of 0.25. Note that the step size here means the proportion of moving length and grid size. We still take two datasets NYC and TKY as examples. The experimental results are showed in Figure 4.19. We can find that the performance peaks at a small step size 0.25 for shifted window mechanism, and dropped until the step size is 0.75. Such phenomenon conforms to the First Law of Geography, which indicates "everything is related to everything else, but near things are more related than distant things." We can also observe that when the step size is 1, there is a performance improvement. When the step size is 1, the model actually degenerates to directly aggregate the neighbor grids of the grid itself at each level. This phenomenon proves the defect of previous methods, that is, they do not fully consider the correlations of adjacent grids.

4.2.4 Conclusions

In this work, we revisit the location recommendation problem. We find that most methods either ignore the prerequisite of knowing the exact time at which the POI needs to be predicted in real world applications, or implicitly fuse temporal infor-

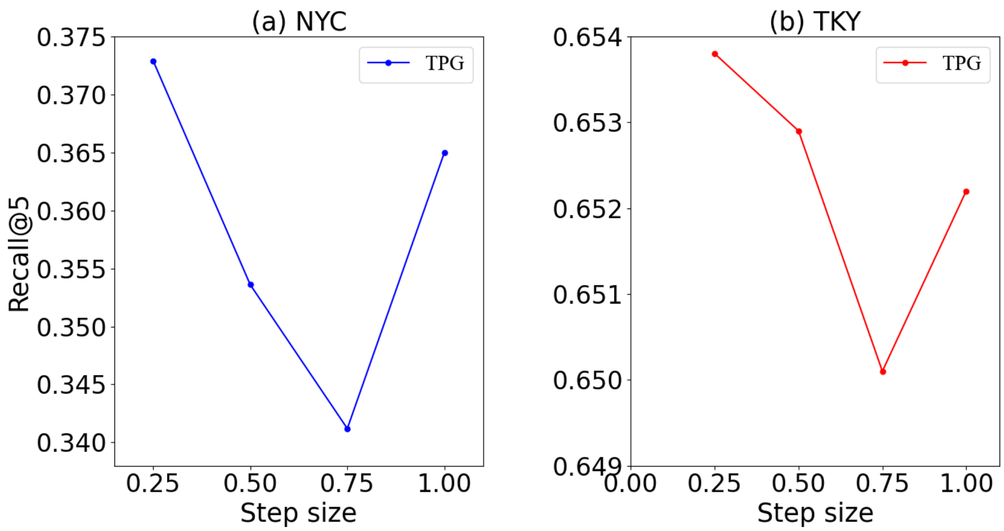

Figure 4.19: The impact of the step size of shifted windows for model performances.

mation with historical check-ins. We propose TPG, a temporal prompt-based and geography-aware framework, for next location recommendations. We show how to use timestamps as prompts to explicitly model time information of locations to be predicted. By proposing a shifted window mechanism, we also show how to avoid the hard boundary problem with regard to geographic coordinates of check-ins. The experimental results on five benchmark datasets demonstrated the superiority of TPG compared with other state-of-the-art methods. The results indicated that temporal signals of locations are of great significance. We also demonstrate through ablation studies that our proposed shifted window mechanism is capable of overcoming defects with regard to geographic information modeling of previous approaches.

As for future work, we plan to design more intelligent prompts. Large-scale pre-trained models from NLP communities have demonstrated unlimited potential of prompt learning. We can consider combining location recommendation with language pre-trained models through prompts.

Chapter 5

Conclusions

In this thesis, we have explored the challenges and limitations of existing methodologies in the field of artificial urban intelligence, focusing on human-land interaction in urban applications as the key to building smarter cities. We have contributed to the understanding of urban environments, human mobility, and location recommendation by proposing novel frameworks and techniques to overcome these challenges. In this concluding chapter, we summarize the main findings and contributions of the research, and discuss the potential implications and future directions of our work.

Our research began with the identification of the importance of urban environment comprehension and human mobility understanding in artificial urban intelligence applications. We highlighted the need to explore human-land interaction in the era of big data, as it paves the way towards more sustainable, efficient, and adaptable urban spaces. The research scope was focused on urban data analytics and applications, which allowed us to delve into the requirements and issues related to human-land interaction in artificial urban intelligence applications.

