

UNIVERSITY OF TARTU
INSTITUTE OF COMPUTER SCIENCE
Software Engineering Curriculum

ALI RAZA

Exploring Group Mobility

Master Thesis (30 EAP)

Supervisor: Huber Flores

TARTU, 2021

Abstract

Exploring Group Mobility

Abstract: Human mobility research is gaining traction, opening up new avenues for research in smart cities infrastructure development. Human mobility can reveal information regarding population dispersion across a large area, human behaviour, and individuals' daily routines. In the meantime, individual mobility is intensively investigated, but group mobility is mostly unexplored. In this thesis, we developed user trajectories based on the dataset collected from their cellular network connectivity. We develop a method for assessing group mobility by examining the similarity in users' trajectories between locations in different users' routines. Our results indicate that while multiple groups can be formed in specific locations (in a static manner), it is difficult to find users that share the same spatial and temporal characteristics while moving.

CERCS: P170 Computer science, numerical analysis, systems, control

Keywords: group mobility, device-to-device communication, opportunistic d2d collaborations, human mobility, mobile crowdsensing

Rühma liikuvuse uurimine

Kokkuvõte: Inimeste liikuvuse uuringud on hakanud jõudsalt tõusma, avades uusi võimalusi teadusuuringuteks tarkade linnade infrastruktuuri arendamisel. Suurtel aladel inimeste liikuvuse andmed avaldavad teavet elanikkonna hajumise, inimeste käitumise ja üksikisikute igapäevase rutiini kohta. Seni on uuritud intensiivselt individuaalset liikuvust, kuid mitte grupi liikuvust. Käesolev lõputöö esitab väljatöötatud kasutajate trajektoore, mis põhinevad

mobiilsidevõrgu kasutajate ühendustel kogutud andmestikul. On loodud grupi liikuvuse hindamise meetod, mis uuris kasutajate trajektooride sarnasust erinevate kasutajate rutiinide asukohtade vahel. Uurimistulemused näitavad, et vaid ühepäevasest liikuvuse andmestikust ei piisa konkreetsete kasutajate trajektooride tuvastamiseks. Tulemuste põhjal on näha, et kuigi teatud asukohtades (staatilisel viisil) saab moodustada mitu rühma, on keeruline leida kasutajaid, kellel on samad ruumilised ja ajalised omadused liikudes.

CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine

Märksõnad: rühmalik liikuvus, suhtlus seadmete vahel, oportunistlik d2d koostöö, inimeste liikuvus, mrahva mobiliseerimine teabe kogumiseks (inglisk crowd sensing)

Contents

List of Figures	vii
List of Tables	ix
1 Introduction	3
1.1 Contributions	4
1.2 Outline	4
2 State of the Art	7
2.1 Human mobility	7
2.2 Mobile crowdsensing	8
2.3 Device-to-device communications	12
2.4 Summary	15
3 Dataset Description and Preparation	17
3.1 Dataset overview	17
3.2 Crowdsensed measurements	18
3.2.1 Spatio-temporal characteristics	20
3.2.2 Validity respecting spatio-temporal behaviour	20
3.3 Data modelling	22
3.3.1 Area of interest	22
3.3.2 Data preprocessing	22
3.4 Summary	25
4 Modelling Mobility with Crowdsensed Data	27
4.1 Grid structure design	27
4.1.1 Constructing user mobility	30

CONTENTS

4.1.2	Tuning user mobility	31
4.1.3	Categorisation of mobility	32
4.2	Summary	35
5	Experimental Setup	37
5.1	Defining group mobility	37
5.2	Experiment setup	38
5.2.1	Selecting locations for each category	39
5.2.2	Users routines based on the locations	40
5.2.3	Further considerations	42
5.3	Summary	42
6	Results	45
6.1	Exploring group mobility	45
6.1.1	Results of routine 1	45
6.1.2	Results of routine 2	47
6.1.3	Comparing results of routine 1 and routine 2	47
6.1.4	Similarity in users mobility (trajectory)	48
6.1.5	Routine 3	49
6.1.6	Similarity in user mobility (trajectory) for routine 3	51
6.2	Exploring group mobility across activities	51
6.3	Summary	54
7	Discussion	57
7.1	On group mobility	57
7.2	Room for improvement	57
7.3	Regularity of groups	58
8	Summary and Conclusion	59
	Bibliography	61
9	Appendix	69
9.1	Licence	69

List of Figures

2.1	Mobile crowdsensing at network level	9
2.2	Visual description of entity resolution tasks. (a) deduplication (b) record linkage and (c) canonicalisation	11
2.3	Illustration of device to device communication in cellular network	12
3.1	Mapping of base station on opencellid database	19
3.2	Spatial (a,b) and temporal (c,d) dataset characteristics, (a,b) land use and base station locations. (c,d) difference in samples collected at different time	21
3.3	Comparison of hourly utilisation of devices (a), applications (b) and base stations(c)	23
3.4	Area of Interest with the highest concentration of the samples. More scatter points describe locations with a higher density of base stations.	24
3.5	Scatter plot of the data after data preprocessing	25
4.1	User trajectory constructed using the grid cells	28
4.2	Base stations scatter plot in the grid: a) grid plot with map; and, b) grid plot without map	29
4.3	User trajectory with the high spatial mobility frequency	31
4.4	Clean mobility pattern for a user with the high spatial mobility frequency	33
4.5	All users mobility patterns	34
4.6	Median value of all users with respect to: (a) distance and (b) mobility	35
5.1	Contextualisation of mobility of the users given different type areas our grid includes	38

LIST OF FIGURES

6.1	A group of people found moving together between locations for routine 1 day time	46
6.2	A group of people found moving together between locations for routine 2 work(morning-afternoon)	48
6.3	Comparison of groups of routine 1 day time (08:00 - 20:59) and routine 2 work (morning-afternoon) (09:00 - 16:59)	49
6.4	Trajectory of users moving from commercial to railway station during work (morning-afternoon)(09:00-16:59)	50
6.5	Comparison of groups of people found moving together between locations for different activities	51
6.6	Trajectory of users moving from commercial to railway station during work (morning)(08:00-11:59)	52
6.7	Users mobility across different time activities of routine 3 a) work (morning) (08:00 - 11:59)/ lunch break (12:00-14:59) and, b) lunch break (12:00-14:59) / work (afternoon) (15:00-17:59)	53

List of Tables

3.1	Statistics summarise the dataset.	18
3.2	Overlapping coordinates by radial distance	19
4.1	Categorisation of mobility of the users given their distance during the time they were mobile	34
5.1	Users and users mobility (activity) in two locations per regions	39
5.2	Number of common users in the among these categories	40
5.3	In different locations, number of available users based on different daily routine characterisation. R - residential, C - commercial, I - industrial, P - park, T - railway station	41

LIST OF TABLES

Acknowledgements

First and foremost, I want to express my gratitude to my family and friends for their unwavering support during this project. I would also want to thank Prof. Dr. Huber Flores for his valuable feedback and direction in my research to complete this thesis. I'd like to express my gratitude to Hamza Sandhu for his invaluable help in deciphering the geographical features of the data; his assistance was critical in completing this project.

1

Introduction

The study of human mobility is important to enable new types of applications in diverse domains ranging from smart cities to infrastructure planning (1, 2). Indeed, human mobility can provide critical insights to design urban environments (3, 4) and regulate the distribution of population over a wide area (5). For instance, capacity metrics from human mobility can be used to optimise transportation systems (6, 7) and mobility patterns can be used to design better public spaces. Besides this, mobility insights can be used to disclose several human behaviors. For instance, the amount of time people spend at work, and even the daily routines of individuals (8). More recently, as the current pandemic has demonstrated, insights of human mobility are fundamental to avoid impacting the local economy of countries. Indeed, mobility patterns can be used to produce smart policies that regulate human mobility instead of restricting it completely (with lockdowns).

Existing efforts to study human mobility have focused on creating trajectories to explain mobility patterns. A trajectory depicts the locations that a person visits over a period of time (9, 10). The creation of trajectories is tied to massive effort in data collection. However, thanks to the adoption of smartphones as tools for personal data management, the collection of user data has become a very simple technical process despite its complex ethical issues in data privacy. Smartphones can be exploited individually or collectively for data collection using opportunistic approaches, e.g., mobile crowdsensing (11) and crowdsourcing (12). The first one implies that individual smartphones are periodically sampling location data to monitor mobility actively (13, 14), while the latter one aggregates samples from different smartphones carried by differ-

1. INTRODUCTION

ent users to model mobility passively (15). While user mobility has been investigated widely, most of the work focuses on mobility of a single user, leaving group mobility (multiple users) partially unexplored.

In this thesis, we conduct a systematic evaluation in which we study group mobility. We rely on mobile operator dataset¹ collected over a period of 24 hrs in the area of Shanghai, China. In our work, we first define individual user mobility in different locations using current state of the art techniques. With this information, we then conduct a similarity analysis to determine users that move together (as a group). We conduct a rigorous evaluation in which different groups are formed by analysing the influx and outflux of users between different locations. Our results indicate that while users may share similar mobility patterns, it is difficult to find multiple users (as a group) sharing the same spatial and temporal characteristics.

1.1 Contributions

- **Novel method:** We develop a grid-like technique that can be used to transform crowdsensed data into trajectories that explain human mobility. Our method is released as open source in GitHub².
- **Novel insights:** We found that a group of users are bonded to a specific location. Moreover, while we observe frequent encounter of familiar users in a location, we did not find familiar encounters of users while moving between locations

1.2 Outline

This thesis is structured as follows:

- Chapter 2 reviews the state-of-the-art about device to device communication, mobile crowdsensing and opportunistic collaborations.
- Chapter 3 explains the preprocessing and cleaning of our crowdsensed dataset.
- Chapter 4 describes the development and design of our grid-like technique to transform crowdsensed data into trajectories.

¹Partially released in AppLens workshop 2019

²<https://github.com/mobile-cloud-computing/D2DGroups.git>

- Chapter 5 describes the experimental setup used to analyse group mobility.
- Chapter 6 presents the results of our analysis.
- Chapter 7 discusses the implications and limitations of our analysis.
- Chapter 8 summarizes and concludes our work.

1. INTRODUCTION

2

State of the Art

In this Chapter, we conduct a literature review about the concepts explored in this work. Specifically, we review the current state in topics related to mobile crowdsensing, device-to-device (D2D) communications and opportunistic collaborations.

2.1 Human mobility

Human mobility is a research area that is widely explored to understand human behavior in society (4). The understanding of human mobility has several implications in different domains. For instance, urban planning uses human mobility to better optimise the construction of roads and regulate traffic of vehicles (16). Typically, human mobility is studied through the monitoring of trajectories. A trajectory depicts the places (locations) that a particular individual explores throughout the day. Multiple information can be derived from trajectories. For example, a trajectory can explain the amount of time that an individual spends in a location, e.g., 5 hrs at work, 10 minutes in a park. Trajectories that are repeated over time become human patterns. These are important as those represent repetitive routines that explain the activities of individuals across time (17).

Human mobility in urban environments has been explored to understand POI (points of interest) (18). This means that the characteristics of human mobility define the urban context. Understanding human mobility within cities, how and where people move helps in urban policy and planning (19), such as estimation of commute flow, forecast of traffic, and urban planning (20). Human mobility is used as a key indicator

2. STATE OF THE ART

of urban connections which interest urban planners and policymakers (21) and it has a significant impact on the economy, society and environment (22).

Human mobility aids in determining how connected or distant different urban regions are, which reflects the underlying urban form (23). The emerging concept of "New Urbanism" strongly encourages the reintegration of spatial form and built-up environments in order to create the perfect neighborhood for active forms of human mobility (24). Fundamental factors of urban mobility are the spatial distribution and temporal dynamics of urban activity density (25). Other research work has also explored human mobility to predict peoples' visited locations (26). Other work has shown human mobility to identify gender groups and communities from different countries (27).

Unlike previous research that has focused on studying individual users and their trajectories and partially focused on the people who moved as a group (group mobility) to identify the mobility patterns. In contrast to other work, we use a dataset of a cellular operator to analyse similar users' trajectories as a group and explore group mobility between locations in a city.

2.2 Mobile crowdsensing

Mobile crowdsensing is a technique that is used to reduce the complexity of a single task among multiple individuals (13). Besides this, mobile crowdsensing also can be used to execute tasks, which a single individual cannot complete by itself (14). A task in mobile crowdsensing depicts a computational operation carried through a smart device. For instance, collecting the GPS coordinates of a place. Multiple devices contribute to the same task, such that in time, all different contributions can be aggregated into a single solution. For example, all the GPS samples captured by multiple devices are used to define the main sightseeing (POI) in a city. While mobile crowdsensing is essential to perform complex tasks, the utilisation of mobile crowdsensing depends on efficient strategies for data collection.

Data collection on smart devices requires instrumentation of applications (aka apps), such that data samples from sensors can be captured and then uploaded to a centralised server for further analysis. This is the most common approach, which is easily hamper when there is not a strong user base for collecting data. Indeed, these apps can collect data, but its representation is sparse (28). Alternatively, crowdsensing data can be

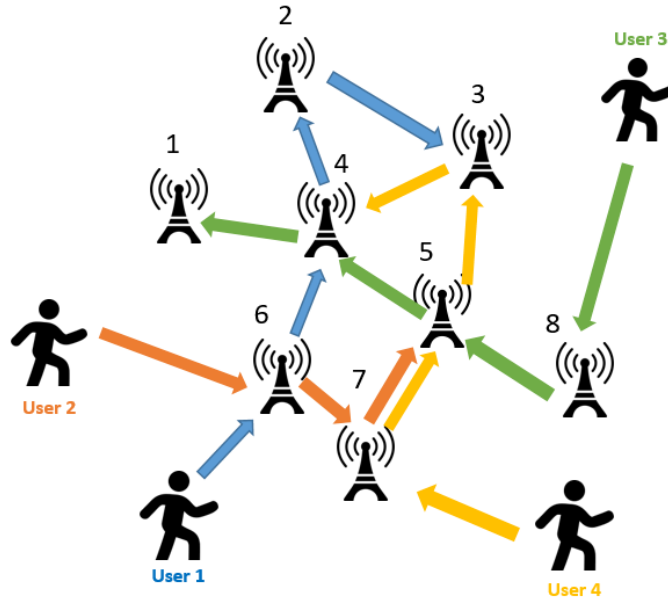


Figure 2.1: Mobile crowdsensing at network level

also captured at network level. However, this requires continuous access to the communication infrastructure that routes the data, e.g., base stations. As a result, cellular operators are the only ones that can access this information easily.

To further understand the data collection through mobile crowdsensing, let's consider Figure 2.1. There are 8 base stations marked with a number identifier and 4 users. User data is collected at each base station and aggregated into a dataset once users send information.

Mobile crowdsensing has been studied to support multiple applications (11). Mobile crowdsensing has been used successfully to construct the indoor mapping of buildings (29). Other work has used mobile crowdsensing for the generation of maps (30). For growing Internet of Things (IoT) applications, mobile crowdsensing is a vital building piece (31) such as, mapping the phenomena of mutual interest (32), smart cities (33) and self-contained indoor navigation system (34). Other research work developed smart loyalty and incentive programs, allowing travellers to gather points and redeem incentives for sites and attractions they visit across a country or region (35). Which helps to develop smart tourism hence, increasing tourism activity (36).

2. STATE OF THE ART

Traffic monitoring is another application of mobile crowdsensing (37), sense driver activities and behavior, sudden traffic events and aggressive driving, which assist in the creation of profiles for bad driving habits (38). Other work has used mobile crowdsensing to monitor health of the patient by monitoring vital signs and enhancing patient care through early detection of adverse health conditions (39). Another example is that individuals can share their exercise data and then compare their levels of exercise to those of the rest of the community (40).

A key disadvantage of mobile crowdsensing is data sparsity (14), which means missing data. Many approaches can be used to fix the missing data, and the simplest one would be to drop the missing data rows but it might not be the best approach because you will also be dropping information (41). In fact, the predictions are often based on new data even though some of the features are missing. However, the missing values could be informative in itself.

