POLITECNICO DI TORINO

Department of Control and Computer Engineering

MASTERS THESIS

Uncovering Latent Patterns In Service-Level Spatiotemporal Mobile Traffic

---

Author:
Prashant Kumar Ray
DAUIN, Politecnico di Torino

Supervisor:
Prof. Marco Fiore
Research Professor, IMDEA Networks Institute
Prof. Luca Vassio
DAUIN, Politecnico di Torino

A thesis submitted in fulfillment of requirements for the degree of Master of Science in Computer Engineering.

Declaration of Authorship

I, Prashant Kumar Ray , declare that this thesis titled, "Uncovering Latent Patterns In Service-Level Spatiotemporal Mobile Traffic" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at the Politecnico di Torino.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signature:

Date:

Abstract

Prashant Kumar Ray

Uncovering Latent Patterns In Service-Level Spatiotemporal Mobile Traffic

Personal mobile communication technologies are amongst the most successful innovations of the 21st century. The widespread adoption of mobile services has resulted in an exponential surge in mobile traffic, which effectively mirrors human behavior. Mobile devices maintain a continuous interaction with network infrastructure, and the associated geo-referenced events can be easily logged by service providers, for different purposes, including billing and resource management. This leads to the implicit potential of monitoring a significant portion of the entire population at a minimal cost which no other technology provides an equivalent coverage.

In this context, analyzing mobile traffic along space, time, and app dimensions can provide actionable insights for improving user experience, optimizing resource allocation, enhancing security, and driving business success in the mobile app and service industry. In this thesis, after performing various data preprocessing procedures, including uplink and downlink traffic aggregation, time and space aggregation as well as scaling, we utilize the Tucker decomposition method to extract latent factors, which represent patterns across the dimensions of space, time, and mobile applications. We apply the technique to real-world mobile traffic data generated by a variety of mobile applications (for example Facebook, Gmail, Skype, Uber etc.) within Paris, France, during 77 consecutive days for every 15 minutes interval.

We obtained 4 temporal factors capturing day-night mobile traffic behavior, working hours, commuting, and weekend patterns. In the dimensions of space and mobile applications, we identified 7 factors each which includes a space factor distinctly differentiating between the city center and the rest of Paris. An exploration of the obtained latent patterns in spatial, temporal, and mobile application dimensions reveals interesting interrelationships in user behaviors.

Keywords: Mobile Traffic Analysis, Pattern Recognition, Time-Series Analysis, Spatio-Temporal Analysis, Tensor Decomposition, Tucker Decomposition.

Acknowledgements

The presented work in this report is the summary of master's thesis work developed at IMDEA Networks Institute - Leganés (Madrid) Spain, to fulfil the requirements of the M.Sc. course at the Politecnico di Torino (PoliTO).

The author extends heartfelt gratitude to Professor Marco Fiore, the supervisor at IMDEA Networks Institute, for offering the invaluable opportunity to engage in this exceptional project. I must acknowledge that his guidance and instruction throughout this thesis journey have greatly enriched my understanding of mobile traffic analysis.

The author would also like to express appreciation to Orlando E. Martínez-Durive and Sachit Mishra, PhD students at IMDEA Networks, for generously offering valuable input and feedback throughout the course of this project.

The author wishes his sincere gratitude to Academic Supervisor from Politecnico di Torino Prof. Luca Vassio for his valuable advice and guidance throughout this work.

Lastly, but most importantly, the author is indebted to his family and friends for their never ending love and support....

Contents

List of Figures

Chapter 1

Introduction

1.1 Background

With the widespread adoption of smartphones, they are no longer a pure communication tool but have gradually become a multimedia interactive platform based on the mobile internet connectivity. This shift has been accompanied by a significant rise in the use of mobile services, resulting in a surge in mobile traffic. Mobile devices are continuously interacting with the network infrastructure, and the associated geo-referenced events are logged by the operators, for different purposes, including billing and resource management. The surge in the usage of mobile devices and Internet services contributes to the generation of enormous amounts of data.

The outcome of this technological success is the substantial presence of mobile subscribers within today's population. These subscribers collectively form a considerable segment, encompassing data from hundreds of thousands, or even millions, of individuals. Moreover, this data spans wide geographical areas, ranging from cities to entire nations, and extends across considerable time spans, encompassing weeks to months. In contrast, conventional data collection methods, such as census surveys, population studies, phone interviews, or volunteer recruitment, are unable to offer even remotely comparable insights into human activities.

The digital traces left by smartphone use have come to provide valuable real-time information concerning the movement, interactions, and mobile service consumption of individuals at unprecedented scales. These digital traces facilitate the study of human behaviors. Various techniques and analytical perspectives can be employed to capture many aspects of human behaviors from mobile phone data, which has resulted in various applications. Mobile traffic data finds application in a spectrum of analyses, including the study of mobility patterns [1, 2, 3, 4] and social interactions [5], explorations of transportation systems [6] estimates of static and dynamic population density [7, 8, 9], predictions of poverty [10, 11], socioeconomic inequality [12, 13] or digital divides [14], and mappings of land usage [15, 16, 17] or urban transformation [18] or pollution [19]. Additionally, the utility of mobile network data extends to assessing the impacts of natural disasters [20] or infectious disease transmission [21, 22, 23], as well as evaluating the efficacy of the corresponding containment strategies [24, 25]. These data can also enable studies aimed at understanding how the mobile network infrastructure is used, improving its management and extending its functionalities.

1.2 Motivation

The study of mobile traffic demand in existing literature falls into two primary categories. Firstly, there are works that adopt a user-centric approach, focusing on individual subscribers. These studies delve into various aspects of user behavior, such as mobility patterns, traffic generation, and mobile service usage. Secondly, there are works that adopt an operator-centric approach, where the interest lies in analyzing demand aggregated across all users within a specific area, often a cell sector or the coverage region of a base station.

In our research, our primary focus is on the classification problem. This entails the identification of concealed regular structures within the aggregated traffic generated by mobile users. While previous research efforts have proposed solutions for detecting either temporal or spatial structures in the data, limited attention has been given to the more intricate task of simultaneously examining both space and time dimensions. In this context, Furno [17] has utilized Exploratory Factor analysis (EFA) on mobile traffic data having space and time dimensions for the joint spatiotemporal classification of the aggregate demand supplied by a mobile network operator.

For our research, we have access to rich and highly detailed mobile traffic data, including traffic information at the level of individual mobile services, encompassing various cities across France. Consequently, our dataset encompasses three fundamental dimensions: space, time, and mobile services. Our objectives for this work are as follows:

1. Uncover latent patterns, often referred to as factors, within each of these dimensions.
2. Determine the optimal number of factors for each dimension.
3. Explore how these factors interplay with one another.

This research focuses on analyzing mobile traffic data to expose hidden patterns, considering the dimensions of space, time, and mobile applications. We employ the tensor decomposition technique, specifically Tucker decomposition, to identify these latent patterns within each dimension and investigate their relationships.

The analysis of mobile traffic data across spatial, temporal, and app dimensions holds the potential to provide actionable insights for enhancing user experience, optimizing resource allocation, strengthening security, and driving success in the mobile app and service industry.

The dataset used for this study encompasses mobile traffic usage data from 68 different mobile applications across 20 urban areas in France, with a spatial resolution of 100 x 100 m2. This data spans 77 consecutive days, roughly equivalent to two and a half months, from March 16, 2019, to May 31, 2019, with a temporal resolution of 15 minutes. The dataset's total size exceeds 2.3 TB and was made available as part of the NetMob23 data challenge in June 2023.

This research specifically focuses on the analysis of mobile traffic data within the city of Paris.

1.3 Thesis Structure

The remainder of this report is structured as follows:

1. **Chapter 2: Mobile Traffic Data Analysis: Evolution & Research Directions -** This chapter offers an extensive overview of prior research in the field of mobile traffic analysis, highlighting the key findings and developments.
2. **Chapter 3: Uncovering Latent Patterns in Mobile Traffic Data -** In this chapter, we delve into the methodology and approach employed for our research, detailing the techniques and tools used to analyze mobile traffic data.
3. **Chapter 4: Results & Findings -** In this chapter, we present the outcomes of our analysis, showcasing the latent patterns uncovered in mobile traffic data along space, time, and mobile services, and how these factors relate to each other.
4. **Chapter 5: Conclusion & Future Directions -** This concluding chapter summarizes the key insights from our study and discusses their broader implications. Additionally, it outlines potential avenues for future research in this field.

Chapter 2

Mobile Traffic Data Analysis: Evolution & Research Directions

2.1 Background

While mobile traffic analysis is a relatively recent research field, it has experienced rapid growth, particularly over the last decade. This field delves into the vast data generated by mobile devices, providing insights into the movement, interactions, and consumption of mobile services by individuals on an unprecedented scale. Mobile traffic provides insights into the movement, interactions, and consumption of mobile services by individuals on an unprecedented scale. Unlike traditional data collection methods such as censuses, population surveys, phone interviews, or volunteer recruitment, mobile traffic data offers a perspective on human activities that was previously unattainable.

Mobile traffic data, irrespective of where it's collected, holds a wealth of information about subscribers' lives, encompassing their activities, interests, schedules, movements, and preferences. The ability to access such extensive information on such a large scale has proven to be critically important for research across diverse fields. However, the accessibility of this rich data source also gives rise to concerns regarding potential violations of mobile customers' privacy rights. Among these concerns are issues like individual identification, tracking of movements, and monitoring of mobile traffic. As a result, regulators have been working on laws aimed at safeguarding the privacy of mobile users. For example, the European Data Protection Directive 95/46/EC [26] mandates that all mobile traffic datasets be anonymized so that no individual is identifiable, before any cross-processing can be run on the data. Moreover, Directive 2002/58/EC states that anonymized data shall be analyzed only for the time necessary to provide the intended value-added service[27].

The nature of mobile traffic datasets, facilitates large-scale research across various disciplines. However, it is not the only one, other aspects have also contributed to the success of mobile traffic analyses.

The first supporting cause behind the surge in research volume is the increasing availability of mobile traffic datasets. Mobile operators have been always monitoring mobile traffic in their networks, for troubleshooting, efficiency, and billing purposes but they were typically cautious about sharing it. However, in recent years, there has been a shift towards greater openness within the research community. This change is partly due to groundbreaking studies demonstrating the value of mobile traffic data for both academic research and operator benefits. As a result, collaborations between academic research groups and network operators, utilizing real-world mobile traffic datasets, have flourished, leading to a significant increase in research outcomes and publications.

The second factor favoring the success of the research field is the increasing quality of the datasets. On one hand, mobile operators, recognizing the potential value of mobile traffic data, are deploying progressively advanced probes in their networks. These probes enable finer measurements of subscribers' activities. On the other hand, mobile services have evolved from simple calls and texts to Cloud-based, always-on applications. This transformation results in much more frequent interactions between users (or their devices) and the network, leading to significantly higher granularity in the recorded activity samples at the operator's end. This enhanced accuracy in mobile traffic datasets allows for more intricate and comprehensive analyses, thereby drawing an even broader research community into the field.

The third key element, originating from the previously mentioned factors, is the formation of a highly active and interdisciplinary community that brings together researchers and industry stakeholders. The synergy between academia and industry has significant implications. For instance, mobile operators are now actively promoting both fundamental and applied research in mobile traffic analysis through targeted challenges. Notable examples include Orange's Data for Development (D4D) Challenges and Telecom Italia's Big Data Challenges. In these initiatives, mobile operators openly provide mobile traffic datasets and task the research community with conducting analyses to address specific societal or technical challenges.

