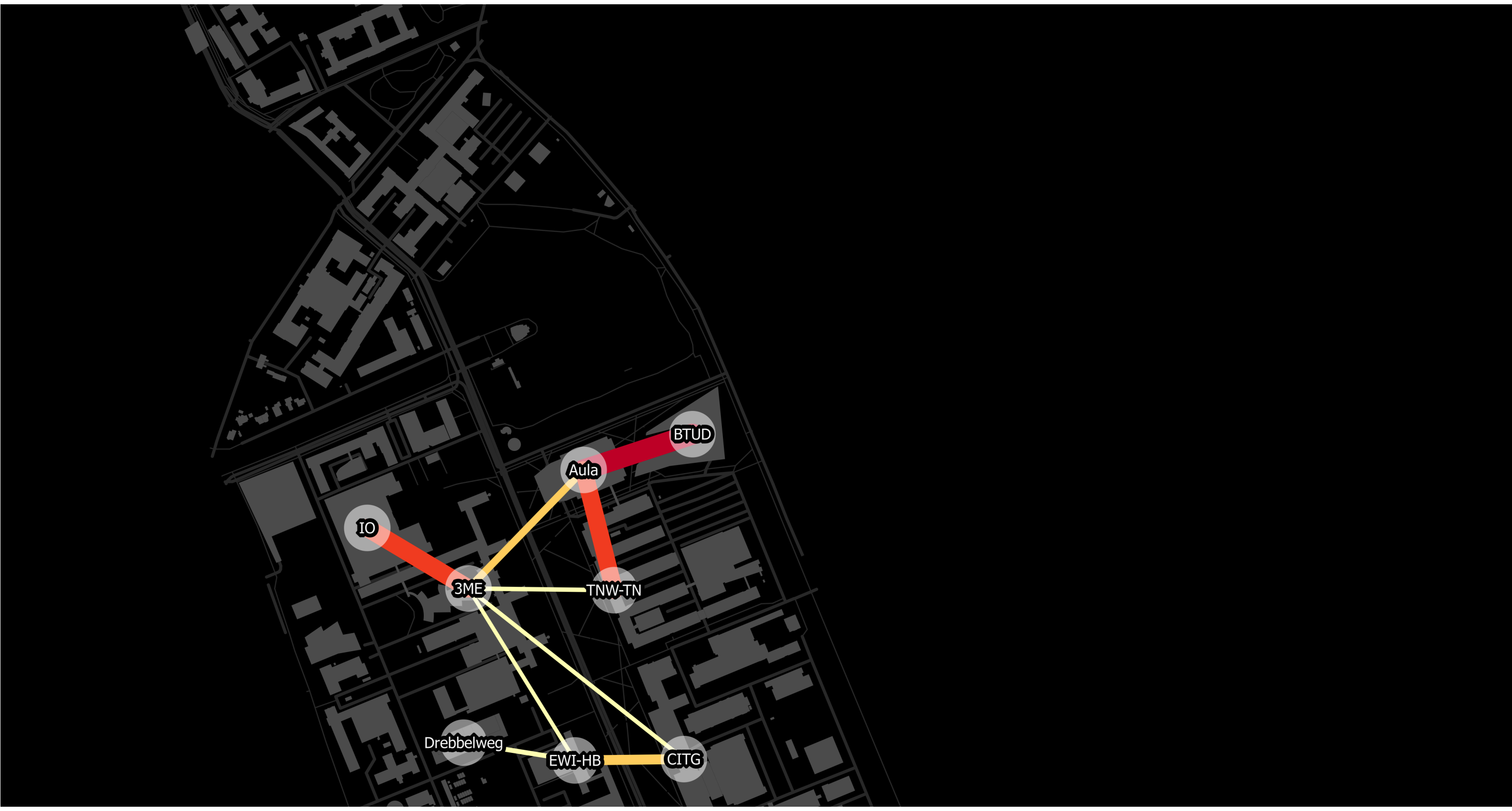
Identifying movement patterns from large scale Wifi-based location data

A case study of the TU Delft Campus

M. P. Bon
X. A. den Duijn
B. Dukai
S. J. Griffioen
Y. Kang
M. Vermeer



TU Delft

Identifying movement patterns from large scale Wifi-based location data

A case study of the TU Delft Campus

by

M. P. Bon
X. A. den Duijn
B. Dukai
S. J. Griffioen
Y. Kang
M. Vermeer

Synthesis project of Geomatics
at the Delft University of Technology,

Supervisor: Ir. E. Verbree

Project duration: April 18, 2016 – June 17, 2016

to be presented on Friday June 17, 2016.

TU Delft

Preface

During the fourth quarter of the first year of the MSc Programme Geomatics for the Built Environment at the TU Delft, the Geomatics Synthesis Project (GSP) takes place. This report is part of this framework and in this project, students will apply all their knowledge they have acquired during the courses while working in groups of five or six students. The students will gain experience throughout the entire process of project management, data processing, data analysis, application and presentation.

This year, the GSP focusses on Wi-Fi tracking data from the eduroam network of the TU Delft. The student will be divided into three groups, each researching one of three different topics:

- Identifying occupancy
- Identifying movement patterns
- Identifying activities

This project is dedicated to the second topic, identifying movement patterns. The project requires 3 main documents: ***1*)** the baseline review; ***2*)** the mid term review, and ***3*)** the final review This document embodies the final review and was created to provide the students, the supervisor(s) and other involved parties with an overview of the project. The document includes the problem description, development process, results, conclusions and recommendations for future work.

Delft, University of Technology

M. P. Bon
X. A. den Duijn
B. Dukai
S. J. Griffioen
Y. Kang
M. Vermeer
Delft, June 2016

Contents

1 Glossary

Used terms and abbreviations:

Building and Faculty names	
AE / LR	Aerospace Engineering
BK / BK City	Faculty of Architecture and the Built Environment
CiTG	Faculty of Civil Engineering and Geosciences
EGM	Thermal Power Plant
EWI / EEMCS	Faculty of Electrical Engineering, Mathematics and Computer Science
FMVG / FMRE	Facility Management & Real Estate
HSL	Hypersonic Wind Tunnel
ID / IO	Faculty Industrial Design Engineering
ISD	International School Delft
LMS	Logistics and Environmental Services
LSL	Low Turbulence Tunnel
O&S	Onderwijs & Studentenzaken
RID	Reactor Institute Delft
SC	Sports & Culture
TNW	Faculty of Applied Sciences
TPM	Faculty of Technology, Policy and Management

Abbreviations	
AP / APs	Access Point(s)
GNSS	Global Navigation Satellite System
GSP	Geomatics Synthesis Project
RFID	Radio frequency identification
RSSI	Received Signal Strength Indicator
SNR	Signal to Noise Ratio
SQL	Structured Query Language
WLAN	Wireless Local Area Network

Commonly used terms

eduroam	Wireless network available at the TU Delft, used internationally.
Movement	A movement is always from the location of one state to the location of another state, where two states can not be the same.
Pattern	Recurring event that helps in the identification of phenomena.
Region / Building-part	Building parts and regions refer to large indoor areas that can be grouped together, i.e. 'Staff area', 'Atelier'.
Sequence	Ordered collection of states.
Spatial level	A spatial level defines the level on which states are aggregated
States / Stay places	A state is defined as a time interval during which a particular device is located in a certain area.
Trajectory	A trajectory is defined for each person by an ordered list of buildings that were visited.
World	Location which, depending on the spatial level, can be either outside a building or outside the campus

2
Executive Summary

2.1. Introduction

Location is a key element of many processes and activities, and the understanding of human movement behaviour is becoming increasingly important. Knowledge of people's locations and related mobility patterns are important for numerous activities, such as urban planning, transport planning and facility management. How to efficiently use the available space, is a common problem in many fields. In the educational sector, universities are struggling to meet the higher expectations of facilities for education and research by students and academic staff. Managing the campus of a university has become a complex and challenging task, including the involvement of many stakeholders. Campus managers are in need for evidence-based information to support their decision making (Heijer den 2012). This includes better location data to detect activities, occupancy and usage of the infrastructure.

To understand the human motion behaviour many studies are conducted based on data collection of GPS receivers. The Global Navigation Satellite System (GNSS) is commonly used to track people in large scale environments. Spek 2008 studied the movement of pedestrians in city centres, where potential participants were asked to carry a GPS receiver. However, the distribution of GPS devices to participants limits the possibilities to collect location data at a large scale. Furthermore, due to poor quality of received signals from satellites in indoor environments, GPS receivers are not suitable in these conditions. Technological developments in the acquisition of location data by smart phones and the use of Wi-Fi networks, enables new opportunities to track users.

Wireless Local Area Networks (WLAN) are widely used for indoor positioning of mobile devices within this network. The use of the Wi-Fi network to estimate the location of people is an attractive approach, since Wi-Fi access points (AP) are often available in indoor environments. Furthermore, smart phones are becoming essential in daily life, making it convincing to track mobile devices. This provides a platform to track people by using WLAN as a sensor network, and study the mobility of users inside buildings or groups of buildings.

At Delft University of Technology (TU Delft) a large scale Wi-Fi network is deployed across all facilities covering the indoor space of the campus. The network is known as an international roaming service for users in educational environments and is called the eduroam network. It allows students and staff members from the university to use the infrastructure throughout the campus for free. This enables the possibility to collect Wi-Fi logs, including individual scans of mobile devices, at a large scale. A continuous collection of re-locations of devices to access points for a long duration will return detailed records of people's movement. This ubiquitous and individual georeferenced data derived from smart phones will present valuable knowledge about the movement on the campus. Several work has been made for studying human mobility patterns in a University's campus. Meneses and Moreira 2012 used the eduroam network to study connectivity between two places, by computing the number of movements between two places within a given observation time period. Previous work has also been made at TU Delft (Kalogianni et al. 2015), where several Wi-Fi monitors were placed to detect occupation and movement between different faculties.