In Chapter 3, we proposed a novel multi-graph framework called Region2Vec for urban region representation learning. This framework is designed to capture inter-region relations through human mobility, geographical contextual information via neighborhood data, and intra-region information using Point of Interest (POI) side information in knowledge graphs. Region2Vec also considers accessibility, vicinity, and functionality correlations among regions. The encode-decode multi-graph fusion module is introduced to jointly learn comprehensive representations, considering the discriminative properties of multiple graphs. Experiments on real-world datasets demonstrate the effectiveness of Region2Vec, as it consistently outperforms state-of-the-art baselines by at least 7.80% in various tasks and metrics. The comprehensive region representation obtained from Region2Vec can be employed in multiple applications, making it a step towards building general-purpose intelligent agents capable of handling diverse urban challenges.

Chapter 4 was divided into two parts: human mobility analysis and location recommendation. For the human mobility analysis part, we used tourist travel patterns as a case study and employed trip chains to model and discover fixed patterns. Through the framework, we found that most patterns can be captured by only 13 key trip chains, and that the principle of least efforts (PLE) affects tourists' structures of trip chains. The results of this analysis not only demonstrated the complex daily travel trip chains from tourism big data, but also filled the gap in tourism literature on multi-destination trips by discovering significant and underlying patterns based on mobile datasets.

In the location recommendation section, we proposed a novel Temporal Prompt-based and Geography-aware (TPG) framework. Temporal information serves as a prompt for our recommendation system, while a shifted window mechanism is devised to augment geographic data and avoid the hard boundary problem when handling longitude and latitude of POIs with grids. Experiments on five real-world datasets (Gowalla, Brightkite, Foursquare-NYC, Foursquare-TKY, and Foursquare-SIN) show that TPG outperforms state-of-the-art counterparts under different settings, and excels in interval prediction. Specifically, the model can predict a user's desired location at a given time, even when the most recent check-in data is masked, or predict a specific future check-in at a given timestamp, not just the next one.

This research has made significant contributions to the field of artificial urban intelligence, providing valuable insights into the complexities of human-land interactions in urban spaces. The proposed frameworks and techniques have the potential to transform the way we understand and manage urban environments, leading to more intelligent, efficient, and sustainable cities.

As urbanization continues to progress, the importance of artificial urban intelligence and the need for smarter cities will only grow. There are several promising avenues for future research in this area, including the development of more advanced models for urban region representation learning, the incorporation of additional data sources to improve the accuracy and comprehensiveness of human mobility analysis, and the refinement of location recommendation algorithms to better account for the diverse needs of urban populations. Furthermore, as artificial urban intelligence continues to evolve, interdisciplinary collaborations between urban planners, data scientists, and policymakers will become increasingly important to ensure that the insights generated by these advanced models are translated into effective urban planning and management strategies.

One potential direction for future research is the incorporation of real-time data sources, such as social media and IoT devices, to improve the accuracy and timeliness of urban environment comprehension and human mobility understanding. This could enable more effective real-time decision-making, allowing for better resource allocation and urban service provision. Additionally, the integration of advanced machine learning techniques, such as deep learning and reinforcement learning, could lead to even more sophisticated models that are better able to capture the complexities of human-land interactions.

Another possible direction is the exploration of the social, economic, and environmental implications of artificial urban intelligence. As these advanced models are increasingly adopted in urban planning and management, it will be important to understand how they impact various stakeholders, including residents, businesses, and the natural environment. This could help guide the ethical and sustainable development of artificial urban intelligence, ensuring that its benefits are shared equitably across society.

In addition to the previously mentioned future research directions, another promising avenue is the integration of artificial intelligence in generative design and urban planning, inspired by the rapid development and popularity of AI in generative content creation (AIGC). This could lead to a paradigm shift in smart city applications, moving away from the conventional "smart brain" model, which primarily focuses on generating indicators based on the current urban situation, towards a more proactive approach that leverages generative algorithms to design neighborhoods with a so-called "perfect" balance. This innovative combination of AIGC and urban planning has the potential to revolutionize the way we envision and create urban spaces, enabling the development of neighborhoods that are optimized for various factors, such as walkability, accessibility, sustainability, and aesthetics. By incorporating generative design algorithms in the planning process, urban planners and policymakers could explore a vast range of possibilities and identify optimal solutions that meet the unique needs and preferences of diverse urban populations. To achieve this ambitious goal, future research should focus on developing advanced generative algorithms that can efficiently model and optimize complex human-land interactions, taking into account the multidimensional nature of urban environments. Moreover, these algorithms should be flexible and adaptable, allowing for customization based on specific local contexts and stakeholder requirements.

In conclusion, this thesis has made important contributions to the understanding of human-land interaction in artificial urban intelligence applications, providing valuable insights and innovative techniques for urban environment comprehension, human mobility analysis, and location recommendation. By furthering our knowledge in these areas, we can better harness the power of big data and advanced analytics to build more intelligent, sustainable, and vibrant urban spaces for the future.