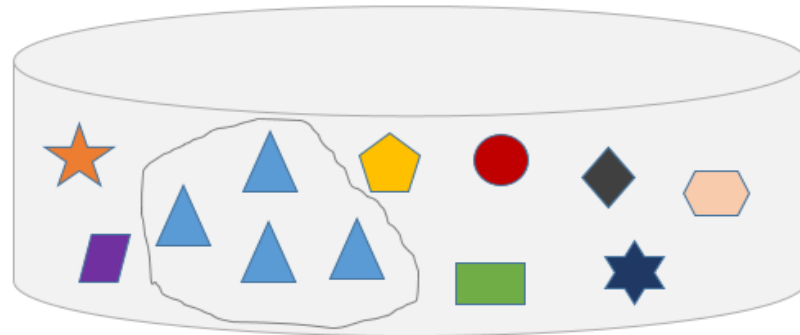
Multiple approaches from interpolation (42) to machine learning can be used to fill missing data (43). A basic solution to impute the missing data point is by filling it with mean/median in the case of numerical data. For categorical data, the model can be used to fill the missing part ¹. However, machine learning models can be used to clean datasets which will be more time effective. For instance, the imputer class from scikit-learns' preprocessing model. Scikit-learn contains amazing machine learning models for data cleaning ². Entity resolution is one example of machine learning in data unification that is based on three tasks. The Figure 2.2 illustrates the visual description of entity resolution tasks.

1. Deduplication i.e. removing redundant data from the dataset.
2. Record linkage which identifies entities that are the same across different sources.
3. Canonicalisation, which means converting the data representation into a standard form.

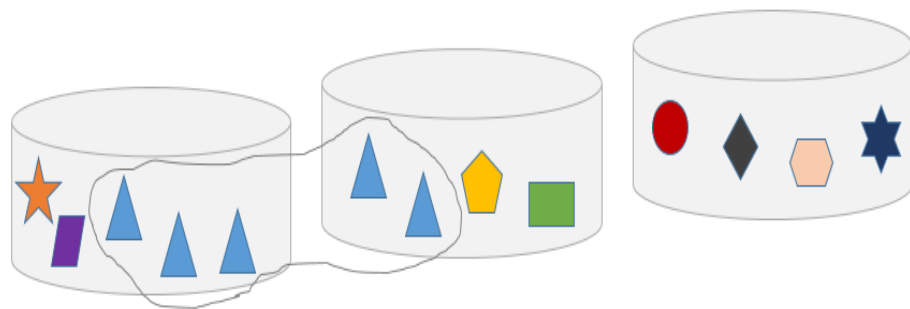
In contrast to other work, we used a mobile operator dataset that has crowdsensed data of multiple users. We used this information to model user trajectory patterns. With this information, we then analyse the mobility of users between locations.

¹<https://towardsdatascience.com/machine-learning-for-data-cleaning-and-unification-b3213bbd18e>

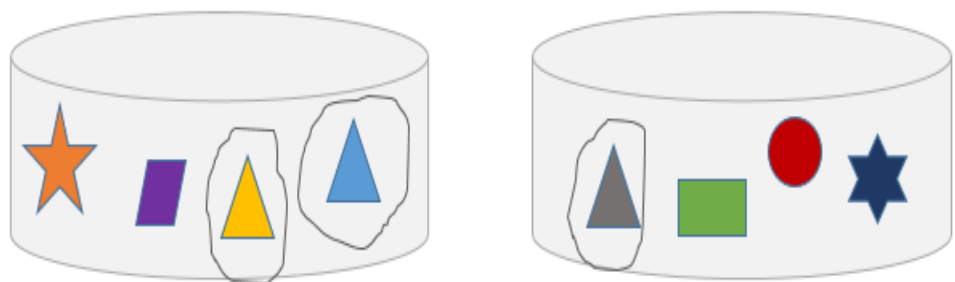
²<https://towardsdatascience.com/the-complete-beginners-guide-to-data-cleaning-and-preprocessing-2070b7d4c6d>



(a)



(b)



(c)

Figure 2.2: Visual description of entity resolution tasks. (a) deduplication (b) record linkage and (c) canonicalisation

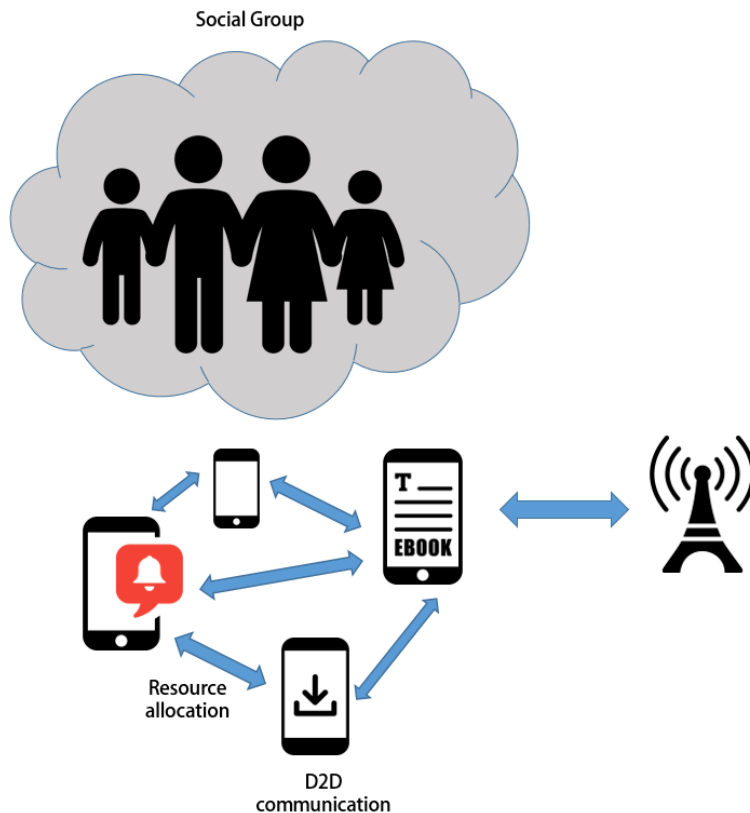


Figure 2.3: Illustration of device to device communication in cellular network

2.3 Device-to-device communications

Another approach to monitor human localisation is device-to-device (D2D) communications (44, 45, 46). Indeed, D2D can be used to characterise specific locations and determine other individuals that are in proximity. D2D relies on opportunistic smartphones carried by users to achieve this type of monitoring. By using short range communication technologies, e.g., Bluetooth and WiFi-Direct, it is possible to obtain detailed information about the context of individuals. While traditional D2D solutions rely on short range technologies, several studies have demonstrated that co-location of individuals also can be detected through the passive collection of data by using long range communication technologies, e.g., LTE/5G (47). Figure 2.3 shows the overall approaches for D2D.

D2D is a technique that is mainly used to produce opportunistic networks (48). In

an opportunistic network, mobile devices become nodes and packages are routed by these devices. The main advantage of this type of network is that it takes advantage of human mobility to propagate and disseminate network information. Similarly, D2D can also be used to engage into opportunistic collaborations. In this type of collaboration, the continuous execution of tasks is divided between multiple devices, such that it is possible to reduce the burden of executing tasks by individual devices. For instance, co-located devices can decide to share the processing load of calculating GPS coordinates. Since co-located devices share the same location, then instead of both devices executing the task, the task is performed by one device and the result of this task is then shared with others (49).

D2D has been studied in different contexts. In 2006, Zhang et al. indicate that the opportunistic collaboration model led to a high level of collective responsibility, more pervasive and distributed collaborations. Also, it gives flexibility and awareness to collaborators, which leads to greater diffusion of information and knowledge (50). These collaborations range from supplying resources to applications. For instance, database server management systems, load balancers and hardware resources (e.g., CPU, storage, and network) for ensuring guaranteed performance for applications (51). Moreover, the applications that share sensor data (49), (52).

Collaboration is essential for accomplishing a task in many applications. For example, edge computing scenarios involving high-velocity processing (e.g., real-time video sharing) that requires collaboration to offset latency requirements (53). Another example is given by Block-chain consensus-based user access strategies, which work on D2D networks (54). D2D communications enable opportunistic collaborations in which input/output resources of devices are combined together as a whole to operate (55). For such collaboration, collaborators are directly dependent on how long they stay connected with each other. In practice, the required number of collaborators, the time frame for devices to stay together in a certain range and even the range of proximity depend on the characteristic of the application. For example, collaborative sensing applications typically required devices to be within each other connectivity range (56), and the collaboration to last for several sensor cycles to prevent the overhead from managing the collaboration to offset its benefits (49), (52). By doing this, we reduce the burden of executing a task on the devices as it is performed in a collaborative manner.

2. STATE OF THE ART

A task represents an operation of a device that requires one or more resources at the same time. For instance, in a sensing task, a sensor takes a sample, and then the CPU is used to analyse the sample, and it can be performed individually or distributed among multiple devices to reduce execution cost. An opportunistic collaboration can be classified into two types: i) synchronous, and ii) asynchronous. In asynchronous collaboration, a task execution needs to take place in one go without any interruption (57). It required devices to stay in close proximity to each other for the whole duration of the execution. On the other hand, asynchronous collaboration is more like an incremental execution, which requires devices to gather multiple times until the task is completed. Every time when devices are in close proximity to each other a chunk of tasks is executed. This process is repeated until the task is completed (58).

D2D communications have been looked at from multiple perspectives. If we consider networking, D2D communications have helped build systems that improved the quality of services and networking experience e.g. wireless sensor networks (59), peer-to-peer (P2P) communications (60), delay-tolerant network and decentralised social systems (61). Another perspective could be load sharing in the sensing applications (62), (49), e.g., GPS, temperature. In recent times, the sudden increase in the capabilities of devices has led to visualisation of these systems as solutions that can contribute computational resources towards deployments of edge and fog infrastructures (57), (53).

The current research on opportunistic collaborations is mainly based on developing methods that help identify the collaborators or demonstrate that devices can find at least individual collaborators. Collaborators are basically devices that share resources over the D2D system. Many approaches analyse collaboration opportunities to identify the collaborators that stay together in the proximity distance for the execution period of the task. One of the approaches is a familiarity-based collaborator selection algorithm to analyse collaboration opportunities in the dataset proposed by Lee et al (49). This approach is based on two parameters:

1. Familiarity
2. Duration

Familiarity refers to the frequency with which the devices meet each other to be considered familiar. In other words, it measures the number of times devices were co-located. Duration refers to the minimum duration the devices needs to be co-located to

be counted as Familiarity. A familiarity-based algorithm indeed suggested that there are collaboration opportunities. The likelihood of finding the collaboration devices is high when familiarity is low. Furthermore, the number of available devices and the duration of collaboration varies when familiarity between devices changes (63). When we apply this to the real world, the more familiar we are with the person increases the likelihood of meeting that person more often. Which means more collaborating opportunities. But the study showed otherwise which means that the characteristics of collaboration opportunities vary significantly depending on the collaboration requirements.

This previous collaboration selection approach had a limitation in that it was incapable of accounting for different tasks and collaboration characteristics. This limitation can be over by using the COSINE collaboration algorithm. COSINE is a collaborator selector service that uses regularity characteristics to recommend the best possible collaborators (47). As a measure of regularity, Markov trajectory entropy (64) is used, which enables our approach to adapt to collaborator requirements. Which allows COSINE to overcome two limitations:

1. Incapability to adapt for different tasks and collaboration characteristics
2. Incapability to account for individual variations in mobility patterns

Collaboration opportunities estimated by regularity is more stable than using familiarity based estimation because selecting collaborators with similar co-location duration is better than checking the frequency of their co-existence.

Unlike previous work, we analyse a cellular operator dataset to extrapolate co-location between devices. While several approaches are available to determine whether users move together using collaboration techniques, we instead used a trajectory similarity approach to establish group mobility.

2.4 Summary

Humans carry mobile devices, and their mobility generates communication opportunities when two or more devices come into contact (48). While tracking mobility of individual devices has been widely explored, group mobility has been less investigated. In this thesis, we explored group mobility by analysing a dataset of crowdsensed measurements. To do this, we first construct individual trajectories of individuals, and then

2. STATE OF THE ART

by looking at the similarities of these trajectories, we then establish users that move together.

The next Chapter explains the data cleaning and preprocessing of the dataset.

3

Dataset Description and Preparation

In this Chapter, we describe the crowdsensed dataset used in our analysis. We describe its key characteristics and its preparation process to analyse it. As highlighted in the previous Chapter, we will use passive modelling of crowdsensed measurements captured by a cellular operator.

3.1 Dataset overview

Our dataset consists in a higher number of samples collected through cellular operator infrastructure. Our dataset captures information from users that connect to base stations in Shanghai, China, covering 24 hours (August 21, 2017). The records in the dataset consist of usage records obtained whenever a mobile application (e.g., multiple different applications) connects to the internet through mobile data. Every mobile application has a unique identifier in the collected dataset, and this information helps us distinguish between different mobile applications the users used. In the collected dataset, every single record consists of the following elements:

User ID: An unique identifier to distinguish users.

Time Stamp: Time when an application connects to the internet.

Traffic Volume: Data usage while a mobile application was connected to the internet.

App ID: An unique identifier to distinguish applications.

Base Station ID: An unique identifier to distinguish base stations.

3. DATASET DESCRIPTION AND PREPARATION

Longitude: Longitude value of a base station

Latitude: Latitude value of a base station.

The dataset is summarised in Table 3.1. The dataset contains a large amount of spatiotemporal close mobile data usage records, which allows us a baseline for estimating the location of users with low error rate.

Table 3.1: Statistics summarise the dataset.

Dataset	Samples	Devices	Apps	Base Stations
Raw	12310705	998	1543	7663
After Data Cleaning	8248775	826	1459	4011
In the area of Interest, X	3425014	548	1289	1198
Validated in X	1655271	512	46	1171

3.2 Crowdsensed measurements

Passive measurements were collected during times devices were connecting to the mobile network. The information collected includes data usage [in bytes], timestamps [date format], an exclusive identifier for applications, devices and base stations (identifiers and GPS coordinates). The raw dataset comprises data from 998 devices and 7663 base stations.

To verify the correctness of base stations in our dataset, we used the Opencellid database. The Opencellid project is the world’s largest collaborative open-data repository of GPS positions of cell towers. This project aims to offer GSM localisation data collected from different sources, including mobile apps and network providers. This project works under open content license, which promotes free use and redistribution of the data¹. We used Opencellid to overlay records of GPS coordinates of our base stations into regions using their GPS coordinates. Records with invalid location or base station identifiers were removed. Figure 3.1 illustrates the mapping result of the base stations and Opencellid database, as it is evident in the two inset maps that base station data is not overlaying geographically over the Opencellid dataset.

¹<https://opencellid.org/>

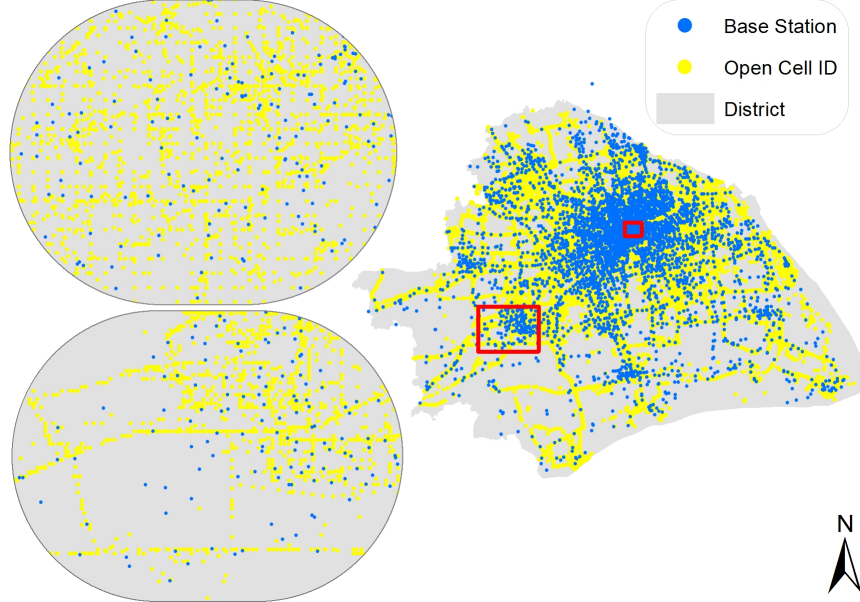


Figure 3.1: Mapping of base station on opencellid database

To tackle the matching between our dataset information and Opencellid, we used radial distance to overlay the GPS coordinates as shown in Table 3.2. We used a minimum 500-meter radial distance to clean our dataset, which gives us the most verified base station and, in our assumption, keeps our records close to real life. After cleaning, the dataset encompasses 8,248,775 records from 826 devices and 4011 base stations.