In this paper, we attempt to identify people's movement patterns from the eduroam network of TU Delft. Other than previous studies, this research-driven project analysed data from more than 30.000 users, and tries to detect movement patterns between buildings, and between large indoor regions. The project is carried out in request of the university's department of Facility Management and Real Estate (FMRE). With this project, we try to illustrate to what extend movement patterns in and between buildings can be identified from anonymised Wi-Fi logs. Firstly, individual states are extracted from the Wi-Fi logs, where users stay for a longer time period. Secondly, movements are detected between a sequence of states. Thirdly, movement patterns can be identified by counting the amount of movement from, to or between certain locations at different time intervals.

The aim of this paper is not to improve a Wi-Fi based positioning technique, but to use the location data to conduct a mobility analysis producing knowledge about the University's campus. Based on the three steps mentioned above, the aim of this project is to provide a method to detect movement patterns from anonymised Wi-Fi logs. This includes the separation of mobile devices (i.e. smart phones) and static devices (e.g. laptops) from the Wi-Fi logs, and detecting movement to and from beyond the spatial extent of the eduroam network by introducing the concept of a 'world' state. Hereby, this paper attempts to contribute with a method to automatically mine people's movement patterns at two spatial levels. First, movement at building level is analysed. Subsequently, indoor movement at building-part level is studied, by constructing a network graph of the underlying building floorplan. The structure of this paper is as follows. Section 2.1.1, describes the case study of TU Delft, the tracking technique and the acquired data that is used in the study. In section 2.2 we present our methodology. Section 2.3 discusses the obtained results. Finally, in section 2.4, we present our concluding remarks and recommendations.

2.1.1. Case description

The project's main area of interest is the campus of Delft University of Technology (TU Delft), used by more than 30.000 students and staff members. The eduroam network of the TU Delft campus consists of 1730 access points, distributed over more than 30 buildings, covering all indoor space. Even large outdoor areas around the buildings have access to the Wi-Fi network, because of the range of APs. Connection to the Wi-Fi eduroam network is free of charge and requires only a NetID (i.e. username and password), which all students and staff get upon registration at the university. Every time a user accesses the network, the connection is logged. When the connected device moves from one AP to another, a new log is done. The location of the AP a mobile device is connected to, will give an estimation of the mobile devices' location, and thus the person. This allows the tracking of devices in space and time by relating buildings and building-parts to an aggregation of APs.

The data is collected for every single AP over a period of almost two months. The logs are stored in a database on a virtual server at regular intervals of 5 minutes. In order to ensure privacy, MAC addresses and NetIDs (i.e. usernames) are hashed. Every log is stored with a start time, session duration, AP name and a description of the AP's location (e.g. System Campus > 20-Aula > 2nd floor). The AP name always contains the ID of the building it is located in. We can use this ID to locate APs at building level. For the Faculty of Architecture and the Built Environment, we also had information about the exact physical position of each AP. This geo-referenced information is used to analyse movement at building-part level.

2.2. Methodology

In this section the data mining methods used to retrieve movement patterns from the Wi-Fi log at two different spatial levels will be described. Figure 2.1 gives an overview of the main workflow, starting with the TU Delft eduroam Wi-Fi log and ending with movement patterns. The two spatial levels for which movement patterns will be derived are 'building' and 'building-part' level. The movement patterns on building level concern the movement from, to and between the buildings on the campus. The movement patterns on building-part level concern the movement from, to and between building-parts of the faculty of Architecture and the Built Environment. First subsection 2.2.1 will describe the extraction of mobile devices. The reason that mobile devices are extracted is that the records of mobile devices (e.g. smart phones) are more complete and representative for the actual movement of the corresponding person then records of static devices (e.g. laptops). subsection 2.2.2 will describe how the raw data of the Wi-Fi log is preprocessed to retrieve clean sessions for both building and building-part level. A session is defined as time interval during which a device is connected to one access point (AP). In subsection 2.2.3 it will be explained how states are created by grouping subsequent sessions that share the same location. A state is defined as a time interval during which a device is located in a certain building or building-part. A key part in the process of state extraction is the creation of a 'world' state which allows the detection of movement from and to campus. subsection 2.2.4 addresses how the resulting states are used to retrieve movements at both spatial levels. A movement is defined by the change from one state to the next subsequent state, where the different states must be at a different locations. Finally subsection 2.2.5 describes how the movements are used to derive and visualize movement patterns.

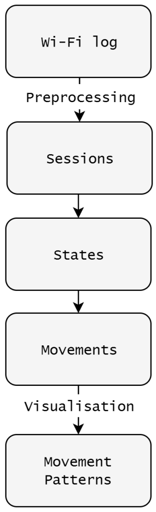


Figure 2.1: Workflow

2.2.1. Mobile device extraction

The Wi-Fi log contains data of different device types, as any device that makes a connection to eduroam will be stored in the log. A distinction can be made between mobile and static devices. Mobile devices, such as smart phones, are usually switched on during the entire day and are usually carried by the user. Static devices, such as laptops, are mostly only switched on during particular periods when a person is stationary for a longer period of time. Furthermore, they are likely to be left by the user for certain time periods. Therefore, the tracking of mobile devices gives more accurate information about the patterns of users than the tracking of static devices. As a result a distinction should be made between the two device types, so that static devices can be filtered out. This distinction can be made based on the knowledge that mobile device are more likely to have very short sessions in the log as they continuously connect to new APs when a person moves around. As a result the mobility of a device can be defined by the ratio between the amount of short, 5 minute, sessions in the Wi-Fi log and the total amount of sessions in the Wi-Fi log.

$$Mobility\ ratio = \frac{number\ of\ short\ sessions}{total\ number\ of\ sessions}$$

Figure 6.7 shows a histogram of the mobility ratio of all devices. The two distinctive peaks corresponding to the static and mobile devices can clearly be identified. The mobility ration of all devices is stored in a separate table, enabling filtering out of static devices at any point during the process.

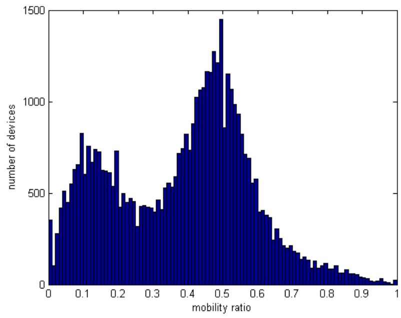


Figure 2.2: Histogram of mobility ratio of all devices in the Wi-Fi log

2.2.2. Preprocessing raw Wi-Fi log sessions

For each session in the Wi-Fi log the name of the access point is stored. By linking this name to a location the data becomes valuable for detecting movement patterns. For some APs however the location is unknown, these are filtered out. The other records are related to a location, both on building and on building-part level. For building level the location can easily be retrieved as the building ID is part of the AP name. These IDs are linked to the corresponding building polygons of a topographical map. The location of each building is the center point of these polygons. For the Faculty of Architecture and the Built Environment the floor plan with the locations of the different APs is available. Building-parts (see Figure 2.4) are defined based on the layout of the APs. Each building-part contains several APs and the relation between them is stored in the database. In this way the AP name can directly be linked to a location.

2.2.3. State extraction

To create states, subsequent sessions at the same location are grouped together. As the location is known for each session on both building and building-part level, states can be created for both spatial levels. Figure 2.3 illustrates how sessions are grouped to create states. For grouping, a time threshold of one hour is used, meaning that subsequent sessions between which the time gap is less than one hour are grouped together. The reason for the one hour threshold is that gaps smaller than an hour are likely to represent a person that was just smoking or lunching outside for a short period of disconnection. If a person is not recorded for more than an hour it is more likely that the person has left the campus. To be able to retrieve this movement away from and back to the campus, 'world' states are added to the data during a time period where the person has not been recorded for more than one hour (see Figure 2.3). Finally, states are present in the data that do not represent real visits, but only people passing by a building. These short states are filtered out (see Figure 2.3).

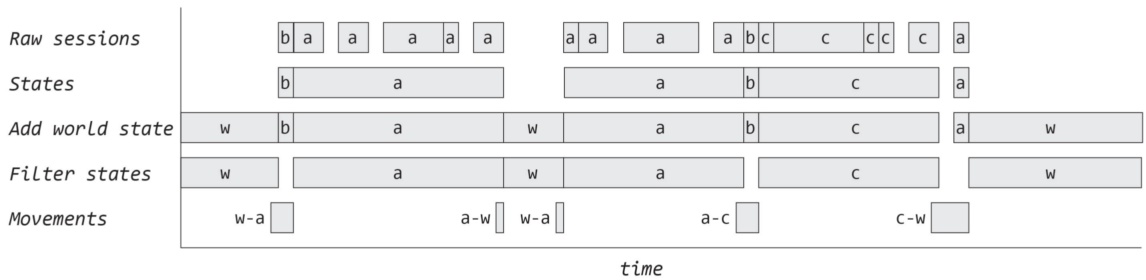


Figure 2.3: Processing steps; from raw data to movements

2.2.4. Movement extraction

The extracted states contain implicit information on the movement of the device. If a device is first located at location A and subsequently at location B it must have moved from location A to B. However, in order to be able to retrieve movement patterns, the movement should be stored explicitly. The origin and destination of the movement are defined by the locations of both states. The timing of the movement is derived by taking the end time of each state minus 5 minutes and the start time of the subsequent state (see Figure 2.3). The reason for the 5 minute subtraction, is that the last moment a device is actually recorded is 5 minutes before the end time of a state.

2.2.5. Movement pattern extraction

The final step is to extract movement patterns from the created movements. These patterns can be derived by counting the amount of movement from, to or between certain buildings and building-parts for different time intervals. To determine if a movement should be counted for a particular time interval, it is checked whether the time between start and end falls within the interval. In this way each movement can only be counted ones when comparing adjacent intervals. The amount movement is both visualized in time profiles and maps with specified time intervals. To visualize the indoor movement on a map, a network graph of the underlying building floor plan is created for the Faculty of Architecture and the Built Environment. For building level, no graph is created as the movement in outdoor space is less constrained, especially considering the spacious character of the TU Delft campus. To determine the route taken from one building-part to another the shortest path is taken using the Dijkstra algorithm.