References

[1] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 2005.

[2] R. Ahas, S. Silm, O. Järv, E. Saluveer, and M. Tiru. Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of Urban Technology*, 17:3–27, 2010.

[3] L. Alessandretti, P. Sapiezynski, V. Sekara, S. Lehmann, and A. Baronchelli. Evidence for a conserved quantity in human mobility. *Nature Human Behaviour*, 2(7):485–491, 2018.

[4] R. M. Anderson, B. Anderson, and R. M. May. *Infectious diseases of humans: dynamics and control*. Oxford University Press, 1992.

[5] Shifra Baruchson-Arbib and Jenny Bronstein. Humanists as information users in the digital age: The case of jewish studies scholars in israel. *Journal of the American society for information science and technology*, 58(14):2269–2279, 2007.

[6] Michael Batty. The new science of cities. *MIT Press*, 2013.

[7] Paul Baumann, Wilhelm Kleiminger, and Silvia Santini. The influence of temporal and spatial features on the performance of next-place prediction algorithms. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, pages 449–458, 2013.

[8] Bruce F Berg. New york city politics: Governing gotham. *New Republic*, page 352, 2007.

[9] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*, 2013.

[10] William Bunge. Theoretical geography. *Lund Studies in Geography, Ser. C, General and Mathematical Geography*, 1:1–63, 1966.

[11] Jian Cao, Qingquan Li, Wei Tu, and Fan Wang. Characterizing preferred motif choices and distance impacts. *PLoS ONE*, 14(4):e0215242, 2019.

[12] Eugenio Cesario, Carmela Comito, and Domenico Talia. An approach for the discovery and validation of urban mobility patterns. *Pervasive and Mobile Computing*, 42:77–92, December 2017.

[13] Buru Chang, Yonggyu Park, Donghyeon Park, Seongsoon Kim, and Jaewoo Kang. Content-aware hierarchical point-of-interest embedding model for successive POI recommendation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 3301–3307, July 2018.

[14] X. Chang, J Wu, Z. He, D. Li, and W. Wang. Understanding user's travel behavior and city region functions from station-free shared bike usage data. *Transportation Research Part F Traffic Psychology and Behaviour*, 72:81–95, 2020.

[15] Carmela Comito. NexT: A framework for next-place prediction on location based social networks. *Knowledge-Based Systems*, 204:106205, September 2020.

[16] C. Cooper. Spatial and temporal patterns of tourist behaviour. *Regional Studies*, 15(5):359–371, 1981.

[17] Alessandro Crivellari and Bernd Resch. Investigating functional consistency of mobility-related urban zones via motion-driven embedding vectors and local POI-type distributions. *Computational Urban Science*, 2(1):19, 2022.

[18] Andrew Crooks, Dieter Pfoser, Andrew Jenkins, Arie Croitoru, Anthony Stefanidis, Duncan Smith, Sophia Karagiorgou, Alexandros Efentakis, and George Lamprianidis. Crowdsourcing urban form and function. *International Journal of Geographical Information Science*, 29(5):720–741, 2015.

[19] B. C. Csáji, A. Browet, V. A. Traag, J.-C. Delvenne, E. Huens, P. Van Dooren, and V. D. Blondel. Exploring the mobility of mobile phone users. *Physica A: Statistical Mechanics and its Applications*, 392(6):1459–1473, 2013.

[20] M. Dijst. Two-earner families and their action spaces: A case study of two dutch communities. *GeoJournal*, 48(3):195–206, 1999.

[21] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. Deepmove: Predicting human mobility with attentional recurrent networks. In *Proceedings of the World Wide Web Conference (WWW)*, pages 1459–1468, 2018.

[22] Vanessa Frías-Martínez, Jesús Virseda, and Enrique Frías-Martínez. On the relation between socio-economic status and physical mobility. *Information Technology for Development*, 18:106 – 91, 2012.

[23] Y. Fu, P. Wang, J Du, L. Wu, and X. Li. Efficient region embedding with multi-view spatial networks: A perspective of locality-constrained spatial autocorrelations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 906–913, 2019.

[24] Yanjie Fu, Pengyang Wang, Jiadi Du, Le Wu, and Xiaolin Li. Efficient Region Embedding with Multi-View Spatial Networks: A Perspective of Locality-Constrained Spatial Autocorrelations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):906–913, July 2019.