Table 3.2: Overlapping coordinates by radial distance

Radial Distance (meters)	Number of Overlapping Coordinates
1	1
500	4011
700	4144
1000	4235
1500	4333
3000	4403
5000	4420
7000	4421

3. DATASET DESCRIPTION AND PREPARATION

3.2.1 Spatio-temporal characteristics

Figure 3.2 describes the spatial and temporal characteristics of the data. Firstly, we validated the records by comparing the land use to the location of the base stations; as shown in Figure 3.2a results are as expected. Base stations are spread out to the whole region, but most importantly in the crowd dense areas (see Figure 3.2b). While this distribution aids in assuring mobile network coverage and availability, it includes spatial sparsity issues as the base stations are not uniformly distributed. Secondly, we measured the variation in the number of samples collected at different hours. As suspected, data collected from the base station locations and the number of samples vary considerably over time (see Figure 3.2c and 3.2d), reflecting the data’s temporal characteristics.

In some cases, the data samples collected are even differing by over 410,000 samples. For example, we considered the time at 04:00 PM as peak afternoon time so that we can expect more mobility and network usage. In contrast, at the time 04:00 AM where most people sleep and several network access will mostly be expected from their home which indicates less mobility. In this case, we got 479,874 samples from 1124 base stations and 580 devices at the time 04:00 PM and significantly fewer samples of 69,042 from 233 base stations and 171 devices at the time 04:00 AM. Hence, it will not be wrong to say that user mobility and app usage patterns result in sample locations and densities being dependent on time.

3.2.2 Validity respecting spatio-temporal behaviour

Figure 3.3 illustrates a histogram of the devices, app and base stations at different times, using hourly bins as a discrete resolution of a measurement concerning time. As shown in the histogram (see Figure 3.3a), the number of devices that are active per hour remained more or less the same with little variation from midnight to 06:00 AM. From the Figure, we can observe an increase in the number of active devices after 06:00 AM. Moreover, the activity of devices also stays more or less the same during the day and gradually decreases in the late afternoon. As demonstrated by several works, this is a common behaviour of users that is expected during the day (65), e.g., most people sleep during the night, which explains less activity. Similarly, people are certainly more

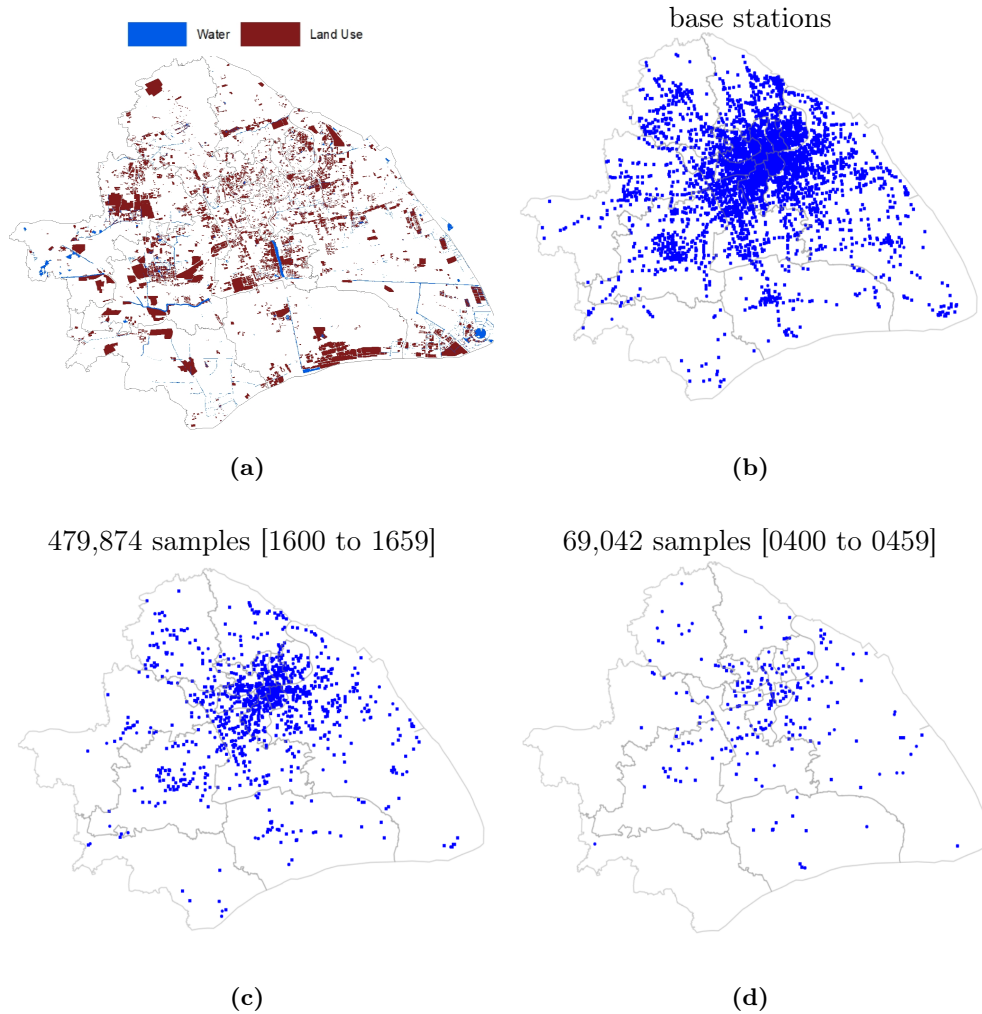


Figure 3.2: Spatial (a,b) and temporal (c,d) dataset characteristics, (a,b) land use and base station locations. (c,d) difference in samples collected at different time

active during the day, such as going to work or school and eating from outside, which explains such a pattern.

Similarly, as shown in Figure 3.3b, the data used by mobile applications are higher even though devices producing data decrease, which shows the temporal trend of data consumed by the devices. This indicates that people tend to consume more data in the evening or before midnight. This is relatable as people use their devices more often once they have free time. The histogram (see Figure 3.3c) explains the connectivity of mobile devices with their base stations that highlight the users' mobility. When we look at the

3. DATASET DESCRIPTION AND PREPARATION

number of base stations from which data is accessible, we can see that more devices are linked to more base stations at a specific hour, resulting in stable coverage across diverse places for approximately half of the day. Lastly, the latest and the earliest hour periods are expected to have the lowest number of records. These observations demonstrate that the characteristics of our data follow regular daily patterns and capture typical urban patterns.

3.3 Data modelling

In the previous section, we validated the spatial and temporal characteristics of the data. This helps us to identify the spatial sparsity issues in the data (as shown in Figure 3.2b - base stations are not uniformly distributed across the region). As group mobility depends on the crowd density, it makes it hard to trace mobility patterns with fewer base stations. Thus, in this section, we focused on modelling sparse data, such that it is possible to mitigate the effect of sparsity when tracking users' mobility.

3.3.1 Area of interest

In our work, we model users' mobility using a particular area. As the area of interest X considered the location with the highest concentration of the base stations as shown in Figure 3.4, which allowed us to evaluate data gathering over various spatio-temporal levels of sparsity. The selected area also known as an area of interest consists of 3,425,104 samples from 1198 base stations, 548 devices and 1289 mobile applications. It is located in the northeast part of the city and overlaps the zone with highest land use (see Figure 3.2a). A grid of 10 x 10 is drawn to aggregate the base stations concentrated in that area. Grid was placed on the district within Shanghai with the highest activity, which means it is the busiest part of the city, hence more mobility pattern while preserving data characteristics. Figure 3.4 clearly illustrate the districts which were considered to extract the data, which has uniform distributions of base stations across the region.

3.3.2 Data preprocessing

Before analysing the mobility of users, we conducted an extensive preprocessing over the crowdsensed measurements. First, we cleaned the dataset in two phases to eliminate apps with lower data contributions and lower amounts of samples. In particular, we

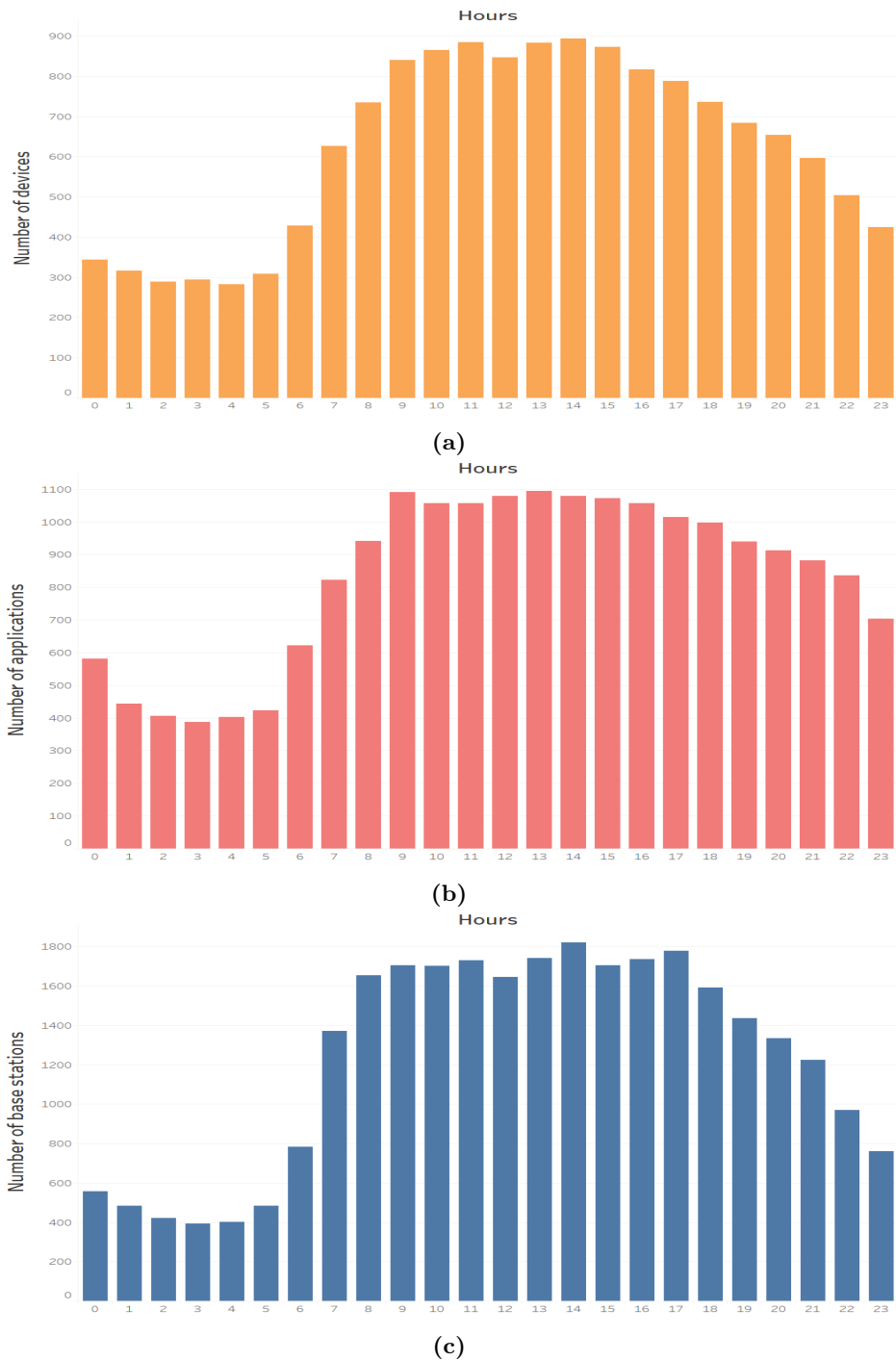


Figure 3.3: Comparison of hourly utilisation of devices (a), applications (b) and base stations(c)

3. DATASET DESCRIPTION AND PREPARATION

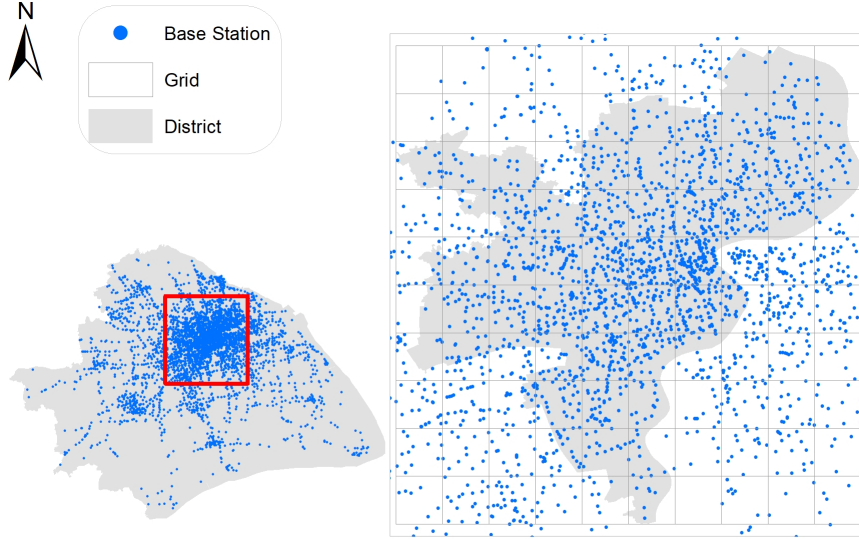


Figure 3.4: Area of Interest with the highest concentration of the samples. More scatter points describe locations with a higher density of base stations.

did not consider data from (i) applications that have less than 10 users use and (ii) applications contributing less than 10 samples per hour. We used these threshold values as it has been effective in other studies (47) In the first phase, we sorted data based on two features: user identifier and application identifier. We prepared the list of apps with the count of the unique users using the particular application. Thus, we removed all the mobile applications with less than 10 users. We applied this approach to the data from Area of Interest X (see Table 3.1). Our number of mobile applications decreased from 548 devices and 1289 apps to 537 devices and 406 apps. Our dataset sample size was reduced to 2,377,912 which is the 70th percentile of the data from the selected area.

Based on this resulting data, we then applied the second phase of the preprocessing pipeline. We created a list of user data collected per hour for each application. Once we have the list, we remove all applications in which the sample collected at any hour is less than 10. By doing this we obtained applications that made analysing mobility patterns better. For Instance, many mobile applications have 1 or 2 samples counted at a particular hour, which cannot be enough to portray the mobility of the users. By removing these, it brought more relevance to the data. This approach reduced the number of samples by 30 percent that encompasses 16,655,271 samples from 512

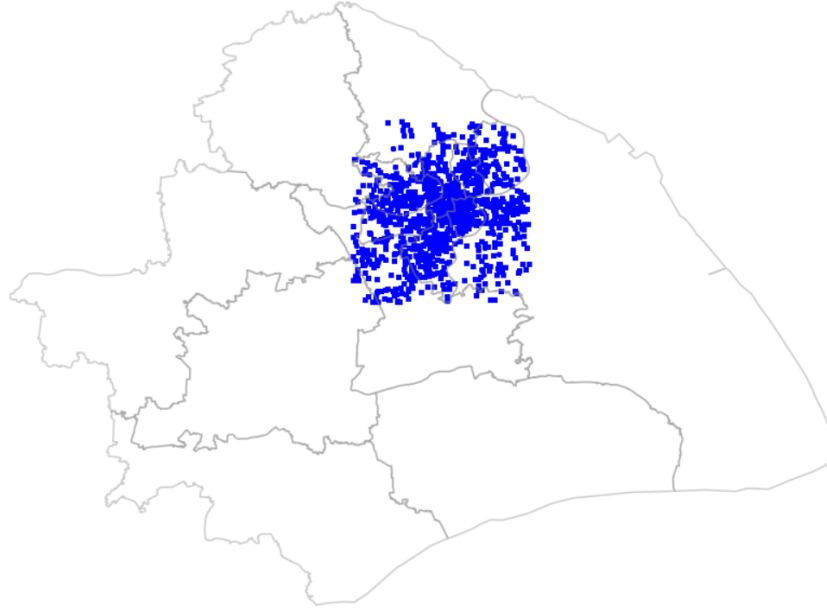


Figure 3.5: Scatter plot of the data after data preprocessing

devices, 46 mobile applications and 1171 base stations. Figure 3.5 illustrates the data after preprocessing, base stations are scattered uniformly in the area of interest with the dataset focusing on the context for analysis.