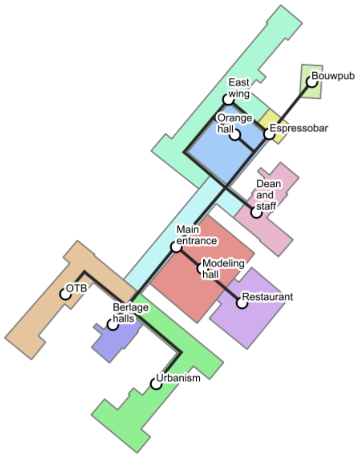


Figure 2.4: Building parts on the ground floor of the Faculty of Architecture and the built environment and its underlying graph.

2.3. Results

First subsection 2.3.1 will give a brief analysis of the data, including some general statistics. In the remainder of the chapter several movement patterns retrieved from the raw Wi-Fi log will be presented. It should be noted that many different movement patterns can be identified by counting the movement for different time intervals and for different routes. This paper aims at giving an overview of the different patterns that can be extracted. Section 2.3.2 will present the outdoor building level patterns, and subsection 2.3.3 the indoor building-part level patterns.

2.3.1. General statistics

Within the dataset 44.952 different users are present that together have 86.413 devices. Of these devices 24.156 are classified as mobile, the remaining devices are either classified as static or had less than 100 sessions in the Wi-Fi log, which was decided to be insufficient for classification. The 100 record threshold is a point of discussion, currently it is based on the reasoning that a person that has less records, is likely to be a visitor with access to eduroam, and not a regular user of the TU Delft campus. Figure 2.5 shows the amount of data during the different processing steps for both spatial levels. The reduction from session to states is mainly due to the grouping. The reduction from states to movements is mainly because the devices with less than 100 records are filtered out at this point.

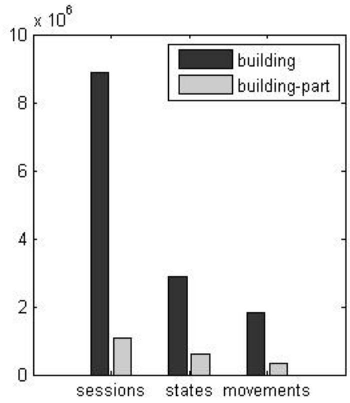


Figure 2.5: Amount of data during processing stages

2.3.2. Outdoor movement patterns

Looking at building level, movement patterns between, from and to buildings can be detected. Figure 2.6 shows the time profile of all movements with and without the ‘world’ state. This graph shows that there is much movement around 8.45, 12.45, 13.45, 15.45 and 17.45, corresponding with lecture hours at TU Delft. With ‘world’ (blue line), the morning and evening rush hours around 8.45 and 17.45 are detected, when students and staff travel between campus and, probably, home. Without introducing the ‘world’ state (red line), these two movement peaks are not detected.

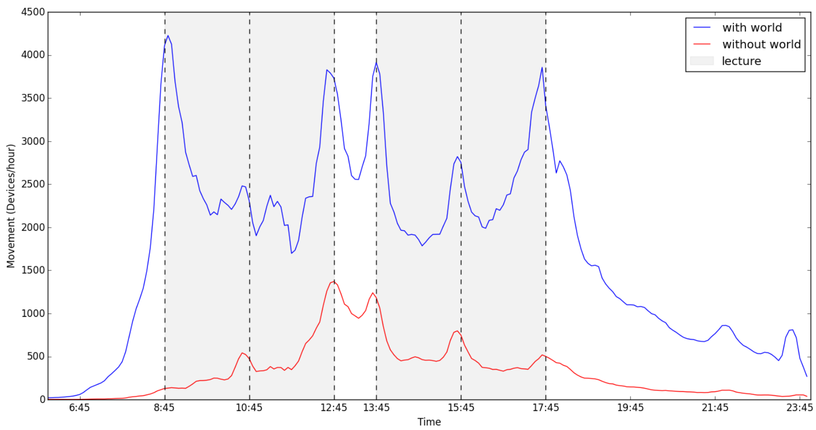


Figure 2.6: Time profile of indoor movement without world (only between building-parts) and with world (including movement from and to building part from ’world’)

Figure 2.7 illustrates the ten most occurring movements between buildings, on a map, where buildings are represented as nodes, and edges represent the number of movements. The number of movement, during the observation period, is illustrated with colour and line width.

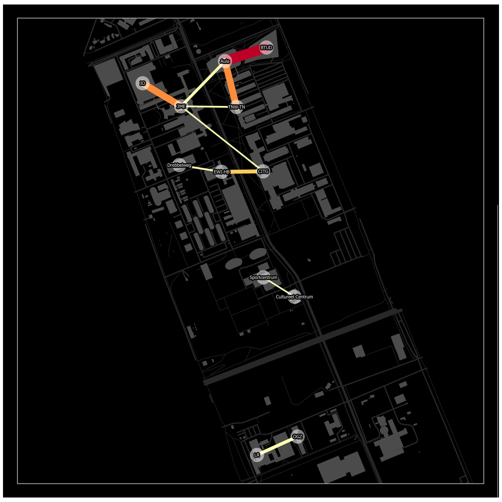


Figure 2.7: Top 10 most occuring movements on building level

In Figure 2.8 the movements of mobile (blue) and static (red) devices are shown. The time profile of static devices, compared with lecture hours, is less explicit than for mobile devices. This supports the assumption that movement of static devices is less related to movement of people, than mobile devices. Therefore, in this paper we only analyse movement of mobile devices, to provide knowledge about human movement behaviour.

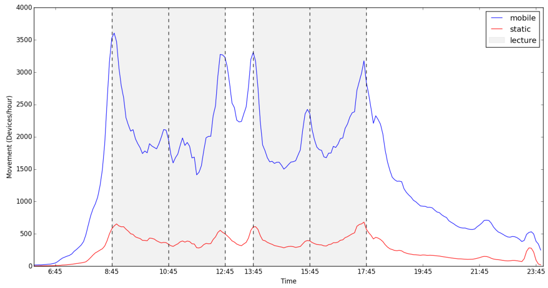


Figure 2.8: Time profile of all outdoor movement for static and mobile devices

The previous figures showed the movement over the entire time span of the research. However, different movement patterns can be identified by querying the data for certain time intervals. Figure 2.9 gives the time profiles of the movement of mobile devices during weekdays and weekends for all buildings including 'world'. Usually, most building at the TU Delft Campus, except for the library, are closed during the weekend. This is reflected by the amount of movement during the weekend. Moreover, compared to weekdays, the number of movement is constant throughout the day, as there are no lectures.

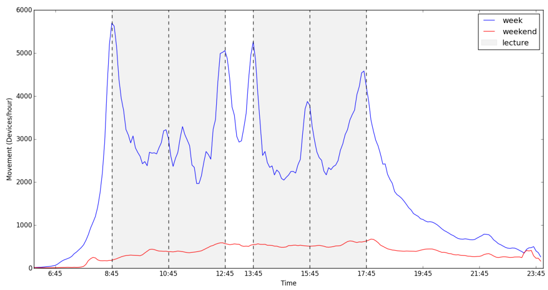


Figure 2.9: Time profile of all outdoor movement of mobile devices for week- and weekend days

Finally, the movement can also be queried based on origin and destination instead of or together with time. This enables a more detailed analysis of specific buildings or events. Figure 2.10 shows the movement from and to the Aula during normal weekdays. The Aula has a lunch facility, explaining the large amount of movement during lunch time. Both the movement to and from the Aula is high at the start and the end of the lunch time, because other facilities are located in the Aula as well. Figure 2.11 gives insight in where people exactly come from and go to, during the interval between 13:15 and 14:00. As expected, many people move to the library at this time, probably to continue studying after lunch.

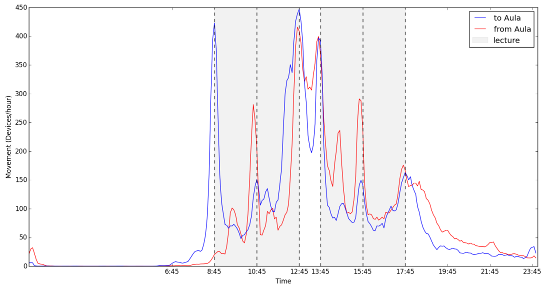


Figure 2.10: Time profile of movement of mobile devices from and to aula during normal weekdays

Figure 2.11: Movement of mobile devices from and to aula between 13:15 and 14:00 during normal weekdays

2.3.3. Indoor movement patterns

Like for building level, the building-part level data can be filtered for mobile devices and queried based on time, and origin and destination. Figure 2.12 shows the time profiles for all movement on building-part level for the Faculty of Architecture and the Built Environment. With 'world' this includes the movement from and to the faculty, without 'world' these movements are excluded. During the working day, the movement is relatively steady, except for two distinct peaks before and after lunch. Furthermore, small peaks can be seen at the start and end of the day when people arrive and leave the building. The Faculty of Architecture and the Built Environment normally closes at 22:00, resulting in a movement peak captured in the graph.

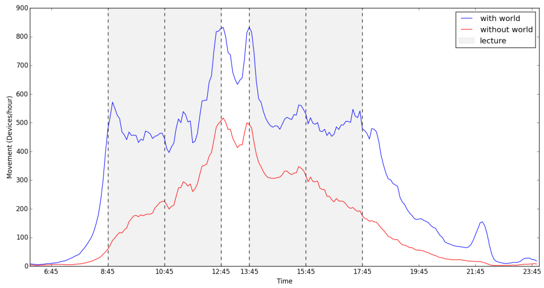


Figure 2.12: Time profile of indoor movement without world (only between building-parts) and with world (including movement from and to building part from 'world')

Figure 2.13 shows all the movement within the Faculty of Architecture and the Built Environment on a map. As the movements follow the shortest route on the graph, it becomes possible to see which particular corridor or staircase is busiest. Here the movement rates are highest in the eastern part of the ground floor. This shows occupation of space by movement instead of the flow between two states.