[25] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Next place prediction using mobility markov chains. In *Proceedings of the Workshop on Measurement, Privacy, and Mobility (MPM)*, pages 1–6, 2012.

[26] Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states. *Proceedings of the National Academy of Sciences*, 114(50):13108–13113, 2017.

[27] Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. Using Deep Learning and Google Street View to Estimate the Demographic Makeup of the US. *Proceedings of the National Academy of Sciences*, 114:201700035, 11 2017.

[28] Thomas F Golob. A simultaneous model of household activity participation and trip chain generation. *Transportation Research Part B: Methodological*, 34(5):355–376, 2000.

[29] Thomas F Golob and David A Hensher. The trip chaining activity of sydney residents: A cross-section assessment by age group with a focus on seniors. *Journal of Transport Geography*, 15(4):298–312, 2007.

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

[31] Michael F Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69:211–221, 2007.

[32] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *the 22nd ACM SIGKDD International Conference*, 2016.

[33] S. J. Guy, S. Curtis, M. C. Lin, and D. Manocha. Least-effort trajectories lead to emergent crowd behaviors. *Physical Review E*, 85(1):016110, 2012.

[34] Christophe Haikal, Pegah Alizadeh, Christophe Rodrigues, and Bi Chongke. Place embedding across cities in location-based social networks. In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, pages 539–546, 2022.

[35] Mazlan Hashim, Norzailawati Mohd Noor, and Maged Marghany. Modeling sprawl of unauthorized development using geospatial technology: Case study in kuantan district, malaysia. *International Journal of Digital Earth*, 4(3):223–238, 2011.

[36] Mattias Höjer and Josefin Wangel. Smart sustainable cities: definition and challenges. In *ICT innovations for sustainability*, pages 333–349. Springer, 2015.

[37] José Holguín-Veras and Gopal R Patil. Observed trip chain behavior of commercial vehicles. *Transportation research record*, 1906(1):74–80, 2005.

[38] Sheng Hu, Song Gao, Liang Wu, Yongyang Xu, Ziwei Zhang, Haifu Cui, and Xi Gong. Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach. *Computers, environment and urban systems*, 87:101619, 2021.

[39] Tianyuan Huang, Zhecheng Wang, Hao Sheng, Andrew Y Ng, and Ram Rajagopal. M3g: Learning urban neighborhood representation from multi-modal multi-graph. In *DeepSpatial 2021: 2nd ACM KDD Workshop on Deep Learning for Spatio-Temporal Data, Applications and Systems*, 2021.

[40] Weiming Huang, Lizhen Cui, Meng Chen, Daokun Zhang, and Yao Yao. Estimating urban functional distributions with semantics preserved POI embedding. *International Journal of Geographical Information Science*, 0(0):1–26, 2022.

[41] R. Hubbard. A review of selected factors conditioning consumer travel behavior. *Journal of Consumer Research (pre-1986)*, 5(1):1, 1978.

[42] L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences*, 101(42):15124–15129, 2004.

[43] Y.-H. Hwang, U. Gretzel, and D. R. Fesenmaier. Multicity trip patterns: Tourists to the united states. *Annals of Tourism Research*, 33(4):1057–1078, 2006.

[44] T. Inzerilli, A. M. Vegni, A. Neri, and R. Cusani. A location-based vertical handover algorithm for limitation of the ping-pong effect. In *2008 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, 2008.

[45] S. Isaacman, R. Becker, R. Cáceres, M. Martonosi, J. Rowland, A. Varshavsky, and W. Willinger. Human mobility modeling at metropolitan scales. pages 239–252, 2012.

[46] Neal Jean, Sherrie Wang, Anshul Samar, George Azzari, David Lobell, and Stefano Ermon. Tile2Vec: Unsupervised Representation Learning for Spatially Distributed Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):3967–3974, July 2019.

[47] James Jeng and Daniel R Fesenmaier. Conceptualizing the travel decision-making hierarchy: A review of recent developments. *Tourism analysis*, 7(1):15–32, 2002.

[48] Porter Jenkins, Ahmad Farag, Suhang Wang, and Zhenhui Li. Unsupervised Representation Learning of Spatial Data via Multimodal Embedding. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1993–2002, November 2019.

[49] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. Knowledge graph embedding via dynamic mapping matrix. In *Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, pages 687–696, 01 2015.

[50] Renhe Jiang, Xuan Song, Zipei Fan, Tianqi Xia, Zhaonan Wang, Quanjun Chen, Zekun Cai, and Ryosuke Shibasaki. Transfer Urban Human Mobility via POI Embedding over Multiple Cities. *ACM/IMS Transactions on Data Science*, 2(1):4:1–4:26, 2021.