Given that all these identified trends are currently strengthening, the future of mobile traffic analysis as a research field appears promising enough.

2.2 Research Directions

The literature on mobile traffic analysis is very heterogeneous. Structuring the relevant works in a comprehensive way is not trivial. The classification is thus organized around research subjects, each of which features multidisciplinary contributions. The global outline of the proposed hierarchy according to Diala Naboulsi (2015) [27] is shown in Fig. 2.1.

Social analyses investigate the connections between mobile traffic and a diverse range of social attributes. The primary research emphasis lies in understanding the social dynamics of

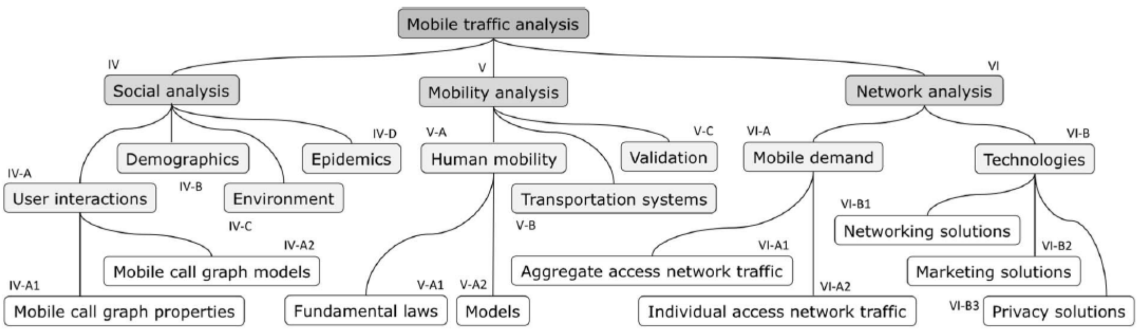


FIGURE 2.1: General classification of the mobile traffic analysis literature

mobile user interactions and exploring how demographic, economic, or environmental factors impact mobile service consumption. This category also encompasses studies that utilize social attributes inferred from mobile traffic data for the characterization and mitigation of disease epidemics.

Mobility analyses deal with the extraction of mobility-related data from mobile traffic. Here, mobility is considered in its widest context, covering general human movements at both individual and aggregate levels, as well as specialized patterns related to specific users, such as travel on transportation systems. This section also provides a comprehensive review of the literature concerning the reliability of mobile traffic data as a source of mobility information.

Network analyses adopt a technical perspective, concentrating on comprehending the dynamics of mobile traffic demand and how to enhance the mobile network infrastructure to meet these demands more effectively. Research in this category encompasses the characterization of mobile service usage patterns and leveraging this knowledge to develop enhanced technological solutions of various kinds.

We will delve into some interesting research domains related to mobile traffic analysis.

2.2.1 Epidemiology

Extensive research over the past year has investigated the impact of the pandemic on internet traffic at various network levels. Most studies have focused on internet traffic as a whole. In Central European Internet Service Providers (ISPs), traffic surged by 15% to 20% during the initial 2020 lockdown, a growth rate much higher than typical years. This increase can be attributed to government-imposed restrictions, which also caused dynamic changes in weekday traffic patterns, resembling weekend patterns [28], [29]. Similar trends were observed in major ISPs in the United States, where peak traffic rates spiked from 30% to 60% during the first quarter of 2020 [30]. Not only network operators but also online service providers experienced significant traffic shifts. For example, Facebook noticed intermittent spikes in their edge network traffic, followed by a sustained increase in load. They also reported changes in user behavior, such as a heightened interest in live streaming services [31].

Andre and Orlando [23] complements the studies mentioned earlier as it delves into the dynamics of individual services, going beyond the broad service categories that previous studies have focused on and it encompasses various containment strategies, allowing us to observe their diverse effects on mobile traffic. it also explores both spatial and temporal behaviors, offering a comprehensive perspective on the phenomenon. This provide insights into how the demands for hundreds of different mobile services are responding to the new environment created by the pandemic.

Earlier numerous studies have explored the potential correlation between patterns observed in mobile traffic data and the spread of infectious diseases. Such correlations, if identified, could offer highly effective yet cost-efficient methods for anticipating and controlling disease outbreaks. In a seminal work, Wesolowski et al. [32] investigated networks of mobile user movements alongside maps of malaria prevalence in Kenya. Their objective was to uncover associations between common patterns of human mobility and parasite infection. The authors successfully identified several routes of infection transmission that contribute to the spread of malaria across different regions of Kenya. A similar approach was adopted by Enns et al. [33] and Gavri ́c et al. [34], both of whom compared mobility and communication networks derived from mobile traffic data with maps depicting the prevalence of diseases, specifically malaria and HIV, respectively.

2.2.2 Urbanization And Land Use

Living in an urban or rural environment yields sociological differences that reflect on mobile traffic. Eagle et al.[35] conducted a comprehensive analysis using four years of mobile traffic data spanning an entire country to study the differences emerging between urban and rural users. Their findings reveal that subscribers in urban areas engage in 50% more communication as compared to those in rural areas. However, rural users, on average, engage in longer conversations among them than with individuals living in cities. Schmitt et al. [36] complement these results by demonstrating a level of segregation between urban and rural regions. Users in rural areas tend to communicate more frequently with each other than with individuals living in urban centers. Eagle et al. [37] also illustrate that the call volume of individuals within urban areas increases, while the call volume directed towards rural regions decreases.

Furthermore, the geographical distribution of mobile traffic hotspots, characterized as high-activity locations, is influenced by land use. Trestian et al. [38] identify daytime, noon, evening, and night time hotspots within a metropolitan region and correlate them with the geographical characteristics of their respective areas. Similarly, Vieira et al. [39] reveal that base stations in downtown areas experience high loads during weekday mornings, while those in commercial and business districts become hotspots for the remainder of weekdays. During weekends, hotspots emerge around commercial and business centers in the morning and afternoon and shift to commercial and nightlife areas in the evening and at night.

The discrepancy in the spatial distribution of mobile traffic between weekdays and weekends has also been documented in other studies. Pulselli et al. [40] employ geographical representations of aggregate daily demand in Milan, Italy, and observe that activity tends to concentrate in the city center on weekdays and in peripheral residential areas on weekends."

2.2.3 Commuting Patterns And Transportation Systems Planning

Mobile data has recently become a valuable source for understanding commuter mobility patterns. For instance, Furletti et al. [41] successfully distinguished commuters from other user categories, such as residents and tourists, in Pisa, Italy. Scepanovic et al. [42] ranked 76 regions in Ivory Coast based on their significance in the country's commuting processes, while Liu et al. [43] identified commuting activity sequences as the primary driver of mobility in the same region. Kung et al. [3] conducted a comprehensive analysis of commuting behaviors across various regions and geographical scales, revealing unique commute time characteristics in different areas.

In the context of improving transportation systems, Berlingerio et al. [44] extracted thirty common mobility patterns from mobile call data in the city of Abidjan. These patterns, primarily related to home-work commuting, were utilized to plan enhancements to the existing public bus transit network. Their findings demonstrated that the addition of 4 new routes could lead to a 10% reduction in overall travel times. Cici et al. [45] leveraged mobile traffic data to investigate the potential for car sharing in Madrid, Spain. Their results indicated that a reduction in the number of cars, by as much as 67%, could be achieved if drivers shared their cars and agreed to detours of no more than 600 meters in their routes. Lastly, Zhang et al. [46] identified under-served routes in Shenzhen, PRC, by comparing trajectories derived from mobile traffic with public transport flows. They proposed a new system of bus lines that could reduce commuter travel times along these routes by approximately 25% on typical days.

As we conclude this chapter, we have explored some of the dynamic evolution and research directions in the field of mobile traffic analysis. The rapid growth of this field has been fueled by the availability of rich mobile traffic datasets, advancements in data quality, and the synergy between academia and industry stakeholders. The diverse research directions, from epidemiology to urbanization and transportation system planning, highlight the potential of mobile traffic analysis in addressing complex societal and technical challenges.

In the next chapter, we delve into the heart of our research methodology. We employ tensor decomposition technique, specifically, Tucker decomposition, to uncover latent patterns in mobile traffic data along spatial, temporal, and mobile service dimensions. We aim to shed light on the intricate interplay between these dimensions, providing valuable insights into mobile traffic behavior.

Chapter 3

Uncovering Latent Patterns in Mobile Traffic Data

3.1 Dataset

The mobile traffic data utilized in this work encompasses usage information from 68 distinct mobile applications across 20 urban areas (cities) in France. For each city, the data is organized into squares with spatial resolution of 100 x 100 square meters with time resolution of 15 minutes for 77 consecutive days. The dataset spans a time frame equivalent to approximately two and a half months, specifically from March 16, 2019, to May 31, 2019. The mobile traffic data value is normalized (due to privacy reasons) such that it captures the proportion of data traffic volumes used by different apps and does not reflect the original traffic data volume. The overall size of the dataset is more than 2.3 terabytes.

This study involves the analysis of mobile traffic data specific to Paris. The traffic map for an app in a city at a particular time instance can be represented using a 2-D matrix. The dimension of this matrix is determined by the location of the extreme square of the city in each direction (east , west, north and south). For example, Paris city traffic map can be represented using a 2-D matrix of size (409 x 346).

To identify the squares that are inside or outside the city, another 2-D matrix of same size is used as a city mask having entries either 0 or 1, where 1 specifies that the square is part of the city and 0 for the squares which are not part of the city. In Fig.3.1, the squares in green are the ones which are inside Paris and all the squares in the white are ones that are outside Paris.

3.2 Statistical Concepts

3.2.1 Median

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. The 3 most common measures of central tendency are the mean, median and mode.

The median of a dataset is the value which falls exactly in the middle when the data is sorted. The median is the best choice when the data has outliers as median is less affected by outliers

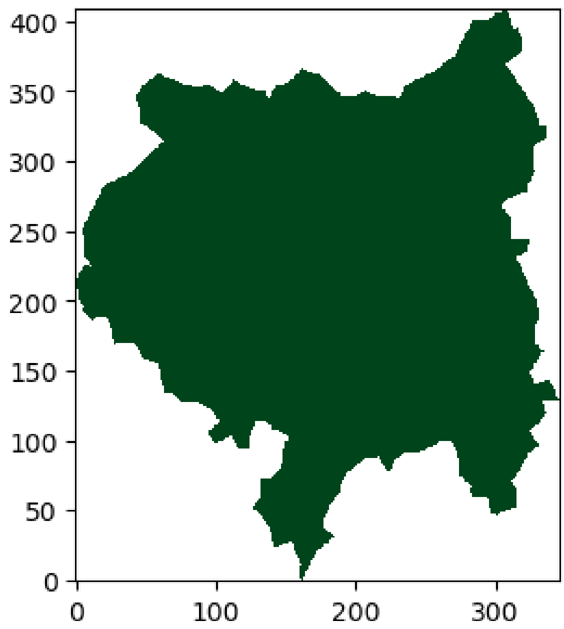


FIGURE 3.1: Paris City Mask

and skewed data. Suppose we have a dataset with n numeric elements. The steps for finding median is as following:

1. Sort the data in either increasing or decreasing order
2. Check if number of element n is even or odd.
3. If n is odd, the median is the value that lies at position $(n+1)/2$ in the sorted dataset.
4. If n is even, the median can be calculated by averaging the two values that lies at position $n/2$ and $((n/2)+1)$in the sorted dataset.