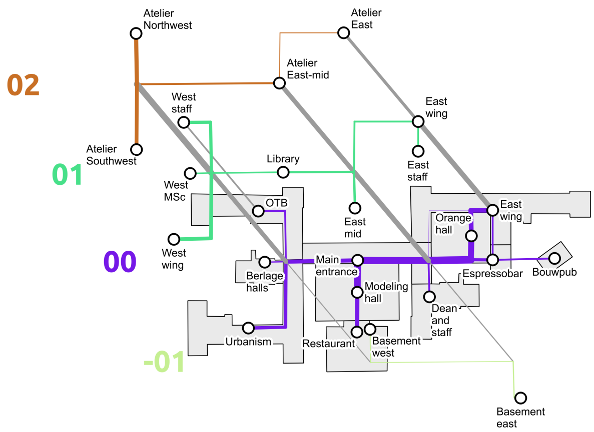


Figure 2.13: Map of all movement within the faculty of Architecture

Finally, Figure 2.14 and Figure 2.15 give a more detailed insight in the movement from and to the Bouwpub (café) and canteen (lunch facility) building-parts. The time profile of the canteen shows the expected peaks before and after lunch. Simarly it can be seen that people move towards the bouwpub after lecture time and leave around 20:00 when it closes.

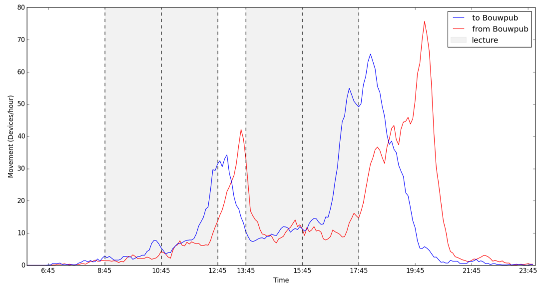


Figure 2.14: All movement of mobile devices during weekdays from and to the bouwpub

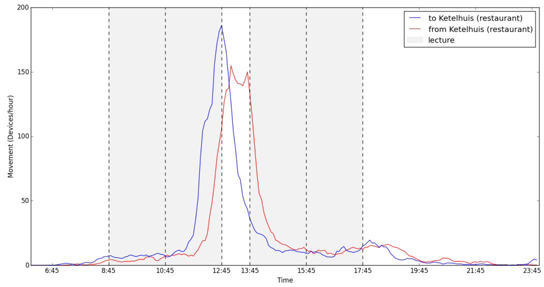


Figure 2.15: All movement of mobile devices during weekdays from and to the canteen

2.4. Conclusion and Recommendations

To understand human motion behaviour for better decision making, many studies have been conducted based on location data collection. Wi-Fi tracking technology is increasingly used due its cost effectiveness and ability to track people at a large scale. For this study, we used the eduroam network of the TU Delft Campus to identify movement patterns. Firstly, states are extracted from the raw Wi-Fi logs. Subsequently, the event of going from one state to another can be detected as movement. Finally, by counting the number of movement for an observation period, movement patterns can be identified. This paper tried to illustrate to what extend movement patterns in and between buildings can be identified from anonymised Wi-Fi logs. We successfully identified movement patterns at two spatial levels.

At building level, the rhythm of the campus is illustrated by time profiles showing the amount of movement for different observation periods. We found that movement at the campus was related to the lecture hours. Flow and direction of aggregated movement can be visualized on a map as edges. At building-part level, similar movement patterns can be identified. An indoor network graph was created of the underlying building floorplan. This successfully illustrates the occupied space for movement. However, the range of APs can extent between building-parts and floors and limits the accuracy of the analysis.

It is possible to identify movement patterns in and between buildings using the eduroam network. The presented method automatically mines movement patterns of large crowds from a dataset with anonymised Wi-Fi logs. However, we also encountered limitations from which several recommendations can be provided for future implementations.

The movement trajectory between two building-part states is computed with a shortest path algorithm, using the constructed network graph. Better models need to be implemented for a more accurate path estimation. No data from APs on the way between two states is used to estimate the path, because the system stores logs at a 5-minute interval. With a shorter log interval, this can be considered. The Faculty of Architecture and the Built Environment has a building lay-out with separate building wings and only three floor levels. It is important to mention that this building lay-out makes it easier to distinguish between building parts. Considering the range of APs, different methods need to be investigated for buildings with more floor levels. This also means that for the identification of movement at room level, other techniques, e.g. including Received Signal Strength (RSS), need to be implemented. In this paper, the road network is not used to estimate a detailed path of outdoor movement between building states. Due to the spacious character of the TU Delft campus and limited constrained space for pedestrians, it is challenging to analyse the usage of the infrastructure. With several strategically placed APs outdoor and logging with a higher frequency, this can be considered. Detailed information about the usage of the infrastructure on the campus can provide valuable knowledge, such as the identification of hotspots at specific time periods.

3
Introduction

3.1. Intro

Wireless Local Area Networks (WLAN) are widely used for indoor positioning of mobile devices within this network. The use of the Wi-Fi network to estimate the location of people is an attractive approach, since Wi-Fi access points (AP) are often available in indoor environments. Furthermore, smart phones are becoming essential in daily life, making it convincing to track mobile devices. This provides a platform to track people by using Wi-Fi monitoring technology. Knowledge of people's locations and related routine activities are important for numerous activities, such as urban planning, emergency rescue and management of buildings.

To understand the human motion behaviour many studies are conducted based on data collection of GPS receivers. The Global Navigation Satellite System (GNSS) is commonly used to track people in large scale environments. However due to poor quality of received signals from satellites in urban or indoor environments, GNSS receivers are not suitable in these environments. Moreover, GNSS receivers are convenient for self-tracking, but for large scale movement analysis, this data should be made available first before others can use it. This led to the development of alternative technologies to track people's locations, including Bluetooth, Dead Reckoning, Radio frequency identification (RFID), ultra-wideband (UWB) and WLAN (Mautz 2012). WLAN has the advantage of widespread deployment, low cost and with the use of a smart phone as a receiver, the possibility to track a large amount of people.

In general, there are four different location tracking techniques by using the Wi-Fi network: Propagation modelling, multilateration, Fingerprinting and Cell of Origin (CoO). Many of these methods rely on Received Signal Strength Indicators (RSSI) and/or previous set of calibration measurements. In comparison, CoO is the most straightforward technique and uses the location of the AP, to locate the mobile device. For, the location of the AP a mobile device is connected to, will give an estimation of the mobile devices' location, and thus the person. For this project, APs related to buildings and building-parts are used to track people's movement.

At the Delft University of Technology (TU Delft) a large scale Wi-Fi network is deployed across all facilities covering the indoor space of the campus. The network is known as an international roaming service for users in educational environments and called the eduroam network. It allows students and staff members from one university to use the infrastructure throughout the campus for free. This allows for large scale collection of Wi-Fi logs including individual scans of mobile devices. A continuous collection of re-locations of devices to access points for a long duration will return detailed records of people's movement. This ubiquitous and individual history location data derived from smart phones will present valuable knowledge on movement on the campus. For this reason, the project is carried out in request of the University's department of Facility Management and Real Estate (FMRE).

In this project, Wi-Fi monitoring technology is used to discover movement patterns on the campus of TU Delft. Based on the relationship between activities and places, location history can be used to discover significant places, movement patterns and hotspots. FMRE can use this information to answer questions such "what is the relation between buildings", "where do people come from" and "how regular a trajectory occurs".

This project will present a method for identification of movement patterns in a large scale indoor environments and between buildings. The method uses concepts of sequential pattern mining. Previous research has been done on sequential pattern mining, such as Zhao et al. 2014 to discover people's life patterns from mobile Wi-Fi scans, Meneses and Moreira 2012 analysed place connectivity using the eduroam network and Radaelli et al. 2013 identifies indoor movement patterns by analysing a sequence of relocations. Individual movement can be identified as a sequence of relocations of a mobile device to different APs. Without any data between two subsequent re-locations, sequential analysis is a convincing way for identifying moving patterns from wifilogs.

3.2. Purpose statement

The project is initiated by the idea that communication technologies can also be used to collect information about connections and connection attempts to Access Points (APs). This geo-referenced information can potentially be used to: ***1***) estimate the number of devices at a location at a certain time, representing presence of people at that place at that time or for a certain duration; ***2***) track unique ID's over several APs, reconstructing individual patterns of movement, resulting in aggregated flows of people and; ***3***) define regular and irregular (temporal, deviating) activities at specific places.

This research will focus on the second matter. Identifying movement patterns has attracted significant interest in recent years. Numerous methods have been explored including Wi-Fi tracking. This report will explain how movement patterns can be identified using large scale Wi-Fi based location data, and tries to contribute with four proposes. ***1***) A method for identifying movement patterns by analysing individual sequences of relocations from a large scale Wi-Fi network; This includes filtering the raw data and automatically create individual trajectories over a time interval as a sequence of relocations; ***2***) Identify spatio-temporal movement patterns of large crowds of people; ***3***) Investigate different visualization methods for showing movement, based on a large scale Wi-Fi network. ***4***) A method for analysing indoor movement using a constructed network graph of the underlying building floorplan.

The contributions can be described in one research question for this project.

- To what extend can movement patterns in and between buildings be identified from large scale Wi-Fi based location data of the eduroam network?

In order to answer the research question, there are three applied subquestions:

- What patterns can be identified moving from and to the TU Delft campus?
- What movement patterns can be identified between buildings on TU Delft campus?
- What movement patterns can be identified between large indoor regions of the Faculty of Architecture and the built environment?

Besides looking at this project from a spatial pattern perspective, this research also aims to investigate the following topics:

- Privacy – how viable is the data for personal concerns?
- Validity & Accuracy – how reliable is the data, how accurate, how robust for errors?
- Representativeness – which amount of the actual users is covered? Is this ratio constant or location dependant?
- System of APs – how well is the system equipped for measuring and tracking, and what is missing /essential to use the system this way?

3.3. Methods

The Geomatics Synthesis Project (GSP) is a small research project that combines a literature study with practical research. This includes a case study of the TU Delft campus, using real-world data. Practical work includes data storing, processing, analysing, interpretation, visualization and validation. The project is carried out in a team of six students with a connection to a supervisor and stakeholders (FMRE). This involves interactive discussions between stakeholders as an important part of the research.

3.4. Top level requirements

To keep track of the progress of the project, it is necessary to monitor to which degree the project is meeting the top level requirements and if the project is still on schedule with these requirements. In the baseline review the requirements are specified using the MoSCoW rules and killer requirements. In this chapter these requirements will be discussed. Each part of the MoSCoW rules that is in grey is a requirement that has not been satisfied.
MUST building level

- Main goal: Identify movement patterns and connectivity between building entrances.
- Relate entrances (place) of buildings to the corresponding APs (location).
- Find the stay places of each individual in order of the scan time.
- Find individual trajectories from a sequence of stay places.
- Find the movement patterns, by deriving a sequence of common places shared by all trajectories.
- Visualize the movement patterns between buildings in static maps.