3.4 Summary

This Chapter mainly focuses on the data cleaning and preprocessing part of the dataset. We verify our base stations with Opencellid and remove all the records from other base stations and validate our verified dataset for spatial-temporal characteristic and behaviour. We found spatial sparsity issues in the data. To tackle this, we modelled the data into an area of interest (using a grid) where base stations are uniformly distributed throughout the region and data from all the base stations outside of this region were dropped. Furthermore, we preprocess the data and discarded all the applications which have less than 10 users and the applications contributing less than 10 samples per hour to make it suitable for analysis.

In the next Chapter we will use this processed data to create user mobility using their connectivity with the base station.

3. DATASET DESCRIPTION AND PREPARATION

4

Modelling Mobility with Crowdsensed Data

This Chapter focuses on estimating mobility patterns and calculating trajectories, which aids in better understanding individual user movement. We analyse these individual user movements that help us in finding their mobility and directional patterns.

4.1 Grid structure design

In the previous Chapter, we extracted from our dataset, the area with the most dense deployment of base stations (also known as an area of interest as shown in Figure 4.2a). As stated previously, high density of these stations is required to capture precise human mobility in an urban area.

To analyse the mobility of individual users, we design a grid-like structure that overlaps this area of interest. The idea of this structure is that trajectories can then be built by monitoring changes between cells. These changes depict the event of devices connecting to different base stations that have available coverage in the area where the users are located. Our grid structure consists of 100 cells and each cell represents a region containing a set of base stations. We used 100 cells for the grid as it provides an optimal cell area to merge base stations while retaining enough descriptive information to differentiate regions. Figure 4.2b illustrates a clear picture of the grid and distribution of the base stations. Each cell has a unique identifier. Thus, a trajectory is constructed by moving between cells. For instance, from the Figure 4.1 also we can observe an

4. MODELLING MOBILITY WITH CROWDSENSED DATA

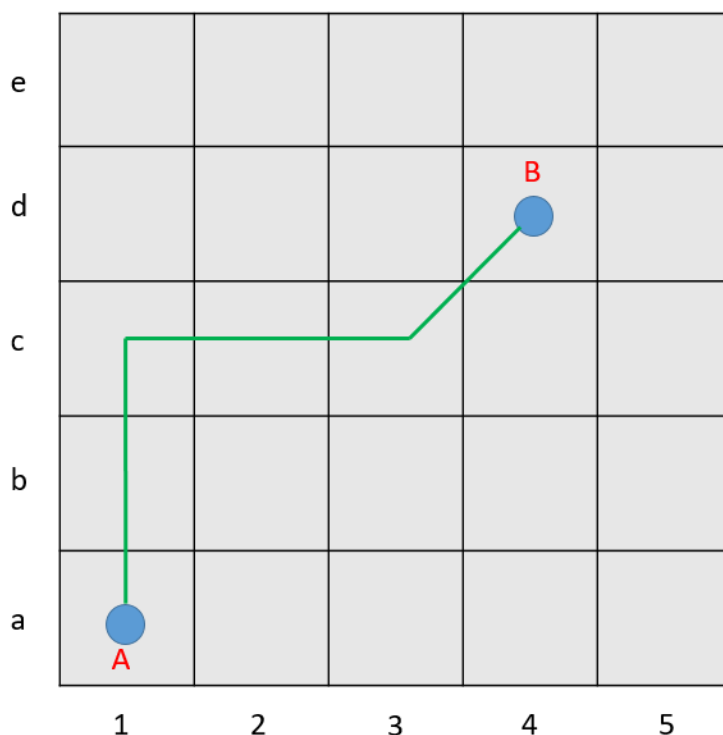


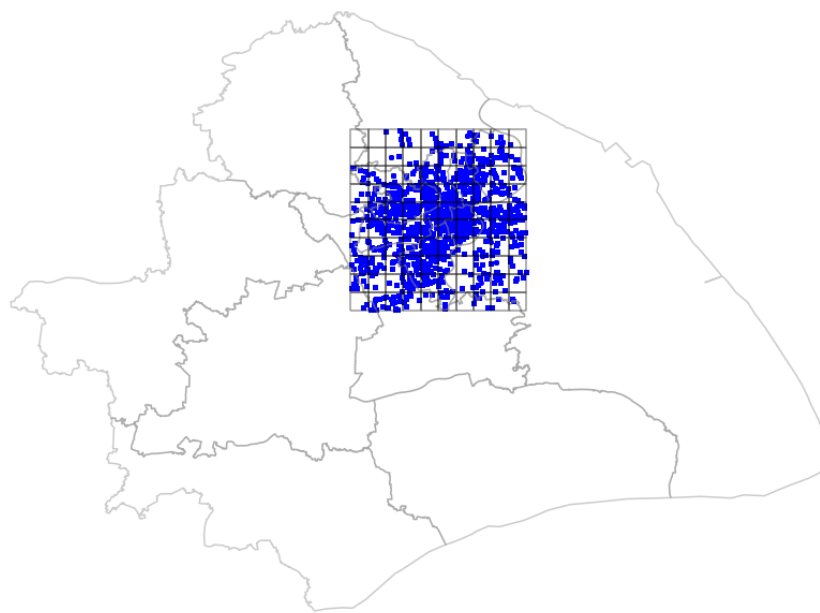
Figure 4.1: User trajectory constructed using the grid cells

emerging trajectory that depicts movement between two points, A and B.

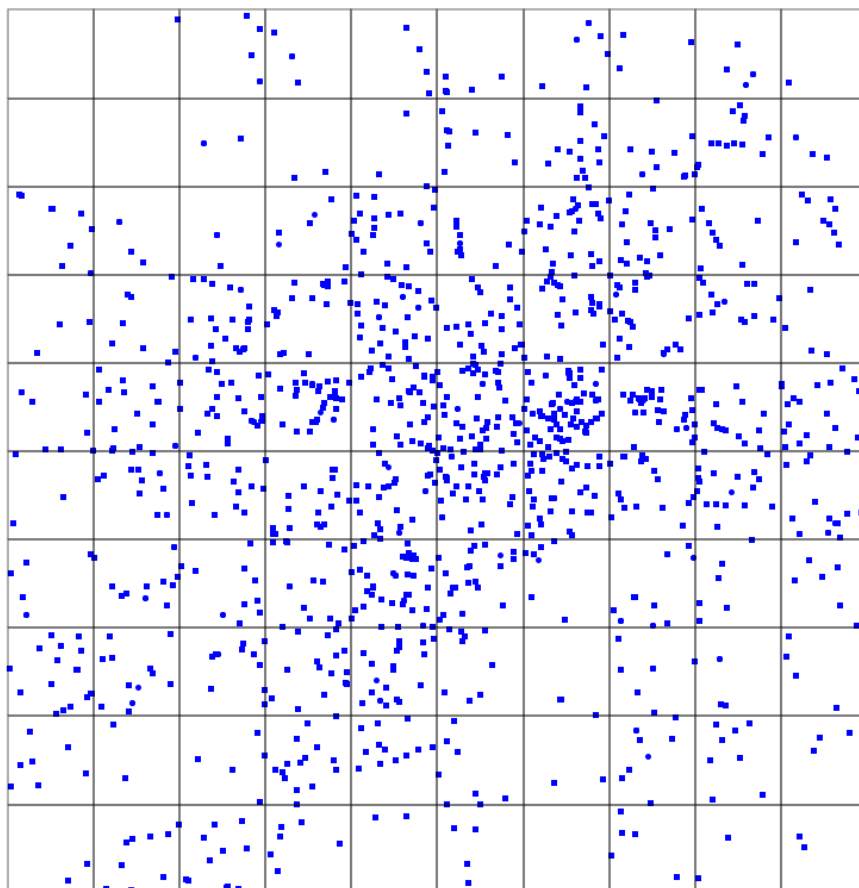
Grid implementation: Our grid is designed and developed based on existing tools and available open source libraries. ArcGIS is a geographic information system (GIS) application. The analysis toolbox includes a robust set of tools for performing the most basic GIS tasks. By using this tool, it is possible to make overlays, construct buffers, calculate statistics, perform proximity analysis, and much more with the features in this toolkit¹. We also used geofabrik, an open source database for freely available geographic data², to extract spatial data of Shanghai, China (Shanghai districts boundary). With this information introduced in our area of interest, we then clipped district border data for the district available only in our area of interest in Shanghai. Moreover, by using the clipped district polygon, we used a fishnet tool available in ArcGIS to draw a 10 x 10 grid, spanning a total area of 631 km².

¹<https://pro.arcgis.com/>

²<https://www.geofabrik.de/>



(a)



(b)

Figure 4.2: Base stations scatter plot in the grid: a) grid plot with map; and, b) grid plot without map

4. MODELLING MOBILITY WITH CROWDSENSED DATA

4.1.1 Constructing user mobility

After developing the grid structure, which merges base station deployments into cells. We next analyse the mobility of individual users. We choose to model initial mobility of individual users as trajectories, such that then it is possible to find similar trajectories to identify groups. Besides overlapping the data with a grid structure, we also created time windows during the day to explore mobility in ranges between hours.

Time windows are produced by combining session information of users to a specific base station. In other words, combining multiple connectivity entries of a device with multiple base stations over a period of time. To better understand the concept of combining session time, let's assume a scenario: A user is connected to base station $B1$ at 09:00 AM. For almost 30 minutes, the user stays static (not moving), but the user then connects multiple times with base station $B1$. Afterwards (some time later), the user moves to another location and connects with base station $B2$ and again having multiple sessions. Later, at 10:00 AM, the user moves again. So, according to the scenario between 09:00 AM and 09:59 AM, our user moved only once. So for the first 30 minutes, we will combine all the session time for the user for base station $B1$, giving us one instance of the users' connectivity and do the same for base station $B2$. This approach provides us with the users' mobility from base stations $B1$ to $B2$, which is effective to depict the time of a mobility pattern as well as to preserve the characteristics of data mobility in our measurements.

After applying this data transformation, which allows us to build trajectories, our dataset contained 6398 samples for 512 devices and 1171 base stations. With this transformation, it is then possible to build trajectories of specific users using the grid. Figure 4.3 visualises the result of the approach we applied to our data. We can observe from the figure a mobility pattern of one of the users with high spatial mobility frequency. This user has 66 samples which got reduced from 4206 after combining sessions. We used the *moving-pandas* library to plot the trajectory of the users, which was sorted by hours. The plot represents the user mobility in the area of interest for the whole day (24 hrs).

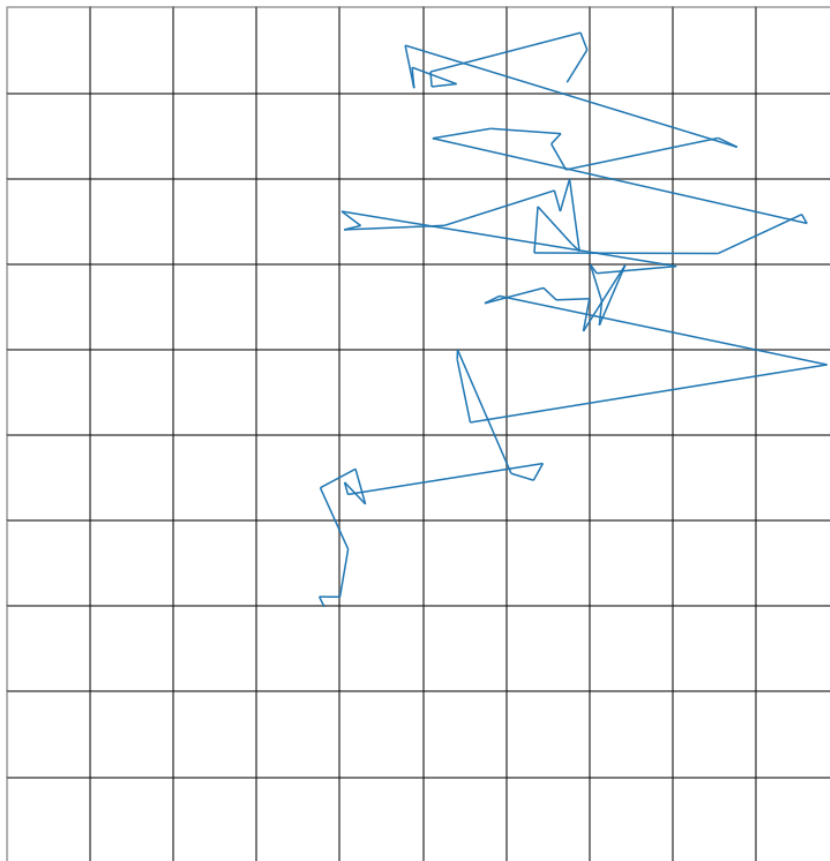


Figure 4.3: User trajectory with the high spatial mobility frequency

4.1.2 Tuning user mobility

In our work, as previously stated, we rely on open source libraries to construct trajectories of individual users. While it is possible to obtain an estimated representation of a trajectory between two areas (Figure 4.3), we can observe that the produced trajectory contains noise, which makes it difficult to explain the mobility of a user. The main reason for this noisy pattern is that devices connect not to one but to different base stations while located in a particular place. This connection can be due to an initiated session by a user (using an application) or just a single heartbeat message that a device sends through the cellular network to notify its active status in the network.

To counter this problem, we reduced the random connectivity of users to base stations, such that it is possible to tune the smoothness of user trajectories. To understand

4. MODELLING MOBILITY WITH CROWDSENSED DATA

this process, let's assume a scenario based on our grid. The 10 x 10 grid consists of 100 cells, and each cell consists of multiple base stations. While a user has been in one of the cells, can randomly connect to any base station in that particular cell. So if the user moves from cell no. 1 to cell no. 2, he connects to almost all the base stations in the cell at least once, which causes zigzagging user movements. As a result, we considered eliminating base stations that are generating randomness in the mobility pattern and only considered the base station from where the mobility originated in cell no. 1 and destination base station in cell no. 2 giving us a single straight mobility line between the cells. Naturally, removing data for calculating trajectories is not an aleatory process as it is necessary to preserve the validity of the dataset and the mobility characteristics of individual users.

To tune trajectories of users, we then apply an incremental construction approach. In this process, we created origin and destination points (OD points) for each hour for a user. For each hour, we will pick the first base station user got connected to and the last one; the rest of the connectivity instances were discarded meaning base stations were removed. Once we get all the OD points for the user, we can put them together and can have a clean mobility pattern that is more meaningful and with less noise. Figure 4.4 shows the clean mobility of the same user represented in the previous (Figure 4.3).

Since the tuning process provides intuitive representation of mobility of users, we then apply this process to the whole data available in our dataset. Figure 4.5 represents the mobility of the users' sample of 6398 from 512 devices. These are the final and clean mobility patterns of all the users found in our dataset. These patterns now can be explored from the perspective of group mobility.