3.2.2 Revealed Comparative Advantage (RCA)

Comparative analysis involves the methodical examination of entities side by side to highlight their distinctions and commonalities, often leading to valuable insights or conclusions. Revealed Comparative Advantage (RCA) extends the concept of comparative advantage and is commonly used as an index in international economics for calculating the relative advantage or disadvantage of a certain country in a certain class of goods or services as evidenced by trade flows.

Extending the concept of RCA to the domain of mobile traffic analysis provides a novel lens through which to examine patterns of mobile service or application consumption. By calculating RCA values based on actual mobile data consumption, we gain insights into the comparative strengths of different mobile services usage patterns across regions.

Mathematically, the RCA for a given application i in a location j can be expressed as:

$$RCA = (T_{ij}/T_j)/(T_i/T) \tag{3.1}$$

where
T_{ij} = mobile traffic generated by app i in location j
T_j = total traffic generated by all apps in location j
T_i = total traffic generated by app i in all locations
T = total traffic generated by all apps in all locations

RCA takes a value between 0 and $+\infty$. A app is said to have a revealed comparative advantage in if the RCA value exceeds 1.

3.2.3 Symmetric Revealed Comparative Advantage

Symmetric Revealed Comparative Advantage (SRCA) is an extension of RCA which generates scores ranging between -1 and +1 and is symmetric around zero. Locations with SRCA scores close to $+1$ have a higher revealed comparative advantage, and locations with scores close to -1 have a lower one. The SRCA score is a function of the RCA and is defined as:

$$SRCA = (RCA - 1)/(RCA + 1) \tag{3.2}$$

3.2.4 Tensor Decomposition

Tensors are multidimensional arrays of numerical values and therefore generalize matrices to multiple dimensions. In the simplest high-dimensional case, such a tensor would be a three-dimensional array, which can be thought of as a data cube. Tensor decomposition operators are of great utility and are used for various purposes such as dimensionality reduction, noise elimination, identification of latent factors, pattern discovery, ranking, recommendation or data completion. They are applied in a wide range of applications, including genomics, analysis of health records, graph mining and identification and evolution of communities in social networks.[47]

The Tucker decomposition is a form of higher-order PCA. It decomposes a tensor into a core tensor multiplied by a matrix along each mode as shown in fig 3.2. Elementwise, the Tucker decomposition can be expressed as shown in equation 3.3.

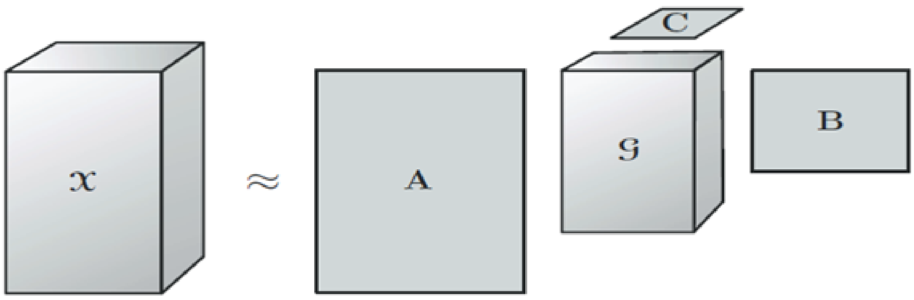


FIGURE 3.2: Tucker decomposition of a three-way array.

$$x_{ijk} \approx \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} a_{ip} b_{jq} c_{kr} \quad \text{for } i = 1, \ldots, I, \; j = 1, \ldots, J, \; k = 1, \ldots, K. \tag{3.3}$$

Here P, Q, and R are the number of components (i.e., columns) in the factor matrices A, B, and C, respectively. If P, Q, R are smaller than I, J, K then the core tensor G can be thought of as a compressed version of X. [48]

3.3 Methodology

3.3.1 Total Traffic Volume Calculation

To begin, the mobile traffic records pertaining to the city of Paris are initially filtered, and records associated with other cities are discarded. The dataset includes traffic volume data for both uplink and downlink traffic generated by each app in every 15-minute intervals for each square within the city. The total traffic volume for an app is computed by summing the uplink and downlink traffic volumes.

3.3.2 Mobile Traffic Aggregation Over Time

As described earlier, the dataset has temporal resolution of 15 minutes for every mobile application in a given square. This time resolution is too detailed for the analysis. When analyzing raw, high-frequency data, isolated outlier events like traffic jams might disproportionately influence the overall patterns and trends. However, by aggregating data over larger time intervals, the effect of these outliers is diluted. For instance, instead of analyzing traffic data every 15 minutes, aggregating the data into hourly intervals or longer can help smooth out the impact of temporary disruptions caused by traffic jams. The resulting aggregated data provides a clearer picture of typical traffic patterns and usage trends, making the analysis more robust and resilient to short-term irregularities.

For this work, the mobile traffic is aggregated for every application for a given square for every 30 minutes. This reduces the time dimension of the data by half and now the time dimension for a particular space and app has length 336 (2 x 24 x 7).

To illustrate, the total mobile traffic volume for Netflix within a specific square during both the 12:00 and 12:15 intervals is added, and the result is recorded as the observed traffic volume at 12:00 for Netflix in that square.

3.3.3 Median Week Mobile Traffic

A typical mobile traffic behaviour is deduced by examining the traffic patterns observed in various locations during different hours throughout a typical week. However, mobile traffic experiences significant impacts due to a wide range of events, each uniquely shaping the communication and data usage patterns among mobile users. For instance, the occurrence of events like football matches or concerts in a specific area often leads to a noticeable surge in observed traffic during those periods, which can be considered outlier behavior.

Employing the mean as a measure of central tendency could be sensitive to these abrupt spikes or extreme values, potentially introducing bias into estimates of typical mobile traffic levels. However, by considering the median of weekly mobile traffic, we can capture the middle value of the data, which represents the central tendency without being influenced by outliers or unusual fluctuations. The median provides a more stable and representative measure, allowing us to identify the typical level of mobile traffic during different hours across a typical week, mitigating the impact of daily or hourly variations.

The median weekly traffic for a specific city 'C', mobile application 'A', and location 'S' can be calculated as follows:

1. Load the time aggregated data in the spark dataframe and filter the records for City C.
2. Identify the day name (e.g., Monday) 'D' and time 'T' for the timestamps and save them into two different columns in the spark dataframe.
3. Group the mobile traffic for city C, day D, time T, and mobile application A within location S.
4. Calculate the median value of the grouped mobile traffic. This value indicates the typical mobile traffic pattern of application A on day D at time T in location S over a typical week.
5. Save the median week traffic data for every app , time and space for the city.

3.3.4 Mobile Traffic Aggregation Over Space

The initial spatial representation of mobile traffic in the dataset consists of individual squares, each measuring 100 x 100 square meters. There are total of 81731 such squares that constitutes Paris city. Similar to time aggregation, the IRIS level mobile traffic aggregation is performed for the Paris. IRIS represents the fundamental unit that respects certain geographic and demographic criteria and have borders which are clearly identifiable and stable in the long term[49]. By aggregating the mobile traffic per IRIS, the space dimension per app and time is reduced from 81731 to 2800.

Square and IRIS are two different geographical representation for space in France. There can be different representation possible for measuring space for a given geographical region. For example, space can be represented in terms of (latitude, longitude) pair or can be represented in terms of meters. Each representation is identified by a unique CRS (Coordinate Reference System) code. A Coordinate Reference System (CRS) refers to the way in which spatial data that represent the earth's surface, are flattened so that one can "Draw" them on a 2-dimensional surface.

An IRIS can have multiple squares of 100 x 100 square meters. Some squares falls completely inside the IRIS while some squares on the IRIS boundary may be shared among multiple IRISes. To compute total mobile traffic volume generated by an app at a particular time in an IRIS, it is required to first compute the list of squares that falls completely or partially inside the IRIS and for each squares what is the portion of area of the square that falls inside the IRIS.

To illustrate this, let's assume that a square falls on the boundary and is shared between two IRISes 'A' and 'B'. The 70% of area of the square falls inside IRIS 'A' and rest 30% area of

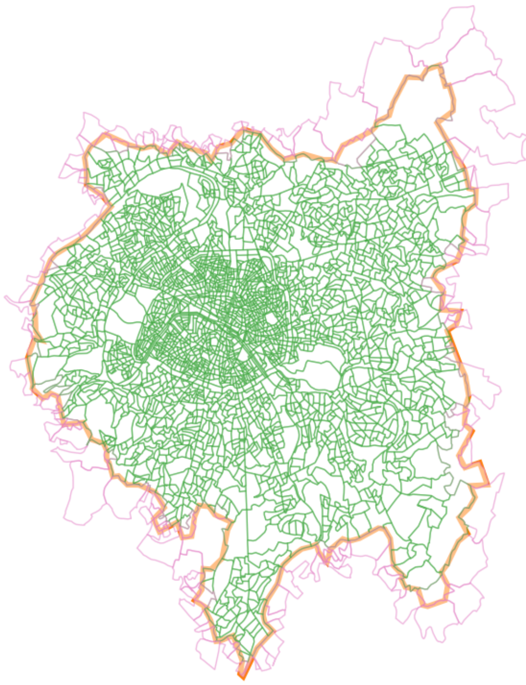


FIGURE 3.3: Paris IRIS Map

the square falls inside IRIS 'B'. The traffic generated by Uber application in the square on Monday 08:00 AM is 'T'. While calculating the total traffic generated by Uber for IRIS 'A' on Monday 08:00 AM, (0.7 x T) amount of traffic volume would be considered for IRIS 'A' and (0.3 x T) amount of traffic volume will be part of total Uber traffic in IRIS 'B'. Therefore, the total traffic volume generated by an app at a particular time in an IRIS is calculated by adding all the traffic shares of the squares that falls inside (fully/partially) the IRIS.

As mentioned earlier, in order to aggregate mobile traffic at IRIS level, it is required to calculate the list of squares along with their portion of area that falls inside the IRIS. This can be calculated as following:

1. Import the shapefile containing IRIS representations for France.
2. Load the shapefile outlining the city's (Paris) boundaries, ensuring it uses the same CRS as in step 1.
3. Determine the intersection of both shapes, yielding the proportional overlap between the two.
4. Exclude IRIS units with a ratio of 0, as they pertain to areas beyond the city limits.
5. Compute the spatial representation for each IRIS located within the city, measuring in meters.
6. Derive bounds (e.g., 'minimum x,' 'maximum x,' 'minimum y,' 'maximum y') in meter-based terms for a specific IRIS. With the knowledge of square size (100 x 100 square meters), deduce shapes measuring 100 x 100 square meters that fit within the boundaries of the given IRIS.

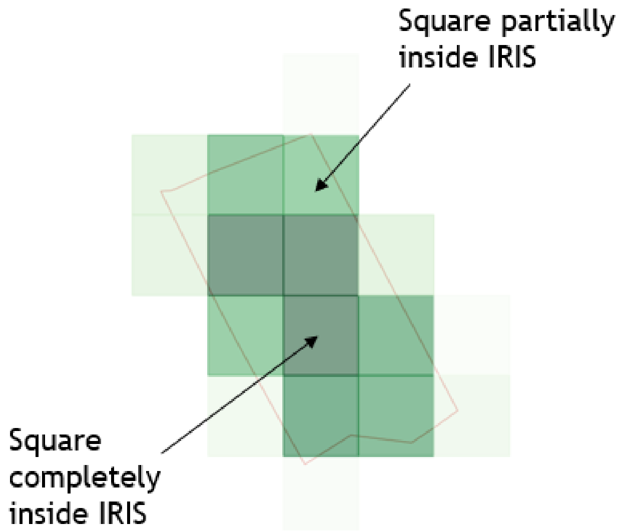


FIGURE 3.4: Example of Square and IRIS intersection

7. Determine the intersection of the shapes obtained in steps 5 and 6. This yields a compilation of squares within a particular IRIS along with their corresponding overlap ratios.