A killer requirement for this level is:

- Identification of APs relating to an entrance of a building

SHOULD building-part level

- Main goal: Identify movement patterns between large indoor regions.
- Create a network graph from the underlying building floorplan for the analysis, where each region is a node.
- Find the movement trajectories between regions as a sequence of stays.
- Find the movement patterns between large indoor regions.
- Visualize the movement patterns between regions of buildings.

The killer requirements for this level are:

- Digital indoor floorplan of the buildings with classified/named regions (e.g. study rooms, canteen, etc.)
- Georeferenced building floorplans with APs.

COULD room level

- Main goal: Classification of movement patterns at room level.

The killer requirements for this level are:

- Digital indoor floorplan of the buildings with classified/named rooms (offices, classrooms, project studios, corridors, etc)
- Location of access points
- Fingerprinting map

The following chapters will reflect on these requirements, indicating how successful the project is.

3.5. Reading guide

This report tries to present our research in 15 chapters. Chapter 4 gives an overview of the project information, including the context, location, privacy issues and data description and representativeness. Chapter 5 provides background information on movement patterns including a literature review. Chapter 6 describes the methodology of our research. After the methodology section 6.1 will describe the pre-processing of the raw Wi-Fi dataset. The identification of movement patterns will be described in the three chapters after pre-processing. Chapter 7 reports on movements, chapter 8 will discuss trajectory patterns and chapter 9 describes movement indoor. Finally, chapter 10 and chapter 11 will conclude the report and provide recommendations.

4
Context

4.1. Use case: TU Delft

This projects main area of interest is the campus of the TU Delft. There are more than 30.000 users using the campus on more than 150 hectares. This emphasises even more the magnitude of this project. The eduroam network logs the devices connected to the access points, which implicitly means logging the (approximate) location of the person carrying the device and more information. This tracking data can be used to derive information about the personality of the person carrying the device, such as the distinction between staff and students, based on the tracked locations. Connection to the Wi-Fi eduroam network is free of charge and requires only a NetID, which all students and staff get upon registration at the university.

It is very important to understand, that 'no data is also data'. This means that a devices that is not being tracked by any access point for a period of time, is either off-campus or disconnected and still on campus. This provides valuable information when researching the movement patterns. This will be further discussed in the section 6.1.

The eduroam network of the TU Delft campus consists of 1730 access points, distributed over more than 30 buildings. The data is collected for each of the access points over a period of little more than 3 months. The logs are stored in a database on a virtual server, where it is accessible to the three project groups and the Geomatics staff. The data that is collected and the storage in the database is further described in section 4.6.

The department of Facility Management and Real Estate (FMRE) is the main client for the entire Synthesis Project. They would like to know how the campus is being used, what the hotspots on campus and in buildings are, when people travel the most from one building to another and which buildings are most visited.

4.2. Previous research: Rhythm of the campus

In the fall of 2014, similar research was conducted during another edition of the Geomatics Synthesis Project. The group "Rhythm of the campus" investigated the use of the Library and the Aula of the TU Delft, to gain insight in patterns the use of the facilities of the Library and Aula. This section will give a short summary of their research (Kalogianni et al. 2015).

During the project, the group used passive Wi-Fi monitoring to detect users of the TU Delft Library and the Aula to gain insight in the occupation, in request of FMRE. They used BlueMark sensors at the Library, Aula and 5 other faculties for a period of one week and collected ground truth data for 2 days. Due to its sheer size, the raw data was difficult to process. The data was filtered from static devices and outliers and the data analysis resulted in identification of the occupation of the Library and the Aula. The end results was a dashboard which visualized the sensor network, data analysis and pattern recognition to help the client in the decision making process.

This research was different from the research conducted in this Synthesis Project, mainly due the larger size of the eduroam network and the ability to track everybody using the Wi-Fi network.

4.3. Privacy

This project focuses on identifying common movement patterns, ignoring the individual, therefore we did not test explicitly whether is possible to identify individuals or not from the data. However, based on our findings about the operation of the eduroam Wi-Fi network and about the methods that are used to identify movement patterns, we can make the following assumptions.

Movement patterns are rather unique, therefore it is possible to match them to individuals even if maybe not in every case. However, in order to do so it is necessary to have additional data available. This additional data itself is often considered private data, e.g. the complete weekly schedule of the person. Provided that timetables are openly accessible and the occupation of the individual is known, then his movement pattern may be identified in the dataset.

The availability of a detailed access point map makes it easier to identify individuals by allowing a more detailed movement analysis (e.g. on building-part level). It reduces the ambiguity that is still present in building level movement analysis.

4.4. Data validation

The spatial accuracy of the Wi-Fi log dataset is defined by the range of the APs. Although we do not have information on the exact range of the different APs, we estimate the range to be a few tens of meters. Therefore, if a user is recorded by a specific AP, in reality he can be anywhere around the AP in its range.

The temporal accuracy of the Wi-Fi log dataset is defined by the five minute campus-wide logging interval of the eduroam system. It means that all APs on the TU Delft campus log all connected devices at the same moment in approximately five minute intervals. Therefore, it is possible that the user is already at a given AP, but he will be first recorded at the next scan round, or the user already left the AP but that also will be only recorded at the next scan round.

4.5. Representativeness

In the GSP a big amount of Wi-Fi logging data is used. The data represents all people that make (active) use of the eduroam network. These are the students and employees of the TU Delft. There is just a small amount of people that are within the spatial scope of the project and cannot connect to eduroam. The data is acquired by the access points, which all are located in a building on the campus. The people that use a building on the campus, but do not make use of the eduroam network, is a very small part. Thus, the main part of actual users is covered by the data used in the GSP. The collection of data is acquired over a continuous time interval of more than 2 months. This time period would be large enough to reflect on all users of the campus to some extend.

4.6. Data description and System of APs

4.6.1. Data description

This section will describe the main data source used within the Geomatics Synthesis Project; a PostgreSQL database containing the logs from the Wi-Fi access points on the TU Delft campus. The wifilog table has several column, with a data value for each row (Table 4.1).

username	mac	asstime	apname	maploc	sesdur	snr	rssi
j85cCQ..	l6iOu+..	14-4-2016 12:30	A-23-0-029	..CITG >4e Verdieping	1:32:02	35	-57
wrBqM..	f2Pw/P..	14-4-2016 7:49	A-23-0-035	..CITG >5e & 6e Verdieping	5:32:16	37	-56
wrBqM..	f2Pw/P..	14-4-2016 13:22	A-23-0-035	..CITG >5e & 6e Verdieping	0:40:20	46	-50
wrBqM..	f2Pw/P..	14-4-2016 14:02	A-23-0-093	..CITG >5e & 6e Verdieping	1:27:13	11	-86
wrBqM..	f2Pw/P..	14-4-2016 15:29	A-23-0-091	..CITG >5e & 6e Verdieping	0:05:08	30	-65
wrBqM..	f2Pw/P..	14-4-2016 15:34	A-23-0-035	..CITG >5e & 6e Verdieping	1:42:32	29	-65
J0IwA+..	HkLY1U..	14-4-2016 11:33	A-23-0-035	..CITG >5e & 6e Verdieping	1:27:40	33	-59
J0IwA+..	HkLY1U..	14-4-2016 13:01	A-23-0-035	..CITG >5e & 6e Verdieping	1:01:01	26	-68
J0IwA+..	HkLY1U..	14-4-2016 14:02	A-23-0-035	..CITG >5e & 6e Verdieping	3:30:19	25	-68
J0IwA+..	HkLY1U..	14-4-2016 17:32	A-23-0-035	..CITG >5e & 6e Verdieping	0:40:05	27	-69

Table 4.1: A segment of the main data source; the wifilog table

The data value for each attribute (column) in the wifilog table will be described in more detail.

Username
The username column provides the username, as a hashed text. Every user has a unique username, but can appear in the data more than once.

Mac
The mac column provides the media access control adress (MAC address), as a hashed text. The MAC address is a unique identifier assigned to a specific piece of hardware, such as the network adapter located in Wi-Fi devices (mobile phones, tablets, laptops etc.). So, it would be possible that a user can have more than one device connected to the Wi-Fi eduroam network at the same time.

Asstime
The asstime is the time of which a connected device is recorded by the system.

Apname
The apname is the name assigned to the access point. Every access point has a unique name.

Maploc
The maploc describes the location of the access point. There could be multiple access points with the same maploc. For instance, there are 31 access points located on the ground floor of the Faculty of Architecture and the built environment.

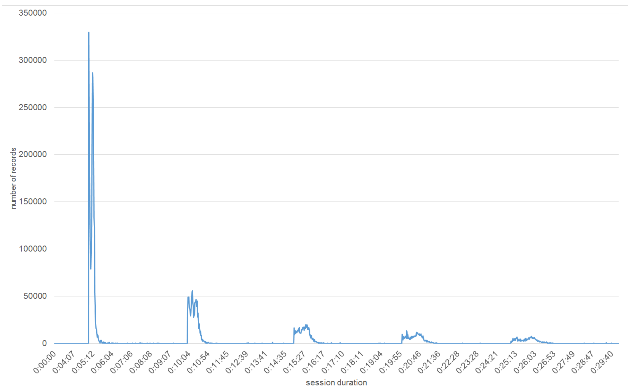


Figure 4.1: The frequency of session durations

Sesdur
The sesdur describes the session duration of which a device is connected to the access point. Because this is not as straightforward as it seems, this will be explained more extensively. Figure 4.1 shows the frequency of

session durations (the peak at exactly 5 minutes is filtered out to make the graph more readable). There is a large peak at exactly 5 minutes, a peak at approximately 5 minutes and decreasing peaks after a time interval of approximately 5 minutes. It looks like it is recording in a certain time interval in which the device is (still) connected.