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

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

[53] A. Kadavankandy, K. Avrachenkov, L. Cottatellucci, and R. Sundaresan. The power of side-information in subgraph detection. *IEEE Transactions on Signal Processing*, PP(99):1–1, 2018.

[54] H. Kamper, W. Wang, and K. Livescu. Deep convolutional acoustic word embeddings using word-pair side information. In *IEEE International Conference on Acoustics*, pages 4950–4954, 03 2016.

[55] Wang-Cheng Kang and Julian McAuley. Self-attentive sequential recommendation. In *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, pages 197–206, 2018.

[56] Sujeong Kim, Stephen J Guy, Wenxi Liu, David Wilkie, Rynson WH Lau, Ming C Lin, and Dinesh Manocha. Brvo: Predicting pedestrian trajectories using velocity-space reasoning. *The International Journal of Robotics Research*, 34(2):201–217, 2015.

[57] Sung-Eun Kim, Kyung Young Lee, Soo Il Shin, and Sung-Byung Yang. Effects of tourism information quality in social media on destination image formation: The case of sina weibo. *Information & management*, 54(6):687–702, 2017.

[58] Thomas N. Kipf and Max Welling. Variational graph auto-encoders, 2016.

[59] Dejiang Kong and Fei Wu. Hst-lstm: A hierarchical spatial-temporal long-short term memory network for location prediction. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2341–2347, 2018.

[60] Ajay Kumar and David Matthew Levinson. Chained trips in montgomery county, maryland. *Institute of Transportation Engineers Journal*, pages 27–32, 1995.

[61] G. Lau and B. McKercher. Understanding tourist movement patterns in a destination: A gis approach. *Tourism and hospitality research*, 7(1):39–49, 2006.

[62] Stephen Law and Mateo Neira. An unsupervised approach to geographical knowledge discovery using street level and street network images. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, pages 56–65, November 2019.

[63] Hee Lee, Basak Denizci Guillet, Rob Law, and Rosanna Leung. Robustness of distance decay for international pleasure travelers: A longitudinal approach. *International Journal of Tourism Research*, 14:409–420, 09 2012.

[64] Alan Lew and Bob McKercher. Modeling tourist movements: A local destination analysis. *Annals of tourism research*, 33(2):403–423, 2006.

[65] Ranzhen Li, Yanyan Shen, and Yanmin Zhu. Next point-of-interest recommendation with temporal and multi-level context attention. In *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, pages 1110–1115, 2018.

[66] Defu Lian, Yongji Wu, Yong Ge, Xing Xie, and Enhong Chen. Geography-aware sequential location recommendation. In *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 2009–2019, 2020.

[67] Defu Lian, Cong Zhao, Xing Xie, Guangzhong Sun, Enhong Chen, and Yong Rui. Geomf: Joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 831–840, 2014.

[68] Yan Lin, Huaiyu Wan, Shengnan Guo, and Youfang Lin. Pre-training Context and Time Aware Location Embeddings from Spatial-Temporal Trajectories for User Next Location Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5):4241–4248, May 2021.

[69] Xin Liu, Yong Liu, and Xiaoli Li. Exploring the context of locations for personalized location recommendations. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pages 1188–1194, 2016.

[70] Yu Liu, Xi Liu, Song Gao, Li Gong, Chaogui Kang, Ye Zhi, Guanghua Chi, and Li Shi. Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105(3):512–530, 2015.

[71] A. L. Lloyd and R. M. May. How viruses spread among computers and people. *Science*, 292(5520):1316–1317, 2001.

[72] Chi-Chuan Lue, John L Crompton, and Daniel R Fesenmaier. Conceptualization of multi-destination pleasure trips. *Annals of tourism research*, 20(2):289–301, 1993.

[73] Yan Luo. Characterizing tourist daily trip chains using mobile phone big data. Technical Report arXiv:2205.14395, 2022.

[74] Yingtao Luo, Qiang Liu, and Zhaocheng Liu. Stan: Spatio-temporal attention network for next location recommendation. In *Proceedings of the World Wide Web Conference (WWW)*, pages 2177–2185, 2021.

[75] Xiaolei Ma, Yao-Jan Wu, Yinhai Wang, Feng Chen, and Jianfeng Liu. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies*, 36:1–12, 2013.

[76] Christopher Manning and Hinrich Schutze. *Foundations of statistical natural language processing*. MIT press, 1999.