4.1.3 Categorisation of mobility

To understand more in detail the type of user mobility that was found in our dataset, we conducted further analysis in which user trajectories are categorised based on two factors, *Distance* and *Mobility*. Here, *Distance* refers to the area covered by the users while moving between locations, whereas *Mobility* refers to the number of times a user moves from one base station to another. We used our grid cells to figure out these two variables for all the users. We track user mobility by using these cells (including multiple base stations) to find the highest and the lowest cell visited by the user which represents the user connected to one of the base stations in these cells. The difference

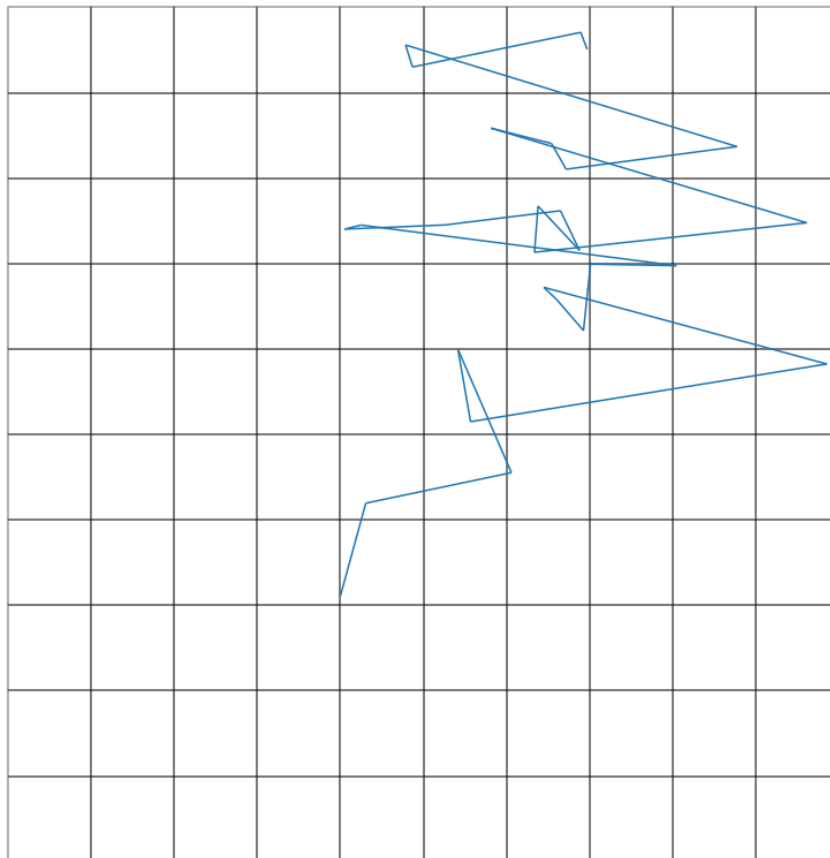


Figure 4.4: Clean mobility pattern for a user with the high spatial mobility frequency

between these cells provides us with the Distance value (number of cells covered by his trajectory). On the other hand, mobility is estimated by the number of cells visited by the users (base station connected to for the period of their trajectory).

Figure 4.6 demonstrates the values retrieved for these variables for all the users in the data, we can see how users' span of distance and mobility varies between the users. The highest distance (difference between the cells) covered by a user is 95. In contrast, approximately 50th percentile of the data is less than 12, which is the median number of cells covered by the user. This suggests that almost 50% of the users have a low distance value. In addition, the highest mobility for a user is 20. However, almost 60% of the users fall under the category of low mobility because of 60th percentile of the data is below the median value of 3. Table 4.1 summaries the type of categories that can be modelled by the distance and mobility factors. From the table, we can observe

4. MODELLING MOBILITY WITH CROWDSENSED DATA

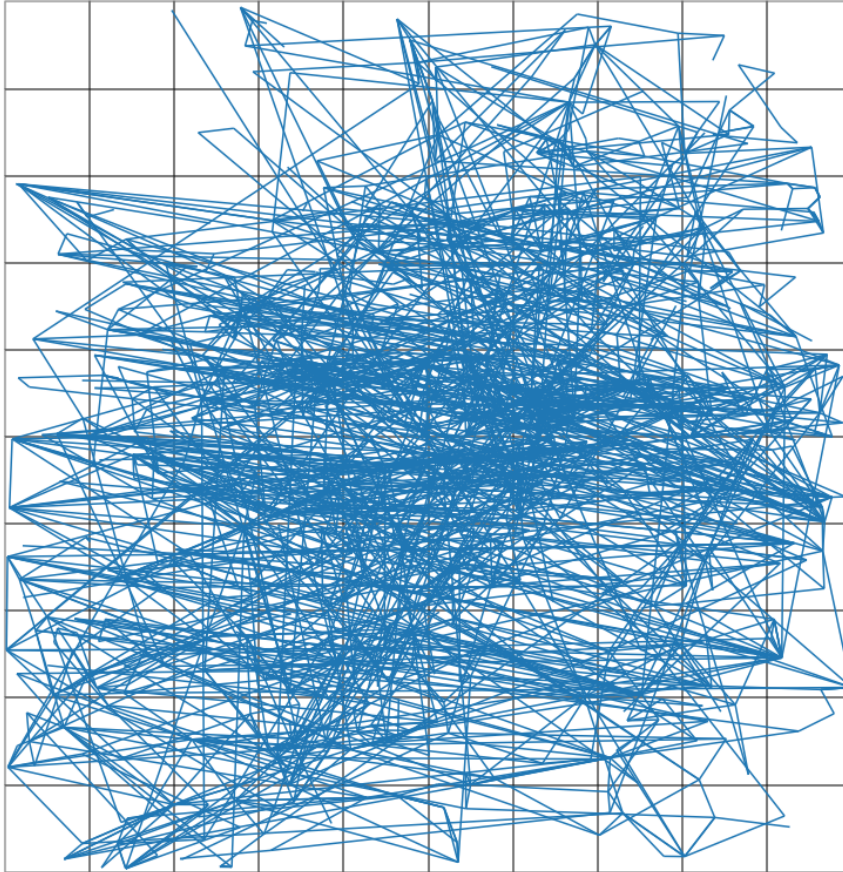


Figure 4.5: All users mobility patterns

that 176 out of 512 devices have high distance and high mobility, whereas most devices have low distance and mobility.

Table 4.1: Categorisation of mobility of the users given their distance during the time they were mobile

Mobility	Distance	
	High	Low
High	176	11
Low	61	264

This analysis also provides us further insights before exploring group mobility, in

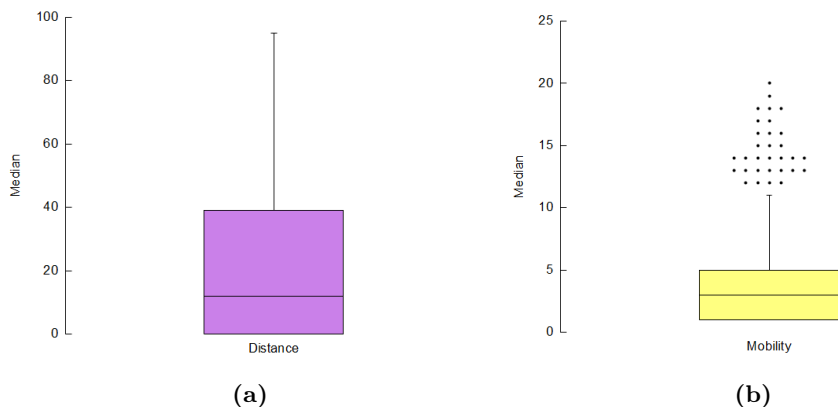


Figure 4.6: Median value of all users with respect to: (a) distance and (b) mobility

particular, the results indicate that 127 users out of 512 are not mobile, and they fall under low distance and low mobility. Considering this particular category, 264 users may have some groups, but they will not be together for a long period and distance. This means that more likely those users just have spontaneous encounters. Moreover, from the rest of available users in our dataset, we can't neglect the time variable, even though it is possible to find similar trajectories (mobility patterns), they also need to match in the same periods of time to be considered as users moving together (as a group).

4.2 Summary

Using our dataset of crowdsensed measurements, we defined trajectories that represent users' mobility across the area of interest (grid) using the moving-pandas library. We found out that defining users' connectivity to base stations to model user mobility introduces noise in the construction of trajectories. To counter this problem, we applied two approaches to clean the data and tune the trajectories of individual users. In the first one, we combined the session time of each base station's user, while in the latter one, we introduced origin and destination points when constructing trajectories of users. These points represent the base station users connected to at the start and end of the hours. By doing this, we obtained meaningful trajectories that describe mobility of users while maintaining data mobility characteristics. After this, we also conducted a further analysis to understand the type of mobility that is found in our dataset when

4. MODELLING MOBILITY WITH CROWDSENSED DATA

constructing trajectories for the whole set of available users.

In the next Chapter, we will describe the experiment that is used to analyse group mobility.

5

Experimental Setup

In this Chapter, we describe the experimental setup used to explore group mobility. We set up the experiment based on regions (locations) and time. The goal of the experiment is to detect a set of users that move together between these regions at different times.

5.1 Defining group mobility

We define group mobility by estimating the amount of users moving together between locations. To analyse group mobility, we introduced (categorical) locations in our area of interest (X) to track mobility of users between specific regions. Figure 5.1 illustrates the distribution of these categories on the area of interest X . We used OpenStreetMap data from Shanghai to extract these regions. Our five regions are commercial areas, residential areas, industrial areas, parks and railway stations. Furthermore, commercial areas represent all the offices, shopping centres, warehouses, or retails. The residential areas refer to all the residency sites like residential buildings, private houses and farm-houses, etc. Park denotes all the available forests, nature reserves and public parks, and the industrial areas represent all the sites that are used for industrial development. Railway stations are the sites that users use to commute.

These regions are scattered all over the grid, which means our users are moving in and out of these regions during 24 hrs. This gives us insight into the users' mobility patterns and their daily routines. For example, a user has a routine to leave his house for work via a train during the day. So, he will leave his house, go to a railway station, and train at his workplace. This means his trajectory will show the user moving from a

5. EXPERIMENTAL SETUP

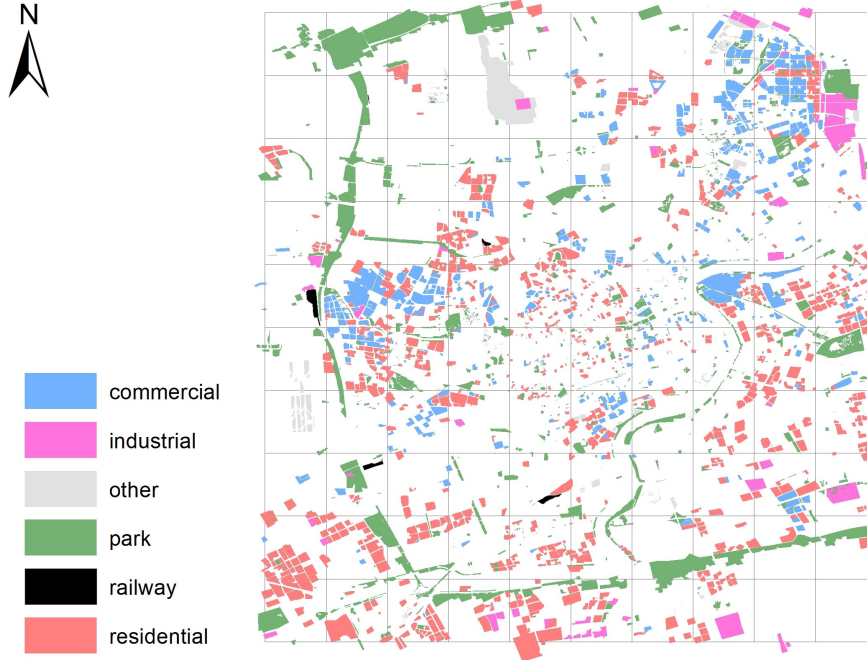


Figure 5.1: Contextualisation of mobility of the users given different type areas our grid includes

residential area to a railway station. This information is captured by our grip structure. Moreover, this demonstrates that users' mobility patterns are not random; but their mobility represents their routine. As we are exploring group mobility, these regions will help us bind groups of people in the context of their location of co-existence. For instance, if we find a group of users in the same residential area and the same group in a commercial area, they likely live and work together, and if they share the same routine, there is a high chance they might move together.

5.2 Experiment setup

We set up an experiment to explore group mobility using the 512 users that were validated in the area of interest X from the whole data available to create users' trajectories. This is a controlled experiment based on locations and users' routines. We select 5 locations, one from each region, and users' routines represent different time frames based on users' activities during the day. Fixed locations allow us to examine user mobility and behaviour in a location, whereas users' routines help us see a change in user mobility

when a user activity changes over a period of time. Similar trajectories of users were used to identify users as a group. The goal of the experiment is to find a group of people that move from one location to another while sharing a similar temporal features and similar trajectory.

5.2.1 Selecting locations for each category

Our five categories are dispersed over the Area of Interest X, we selected two locations per category. These locations were picked based on dominating characteristics of the location itself. This means that if an area has 80% of residential buildings and 20% of commercial area, then the whole location will be considered residential. For example, if we consider residential areas, we picked two different locations where the most areas were covered by the residential building/houses. After getting these locations, we extracted numbers of users and users' mobility (Activity) in these locations - Activity depicts the connectivity behavior to the base stations and it is quantified based on the number of samples that are monitored. Base stations in these regions are the common interaction point for the users, which help us define similarity between users in their locations.

Table 5.1: Users and users mobility (activity) in two locations per regions

Category	Location 1		Location 2	
	Users	Activity	Users	Activity
Residential	45	158	29	119
Commercial	27	155	6	274
Industrial	16	60	18	59
Railway Station	53	220	6	6
Park	21	70	29	70

Table 5.1 illustrates the number of users available in the locations for each category and the number of their mobility activities within the particular location (users connectivity). This table shows a change in the number of users based on location and category. For example, for the commercial category, we see that there are 27 users in location 1 and location 2 has 6 users. But, if we see their activity in this category, location 2 users have more activity although they are less in number. The change in users'

5. EXPERIMENTAL SETUP

number by category is also visible; we can observe most users were found at railway stations. In contrast, the industrial area location had the least number of users.

For our experiment, we define a single location per category. We picked the location which has most users as it provides us with a bigger sample of users in a group for each location. We selected locations with these numbers of users 45, 27, 18, 53 and 29 for Residential, Commercial, Industrial, Railway Station, and Park. To verify these locations, we examined these locations for user mobility across the categories in the duration of 24 hrs and found groups of people common in these regions.

Furthermore, we also ensured the existence of users' mobility across these regions. Table 5.2 illustrates that all the locations have at least one user that is common between two categories which is true for all the combinations of these five categories. It indicates users mobility in these locations, and from the perspective of similar users, we can say that 6 users moved from a railway to residential during the time frame of 24 hrs. Still, the question is, did they have the same temporal characteristics and similar trajectories?

Table 5.2: Number of common users in the among these categories

Category	Residential	Commercial	Railway Station	Park	Industrial
Residential	0	3	6	2	3
Commercial	3	0	6	3	1
Railway Station	6	6	0	2	2
Park	2	3	2	0	3
Industrial	3	1	2	3	0

5.2.2 Users routines based on the locations

After defining the locations to analyse and verifying the existence of the same users between locations, we then explore whether these users share the same temporal characteristics. Thus, we then explored group mobility in the context of duration and time. In this case, the temporal variable plays an important role in finding groups that move across different locations together. The availability of users in the locations oscillates considerably over time. As a result, we characterised the users available in our 5 locations in such a way that is indicative of the users' routine (66). We rely on descriptive

5.2 Experiment setup

routines to make our results more intuitive to understand. For this experiment, we relied on three daily routines to demonstrate change in the size of a group of people moving together in and out of our 5 locations (Residential, Commercial, Industrial, Railway Station and Park). Routines are described in Table 5.3. The first one divides the day into DayTime and NightTime; the second one splits the day into three categories, Rest, Work and Leisure; and the last one splits up the day into more granular user routines consisting of 7 activities.

Table 5.3: In different locations, number of available users based on different daily routine characterisation. R - residential, C - commercial, I - industrial, P - park, T - railway station

Routine	Category	Time range	R	C	I	P	T
1	Daytime	08:00-20:59	37	23	17	27	51
	Nighttime	21:00-07:59	19	14	3	5	15
2	Rest (Early Morning)	01:00-08:59	15	14	1	8	12
	Work (Morning-Afternoon)	09:00-16:59	23	17	12	19	37
	Leisure (Afternoon-Evening)	17:00-00:59	22	15	11	13	24
3	Rest (Early Morning)	01:00-05:59	7	3	1	0	2
	Rush hours (Morning)	06:00-07:59	9	9	0	2	6
	Work (Morning)	08:00-11:59	14	15	6	10	26
	Lunch break	12:00-14:59	11	12	7	11	24
	Work (Afternoon)	15:00-17:59	14	12	4	9	21
	Rush hours (Evening)	18:00-20:59	15	11	10	10	17
	Leisure (Evening)	21:00-00:59	10	15	11	3	9
Total			45	27	18	29	53

Furthermore, Table 5.3 also demonstrates the effect of location and time on the number of users. It illustrates the partition of the 24 hrs into such a manner that we see three different routines for the users. For instance, Routine 1 (Daytime) shows more users available in these locations between 08:00-20:59. If we compare routine 1 for Railway Station we can see 51 users out of 53 were at this location during the daytime, whereas 15 were available during the night time (21:00-07:59). This also indicates that few users were at Railway Station during DayTime and NightTime.