In order to justify this, the query below is used to see the asstimes (and time to next asstime)

```
select *, asstime_next-asstime as difference
from (
        select count(*),asstime, lead(asstime) over (order by asstime) asstime_next
        from wifilog
        where extract(day from asstime) = 4
        and extract(month from asstime) = 4
        and extract(year from asstime) = 2016
        group by asstime
        order by asstime) as subquery
```

count	asstime	asstime_next	difference
2578	4-4-2016 13:04	4-4-2016 13:09	0:05:10
2435	4-4-2016 13:09	4-4-2016 13:15	0:05:11
2486	4-4-2016 13:15	4-4-2016 13:20	0:05:11
2530	4-4-2016 13:20	4-4-2016 13:25	0:05:11
2471	4-4-2016 13:25	4-4-2016 13:30	0:05:11
2444	4-4-2016 13:30	4-4-2016 13:35	0:05:11
2524	4-4-2016 13:35	4-4-2016 13:40	0:05:11
2588	4-4-2016 13:40	4-4-2016 13:46	0:05:12
2690	4-4-2016 13:46	4-4-2016 13:51	0:05:11
2560	4-4-2016 13:51	4-4-2016 13:56	0:05:11

Table 4.2: The time and time to next scan at a random day

Table 4.2 shows that the time to the next scan is 5 minutes and several seconds in all cases. Most important to know is that all access points are logging the connected device(s) at the same time campus wide.

Table 4.3 will be used to explain the way the time interval of approximately 5 minutes is coming back in the session duration.

The first record shows the device is not connected to any of the access points on the campus in the subsequent moment of recording, resulting in a session duration of exactly 5 minutes. The last record in Table 4.3 shows the result of a device that is still connected to the same access point at the subsequent moment of recording. In this case the session duration will be 10 minutes and 21 seconds. This is the time interval between the first moment the device is recorded and the first time the device is not recorded by the same access point anymore. The record with id number 6 describes a situation in which the device is connected to an access point at the moment of recording and connected to another access point at the subsequent moment of recording, the session duration is 5 minutes and 18 seconds in this case. This is the time interval between the two moments of recording. This time interval can vary, but is always approximately 5 minutes.

id	username	mac	asstime	apname	maploc	sesdur
1	oHh0Sz..	WWW0Cd..	1-4-2016 10:13	A-12-0-104	..& Proeffabriek >!e Verdieping	0:05:00
2	oHh0Sz..	WWW0Cd..	1-4-2016 10:18	A-132-0-064	..32-OCP-IO >1e Verdieping	0:20:27
3	oHh0Sz..	WWW0Cd..	1-4-2016 11:36	A-132-0-105	Root Area	0:15:22
4	oHh0Sz..	WWW0Cd..	1-4-2016 11:51	A-132-0-066	..32-OCP-IO >1e Verdieping	0:20:35
5	oHh0Sz..	WWW0Cd..	1-4-2016 14:01	A-132-0-069	..32-OCP-IO >1e Verdieping	0:05:43
6	oHh0Sz..	WWW0Cd..	1-4-2016 14:06	A-132-0-133	..32-OCP-IO >4e Verdieping	0:05:18
7	oHh0Sz..	WWW0Cd..	1-4-2016 14:12	A-132-0-066	..32-OCP-IO >1e Verdieping	0:05:10
8	oHh0Sz..	WWW0Cd..	1-4-2016 14:17	A-132-0-104	..32-OCP-IO >2e Verdieping	0:05:10
9	oHh0Sz..	WWW0Cd..	1-4-2016 14:22	A-132-0-067	..32-OCP-IO >1e Verdieping	0:05:10
10	oHh0Sz..	WWW0Cd..	1-4-2016 14:27	A-132-0-066	..32-OCP-IO >1e Verdieping	0:10:21

Table 4.3: Varying session durations

SNR

The signal to noise ratio(SNR) describes a measurement that compares the signal strength to the level of background noise (in dB).

RSSI

The received signal strength indicator (RSSI) describes the received signal strength (in dB).

4.6.2. System of APs

This section will describe the current layout of access points (APs) on the TU Delft campus. The location of APs in a building is not known, but for the Faculty of Architecture and the built environment a paper map was available. Therefore the system of APs in the Faculty of Architecture and the built environment will be described in more detail.

In total there are 1730 access points, distributed over more than 30 buildings on the campus. The access points are mostly placed on walls or ceilings. The data describes that every access point is linked to a certain location. Due to the (wide) signal range of the access point, the device can be located at a different floor level than the access point it is connected to. Moreover, there could be access points located at the first floor while serving people at ground floor as well. This is the case in rooms with high ceilings, such as the orange hall in the Faculty of Architecture and the built environment.

As said, the Faculty of Architecture and the built environment is the only building of which the location of the access points are known. The floor plans are enriched with the location of the access point (see Figure 4.2). Next to that, a table is provided with additional information regarding the access points, although this table does not contain all present access points. This table includes the MAC address of the access point. This could be used to look up to what access point the device is connected.

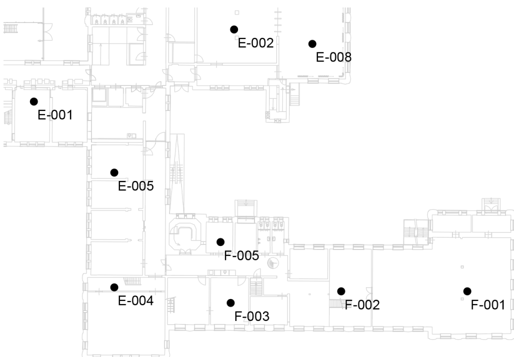


Figure 4.2: Ground floor plan with the location of the access points

5

Movement patterns

5.1. Introduction

The objective of this project is to identify movement patterns. To have a better understanding of this concept, it is important to describe relevant types of movement patterns in a systematic and comprehensive way. A classification of different patterns will provide guidelines for development of different mining algorithms and identify patterns. This chapter will first approach the definition of movement patterns. Subsequently, the theory is demonstrated with the research case of TU Delft in chapter 7, chapter 8 and chapter 9. This illustrates what type of pattern mining methods can be used on a movement dataset.

5.2. Movement identification

By definition, moving objects are entities whose positions of geometric attributes change over time (Dodge, Weibel, and Lautenschütz 2008). People always move in geographic space, this means that human movement is geo-referenced. When the start and end time of one movement is specified, its trajectory can be constructed by ordering several movements of one individual. These trajectories can be visualized and analysed.

In order to identify movement patterns, it is important to understand what types of patterns may exist in the data. Besides, there are many types of patterns and not everything is relevant for this project. Therefore, this section will organize various categories. This project aims to identify three different movement patterns: ***1***) Spatio-temporal movement patterns; ***2***) ordered co-location in space; ***3***) unordered co-location in space.

Individual and group movement

Patterns can occur in individual movements or in movements of a larger group. Typical movements of individuals will be different from typical movements of a larger group. For analyzing movement in a larger area with more than 25.000 users, we are interested in typical movement at the larger aggregate level of crowds.

5.2.1. Spatio-temporal movement patterns

As described previous in this section, movement is from one location, or state, to another state, i.e. A to B. These movements can be analysed from movement data to detect the direct connectedness and flow between two locations in a time interval. Questions such as "where do people come from" and "how many people move between two locations" can be answered. Several patterns can be identified from this analysis. Firstly, the number of movements over time can be detected. This will provide insight in the behaviour of humans, e.g. when people go home or at what time people have lunch. Secondly, the flow and direction between two states, i.e. the analysis of the direction of the flows provides information on the symmetry of movement between two locations. For example, if 100 people move from A to B within a time interval and 100 people move from B to A in the same time interval, the movement pattern is perfectly symmetrical. Besides analysing movements between two states, consecutive movements of one individual can be used to identify movement patterns. These trajectories will be the basis for the next section to identify co-locations of several trajectories.

5.2.2. Co-location in space

When moving individuals share some locations in their trajectory, you can speak of co-location in space. According to Dodge, Weibel, and Lautenschütz 2008 there are three types of co-location in space: ***1)*** ordered co-location occurs when some locations are shared by multiple trajectories in the same order; ***2)*** unordered co-location when shared locations are attained in different orders; ***3)*** symmetrical co-location when the shared locations are in opposite order. This means that co-location in space, helps to identify movement patterns in the sense of frequently visited locations in one trajectory. For example buildings A, B, C can be visited in the same order by multiple trajectories, and the same buildings can be visited by multiple trajectories, but in different orders. Figure 5.1 illustrates the concept of ordered co-location in space.

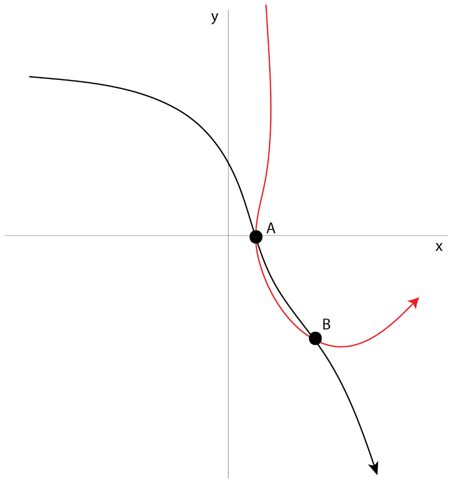


Figure 5.1: Ordered co-location in space

Ordered co-location is space can be analysed with the concept of sequences. A sequence is an ordered list of visited locations. Sequential pattern mining algorithm help to understand a what order common locations are visited. In this report, trajectories of a sequence of locations are analysed to identify ordered co-location in space movement patterns. Unordered co-location in space analyses the same trajectories, but does not consider direction or order of the movement. This means that common locations visited together in one trajectory can be identified. In other words, the association between buildings is detected. A commonly used method to detect groups of objects in a list (i.e. a trajectory), an association rule mining algorithm can used. Our research did not include this in the final results, but this report will elaborate on the concept of this algorithm in section 11.3.

6

Methodology

In this chapter the data mining methods used to retrieve movement patterns from the TU Delft eduroam Wi-Fi log data will be described in detail. Figure 6.1 gives an overview of the main workflow to derive movement patterns from the Wi-Fi log. First the raw Wi-Fi log is preprocessed to get states at two different spatial levels (building- and building-part level). A state is defined as a time interval during which a particular device is located in a certain area. An example of a state on building level is: device A is located at Library from 11:00 to 12:00. An example on building-part level is: device A is located at canteen from 11:00 to 12:00.