3.3.5 Symmetric RCA Calculation For Mobile Traffic

Revealed Comparative Advantage (RCA) highlights the comparative aspect of different app usage and serves as an indicator of how various mobile apps are utilized within a geographical location and time frame. However, when an app's generated traffic is minimal or even zero compared to the total traffic generated by all apps across the city during a time frame, the RCA value can become ∞ because of the denominator of the RCA formula. As a result, RCA mobile traffic values spans from 0 to ∞. Symmetric Revealed Comparative Advantage (SRCA), derived from RCA, offers the benefit of constraining the values within the range of -1 to 1. This adjustment maintains the comparative app usage information while confining the values to a more interpretable and standardized scale.

The SRCA traffic can be calculated as following:

1. Load the median week mobile traffic data which is described in section 3.3.3.
2. Calculate the total mobile traffic generated by a specific app across the entire city, as well as the cumulative mobile traffic generated by all apps citywide, for each day of the week and time interval. These values are essential for computing the RCA mobile traffic for individual apps within a particular geographical location (IRIS), on a specific day of the week, and during a particular time interval.
3. Calculate the cumulative mobile traffic generated by all apps within an IRIS for each day of the week and time.
4. Compute the RCA value for each app, across every day of the week and time, for every IRIS as described in eq. 3.1. Subsequently, calculate the SRCA value using eq. 3.2.
5. Save the SRCA values of mobile traffic for every IRIS, app, day of the week and time. An interesting observation emerged during the experiment that the SRCA values are

saved as 'float32' values by default. However, by saving these values as 'float16' results is saving considerable amount of memory and it is fast to load as well.

Fig. 3.4 shows the distributions of SRCA values for mobile traffic generated by individual apps over the course of a typical week. To streamline the analysis, we excluded less popular mobile services based on their SRCA distribution. Additionally, several Apple-related services were omitted due to their abundance. In summary, the following mobile services were discarded: Apple App Store, Apple Siri, Apple Web Services, Apple iCloud, Fortnite, Microsoft Skydrive, Microsoft Store, Microsoft Web Services, Molotov TV, Tor.

Following this step, we compiled the data into a tensor structure with space, app, and time as its dimensions. The tensor dimension for Paris is represented as $(2800 \times 58 \times 336)$.

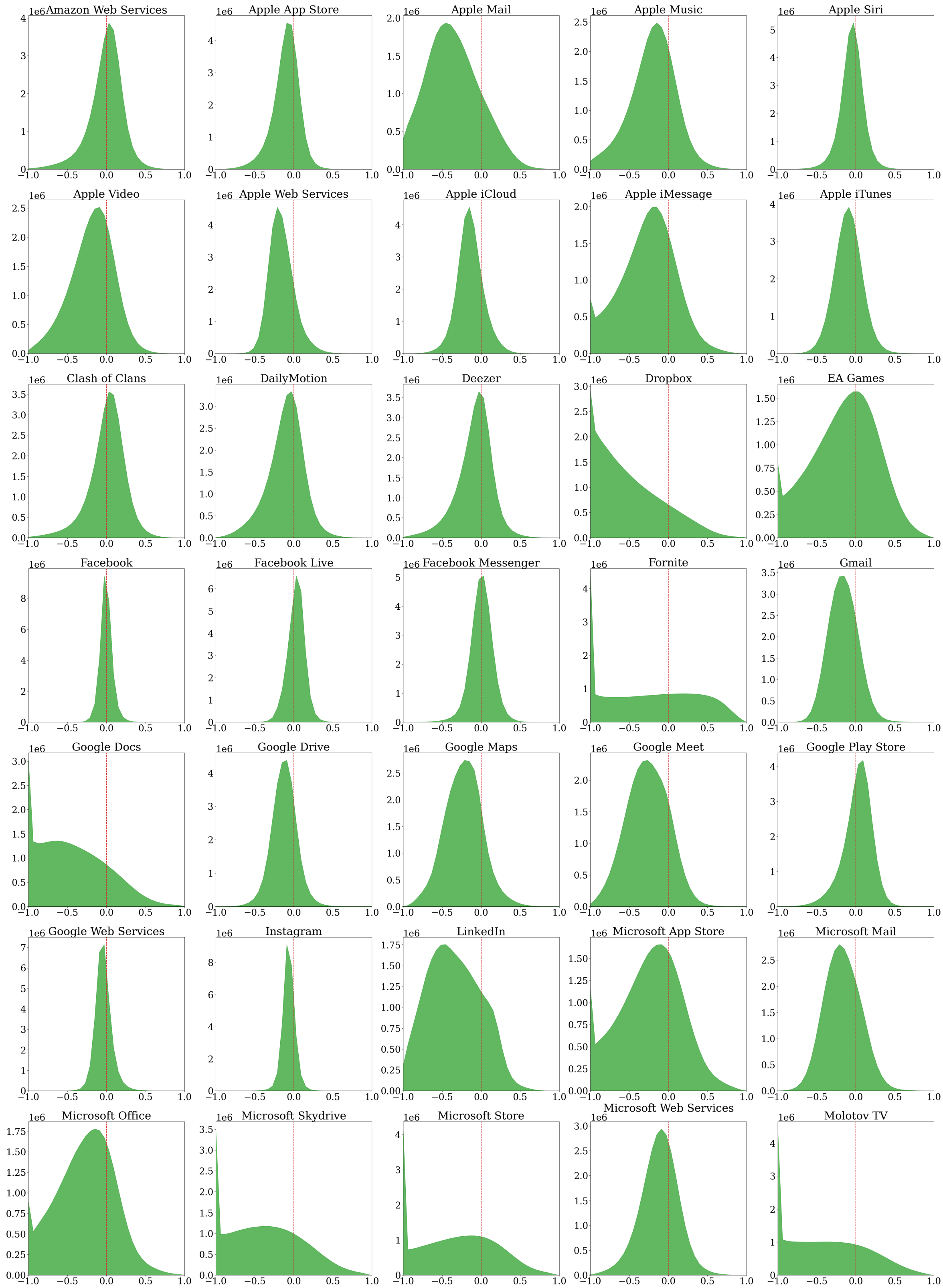
3.3.6 Tucker Decomposition

For performing Tucker decomposition, we have utilized TensorLy library. TensorLy is an open-source Python library designed for tensor algebra and decomposition. Tensors, in the context of mathematics and data analysis, extend the concepts of scalars, vectors, and matrices to higher dimensions, making them suitable for representing complex multi-dimensional data structures. TensorLy provides a collection of tools and functions for working with tensors, including tensor operations, decomposition methods, and tensor algebra. To enhance computational performance, we leveraged GPUs for these matrix operations.

To perform the Tucker decomposition, the library requires two essential inputs: the tensor and a list specifying the desired rank along each dimension. In practical terms, this rank represents the number of factors or prevalent patterns we aim to extract along a particular dimension. For instance, if we have a tensor with dimensions $(N_1 \times N_2 \times N_3)$ representing space, mobile services, and time respectively, and our objective is to uncover R_s most prominent patterns in space, R_a most prominent patterns in mobile apps, and R_t prominent patterns in time, we supply a list $[R_s, R_a, R_t]$ as input to define the desired ranks. This, along with the tensor, is then processed using the TensorLy Library. The outcome of the Tucker decomposition includes a space factor matrix with dimensions $(N_1 \times R_s)$, a mobile services factor matrix with dimensions $(N_2 \times R_a)$, a time factor matrix with dimensions $(N_3 \times R_t)$ and a core tensor with dimension $(R_s \times R_a \times R_t)$. Each of these factor matrices captures the most prevalent patterns within its respective dimension, while the core tensor establishes the relationships and interactions between these factors. Selecting the appropriate ranks, i.e., determining how many patterns we want along each dimension, is a challenge we will address in the next chapter.

3.4 Libraries & Frameworks

1. **Apache Spark** is an open-source data processing framework for large-scale data processing.
2. **TensorLy** is an open-source Python library designed for tensor algebra and decomposition.
3. **NumPy** (Numerical Python) is an open source Python library that's used in almost every field of science and engineering. It's the universal standard for working with


Amazon Web Services
Apple App Store
Apple Mail
Apple Music
Apple Siri
Apple Video
Apple Web Services
Apple iCloud
Apple iMessage
Apple iTunes
Clash of Clans
DailyMotion
Deezer
Dropbox
EA Games
Facebook
Facebook Live
Facebook Messenger
Fornite
Gmail
Google Docs
Google Drive
Google Maps
Google Meet
Google Play Store
Google Web Services
Instagram
LinkedIn
Microsoft App Store
Microsoft Mail
Microsoft Office
Microsoft Skydrive
Microsoft Store
Microsoft Web Services
Molotov TV

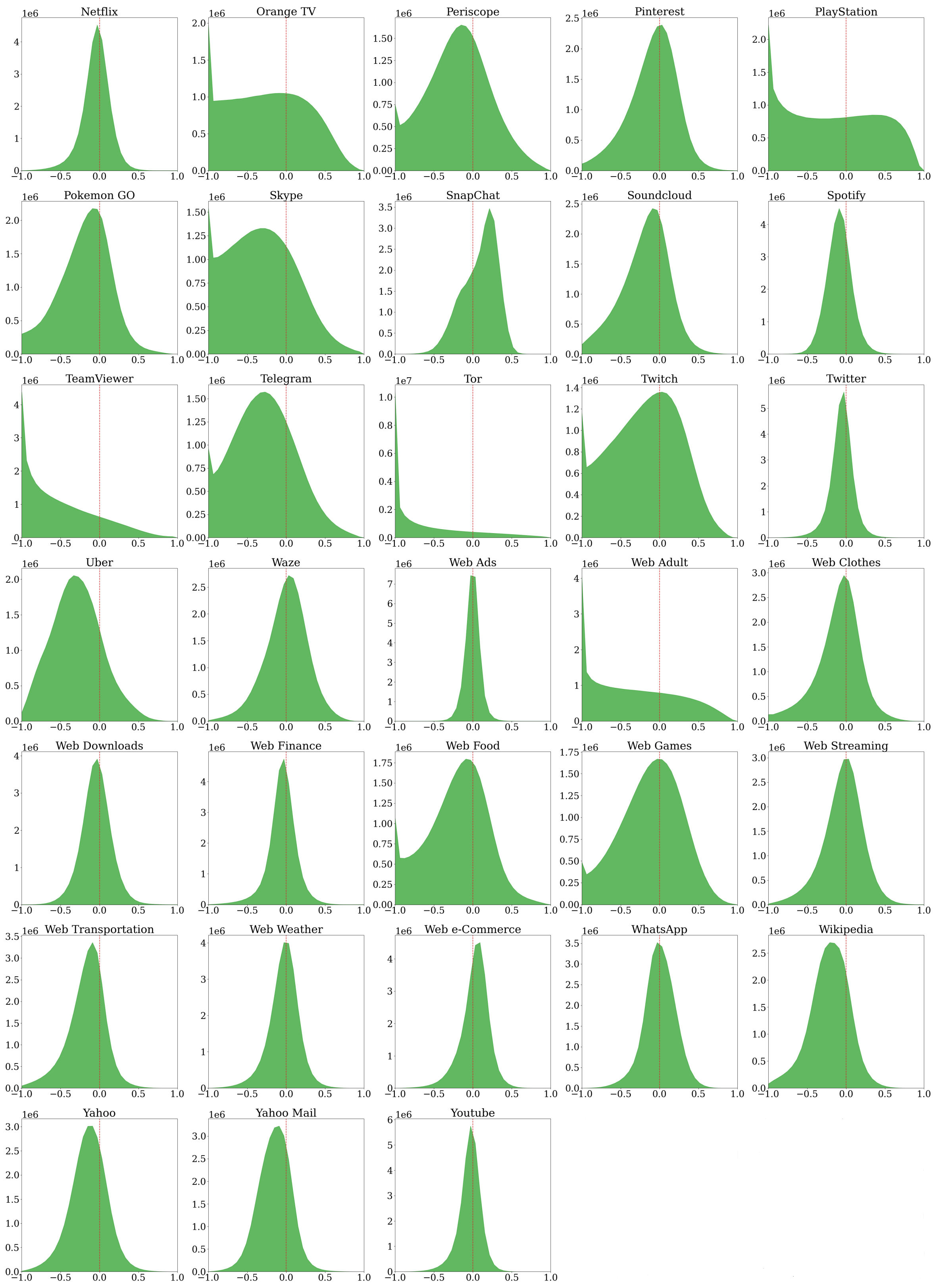


FIGURE 3.4: SRCA Distribution of Mobile Services

numerical data in Python. It provides support for large, multi-dimensional arrays and matrices, along with an extensive collection of mathematical functions to operate on these arrays.