[77] Jarana Manotumruksa, Craig Macdonald, and Iadh Ounis. A contextual attention recurrent architecture for context-aware venue recommendation. In *Proceedings of the ACM International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 555–564, 2018.

[78] Takuya Maruyama and Noboru Harata. Difference between area-based and cordon-based congestion pricing: Investigation by trip-chain–based network equilibrium model with nonadditive path costs. *Transportation Research Record*, 1964(1):1–8, 2006.

[79] Adolfo Paolo Masucci, Duncan Smith, Andrew Crooks, and Michael Batty. Random planar graphs and the london street network. *The European Physical Journal B*, 71:259–271, 2009.

[80] N. McGuckin and E. Murakami. Examining trip-chaining behavior: Comparison of travel by men and women. *Transportation Research Record*, 1693(1):79–85, 1999.

[81] N. McGuckin and Y. Nakamoto. Trips, chains and tours-using an operational definition. In *National Household Travel Survey Conference*, 2004.

[82] N. McGuckin, J. Zmud, and Y. Nakamoto. Trip-chaining trends in the united states: understanding travel behavior for policy making. *Transportation Research Record*, 1917(1):199–204, 2005.

[83] Nancy McGuckin and Elaine Murakami. Examining trip-chaining behavior: Comparison of travel by men and women. *Transportation research record*, 1693(1):79–85, 1999.

[84] B. Mckercher and G. Lau. Movement patterns of tourists within a destination. *Tourism Geographies*, 10(3):355–374, 2008.

[85] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, pages 3111–3119, December 2013.

[86] Clive Morley, Jaume Rosselló, and Maria Santana-Gallego. Gravity models for tourism demand: theory and use. *Annals of Tourism Research*, 48:1–10, 2014.

[87] Pamela M Murray-Tuite and Hani S Mahmassani. Model of household trip-chain sequencing in emergency evacuation. *Transportation Research Record*, 1831(1):21–29, 2003.

[88] G. Olsson. Distance and human interaction a migration study. *Geografiska Annaler: Series B, Human Geography*, 47(1):3–43, 1965.

[89] Silvia Paldino, Iva Bojic, Stanislav Sobolevsky, Carlo Ratti, and Marta C González. Urban magnetism through the lens of geo-tagged photography. *EPJ Data Science*, 4:1–17, 2015.

[90] Patrick Cheong-Iao Pang, Shanton Chang, Karin Verspoor, and Jon Pearce. Designing health websites based on users' web-based information-seeking behaviors: A mixed-method observational study. *Journal of medical Internet research*, 18(6):e145, 2016.

[91] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, page 701–710, New York, NY, USA, 2014. Association for Computing Machinery.

[92] Qi Pi, Guorui Zhou, Yujing Zhang, Zhe Wang, Lejian Ren, Ying Fan, Xiaoqiang Zhu, and Kun Gai. Search-based user interest modeling with lifelong sequential behavior data for click-through rate prediction. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*, pages 2685–2692, 2020.

[93] Steward TA Pickett and Mary L Cadenasso. Landscape ecology: spatial heterogeneity in ecological systems. *Science*, 269(5222):331–334, 1995.

[94] Frank Primerano, Michael AP Taylor, Ladda Pitaksringkarn, and Peter Tisato. Defining and understanding trip chaining behaviour. *Transportation*, 35:55–72, 2008.

[95] L. K. Rasmussen and I. J. Oppermann. Ping-pong effects in linear parallel interference cancellation for cdma. *IEEE Transactions on Wireless Communications*, 2(2):357–363, 2003.

[96] Ronald R Rindfuss, Stephen J Walsh, Billie L Turner, Jefferson Fox, and Vinod Mishra. Developing a science of land change: challenges and methodological issues. *Proceedings of the National Academy of Sciences*, 101(39):13976–13981, 2004.

[97] C. M. Schneider, V. Belik, T. Couronne, Z. Smoreda, and M. C. Gonzalez. Unravelling daily human mobility motifs. *J R Soc Interface*, 10(84):20130246, 2013.

[98] S. Schönfelder and K. W. Axhausen. Activity spaces: measures of social exclusion? *Transport policy*, 10(4):273–286, 2003.

[99] Toru Shimizu, Takahiro Yabe, and Kota Tsubouchi. Enabling Finer Grained Place Embeddings using Spatial Hierarchy from Human Mobility Trajectories. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, pages 187–190, November 2020.

[100] Noam Shoval and Rein Ahas. The use of tracking technologies in tourism research: the first decade. *Tourism Geographies*, 18(5):587–606, 2016.