In the preprocessing phase the data is enriched with 'world' states, reduced by grouping states and cleaned by filtering out 5 minute states representing people that only pass by a building without actually entering it. The insertion of world states enables the detection of movement from and to the campus in the case of building level, and movement from and to the building in the case of building-part level. The assumption made here is that the device is not switched off. Especially in the case of laptops it is likely that the device is switch of for some hours during a lecture for example, this could be interpreted as a movement off campus. Therefore a mobility analysis is conducted attempting to distinguish between mobile phones and laptops, which are the two main device categories present in the dataset.

The states resulting from the preprocessing are used to retrieve movements at both spatial levels. A movement is defined by the change from one state to the next subsequent state, where the different states must be at a different locations. Furthermore, the building level states are used to retrieve trajectories for each device. A trajectory is defined as an ordered list of states. The trajectory thus stores the entire route or trajectory the particular person travelled. For the building-part level no trajectories are retrieved. For building-part level a graph is made for BK-city. In this graph the nodes represent the different building-parts and the edges follow movement space, such as corridors and stairs. Using the shortest path in the graph, the route of the movements within BK-city can be visualized in more detail. For building level no graph is created, therefore the trajectories and movements are visualized simply as a straight line. In addition to the maps at both the building and building-part spatial level, movement time series are created for both spatial levels. Together these maps and time series are used to identify different types of movement patterns.

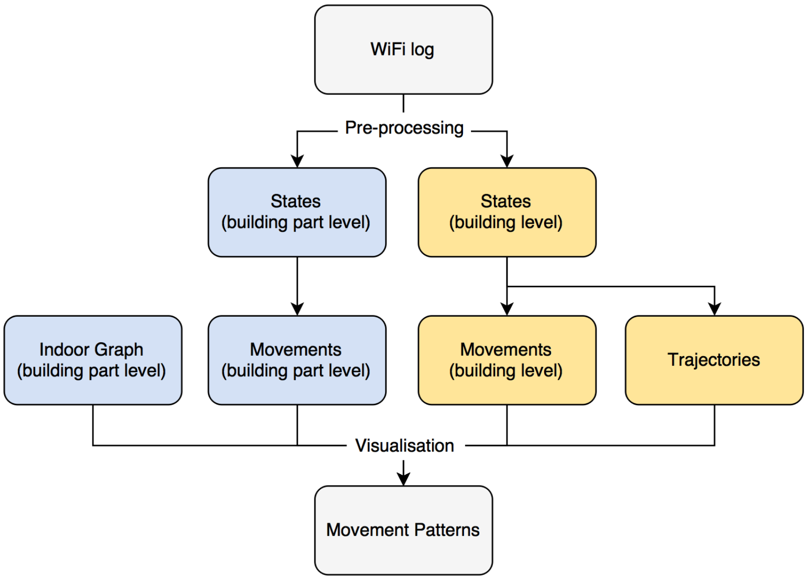


Figure 6.1: Grouping

In the following sections all steps to derive movements and trajectories from the Wi-Fi log will be described in more detail. First section 6.1 describes the various pre-processing steps to clean, reduce and enrich the raw data. Section 6.2 will describe the mobility analyses which aims to distinguish between laptops and mobile phones. Subsequently section 6.3 addresses how the movements are retrieved from the states for both spatial levels. Finally section 6.4 describes how the trajectories are created from the building level states. In chapter 7 to chapter 9 the visualisation and results of the movements and trajectories are addressed, for building-part level this also includes creation of the graph.

6.1. Preprocessing

Before movement patterns between buildings and building-parts can be retrieved, pre-processing of the raw data is required. In this chapter the different pre-processing steps will be described in detail. First subsection 6.1.1 addresses the initial data filtering. subsection 6.1.2 concerns the grouping of records with the same mac address and location that are subsequent in time. Section 6.1.3 describes the filling of the dataset with a ‘world' location. This enables detection of movement from and to the campus on the building level, and movement from and to the building on building-part level. Finally subsection 6.1.4 is about the filtering of records of people only passing by a building or building-part.

6.1.1. Initial filtering

Each record in the wifilog represents the scanning of a certain device at a certain time by a certain access point. In order to detect the movement patterns of these devices between buildings it should be known for each access point in which building it is located. The data in the table 'wifilog' contains information about the location of the Access Point (AP) in two columns. The first column which contains information about location, is the column 'maploc'. This column contains strings, which look as follows:
'System Campus >[buildingid] >[specific location]'. An example of such a string is 'System Campus >21-BTUD >1e verdieping'. In such a string, the middle part can be linked to a building, so to a real-world location. But there are some other values for maploc, which can less clearly be linked to a real-world location. Such a value is 'Root Area', it is unclear what this value means and it contains no information about a building or area it might be in. This makes it impossible to link it to a location in the world. Then there is the value 'Unknown', a value that indicates that there was no name attached to the Access Point that user was connected to. Again in this case, it is impossible to link this value to a real-world location. As both 'Root Area' and 'Unknown' are in the minority of records, they could be left out of the queries, but this would mean removing many records from the dataset, which is not desired.

The second one is the column 'apname', which is a string with the symbolic name of the AP, for example 'A-08-G-010'. The two numbers in the second part of the string, in this case '08', represent the building number. This building number can be linked to a location in the world. In some cases, the column 'apname' did provide information about the location, while the 'maploc' column value was unclear. In most of these cases however, the building number, the second part of the string, was a three digit number. But there are no buildings on the TU Delft campus with a building number that high. When consulting Wilko Quack about this, he explained that these building numbers had an arbitrary 1 in front of the building number. So 'A-134-A-001' was not building 134, but building 34, which was an actual building number on the campus. This would mean that using the column 'apname' for getting the building number would mean a higher number of results and therefore a more realistic visualization of the movements. Two other special cases are present in which the building id in the apname is 102 or 104. These ids corresponds to the legermuseum and the VLL-LAB respectively. As the legermuseum in not located on the campus it was decided to omit this building. The VLL-LAB did have no building number according to the TU Delft Campus maps, which was why the building was not identified before. After finding this, the building was manually added to the buildings table.

The information about the location is linked to the actual locations of the buildings using the buildings table. Each building has an id, name and geometry. The id is taken from each record using a Python function and linked to the id in the building table. The building-part tables works in the same way, but then every full apname from a record is linked to the corresponding building-part.

6.1.2. Grouping of states

In order to reduce the data and to be able to filter out records of people only passing by a building, the data needs to be grouped. The overall goal is to identify movement patterns between different buildings or building-parts. As a result records of subsequent states of the same device in the same building or building-part can be grouped together into one single record. Namely, if two subsequent states are at the same location they do not represent a movement and can thus be grouped. When looking at building level, the mobile of someone who studies the whole day at architecture might have 20 records (states) in the database for that day. This can be reduced to one record (state) that contains the time the device arrived at Architecture and left again. For building-part level the same applies for someone that has multiple subsequent states in the same building-part. To determine whether two records are subsequent in time, and therefore should be grouped together, a threshold for the time gap between two records needs to be defined. It was decided to set the gap threshold for grouping states to 1 hour. The reasoning behind this is that someone who is not scanned for a period of more than 1 hour has likely left the building. However, if someone is away for less than an hour it more likely that that person was just smoking or lunching outside or just disconnected for from the system for a while. Figure 6.2 gives an example of how the records are grouped on building level for one device for one single day. It should be noted that two states remain at faculty A as the gap between 12:30 and 13:45 is bigger than an hour. In the other other cases the gap is smaller than an hour and the records are grouped together. Only one records is present at faculty B so this record can not be grouped. The grouped records still contain all the information that is required to know that the person moved from faculty A to B to C during the day.

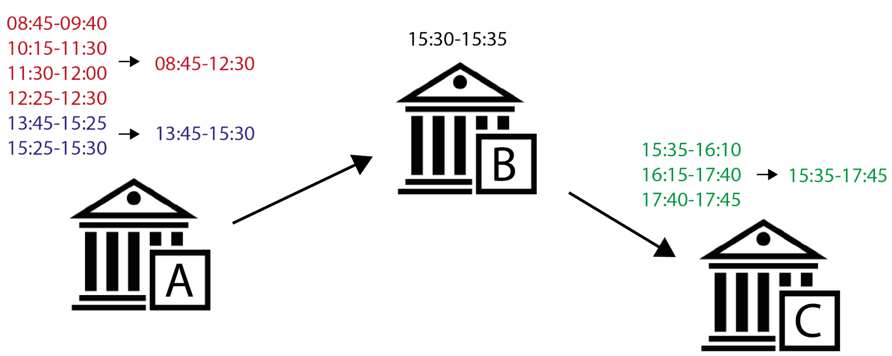


Figure 6.2: Grouping

6.1.3. Adding world state

Because the dataset contains all records of when certain devices are scanned, it also Implicitly stores information on when the device is not located at the campus. These time gaps in which a particular device is not scanned at the campus give information on when the corresponding person is not at the campus. This information is valuable for detecting movement patterns from and to the campus in addition to the movement between buildings at the campus. Considering the fact that many student only visit one faculty each day. It becomes especially clear, that the movement from and to the campus plays an important role in the overall movement pattern of a person. In order to be able to directly derive movement from and to the campus from the dataset, the time gaps present in the data should be stored explicitly. Therefore each time gap larger than an hour is filled with a outside campus or 'world' record. The word 'world' is used to indicate that the device could be located at any place in the world outside the campus during the time spans that it is not scanned at the campus. It should be noted the reason that a device is not connected to one of the access points could also be that the device is simply switched of, in this case however the assumption is made that the device moves off campus. The begin and end time of a world record is defined by the end of the previous record and the start of the next record in time. In case there is no previous or next record the boundaries are defined by the starting time of the whole dataset and the current time. Figure 6.3 visualizes the explicit storing of world states that fill time gaps during which a device is not recorded on campus. In can be seen that three 'world' states are added in the example. First during the start of the day before the person goes to faculty A, second during the lunch break, and finally in the end of the day starting from the moment when the person leaves faculty C. Storing these world states explicitly enriches the data as much more movement can be defined. The grouping of records described in subsection 6.1.2 and adding of a world state are complementary. If the gap between two states is smaller than an hour they are grouped if the gap is bigger than an hour a world state is inserted. For the building-part level the insertion of world states works exactly the same. In this case however the world represent the outside building area instead of outside campus area.