4. **Pandas** is an open-source Python library that provides powerful data manipulation and analysis capabilities.

5. **GeoPandas** is an open-source Python library that extends the capabilities of pandas, to include geospatial data. It provides tools for working with geospatial data, such as geographical shapes, maps, and spatial attributes, making it easier to analyze, visualize, and manipulate geographic information within the Python programming ecosystem.

6. **Matplotlib** is a widely-used Python library for creating static, interactive, and animated visualizations in various formats. It provides a comprehensive set of tools for generating high-quality plots, charts, graphs, and other visual representations of data.

7. **Folium** is a Python library used for visualizing geospatial data.

In the next chapter, we will delve into the outcomes of our research, focusing on the results obtained through the application of Tucker decomposition on mobile traffic data. Specifically, we will discuss the process of selecting the optimal rank for Tucker decomposition and present the distinct space, time, and app factors that have emerged from our analysis. Furthermore, we will explore the interrelationships between these factors and the possible explanation for the related behaviors.

Chapter 4

Results & Findings

4.1 Rank Selection For Tucker Decomposition

The determination of ranks for tensor decomposition is not straight forward task, particularly at the outset, as we lack prior knowledge regarding the optimal number of factors or predominant patterns to extract via Tucker decomposition across the spatial, mobile app, and time dimensions. To measure the efficacy of this decomposition, we employed the mean square error (MSE) metric, comparing the original tensor with its reconstructed counterpart utilizing the factor matrices and core tensor. In essence, when we request an increasing number of factors, the MSE naturally diminishes. Conversely, a higher number of factors along the spatial, app, and time dimensions may lead, beyond a certain point, to the emergence of noisy patterns within the factors. Thus, it becomes imperative to strike a balance between the number of factors along each dimension and the associated MSE. Our objective hinges on achieving the minimal requisite factors along each dimension while concurrently minimizing the MSE, thereby ensuring the robustness and fidelity of our decomposition outcomes.

As previously mentioned, we lacked prior information regarding the optimal rank for the decomposition. To address this, we systematically generated a range of rank values, spanning from 3 to 12 along each dimension. This resulted in a total of 1,000 combinations, encompassing possibilities from [3,3,3] to [12,12,12]. For each of these combinations, we meticulously logged the Mean Square Error (MSE) associated with the decomposition process.

To determine the optimal rank for the space dimension, we began by filtering the Mean Square Error (MSE) values for each combination of time and app ranks, while keeping the

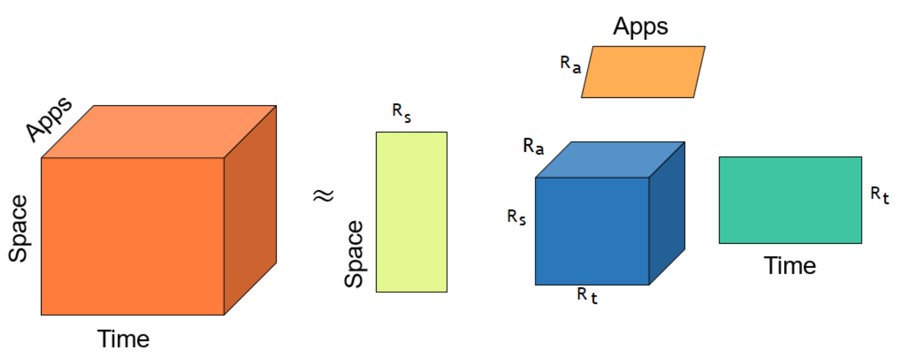


FIGURE 4.1: Tucker Decomposition

space rank constant. For each of these fixed space ranks, we calculated the median MSE and then plotted these median MSE values against their corresponding space ranks.

Our observation from the Fig. 4.2a indicated that, beyond a space rank of 7, the reduction in error became negligible with further increases in the space rank. As a result, we determined the optimal space rank to be 7. It's important to note that we chose to utilize median MSE rather than minimum MSE. This choice was made because, if we were to fix the space rank and progressively increase the factors related to time and mobile apps, the MSE would invariably decrease. Thus, median MSE provided a more reliable measure in this context.

We followed a similar method to determine the optimal ranks for both the time and app dimensions. Firstly, for the time rank, we conducted an analysis and plotted the Mean Square Error (MSE) against different time ranks. Our findings, as depicted in Fig 4.2b, reveal that the MSE does not exhibit significant changes beyond a time rank of 4. Additionally, during our practical observations, we noticed that patterns obtained beyond time factor 4 tend to lose their interpretability.

Subsequently, we applied the same approach to determine the optimal rank for the app dimension. Fig 4.2c illustrates the results of this analysis, which indicate that the optimal rank for the app dimension is 7. In summary, the selected ranks for the Tucker decomposition in the dimensions of space, app, and time are (7, 7, 4).

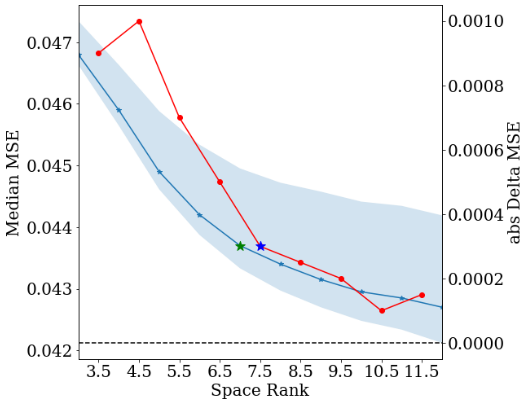
4.2 Space Factors

As mentioned previously, Paris is composed of 2800 IRIS, which corresponds to the spatial dimension in the tensor. We opted for a space rank of 7, which implies that the space factor matrix resulting from the Tucker decomposition has dimensions of 2800 x 7. Each column in this matrix represents a distinct space factor. When these factors are individually plotted against the IRIS, we obtain visual representations of the 7 most prevalent spatial patterns, as illustrated in Fig 4.3.

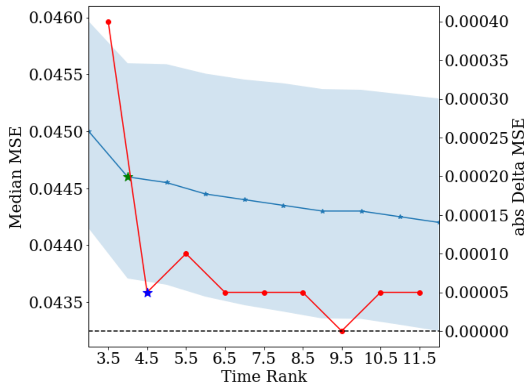
The Space factor 0 as shown in Fig.4.3a, represents the general usage profile of mobile traffic across the city of Paris. A closer examination of the actual map of Paris reveals that the areas highlighted in blue, signifying the lowest traffic volume compared to other parts of the city, correspond to parks. While Space factor 1 as shown in fig 4.3b clearly separates mobile traffic behaviour of city center from that of the rest of Paris. Space factor 2 is indicative of higher mobile traffic volume behavior in areas encompassing the airports (Paris-Charles de Gaulle (CDG), Paris-Orly Airport, and Paris Airport-Le Bourget), as well as certain portions of the city having tourist attractions. While, Space factor 3 primarily emphasizes the suburban areas in Paris that are located at a distance from the city center. Likewise, other Space factors highlight specific areas within the city that exhibit higher traffic volumes in comparison to other parts of the city.

4.3 Time Factors

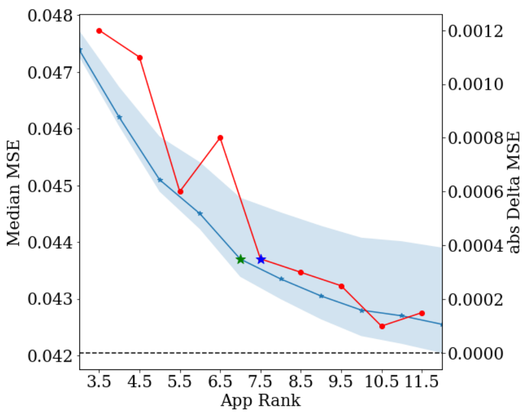
For our analysis, we aggregated time intervals into 30-minute segments over a median week. Consequently, the time dimension of our tensor is 336 ($2 \times 24 \times 7$). Since we asked for 4 time factors, the resulting factor matrix related to time has a dimension of 336×4. Each column



(A) Space Rank vs MSE



(B) Time Rank vs MSE

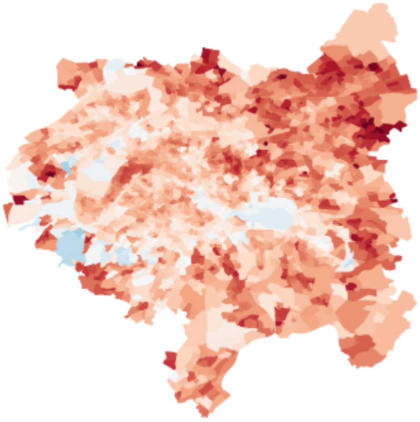


(C) App Rank vs MSE

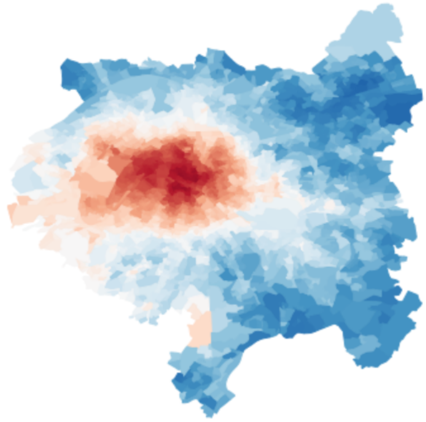
FIGURE 4.2: Rank Selection for Tucker Decomposition

in this matrix represents a unique time factor. When we reshape each column into a (48 × 7) format, corresponding to the 24-hour period over 7 days, we obtain visual representations of the 4 most prevalent temporal patterns, as illustrated in Fig4.4.