However, Routine 2 Further splits the day and displays the characteristics of users at a different location. We observe that all locations have more users during working

5. EXPERIMENTAL SETUP

hours (09:00-16:59) than any other time of the day. If we look at the Industrial area we see a similar number of users available during Work and Leisure but only one user available during Rest (01:00-08:59), which describes the nature of the industrial area because this area usually has long working hours.

Routine 3, shrinks the time frames more to give us 7 activities to examine users' mobility across the locations in limited temporal constraints. If we compare routine 1 DayTime with Routine 3, we can see the DayTime is distributed into 4 activities (Work (Morning), Lunch Break, Work (Afternoon), Rush Hours (Evening)). Which also breaks down the number of users available in different locations in a fixed window of time. There were 23 users during DayTime (08:00-20:59) within the commercial location, which were broken down into 15,12,12 and 11 number of users in Work (Morning), Lunch Break, Work (Afternoon), Rush Hours (Evening) time frame respectively. which also tells us that some users stay in this location for multiple activities.

5.2.3 Further considerations

In the case of routines, for Routine 1 and Routine 2, we run our experiment and check for the common group of users between locations during all the activities. For example, we examine DayTime from Routine 1 and see how many groups of people moved together between two locations during the same temporal characteristics and check similarity in their trajectories. In the case of Routine 3, since the activities have smaller time frames, we examine group mobility within and across the activities and use their mobility patterns between these activities. For example, we checked group mobility within Work(Morning) and between Work(Morning) and Lunch Break.

5.3 Summary

We identified different regions in our area of interest overlapped by our grid structure. Based on these regions we selected 5 locations with the most users and set up a control experiment to explore users' mobility in and out of these locations. We used different user routines (time frames) to analyse effect on user mobility as a group and used their trajectories to detect similarity in their mobility, such that it is possible to characterise a group moving between specific locations.

In the next Chapter, we present the results of our described experiment.

5. EXPERIMENTAL SETUP

6

Results

This Chapter presents the results of our experimental setup described previously. As stated before, our experiment accounts for 5 different locations and uses 3 users' routines with multiple time frames. Our results provide key insights about group mobility between different locations.

6.1 Exploring group mobility

We now proceed to interpret the results of our analysis when using different routines, consisting of multiple activities. Kruskal Wallis test¹ to detect whether there are users in our data that share same data distribution showed that indeed, some users have similar data distribution ($p < 0.05$), but the their trajectories constructed with our grid do not match each other. This means that data distribution is not enough to detect similar users moving together, and instead, we rely exclusively on similarity of trajectories described by our grid to find users moving together.

6.1.1 Results of routine 1

Routine 1 indicates that there are groups of people moving between two locations during DayTime (08:00-20:59). However, NightTime (21:00-07:59) activity shows no group mobility, indicating that users were static and together for a long time. Figure 6.1 shows the results obtained from DayTime activity of Routine 1. Results show that two groups of 5 users moved from Residential to Railway and Commercial to Railway during

¹Statistical significance analysis

6. RESULTS

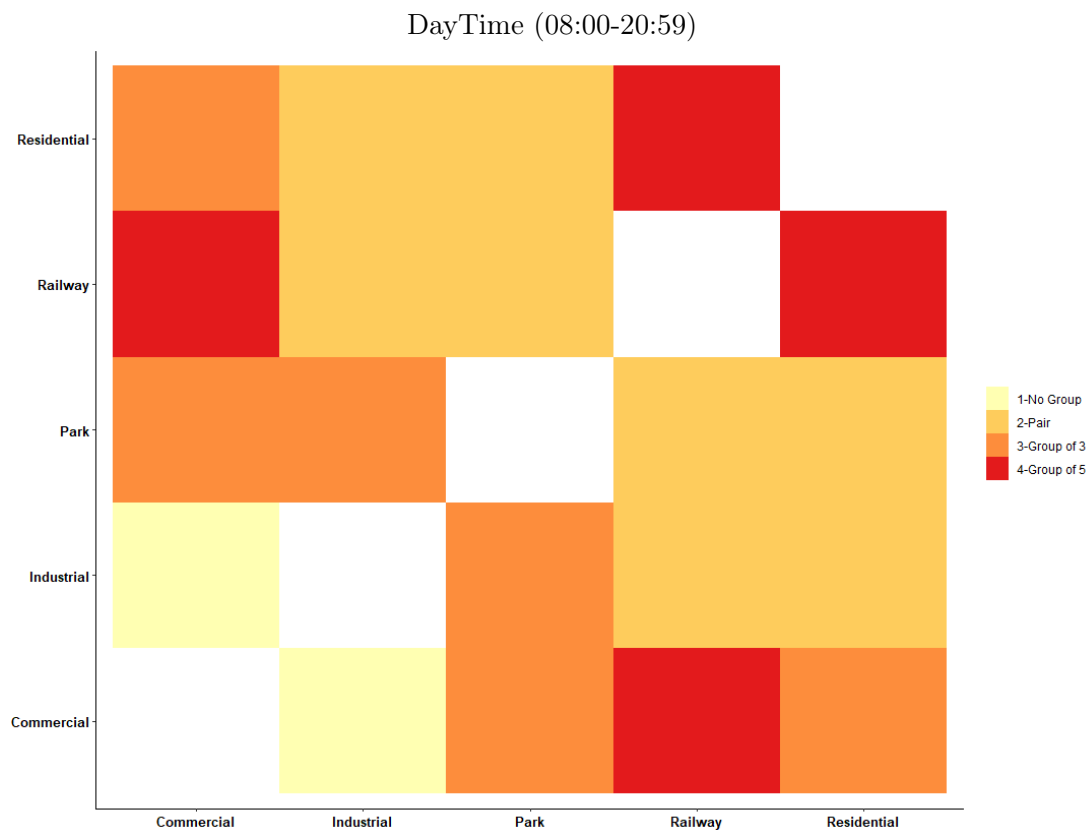


Figure 6.1: A group of people found moving together between locations for routine 1 day time

this time frame. Moreover, there are 3 groups of 3 users that moved from Commercial to Residential, Commercial to Park and Industrial to Park.

Our analysis also shows the users which are moving in groups of 2 users, otherwise in general users move alone. If a single user is moving from one location to another, we categorise it as no group. For example, we found one user moving from Industrial to Commercial during the DayTime activity, so there was no group mobility. In comparison, the users moving in groups of 2 users will be considered as pairs. Our results show 4 pairs each moving from Residential to Industry, Residential to Park, Railway to Industrial and Railway to Park respectively.

6.1.2 Results of routine 2

Results obtained for Routine 2 for all three activities (Rest (Early Morning), Work (Morning-Afternoon) and Leisure (Afternoon-Evening)) indicates that there was no group of users found during Rest (01:00-08:59) and Leisure (17:00-00:59) activity. There was one user in each activity that moved between locations. Still, one user cannot be categorised as a group, so no group mobility was found in these activities of Routine 2. Still, we can say we found a group of static users that stays together for a longer time at fixed proximity.

However, our analysis for Routine 2 activity Work(Morning-Afternoon) indicates that during the time frame of 09:00-16:59, there are groups of people that are moving between two locations. Figure 6.2 demonstrates the results obtained for the working hour of Routine 2. Results show that there are two groups of 3 users and two pairs. Groups moved from Residential to Railway and Commercial to Railway whereas pairs mobility was from Residential to Park and Industrial to Park.

Our results also show the user who is moving alone between locations during this activity. For example, we found one user moving from Railway to Industrial and another user moving from Railway to Park. This user will be considered as no group, and so no group mobility can be determined in these locations.

6.1.3 Comparing results of routine 1 and routine 2

We compared the results of Routine 1 and Routine 2; we can see how the number of users changed in a group by reducing the time frame. We draw a comparison between Routine 1 DayTime (08:00-20:59) and Routine 2 Work (Morning-Afternoon) (09:00-16:59). Figure 6.3 demonstrates the numbers of users in a group changed over this activity. We observed the groups moving from Residential to Railway and Commercial to Railway; their group size reduces from 5 to 3 users. Industrial to Park users group size dropped from 3 to 2 whereas Commercial to Park group size decreased to 1 user from 3. However, there was no change in the mobility behaviour of the pair that moved from Residential to Park between both DayTime and Work (Morning and Afternoon) activity.

Our Comparison indicates that all users moving together in DayTime (08:00-20:59) don't share the same temporal characteristic except for the pair of users that holds their

6. RESULTS

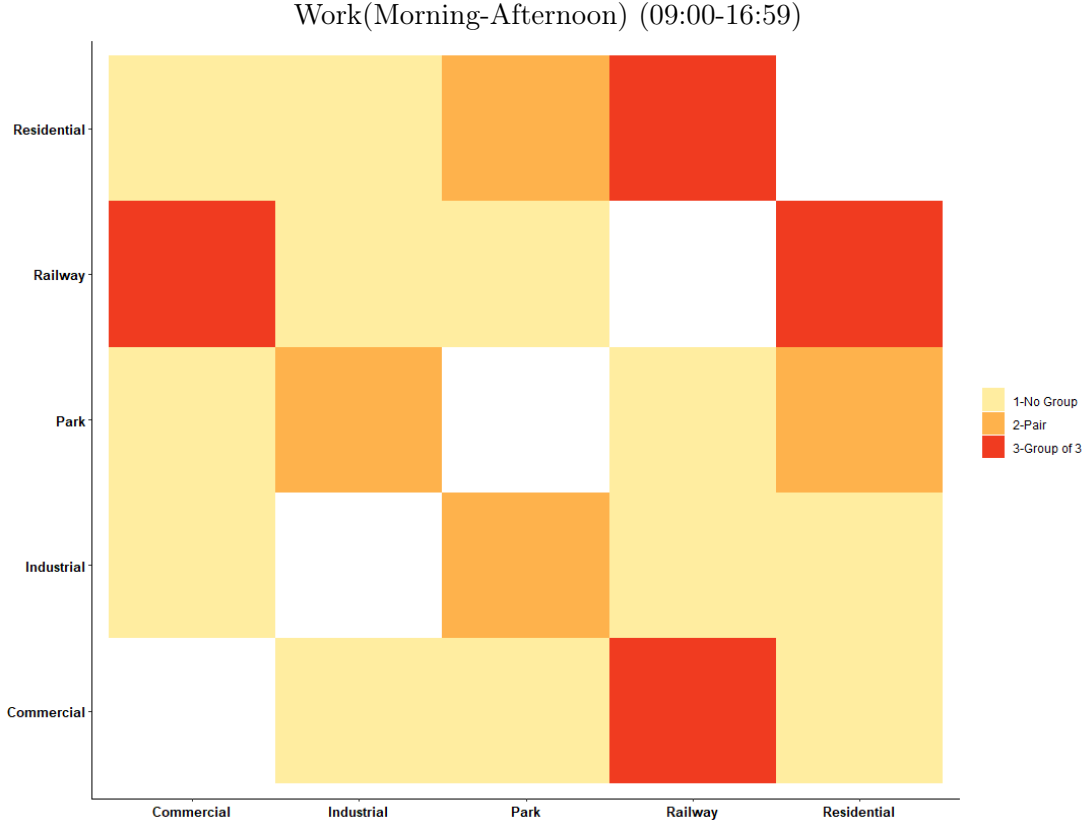


Figure 6.2: A group of people found moving together between locations for routine 2 work(morning-afternoon)

mobility together even in a reduced time frame of Work (09:00-16:59). Moreover, we deduce the locations that have groups and pairs moving among them.

6.1.4 Similarity in users mobility (trajectory)

We consider users to have similar trajectories if they share the same path defined by the cells in the grid. Similarity comparison was drawn for the users that are moving from Residential to Railway and Commercial to Railway. The following group trajectories were drawn to examine if they have similar mobility patterns or not. Our results indicate users' mobility pattern was different but overlapping each other for some duration when they were moving between the locations.

Figure 6.4 illustrates the users' trajectories moving from Commercial to Railway during the Work (09:00-16:59). Users 1, 2, and 3 moved from Commercial to Railway,

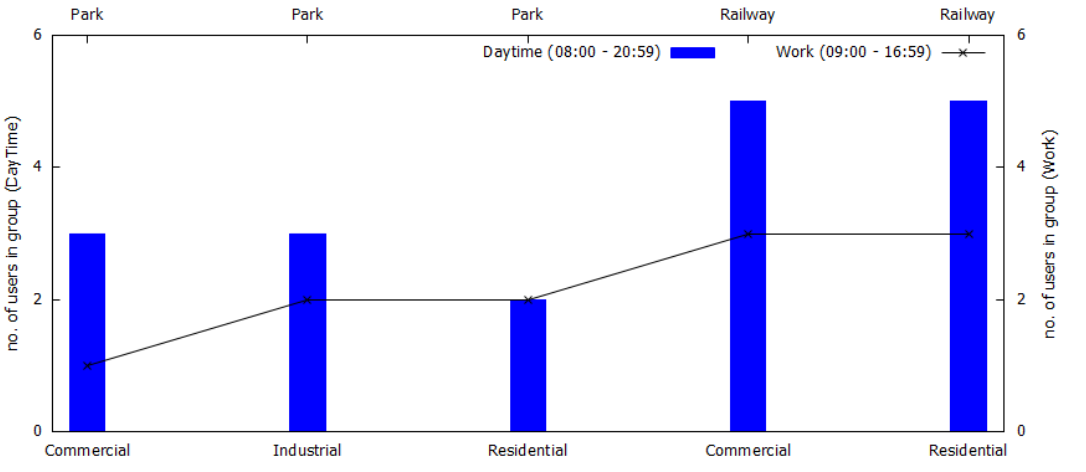


Figure 6.3: Comparison of groups of routine 1 day time (08:00 - 20:59) and routine 2 work (morning-afternoon) (09:00 - 16:59)

and all the other users are moving to different locations from Commercial. The similarity in users’ trajectories can be seen for multiple users when they leave the Commercial area. Even users who did not reach the Railway had a similar trajectory with the users who moved from Commercial to Railway for some part of their whole mobility period. Focusing on User 1, 2, and 3, trajectories show them leaving the Commercial area together and using a different path to reach the Railway while having similar trajectory at different intervals.

6.1.5 Routine 3

Our Routine 3 are based on the groups that were moving from Residential to Railway and Commercial to Railway and pairs moving from Industrial to Park and Residential to Park we found from our analysis of Routine 1 and Routine 2. We used Routine 3 activities Work (Morning) (08:00-11:59), Lunch break (12:00-14:59) and Work (Afternoon)(15:00-17:59) to break down the DayTime and Work (Morning-Afternoon) into the more granular time frame.