[101] Jungyul Sohn. Are commuting patterns a good indicator of urban spatial structure? *Journal of Transport geography*, 13(4):306–317, 2005.

[102] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

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

[104] Susan I. Stewart and Christine A. Vogt. Multi-destination trip patterns. *Annals of Tourism Research*, 24:458–461, 1996.

[105] Susan I Stewart and Christine A Vogt. Multi-destination trip patterns. *Annals of Tourism Research*, 24(2):458–461, 1997.

[106] Shiliang Su, Chen Wan, Yixuan Hu, and Zhongliang Cai. Characterizing geographical preferences of international tourists and the local influential factors in china using geo-tagged photos on social media. *Applied Geography*, 73:26–37, 2016.

[107] Ke Sun, Tieyun Qian, Tong Chen, Yile Liang, Quoc Viet Hung Nguyen, and Hongzhi Yin. Where to go next: Modeling long- and short-term user preferences for point-of-interest recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 214–221, 2020.

[108] Zhihao Sun, Hongzan Jiao, Hao Wu, Zhenghong Peng, and Lingbo Liu. Block2vec: An Approach for Identifying Urban Functional Regions by Integrating Sentence Embedding Model and Points of Interest. *ISPRS International Journal of Geo-Information*, 10(5):339, 2021.

[109] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. Line: Large-scale information network embedding. In *International Conference on World Wide Web*, 2015.

[110] Robert Tibshirani. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3):273–282, 2011.

[111] Carmen Tideswell and Bill Faulkner. Multidestination travel patterns of international visitors to queensland. *Journal of Travel research*, 37(4):364–374, 1999.

[112] W. R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234–240, 1970.

[113] Waldo R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234–240, 1970.

[114] W. G. van der Knaap. Gis-oriented analysis of tourist time-space patterns to support sustainable tourism development. *Tourism Geographies*, 1(1):56–69, 1999.

[115] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc.

[116] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 5998–6008, 2017.

[117] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2018.

[118] Alexander von Humboldt and Carl Ritter. *The founders of human geography: Alexander von Humboldt and Carl Ritter*. Berlin, 1856.

[119] Hongjian Wang and Zhenhui Li. Region representation learning via mobility flow. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 237–246, 2017.

[120] Hongjian Wang and Zhenhui Li. *Region Representation Learning via Mobility Flow*, page 237–246. Association for Computing Machinery, New York, NY, USA, 2017.

[121] Jingyuan Wang, Jiawei Jiang, Wenjun Jiang, Chao Li, and Wayne Xin Zhao. Libcity: An open library for traffic prediction. In *Proceedings of the International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, pages 145–148, 2021.

[122] Zhaobo Wang, Yanmin Zhu, Qiaomei Zhang, Haobing Liu, Chunyang Wang, and Tong Liu. Graph-enhanced spatial-temporal network for next poi recommendation. *ACM Transactions on Knowledge Discovery from Data*, 16(6):1–21, 2022.

[123] Zhecheng Wang, Haoyuan Li, and Ram Rajagopal. Urban2Vec: Incorporating Street View Imagery and POIs for Multi-Modal Urban Neighborhood Embedding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01):1013–1020, April 2020.

[124] Nam woo Kim and Yoonjin Yoon. Representation learning of urban regions via mobility-signature-based zone embedding: A case study of Seoul, South

Korea. In *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, pages 1–4, 2021.

[125] Arch G Woodside and Chris Dubelaar. A general theory of tourism consumption systems: A conceptual framework and an empirical exploration. *Journal of travel research*, 41(2):120–132, 2002.

[126] Arch G Woodside and Robert I King. An updated model of travel and tourism purchase-consumption systems. *Journal of Travel & Tourism Marketing*, 10(1):3–27, 2001.

[127] Min Xie, Hongzhi Yin, Hao Wang, Fanjiang Xu, Weitong Chen, and Sen Wang. Learning graph-based poi embedding for location-based recommendation. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*, pages 15–24, 2016.

[128] Philippe Xu, Franck Davoine, Jean-Baptiste Bordes, Huijing Zhao, and Thierry Denœux. Multimodal information fusion for urban scene understanding. *Machine Vision and Applications*, 27(3):331–349, 2016.

[129] Hao Xue, Flora Salim, Yongli Ren, and Nuria Oliver. Mobtcast: Leveraging auxiliary trajectory forecasting for human mobility prediction. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 30380–30391, 2021.

[130] Takahiro Yabe, Kota Tsubouchi, Toru Shimizu, Yoshihide Sekimoto, and Satish V. Ukkusuri. City2City: Translating Place Representations across Cities. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 412–415, 2019.