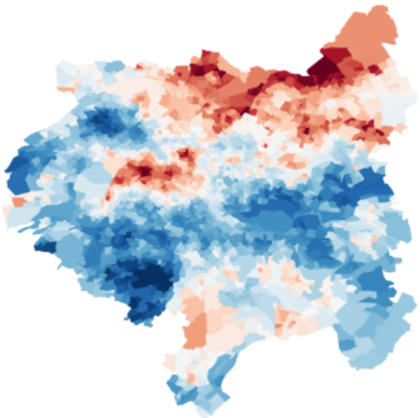
When examining Time Factor 0, as shown in Fig 4.4a, we can notice that the factor loadings exhibit distinct patterns. Notably, these loadings are prominently higher during the night, particularly from 12:30 AM to approximately 06:00 AM, with a slight extension observed during the weekends. Following this period, the loadings gradually diminish during the waking hours and subsequently remain relatively consistent throughout the remainder of the day. This time factor 0 represents the typical mobile traffic behavior, distinguishing between day and night mobile traffic patterns. For Time factor 1, as shown in Fig 4.4b, it clearly separates the sleep hours from the waking hours with high loadings during the sleep hours and very low during the awake hours. This behavior indicates that the mobile traffic during the the day is very less in comparision to the night (sleep hours). Time Factor 2, as illustrated in Fig. 4.4c, characterizes the mobile traffic behavior during early morning and working hours, effectively distinguishing these patterns from those occurring during other times of the day. While, Time Factor 3, displayed in Fig. 4.4d, highlights mobile traffic patterns associated with commuting and weekends, effectively setting them apart from patterns observed during other time periods.



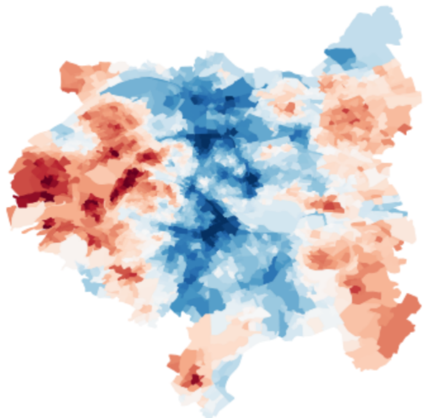
(A) Space Factor 0



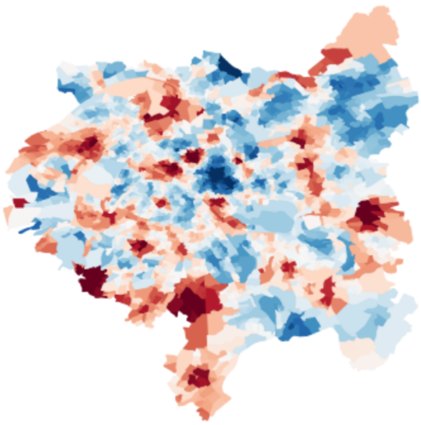
(B) Space Factor 1



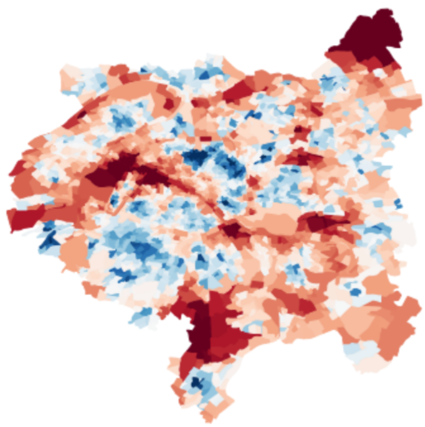
(C) Space Factor 2



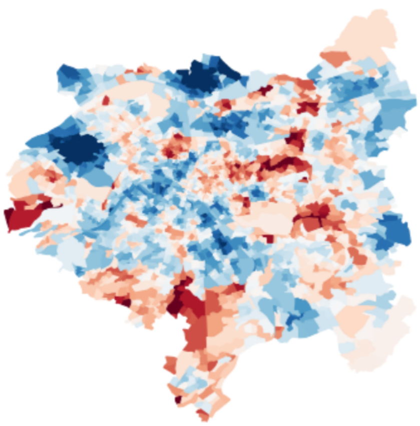
(D) Space Factor 3



(E) Space Factor 4



(F) Space Factor 5



(G) Space Factor 6

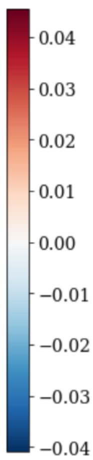
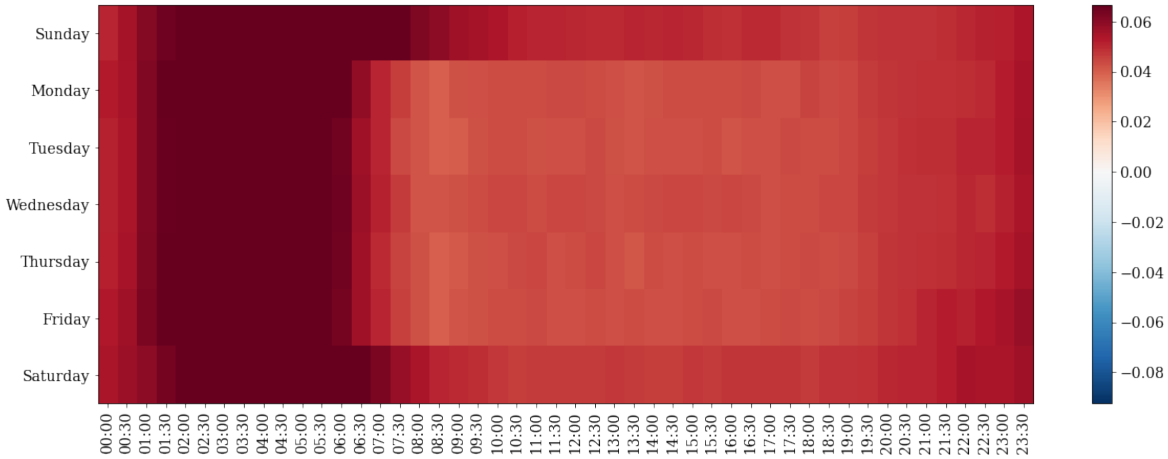
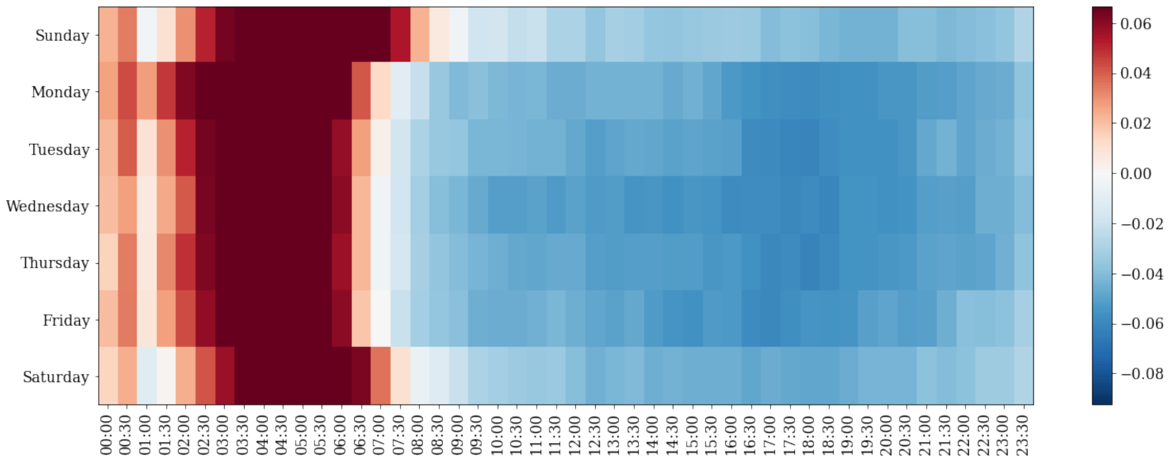


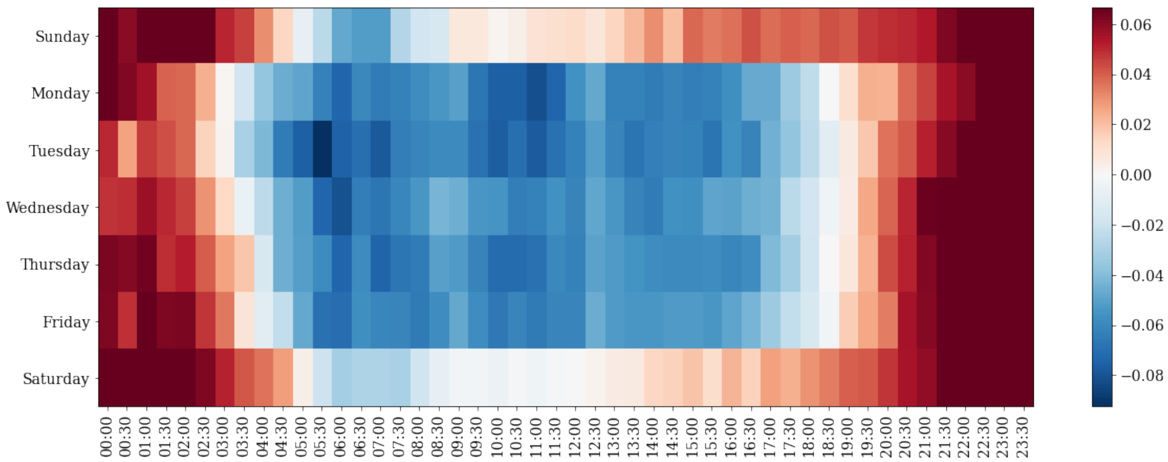
FIGURE 4.3: Space Factors



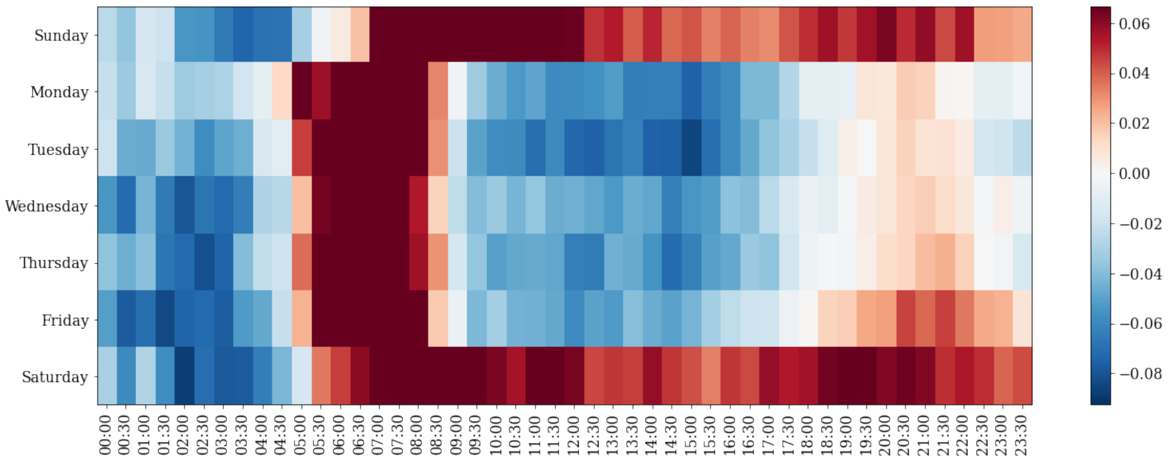
(A) Time Factor 0



(B) Time Factor 1



(C) Time Factor 2



(D) Time Factor 3

FIGURE 4.4: Time Factors

4.4 App Factors

The app factor matrix has a dimension of 58 x 7, as we sought to identify 7 distinct patterns along the app dimension for the 58 mobile services. As initially the apps were arranged in the tensor according to the name starting from A to Z, the obtained factor matrix after Tucker decomposition is shown in fig 4.5.