Figure 6.5 illustrates the results of our analysis. It shows that the number of users in a group decreases to one or no users in a group as we decrease the temporal window. For Commercial to Railway, we had 5 users in a group for DayTime and a group of 3 users for Work. We saw only one users’ mobility between these locations during work

6. RESULTS

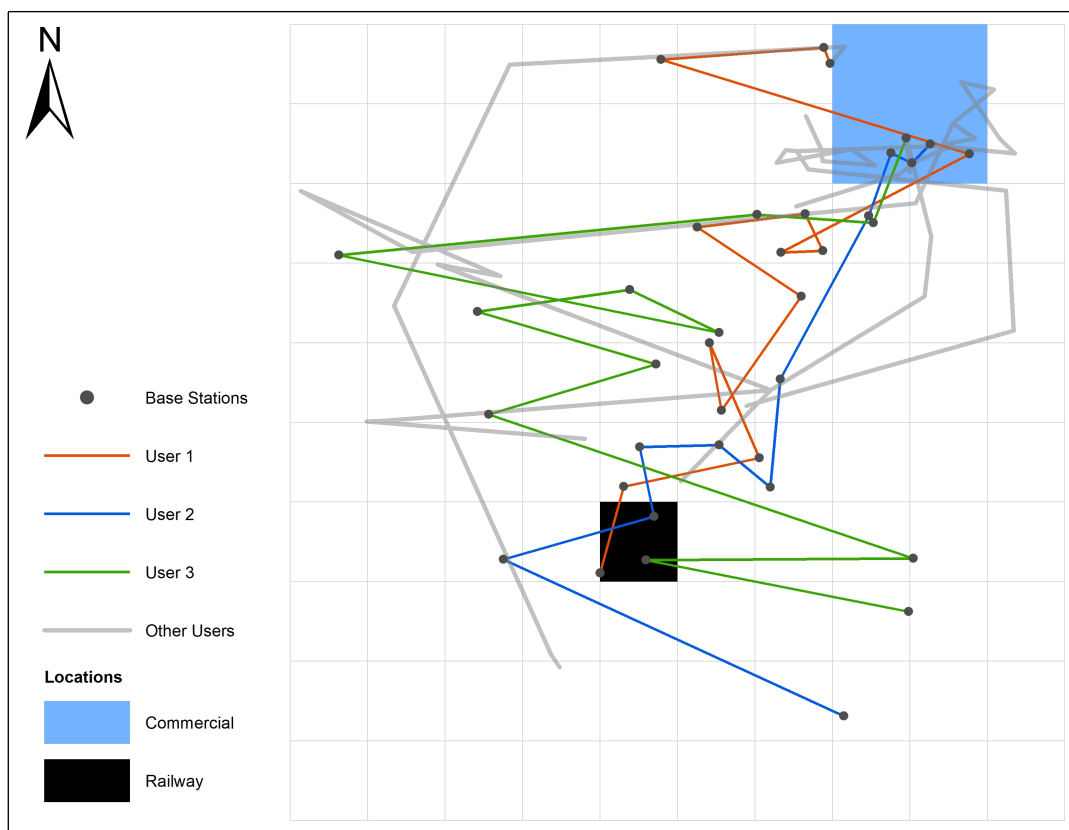


Figure 6.4: Trajectory of users moving from commercial to railway station during work (morning-afternoon)(09:00-16:59)

morning hours of 08:00 to 11:59, and likewise single users' mobility in other activities. However, no group mobility was observed in these granular activities. This tells us a group of 3 users were together at Commercial and Railway during the Work activity of Routine 2 but they did not move together. One user moved between these locations, the remaining two moved to different locations and then moved from those locations to the Railway clearly indicating no group mobility. Similar was the case with the group of users moving from Residential to Railway. Furthermore, for a group that moved from Industrial to Park, our results demonstrate that 2 users moved as a group during the time frame of 09:00 to 16:59. They both moved from Industrial to Park, but 1 user moved during Lunch Break (12:00-14:59), and 1 user moved during Work (Afternoon) (15:00-17:59). Users that moved from Residential to Park also moved to different locations and then united at one location.

6.2 Exploring group mobility across activities

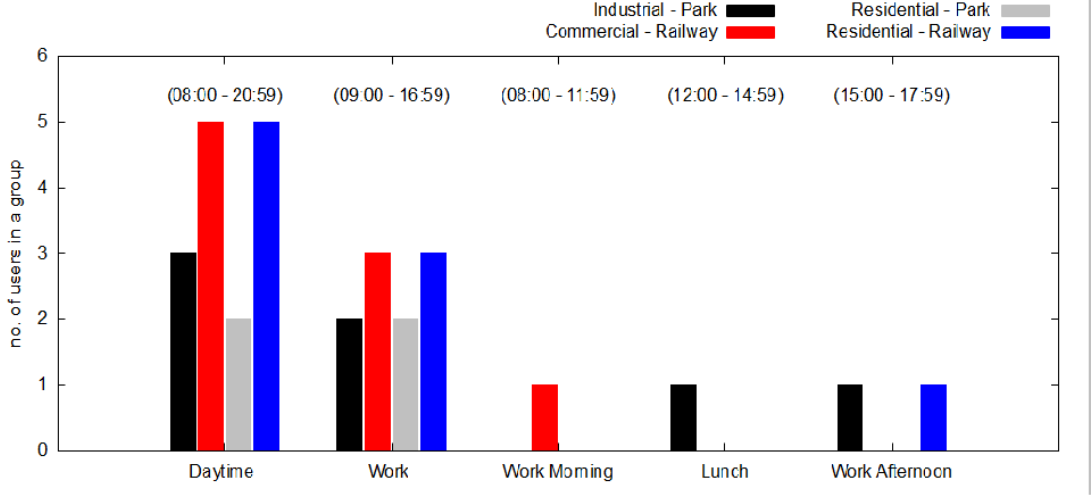


Figure 6.5: Comparison of groups of people found moving together between locations for different activities

6.1.6 Similarity in user mobility (trajectory) for routine 3

By using our grid similarity to define group mobility, we analysed users' trajectories to see the effects of time on their mobility patterns. Figure 6.6 illustrates the users' trajectories moving from Commercial to Railway during the Work (Morning) (08:00-11:59) and demonstrate that 3 users that were moving together during (09:00-16:59) (see Figure 6.4) do not maintain the same behaviour in Work (Morning) (08:00-11:59). Only one user out of 3 users is moving from Commercial to Railway. Hence, our results show no similarity in users' mobility between these locations.

Our results show no group mobility between these locations but the upside results indicate that there are groups of users that co-existed at some point of the day for a particular duration.

6.2 Exploring group mobility across activities

We now know that no group of users moved between locations within Work (Morning) (08:00-11:59), Lunch break (12:00-14:59), Work (Afternoon) (15:00-17:59), Rush Hours (Afternoon) (18:00-20:59) activities of Routine 3 which are granular time frames of DayTime (08:00-20:59) activity of Routine 1. However, so far, we also know there are

6. RESULTS

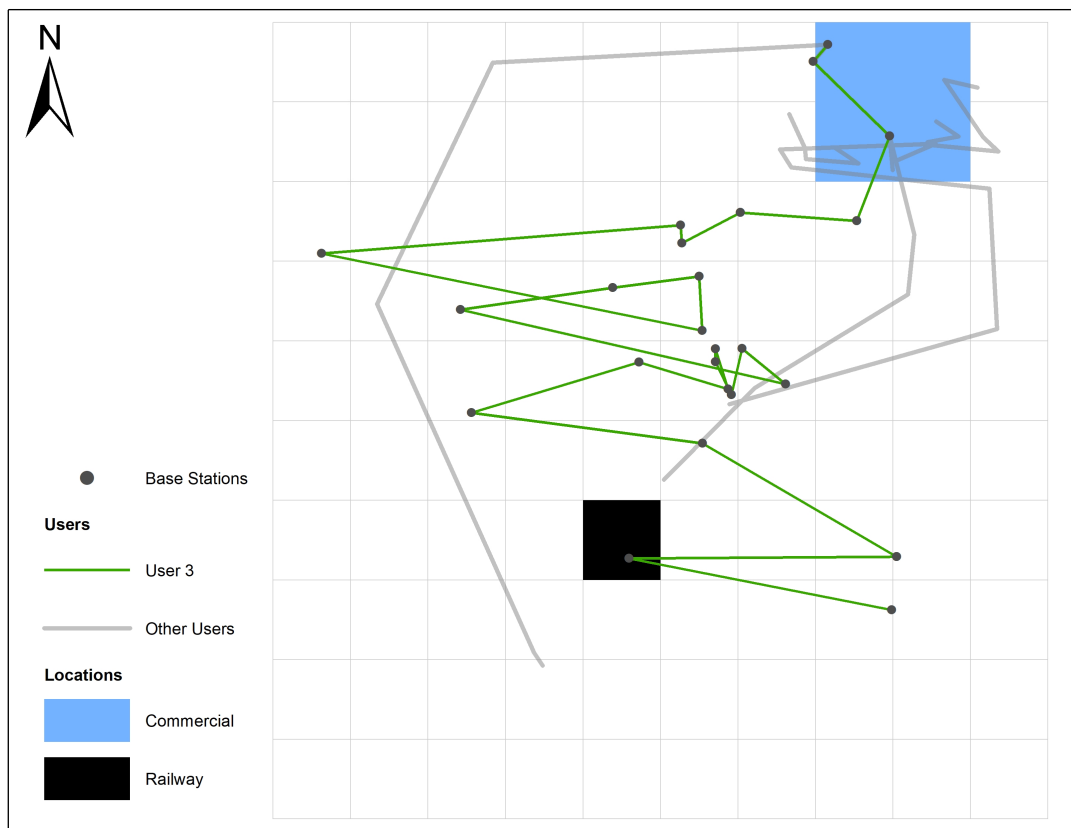


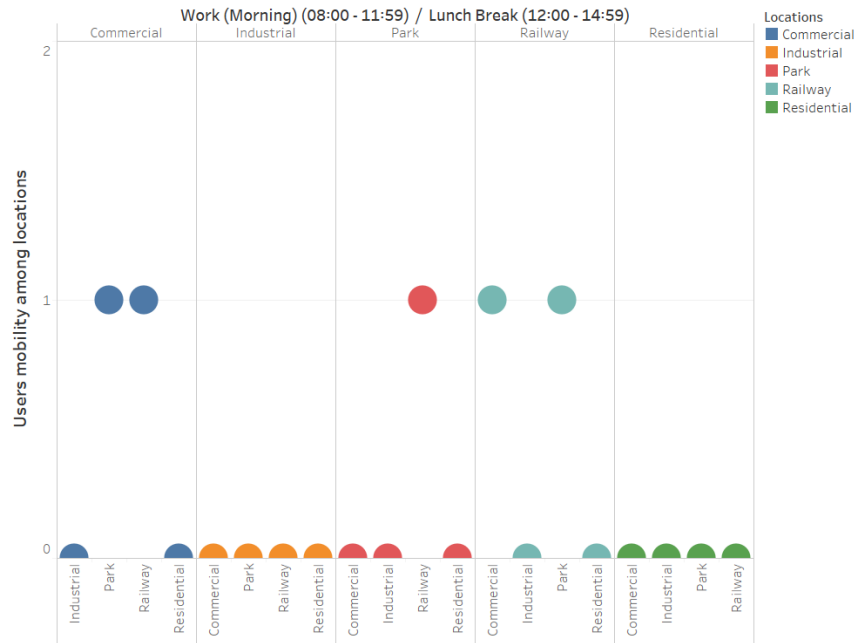
Figure 6.6: Trajectory of users moving from commercial to railway station during work (morning)(08:00-11:59)

groups of users that stay static in each location during granular time frames of Routine 3.

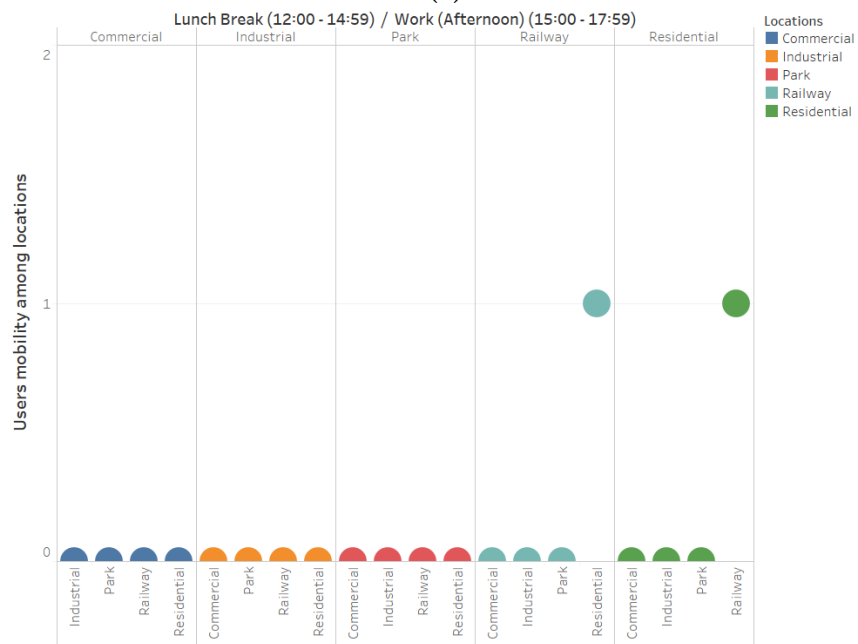
Based on our findings, we look at group mobility from the preservative of the users' routine and examine group mobility. For example, there are groups of users in the commercial area during the work morning hours, and they are static and together, which indicates that they work in the same place. We explored if they work together how their mobility will change during lunch hours. This means we explored group mobility across activities of Routine 3. Figure 6.7 illustrates the results of group mobility across Work (Morning) (08:00-11:59) to Lunch Break (12:00-14:59) and Lunch Break (12:00-14:59) to Work (Afternoon) (15:00-17:59) showing the number of users moved between locations.

Figure 6.7a demonstrates that one user moved from Park to Commercial and another user moved from Railway to Commercial between work morning and lunch break,

6.2 Exploring group mobility across activities



(a)



(b)

Figure 6.7: Users mobility across different time activities of routine 3 a) work (morning) (08:00 - 11:59)/ lunch break (12:00-14:59) and, b) lunch break (12:00-14:59) / work (afternoon) (15:00-17:59)

6. RESULTS

formulating a pair at Commercial. Likewise, one user from Park and Commercial each moved to Railway to form another pair. One user moved from Railway to Park. Our results indicate two users are moving from two different locations and reach the same location to form a pair. Figure 6.7b demonstrates the group users mobility between lunch break to working evening hours. The results we obtained indicate no group mobility. One user mobility was found in between Residential to Railway and Railway to Residential each. We compared the results from both Work (Morning) (08:00-11:59) to Lunch Break (12:00-14:59) and Lunch Break (12:00-14:59) to Work (Afternoon) (15:00-17:59) results to see if the pair found at Commercial and Railway moved together from these locations. Our analysis indicates no mobility, which means they stay together at one location for the whole period of Lunch break and Work Afternoon.

Our analysis gives insight into the user mobility patterns between the locations across different time frames, representing users' daily activities. Our results show no group mobility between locations across different activities. However, it gives us insight into how users moved from different locations to one location. Results also indicate that users' mobility or staying at one place together is dependent on locations and temporal characteristics.

While groups were identified statically in specific locations, our results indicate that group mobility is limited to pairs of users at most when moving between locations. No group mobility was found, where a large number of users were identified. Our observations showed the effects on users' mobility in the context of locations and time.

6.3 Summary

We evaluated users moving as a group from one location to another in different time frames by conducting a controlled experiment. We analysed the results of our experiment to explore group mobility in 5 specific locations. While groups were identified statically in specific locations, our results indicate that group mobility is limited to pairs of users at most when moving between locations. No group mobility was found, where a large number of users were identified. Our observations showed the effects on users' mobility in the context of locations and time.

In the next Chapter, we discuss the limitations and possible implications of our work.

6. RESULTS

7

Discussion

This Chapter discusses the implications and limitations of our work.

7.1 On group mobility

Our results suggest that monitoring group mobility is difficult using our current methodology. Our analysis revealed the effects on users' mobility in terms of location and time. However, our findings reveal that groups of users are easy to track in static locations rather than on the move (moving between locations).

7.2 Room for improvement

We analyse mobility of individual users using default routines that can be found freely in available software packages. As a result, the accuracy of the trajectories produced is low. However, it is possible to increase the accuracy and precision in these trajectories by using more sophisticated algorithms found in the art (67), e.g., Kalman filter. Since our work was about tracking a group of users moving together, we left the implementation of these algorithms for future work.

In addition, since our dataset contains only one-day mobility measurements, it is difficult to validate the accuracy of repetitive routines over time. Moreover, during the mobility analysis, we found that almost half of users' mobility was either between two locations or were static. To verify this, data from different days could lead to more conclusive results.

7.3 Regularity of groups

Our present strategy aims to investigate user group mobility based on locations, mobility, and time. All of these components are combined in our method to grasp users' mobility by shifting their locations and times. These metrics, however, can be enhanced in the future. Regularity of users can help improve their mobility. By verifying users' mobility on a day-to-day basis, we can create stable trajectories for the users and based on the commonly visited sites, we can have a greater chance at defining better users' routines. By generalising group mobility of users, it is possible to envision new applications that rely on group size for effective processing of tasks in a collaborative manner, e.g., processing and sensing tasks.