[131] Bo Yan, Krzysztof Janowicz, Gengchen Mai, and Song Gao. From ITDL to Place2Vec: Reasoning About Place Type Similarity and Relatedness by Learning Embeddings From Augmented Spatial Contexts. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–10, November 2017.

[132] Song Yang, Jiamou Liu, and Kaiqi Zhao. Getnext: Trajectory flow map enhanced transformer for next poi recommendation. In *Proceedings of the ACM International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 1144–1153, 2022.

[133] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, and Zhenhui Li. Deep multi-view spatial-temporal network for taxi demand prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[134] Zijun Yao, Yanjie Fu, Bin Liu, Wangsu Hu, and Hui Xiong. Representing urban functions through zone embedding with human mobility patterns. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 3919–3925, July 2018.

[135] Zijun Yao, Yanjie Fu, Bin Liu, Wangsu Hu, and Hui Xiong. Representing urban functions through zone embedding with human mobility patterns. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligenc*, pages 3919–3925. International Joint Conferences on Artificial Intelligence Organization, 7 2018.

[136] Jiangye Yuan. Automatic building extraction in aerial scenes using convolutional networks. *arXiv preprint arXiv:1602.06564*, 2016.

[137] Quan Yuan, Gao Cong, and Aixin Sun. Graph-based point-of-interest recommendation with geographical and temporal influences. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*, pages 659–668, 2014.

[138] Helga Zeiher. Shaping daily life in urban environments. In *Children in the city*, pages 84–99. Routledge, 2003.

[139] Wei Zhai, Xueyin Bai, Yu Shi, Yu Han, Zhong-Ren Peng, and Chaolin Gu. Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. *Computers, Environment and Urban Systems*, 74:1–12, March 2019.

[140] Chao Zhang, Keyang Zhang, Quan Yuan, Haoruo Peng, Yu Zheng, Tim Hanratty, Shaowen Wang, and Jiawei Han. Regions, Periods, Activities: Uncovering Urban Dynamics via Cross-Modal Representation Learning. In *Proceedings of the 26th International Conference on World Wide Web*, pages 361–370, 2017.

[141] Chengkun Zhang, Liuchang Xu, Zhen Yan, and Sensen Wu. A GloVe-Based POI Type Embedding Model for Extracting and Identifying Urban Functional Regions. *ISPRS International Journal of Geo-Information*, 10(6):372, 2021.

[142] Liangpei Zhang, Lefei Zhang, and Bo Du. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and remote sensing magazine*, 4(2):22–40, 2016.

[143] Mingyang Zhang, Tong Li, Yong Li, and Pan Hui. Multi-View Joint Graph Representation Learning for Urban Region Embedding. In *Twenty-Ninth International Joint Conference on Artificial Intelligence*, volume 5, pages 4431–4437, July 2020.

[144] Mingyang Zhang, Tong li, Yong Li, and Pan Hui. Multi-view joint graph representation learning for urban region embedding. In *International Joint Conference on Artificial Intelligence*, 2020.

[145] Yunchao Zhang, Yanjie Fu, Pengyang Wang, Xiaolin Li, and Yu Zheng. Unifying Inter-region Autocorrelation and Intra-region Structures for Spatial Embedding via Collective Adversarial Learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1700–1708, July 2019.

[146] Zheng ZHANG, MAO Baohua, LIU Mingjun, CHEN Jinchuan, and GUO Jifu. Analysis of travel characteristics of elders in beijing. *Journal of Transportation Systems Engineering and Information Technology*, 7(6):11–20, 2007.

[147] J. Zhao, J. Wang, and W. Deng. Exploring bikesharing travel time and trip chain by gender and day of the week. *Transportation Research Part C: Emerging Technologies*, 58:251–264, 2015.

[148] Vincent W Zheng, Yu Zheng, Xing Xie, and Qiang Yang. Collaborative location and activity recommendations with gps history data. In *Proceedings of the World Wide Web Conference (WWW)*, pages 1029–1038, 2010.

[149] Yu Zheng. Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3):1–41, 2015.

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

[151] Yang Zhou and Yan Huang. DeepMove: Learning Place Representations through Large Scale Movement Data. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 2403–2412, 2018.

[152] Yu Zhu, Hao Li, Yikang Liao, Beidou Wang, Ziyu Guan, Haifeng Liu, and Deng Cai. What to do next: Modeling user behaviors by time-lstm. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3602–3608, 2017.

[153] G. K. Zipf. *Human behavior and the principle of least effort*. 1949.