We employ hierarchical clustering, grouping apps with similar loading. The resulting app factor matrix is illustrated in Fig. 4.6, facilitating a more comprehensive visualization of the outcomes. This clustering approach effectively groups similar apps together; for instance, streaming apps like Web Streaming, Dailymotion, Netflix, and Youtube are clustered together. Similarly, social media apps such as Instagram, Facebook, Facebook Live, and Facebook Messenger are grouped together, as are Telegram, Whatsapp, and Snapchat.

To determine the relationship between an app and its corresponding app factor, we consider two key elements: the sign of the loading, which indicates the direction of the relationship, and the magnitude of the loading, which signifies the strength of the relationship.

To determine the app factor most closely related to a specific app, we examine the factor

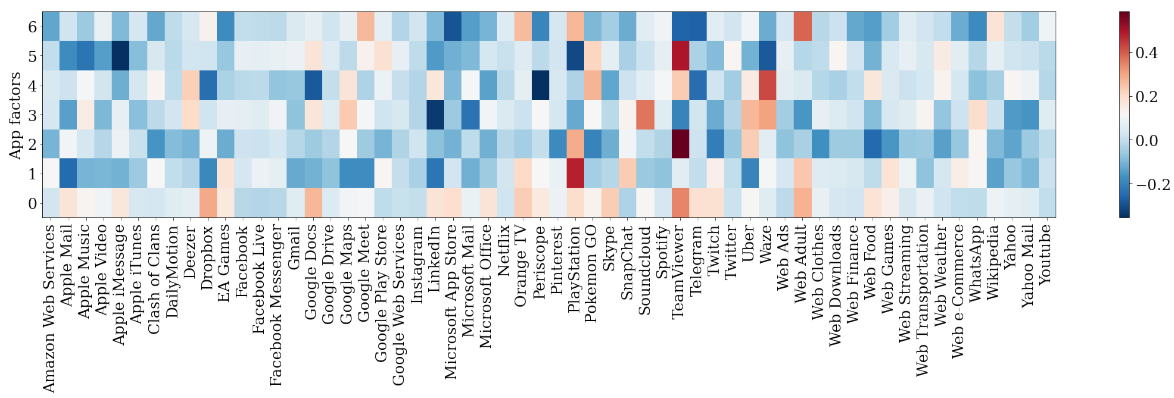


FIGURE 4.5: App Factors

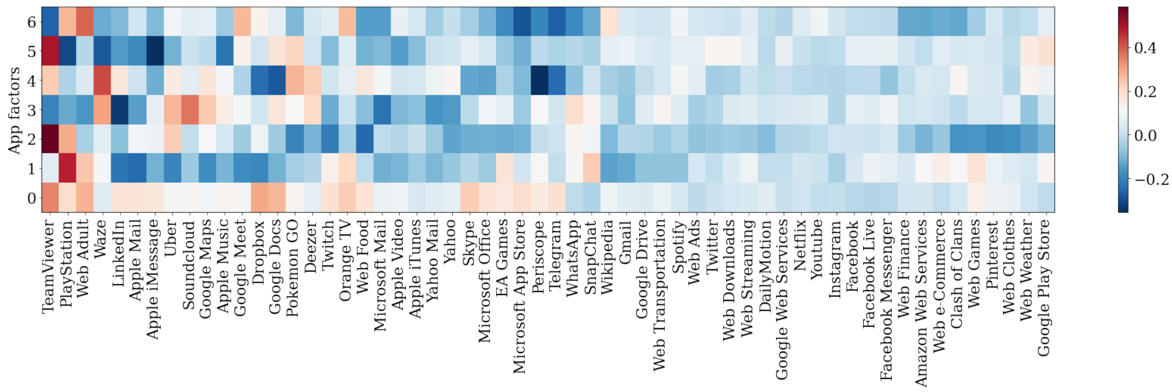


FIGURE 4.6: Clustered App Factors

with the highest loading for that app. For example, Uber demonstrates its most pronounced positive correlation with app factor 3. This suggests that users who frequently use the Uber app also tend to use other apps or engage in activities associated with App Factor 3. Notably, Google Maps is also predominantly associated with app factor 3. This alignment between Uber and Google Maps aligns with real-world usage, as Uber drivers often rely on Google Maps for navigation when picking up and dropping off passengers.

Conversely, the high-magnitude negative loading of LinkedIn for App Factor 3 suggests an inverse or negative relationship between LinkedIn usage and the apps or characteristics represented by this factor. Users who use LinkedIn are less likely to use apps or engage in activities associated with App Factor 3, and vice versa. This could imply that LinkedIn serves a different purpose or targets a different audience compared to the apps within App Factor 3.

4.5 Core Tensor and Factor Relationships

The factor matrices reveal the most common patterns within mobile traffic data concerning space, apps, and time. Meanwhile, the core tensor helps us understand how these patterns are connected and how they influence each other. For instance, if we wish to examine how apps correspond to a specific app factor in both space and time, we can create a slice from the core tensor by keeping the app factor constant. This slice would form a matrix that illustrates the connection between space factors and time factors for that particular app factor. Fig 4.7 shows the relationship matrices between the space factors and time factors across different app factors.

To illustrate the relationship with an example, lets consider the mobile app Uber. From the fig 4.6, we can see that Uber is more related to app factor 3. To find out how the mobile traffic profile typically seen for Uber (i.e. app factor 3) in space and time, we can find this relationship in core tensor. For that, we can take the slice of the core tensor by fixing app factor equals to 3, we get a matrix having relation between space factors and time factors for app factor 3 (as shown in fig 4.7 for App Factor 3).

Now, the importance of loadings in the core tensor can be assessed based on their magnitude and sign. The magnitude of the loadings reflects the strength of the relationship or interaction

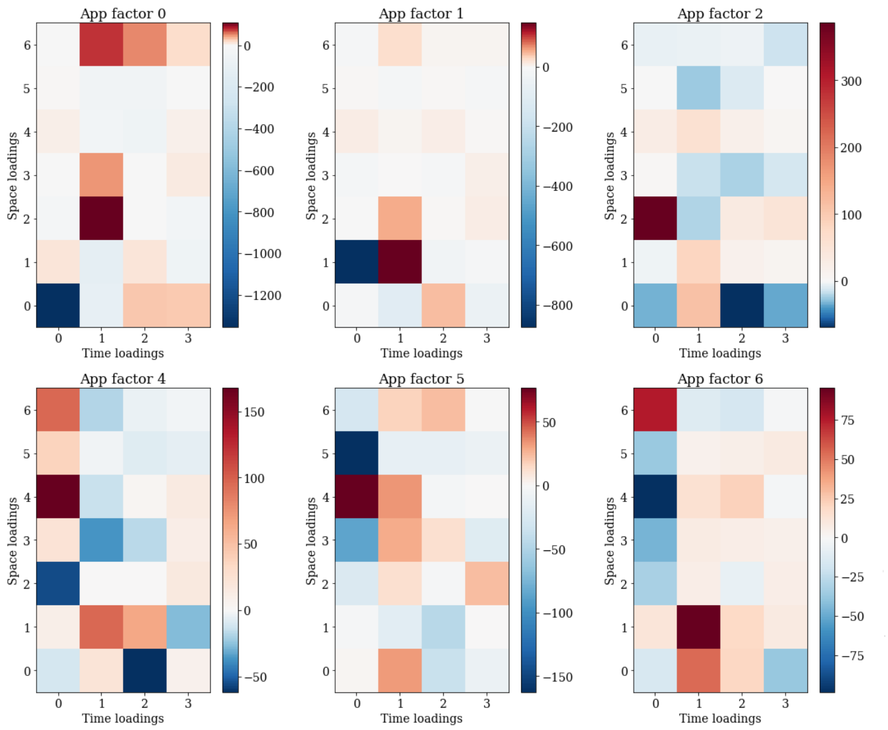


FIGURE 4.7: Space Factors vs. Time Factors for Different App Factors

between the corresponding factors. The sign of the loadings indicates the direction of the relationship.

If we observe the core tensor slice for app factor 3, the loading associated with the space factor 3 and Time factor 0 has the largest magnitude with negative sign. It implies that the space factor 3 has inverse relation with time factor 0 for Uber (App factor 3). In simpler terms, it suggests that for Uber (App factor 3), space factor 3 is linked to Time factor 0 in such a way that during late-night to early-morning hours, Uber experiences reduced demand in areas characterized by high loadings for space factor 3 (highlighted in red). This suggests that users in these regions tend to request fewer Uber rides during these specific time periods compared to other locations or times of the day. One possible explanation for this behavior could be related to the fact that space factor 3 highlights suburban parts of Paris, where there may be fewer late-night entertainment options like bars, clubs, and restaurants compared to urban areas. Consequently, this could result in reduced late-night travel for leisure purposes.

From Fig. 4.6, we can also see that Uber has a significant relationship with app factor 2. Following the same procedure as previously explained, we can extract a slice from the core tensor by fixing the app factor to 2. In Fig. 4.7, when observing the matrix representing the relationship between space factors and time factors for app factor 2, we notice that the loading at space factor 2 and time factor 0 exhibits the highest magnitude with a positive sign. This suggests that the regions highlighted in red for space factor 2 exhibit a behavior similar to that of time factor 0, indicating that Uber experiences high demand during the nighttime hours until the early morning compared to the rest of the day. As mentioned earlier, space factor 2 includes the regions near the airports of Paris. This mobile traffic behavior may be attributed to the fact that public transport in Paris is not available from midnight until early morning, roughly between 01:00 AM and 6:00 AM, which leads to increased demand for

Uber during these hours.

If we observe highly popular apps like streaming apps (Netflix, YouTube, etc.) and social media apps (Instagram, Facebook, Twitter etc.) in fig 4.6, they do not exhibit a strong relationship with any specific app factor. This behavior can likely be attributed to the widespread use of these apps across different regions and times, making them less influenced by specific patterns.

If we focus on mobile services which relates mostly to app factor 0, for example, Google Docs, Dropbox, Skype, and more, and analyze their relationship with the space and time factors in the core tensor for app factor 0, we observe that the loading associated with space factor 0 and time factor 0 exhibit the highest magnitude with a negative sign. This suggests that mobile services attributed to app factor 0 experienced reduced traffic during nighttime as opposed to daytime. This phenomenon is particularly pronounced in the regions highlighted in red within space factor 0, which covers a significant portion of the Paris city.

In summary, our comprehensive exploration of mobile traffic data using Tucker decomposition has yielded invaluable insights into the complex interplay of factors influencing mobile usage in Paris. Employing the methodology outlined earlier, one can delve deeper into understanding how these three factors are interconnected and assess the intensity of their relationships by examining the loading within the core tensor.

Chapter 5

Conclusion & Future Directions

In this thesis, we started a comprehensive exploration of mobile traffic analysis in the context of Paris, examining its complex dimensions of space, time, and mobile applications. Leveraging the powerful Tucker decomposition technique, we aimed to uncover hidden patterns within the mobile traffic data along space, time and mobile services, revealing valuable insights into the workings of this complex ecosystem.

Our research yielded several noteworthy accomplishments. The Tucker decomposition method effectively identified latent factors within the three critical dimensions of space, time, and mobile applications. This allowed us to not only unveil these patterns but also establish connections among them through the core tensor. This capability proved invaluable, enabling us to conduct in-depth analyses of the relationships between any two factors while considering the third.