8

Summary and Conclusion

In this thesis, we used a crowdsensed dataset (from a cellular operator) to analyse group mobility (aka multiple users moving together). We designed a grid-like technique that can transform crowdsensed data into trajectories that explain human mobility of individual users. With this information, we then analyse the mobility patterns of multiple users between locations, e.g., Residential and Commercial locations. By contextualising our results into daily routines to explain mobility patterns, we found that while multiple groups can be formed in specific locations (in a static manner), it is difficult to find users that share the same spatial and temporal characteristics while moving.

8. SUMMARY AND CONCLUSION

Bibliography

- [1] S. Höffken, B. Streich, Mobile participation: Citizen engagement in urban planning via smartphones, in: *Citizen E-Participation in urban governance: crowdsourcing and collaborative creativity*, IGI Global, 2013, pp. 199–225. 3
- [2] D. Shin, D. Aliaga, B. Tunçer, S. M. Arisona, S. Kim, D. Zünd, G. Schmitt, Urban sensing: Using smartphones for transportation mode classification, *Computers, Environment and Urban Systems* 53 (2015) 76–86. 3
- [3] Y. Zhou, B. P. L. Lau, C. Yuen, B. Tunçer, E. Wilhelm, Understanding urban human mobility through crowdsensed data, *IEEE Communications Magazine* 56 (11) (2018) 52–59. 3
- [4] M. C. Gonzalez, C. A. Hidalgo, A.-L. Barabasi, Understanding individual human mobility patterns, *nature* 453 (7196) (2008) 779–782. 3, 7
- [5] J. Weppner, P. Lukowicz, Bluetooth based collaborative crowd density estimation with mobile phones, in: *2013 IEEE international conference on pervasive computing and communications (PerCom)*, IEEE, 2013, pp. 193–200. 3
- [6] G. Bajaj, G. Bouloukakis, A. Pathak, P. Singh, N. Georgantas, V. Issarny, Toward enabling convenient urban transit through mobile crowdsensing, in: *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, IEEE, 2015, pp. 290–295. 3
- [7] F. Kalim, J. P. Jeong, M. U. Ilyas, Crater: A crowd sensing application to estimate road conditions, *Ieee access* 4 (2016) 8317–8326. 3
- [8] K. Stawarz, A. L. Cox, A. Blandford, Don't forget your pill! designing effective medication reminder apps that support users' daily routines, in: *Proceedings of*

BIBLIOGRAPHY

- the SIGCHI Conference on Human Factors in Computing Systems, 2014, pp. 2269–2278. 3
- [9] M. W. Traunmueller, N. Johnson, A. Malik, C. E. Kontokosta, Digital footprints: Using wifi probe and locational data to analyze human mobility trajectories in cities, *Computers, Environment and Urban Systems* 72 (2018) 4–12. 3
- [10] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, X. Zhou, Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm, *ACM computing surveys (CSUR)* 48 (1) (2015) 1–31. 3
- [11] R. K. Ganti, F. Ye, H. Lei, Mobile crowdsensing: current state and future challenges, *IEEE communications Magazine* 49 (11) (2011) 32–39. 3, 9
- [12] Y. Wang, X. Jia, Q. Jin, J. Ma, Mobile crowdsourcing: framework, challenges, and solutions, *Concurrency and Computation: Practice and experience* 29 (3) (2017) e3789. 3
- [13] H. Ma, D. Zhao, P. Yuan, Opportunities in mobile crowd sensing, *IEEE Communications Magazine* 52 (8) (2014) 29–35. 3, 8
- [14] L. Wang, D. Zhang, Y. Wang, C. Chen, X. Han, A. M'hamed, Sparse mobile crowd-sensing: challenges and opportunities, *IEEE Communications Magazine* 54 (7) (2016) 161–167. 3, 8, 10
- [15] A. Zuniga, H. Flores, E. Lagerspetz, P. Nurmi, S. Tarkoma, P. Hui, J. Manner, Tortoise or hare? quantifying the effects of performance on mobile app retention, in: *The World Wide Web Conference*, 2019, pp. 2517–2528. 4
- [16] G. Sagl, M. Loidl, E. Beinat, A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic, *ISPRS International Journal of Geo-Information* 1 (3) (2012) 256–271. 7
- [17] K. Farrahi, D. Gatica-Perez, Discovering routines from large-scale human locations using probabilistic topic models, *ACM Transactions on Intelligent Systems and Technology (TIST)* 2 (1) (2011) 1–27. 7

- [18] J. Yuan, Y. Zheng, X. Xie, Discovering regions of different functions in a city using human mobility and pois, in: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012, pp. 186–194. 7
- [19] P. Galpern, A. Ladle, F. Alaniz Uribe, B. Sandalack, P. T. Doyle-Baker, Assessing urban connectivity using volunteered mobile phone gps locations, Applied Geography 93 (2018) 37–46. doi:10.1016/j.apgeog.2018.02.009. 7
- [20] H. Barbosa, M. Barthelemy, G. Ghoshal, C. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, M. Tomasini, Human mobility: Models and applications, Physics Reports 734. doi:10.1016/j.physrep.2018.01.001. 7
- [21] Z. Ebrahimpour, W. Wan, J. García, O. Cervantes Villagómez, L. Hou, Analyzing social-geographic human mobility patterns using large-scale social media data, ISPRS International Journal of Geo-Information 9 (2020) 125. doi:10.3390/ijgi9020125. 8
- [22] W. Li, S. Wang, Z. Xiaoyi, Q. Jia, Y. Tian, Understanding intra-urban human mobility through an exploratory spatiotemporal analysis of bike-sharing trajectories, International Journal of Geographical Information Science 34 (2020) 1–24. doi:10.1080/13658816.2020.1712401. 8
- [23] Shi, Chi, Liu, Human mobility patterns in different communities: a mobile phone data-based social network approach, Annals of GIS 21. doi:10.1080/19475683.2014.992372. 8
- [24] S. Lu, Y. Huang, C. Shi, X. Yang, Exploring the associations between urban form and neighborhood vibrancy: A case study of chengdu, china, ISPRS International Journal of Geo-Information 8 (2019) 165. doi:10.3390/ijgi8040165. 8
- [25] A. Psaltoglou, E. Calle, Enhanced connectivity index – a new measure for identifying critical points in urban public transportation networks, International Journal of Critical Infrastructure Protection 21. doi:10.1016/j.ijcip.2018.02.003. 8
- [26] S. Hasan, C. M. Schneider, S. V. Ukkusuri, M. C. González, Spatiotemporal patterns of urban human mobility, Journal of Statistical Physics 151 (1) (2013) 304–318. 8

BIBLIOGRAPHY

- [27] L. Alessandretti, U. Aslak, S. Lehmann, The scales of human mobility, *Nature* 587 (7834) (2020) 402–407. 8
- [28] A. J. Oliner, A. P. Iyer, I. Stoica, E. Lagerspetz, S. Tarkoma, Carat: Collaborative energy diagnosis for mobile devices, in: *Proceedings of the 11th ACM conference on embedded networked sensor systems*, 2013, pp. 1–14. 8
- [29] R. Gao, M. Zhao, T. Ye, F. Ye, Y. Wang, K. Bian, T. Wang, X. Li, Jigsaw: Indoor floor plan reconstruction via mobile crowdsensing, in: *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, pp. 249–260. 9
- [30] Z. Peng, S. Gao, B. Xiao, S. Guo, Y. Yang, Crowdgis: Updating digital maps via mobile crowdsensing, *IEEE Transactions on Automation Science and Engineering* 15 (1) (2017) 369–380. 9
- [31] J. Liu, H. Shen, X. Zhang, A survey of mobile crowdsensing techniques: A critical component for the internet of things, 2016. doi:10.1109/ICCCN.2016.7568484. 9
- [32] R. Ganti, F. Ye, H. Lei, Mobile crowd sensing: Current state and future challenges, *IEEE Communications Magazine* 49 (2011) 32–39. doi:10.1109/MCOM.2011.6069707. 9
- [33] A. M. Ali Al-muqarm, F. Rabee, Iot technologies for mobile crowd sensing in smart cities, *Journal of Communications* 14 (2019) 745–757. doi:10.12720/jcm.14.8.745-757. 9
- [34] C. Zhang, K. pathapati subbu, J. Luo, J. Wu, Groping: Geomagnetism and crowdsensing powered indoor navigation, *IEEE Transactions on Mobile Computing* 14 (2015) 387–400. doi:10.1109/TMC.2014.2319824. 9
- [35] A. Zopiatis, A. Theocharous, P. Kosmas, C. Webster, Y. Melanthiou, Developing a country-wide tourist loyalty scheme: A barren landscape, *International Journal of Tourism Research* 18. doi:10.1002/jtr.2075. 9
- [36] Y. Sun, H. Song, A. J. Jara, R. Bie, Internet of things and big data analytics for smart and connected communities, *IEEE Access* 4 (2016) 1–1. doi:10.1109/ACCESS.2016.2529723. 9

- [37] K. Abualsaud, T. El-Fouly, T. Khattab, E. Yaacoub, L. Ismail, M. Ahmed, M. Guizani, A survey on mobile crowd-sensing and its applications in the iot era, *IEEE Access PP* (2018) 1–1. doi:10.1109/ACCESS.2018.2885918. 10
- [38] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, J. Eriksson, Vtrack: Accurate, energy-aware road traffic delay estimation using mobile phones, 2009, pp. 85–98. doi:10.1145/1644038.1644048. 10
- [39] G. Virone, A. Wood, L. Selavo, Q. Cao, L. Fang, T. Doan, Z. He, R. Stoleru, S. Lin, J. Stankovic, An advanced wireless sensor network for health monitoring, *Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare (D2H2)*. 10
- [40] F. Laws, C. Scheible, H. Schütze, Active learning with amazon mechanical turk, in: *EMNLP*, 2011. 10
- [41] A. Acock, Working with missing values, *Journal of Marriage and Family* 67 (2005) 1012 – 1028. doi:10.1111/j.1741-3737.2005.00191.x. 10
- [42] P. Allison, Multiple imputation for missing data: A cautionary tale, *Sociological Methods Research* 28. doi:10.1177/0049124100028003003. 10
- [43] P. Lodder, To Impute or not Impute: That’s the Question, 2014. 10
- [44] U. Kar, D. K. Sanyal, An overview of device-to-device communication in cellular networks, *ICT Express* 4 (2018) 203–208. 12
- [45] A. Asadi, Opportunistic device-to-device communication in cellular networks: From theory to practice, *CoRR abs/1610.08293*. arXiv:1610.08293. URL <http://arxiv.org/abs/1610.08293> 12
- [46] A. Asadi, V. Mancuso, A survey on opportunistic scheduling in wireless communications, *IEEE Communications Surveys Tutorials* 15 (4) (2013) 1671–1688. doi:10.1109/SURV.2013.011413.00082. 12
- [47] H. Flores, A. Zuniga, F. Faghihi, X. Li, S. Hemminki, S. Tarkoma, P. Hui, P. Nurmi, Cosine: Collaborator selector for cooperative multi-device sensing and computing,

BIBLIOGRAPHY

- in: 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom), 2020, pp. 1–10. doi:10.1109/PerCom45495.2020.9127364. 12, 15, 24
- [48] M. Conti, S. Giordano, M. May, A. Passarella, From opportunistic networks to opportunistic computing, *Communications Magazine, IEEE* 48 (2010) 126 – 139. doi:10.1109/MCOM.2010.5560597. 12, 15
- [49] Y. Lee, Y. Ju, C. Min, S. Kang, I. Hwang, J. Song, Comon: Cooperative ambience monitoring platform with continuity and benefit awareness, *MobiSys'12 - Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services* doi:10.1145/2307636.2307641. 13, 14
- [50] J. Zhang, M. Scardamalia, R. Reeve, R. Messina, Designs for collective cognitive responsibility in knowledge-building communities, *Journal of the Learning Sciences* 18 (2009) 44 – 7. 13
- [51] H. S. Guruprasad, b. H, Resource provisioning techniques in cloud computing environment: A survey, *International Journal of Research in Computer and Communication Technology* 3 (2014) 395–401. 13
- [52] J. Leppanen, M. Pelkonen, H. Guo, S. Hemminki, P. Nurmi, S. Tarkoma, Collaborative and energy-efficient speech monitoring on smart devices, *Computer* 49 (2016) 22–30. doi:10.1109/MC.2016.380. 13
- [53] E. Lagerspetz, J. Hamberg, X. Li, H. Flores, P. Nurmi, N. Davies, S. Helal, Pervasive data science on the edge, *IEEE Pervasive Computing* 18 (3) (2019) 40–48, ©2019 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. doi:10.1109/MPRV.2019.2925600. 13, 14
- [54] D. Lin, Y. Tang, Blockchain consensus based user access strategies in d2d networks for data-intensive applications, *IEEE Access PP* (2018) 1–1. doi:10.1109/ACCESS.2018.2881953. 13

- [55] A. Amiri Sani, K. Boos, M. Yun, L. Zhong, Rio: A system solution for sharing i/o between mobile systems, *MobiSys 2014 - Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services* doi: 10.1145/2594368.2594370. 13
- [56] C. Julien, C. Liu, A. L. Murphy, G. P. Picco, Blend: Practical continuous neighbor discovery for bluetooth low energy, in: *2017 16th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2017, pp. 105–116. 13
- [57] H. Flores, R. Sharma, D. Ferreira, V. Kostakos, J. Manner, S. Tarkoma, P. Hui, Y. Li, Social-aware hybrid mobile offloading, *Pervasive Mob. Comput.* 36 (2017) 25–43. 14
- [58] K. Habak, M. Ammar, K. Harras, E. Zegura, Femtoclouds: Leveraging mobile devices to provide cloud service at the edge, 2015. doi:10.1109/CLOUD.2015.12. 14
- [59] M. Vuran, O. Akan, I. Akyildiz, Spatio-temporal correlation: Theory and applications for wireless sensor networks, *Computer Networks* 45 (2004) 245–259. doi:10.1016/j.comnet.2004.03.007. 14
- [60] J. Scott, P. Hui, J. Crowcroft, C. Diot, C.: Hagggle: A networking architecture designed around mobile users, *Proceedings of the Third Annual Conference on Wireless On-demand Network Systems and Services*. 14
- [61] D. Chatzopoulos, C. Bermejo, E. u. Haq, Y. Li, P. Hui, D2d task offloading: A dataset-based q amp;a, *IEEE Communications Magazine* 57 (2) (2019) 102–107. doi:10.1109/MCOM.2018.1700873. 14
- [62] J. Kwak, J. Kim, S. Chong, Proximity-aware location based collaborative sensing for energy-efficient mobile devices, *IEEE Transactions on Mobile Computing* 18 (2) (2019) 417–430. doi:10.1109/TMC.2018.2833842. 14
- [63] H. Flores, A. Zuniga, L. Tonetto, T. Braud, P. Hui, Y. Li, S. Tarkoma, M. Ammar, P. Nurmi, Collaboration stability: Quantifying the success and failure of opportunistic collaboration 1–8. 15

BIBLIOGRAPHY

- [64] L. Ekroot, T. Cover, The entropy of markov trajectories, *IEEE Transactions on Information Theory* 39 (4) (1993) 1418–1421. doi:10.1109/18.243461. 15
- [65] N. Correia, C. Lopes, J. Hawkey, S. Oliveira, O. Perriquet, Personal routine visualization using mobile devices, in: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, 2012, pp. 1–3. 20
- [66] F. Xia, J. Wang, X. Kong, Z. Wang, J. Li, C. Liu, Exploring human mobility patterns in urban scenarios: A trajectory data perspective, *IEEE Communications Magazine* 56 (3) (2018) 142–149. doi:10.1109/MCOM.2018.1700242. 40
- [67] Q. Cheng, C. Wang, A method of trajectory prediction based on kalman filtering algorithm and support vector machine algorithm, in: *Chinese Intelligent Systems Conference*, Springer, 2017, pp. 495–504. 57

9

Appendix

9.1 Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, Ali Raza,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

Exploring Group Mobility,

supervised by Huber Flores.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.
3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.
4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

9. APPENDIX

Ali Raza

04/8/2021