In response to research question:

1. we successfully uncovered latent patterns, often referred to as factors, within each of the three dimensions: space, time, and mobile applications. Specifically, we identified 7 latent patterns within both the space and mobile app dimensions, as well as 4 latent patterns within the time dimension.
2. we determined the optimal number of factors for each dimension. For the space and mobile app dimensions, this optimal number was found to be 7 latent patterns, while for the time dimension, it was 4 latent patterns.
3. Our research successfully explored the interplay among these latent factors using the core tensor. This comprehensive analysis allowed us to gain valuable insights into how these factors relate and influence each other within the mobile traffic data.

However, it's crucial to acknowledge the challenges and limitations we encountered during our research. One notable challenge was determining the optimal ranks for the Tucker decomposition. Without prior knowledge of these ranks, we adopted a brute-force approach, which, while effective, can be computationally demanding and time-consuming. Additionally, the Tucker decomposition method, unlike Exploratory Factor Analysis (EFA), takes all dimensions into account simultaneously. This sometimes led to the merging of multiple patterns into a single factor. For instance, we observed instances where commuting and weekend time patterns were combined within a single time factor, potentially masking specific insights.

Looking forward, there are several avenues for future research in this domain. One critical area of focus should be the development of more efficient techniques for determining the optimal rank in Tucker decomposition. Streamlining this process would enhance the method's applicability and reduce computational overhead. Moreover, it would be beneficial to explore advanced algorithms that can disentangle intricate patterns and further refine the granularity of our analyses.

In conclusion, our investigation into mobile traffic analysis has revealed a wealth of information, shedding light on the complex interplay between space, time, and mobile applications. While challenges remain, our work has laid a solid foundation for future research and has the potential to contribute significantly to our understanding of mobile traffic behavior in urban environments.

Bibliography

1. Gonzalez MC, Hidalgo CA, and Barabasi AL. Understanding individual human mobility patterns. nature 2008; 453:779–82
2. Csáji BC, Browet A, Traag VA, Delvenne JC, Huens E, Van Dooren P, Smoreda Z, and Blondel VD. Exploring the mobility of mobile phone users. Physica A: statistical mechanics and its applications 2013; 392:1459–73
3. Kung KS, Greco K, Sobolevsky S, and Ratti C. Exploring universal patterns in human home-work commuting from mobile phone data. PloS one 2014; 9:e96180
4. Louail T, Lenormand M, Picornell M, Garcia Cantu O, Herranz R, Frias-Martinez E, Ramasco JJ, and Barthelemy M. Uncovering the spatial structure of mobility networks. Nature communications 2015; 6:6007
5. Miritello G, Lara R, Cebrian M, and Moro E. Limited communication capacity unveils strategies for human interaction. Scientific reports 2013; 3:1950
6. Seppecher M, Leclercq L, Furno A, Lejri D, and Rocha TV da. Estimation of urban zonal speed dynamics from user-activity-dependent positioning data and regional paths. Transportation Research Part C: Emerging Technologies 2021; 129:103183
7. Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE, Blondel VD, and Tatem AJ. Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences 2014; 111:15888–93
8. Khodabandelou G, Gauthier V, Fiore M, and El-Yacoubi MA. Estimation of static and dynamic urban populations with mobile network metadata. IEEE Transactions on Mobile Computing 2018; 18:2034–47
9. Silva F Batista e, Freire S, Schiavina M, Rosina K, Marín-Herrera MA, Ziemba L, Craglia M, Koomen E, and Lavalle C. Uncovering temporal changes in Europe's population density patterns using a data fusion approach. Nature communications 2020; 11:4631
10. Steele JE, Sundsøy PR, Pezzulo C, Alegana VA, Bird TJ, Blumenstock J, Bjelland J, Engø-Monsen K, De Montjoye YA, Iqbal AM, et al. Mapping poverty using mobile phone and satellite data. Journal of The Royal Society Interface 2017; 14:20160690
11. Pokhriyal N and Jacques DC. Combining disparate data sources for improved poverty prediction and mapping. Proceedings of the National Academy of Sciences 2017; 114:E9783–E9792
12. Moro E, Calacci D, Dong X, and Pentland A. Mobility patterns are associated with experienced income segregation in large US cities. Nature communications 2021; 12:4633
13. Ucar I, Gramaglia M, Fiore M, Smoreda Z, and Moro E. News or social media? Socioeconomic divide of mobile service consumption. Journal of the Royal Society Interface 2021; 18:20210350

14. Mishra S, Smoreda Z, and Fiore M. Second-level digital divide: A longitudinal study of mobile traffic consumption imbalance in france. *Proceedings of the ACM Web Conference 2022*. 2022 :2532–40
15. Toole JL, Ulm M, González MC, and Bauer D. Inferring land use from mobile phone activity. *Proceedings of the ACM SIGKDD international workshop on urban computing*. 2012 :1–8
16. Lenormand M, Picornell M, Cantú-Ros OG, Louail T, Herranz R, Barthelemy M, Frías-Martínez E, San Miguel M, and Ramasco JJ. Comparing and modelling land use organization in cities. Royal Society open science 2015; 2:150449
17. Furno A, Fiore M, and Stanica R. Joint spatial and temporal classification of mobile traffic demands. *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE. 2017 :1–9
18. De Nadai M, Staiano J, Larcher R, Sebe N, Quercia D, and Lepri B. The death and life of great Italian cities: a mobile phone data perspective. *Proceedings of the 25th international conference on world wide web*. 2016 :413–23
19. Chen W, He Y, and Pan S. Impact of air pollution on human activities: Evidence from nine million mobile phone users. Plos one 2021; 16:e0251288
20. Yabe T, Jones NK, Rao PSC, Gonzalez MC, and Ukkusuri SV. Mobile phone location data for disasters: A review from natural hazards and epidemics. Computers, Environment and Urban Systems 2022; 94:101777
21. Vazquez Brust A, Olego T, Rosati G, Lang C, Bozzoli G, Weinberg D, Chuit R, Minnoni M, and Sarraute C. Detecting Areas of Potential High Prevalence of Chagas in Argentina. *Companion Proceedings of The 2019 World Wide Web Conference*. 2019 :262–71
22. Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, De Nadai M, Letouzé E, Salah AA, Benjamins R, Cattuto C, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. 2020
23. Zanella AF, Martínez-Durive OE, Mishra S, Smoreda Z, and Fiore M. Impact of later-stages COVID-19 response measures on spatiotemporal mobile service usage. *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE. 2022 :970–9
24. Heroy S, Loaiza I, Pentland A, and O'Clery N. COVID-19 policy analysis: labour structure dictates lockdown mobility behaviour. Journal of the Royal Society Interface 2021; 18:20201035
25. Pullano G, Valdano E, Scarpa N, Rubrichi S, and Colizza V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. The Lancet Digital Health 2020; 2:e638–e649
26. Protection Directive 95/46/EC. 1995. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046
27. Naboulsi D, Fiore M, Ribot S, and Stanica R. Large-scale mobile traffic analysis: a survey. IEEE Communications Surveys & Tutorials 2015; 18:124–61
28. Feldmann A, Gasser O, Lichtblau F, Pujol E, Poese I, Dietzel C, Wagner D, Wichtlhuber M, Tapiador J, Vallina-Rodriguez N, et al. The lockdown effect: Implications of the COVID-19 pandemic on internet traffic. *Proceedings of the ACM internet measurement conference*. 2020 :1–18

29. Feldmann A, Gasser O, Lichtblau F, Pujol E, Poese I, Dietzel C, Wagner D, Wichtlhuber M, Tapiador J, Vallina-Rodriguez N, et al. A year in lockdown: how the waves of COVID-19 impact internet traffic. Communications of the ACM 2021; 64:101–8
30. Liu S, Schmitt P, Bronzino F, and Feamster N. Characterizing service provider response to the covid-19 pandemic in the united states. *Passive and Active Measurement: 22nd International Conference, PAM 2021, Virtual Event, March 29–April 1, 2021, Proceedings 22*. Springer. 2021 :20–38
31. Böttger T, Ibrahim G, and Vallis B. How the Internet reacted to Covid-19: A perspective from Facebook's Edge Network. *Proceedings of the ACM Internet Measurement Conference*. 2020 :34–41
32. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, and Buckee CO. Quantifying the impact of human mobility on malaria. Science 2012; 338:267–70
33. Enns E and Amuasi J. Human mobility and communication patterns in Cote d'Ivoire: A network perspective for malaria control. NetMob D4D Challenge 2013 :1–14
34. Gavric K, Brdar S, Culibrk D, and Crnojevic V. Linking the human mobility and connectivity patterns with spatial HIV distribution. NetMob D4D Challenge 2013 :1–6
35. Eagle N, Montjoye YA de, and Bettencourt LM. Community computing: Comparisons between rural and urban societies using mobile phone data. *2009 international conference on computational science and engineering*. Vol. 4. IEEE. 2009 :144–50
36. Schmitt P, Vigil M, Zheleva M, and Belding E. Egocentric and population-density patterns of cellphone communication in ivory coast. NetMob D4D Challenge 2013 :1–14
37. Krings G, Calabrese F, Ratti C, and Blondel VD. Urban gravity: a model for inter-city telecommunication flows. Journal of Statistical Mechanics: Theory and Experiment 2009; 2009:L07003
38. Trestian I, Ranjan S, Kuzmanovic A, and Nucci A. Measuring serendipity: connecting people, locations and interests in a mobile 3G network. *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement*. 2009 :267–79
39. Vieira MR, Frias-Martinez V, Oliver N, and Frias-Martinez E. Characterizing dense urban areas from mobile phone-call data: Discovery and social dynamics. *2010 IEEE Second International Conference on Social Computing*. IEEE. 2010 :241–8
40. Pulselli R, Romano P, Ratti C, and Tiezzi E. Computing urban mobile landscapes through monitoring population density based on cell-phone chatting. International Journal of Design & Nature and Ecodynamics 2008; 3:121
41. Furletti B, Gabrielli L, Renso C, and Rinzivillo S. Identifying users profiles from mobile calls habits. *Proceedings of the ACM SIGKDD international workshop on urban computing*. 2012 :17–24
42. Scepanovic S, Hui P, and Yla-Jaaski A. Revealing the pulse of human dynamics in a country from mobile phone data. NetMob D4D Challenge 2013 :1–15
43. Liu F, Janssens D, Wets G, and Cools M. Profiling workers' activity-travel behavior based on mobile phone data. 2013
44. Berlingerio M, Calabrese F, Di Lorenzo G, Nair R, Pinelli F, and Sbodio ML. AllAboard: a system for exploring urban mobility and optimizing public transport using cellphone data. *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part III 13*. Springer. 2013 :663–6

45. Cici B, Markopoulou A, Frías-Martínez E, and Laoutaris N. Quantifying the potential of ride-sharing using call description records. *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*. 2013 :1–6
46. Zhang D, Huang J, Li Y, Zhang F, Xu C, and He T. Exploring human mobility with multi-source data at extremely large metropolitan scales. *Proceedings of the 20th annual international conference on Mobile computing and networking*. 2014 :201–12
47. Gillet A, Leclercq E, and Sautot L. The Tucker tensor decomposition for data analysis: capabilities and advantages. *38ème Conférence sur la Gestion de Données (BDA)*. 2022
48. Kolda TG and Bader BW. Tensor decompositions and applications. SIAM review 2009; 51:455–500
49. Insee Website - https://www.insee.fr/en/metadonnees/definition/c1523. 2016. Available from: https://www.insee.fr/en/metadonnees/definition/c1523