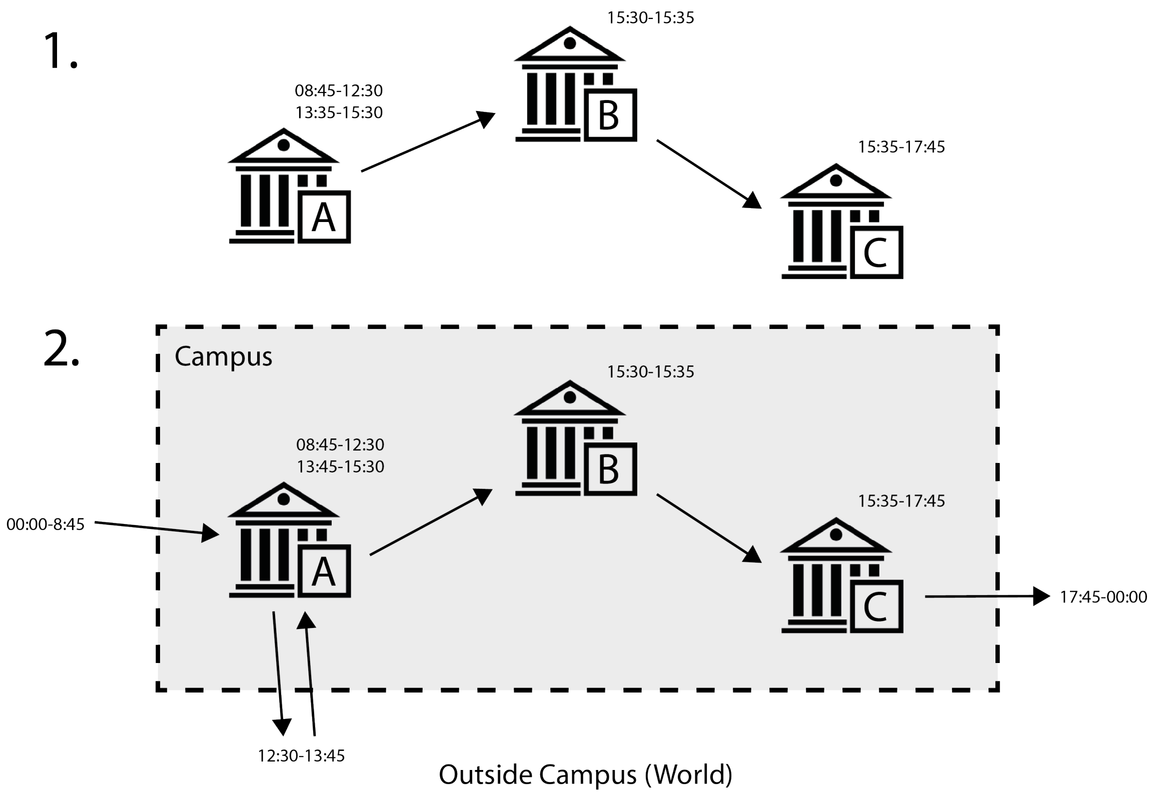


Figure 6.3: Adding World

6.1.4. Passing by

For the detection of movement patterns between buildings, records of people that only pass by a building without actually visiting it should be excluded. The reason for this is that records of people only passing by a building could result in misinterpretation of the movement patterns. This is illustrated by the example in Figure 6.4. In this case faculty B is located on the route from faculty A to faculty C. Therefore it is likely that people moving from faculty A to faculty C are picked up by a scanner located at faculty B. The eduroam system records all devices at intervals of approximately 5 minutes as explained in subsection 4.6.1. Such a recording by the eduroam system can happen during the short time period that the device, which is on its way from faculty A to C, is connected at faculty B. This will result in records of approximately 5 minutes at faculty B, whilst the person has not been inside faculty B. As the movement is the change between two states, the movement to faculty C will origin from faculty B. Someone that is not aware of the 'passing by' problem might conclude that people from faculty B often go to faculty C. In reality however, people from faculty A go often to faculty C. By filtering out the records of people only passing by buildings the correct movement can be visualized (see Figure 6.4 bottom). It should be noted that filtering out 'passing by' records can only be done after the grouping process. The reason for this is that 5-minute records that would individually be classified as someone passing by might be grouped together into one record with a longer duration. After grouping the combined record is not classified as someone who passes by anymore. Furthermore it should be noted that the filtering of 'passing by' records occurs after filling the data with 'world' states. The reason for this is that a passing by event does mean that the device was located on the campus. The world records are meant to represent the time the device is not on the campus. Filtering passing by events works exactly the same for the building-part level. If a person only passes by a particular building-part without staying in it, it is filtered out. For building-parts this filtering is especially important as the route from one building-part to another often leads through several other building-parts.

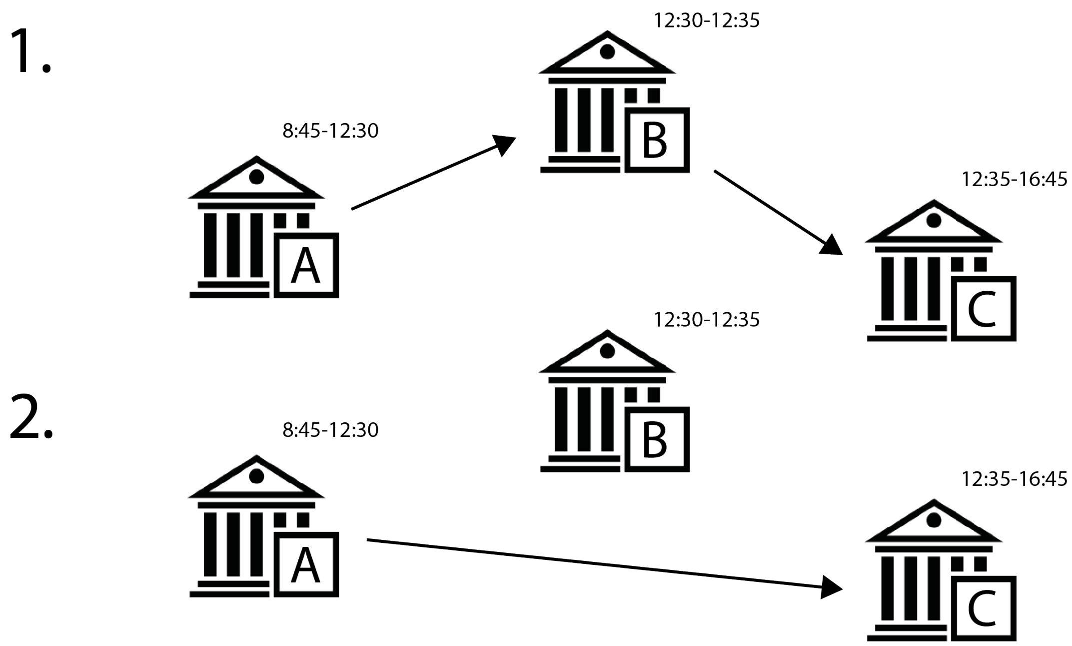


Figure 6.4: Passing by

6.1.5. Implementation

The filling (world), grouping and filtering (passing by) steps described above are implemented in an integrated way. The pseudo code for the implementation for building level is shown in Figure 6.5, for building-part level the implementation is exactly the same only grouping is done for building-parts instead of buildings. As can be seen in the code there is communication with the database at several points. The table from which the records are retrieved for each mac address is already processed as described in the general filtering section. Furthermore the format of the table is slightly different compared to the initial wifilog. The session duration is exchanged for an end time column which is derived by adding the session duration to the asstime (start time of a record).

```
macs = get distinct macs from db
create new empty table with 4 columns(mac, building, start, end)
min_time = minimum time in entire db
max_time = current time
for mac in macs:
        records = get all records for mac from db
        cur_rec = first record from records
        insert world at start (mac, world, min_time, cur_rec[start])                    # fill
        for next_rec in records[1:-1]:
                gap = next_rec[start] - cur_rec[end]
                if gap > hour:
                        insert world (mac,'world',cur_rec[end],next_rec[start])        # fill
                if gap < 15 minutes and cur_rec[building] == next_rec[building]:
                        cur_rec = (mac,cur_rec[building],cur_rec[start],next_rec[end])  # group
                elif cur_rec[end]-cur_rec[start] > 6 minutes:                          # filter passing by
                        insert cur_rec
                cur_rec = next_rec
        if cur_rec[i_end]-cur_rec[i_start] > 6 minutes:                                # filter passing by
                insert cur_rec
        insert world at end (mac, world, cur_rec[end],max_time)                         # fill
```

Figure 6.5: Pseudocode preprocessing

Figure 6.6 shows an example of the records of one device over a time span of one day during the different pre-processing steps. From the raw data it can be seen that this person spends most of the day in building B. The person is scanned once at building A before he arrives in the morning and after what is likely to be his lunch break. The last two hours the person is scanned in building C. After filling three world records are added, at the beginning of the day, during the lunch break, and at the end of the day. The grouped records show that the subsequent scans in building B and C are grouped together. Finally the scans at building A are removed from the dataset as they are likely to indicate passing by events.

Raw

Bld.	Start	End
A	09:30:00	09:35:07
B	09:35:07	09:40:07
B	09:50:28	10:41:21
B	10:41:21	12:08:40
B	12:08:40	12:13:51
A	13:30:03	13:35:12
B	13:35:12	13:40:16
B	13:40:16	15:34:22
B	15:34:22	15:39:26
B	15:44:34	15:49:34
C	15:59:47	18:06:54
C	18:06:54	18:11:54

Filled

Bld.	Start	End
W	00:00:00	09:30:00
A	09:30:00	09:35:07
B	09:35:07	09:40:07
B	09:50:28	10:41:21
B	10:41:21	12:08:40
B	12:08:40	12:13:51
W	12:13:51	13:30:03
A	13:30:03	13:35:12
B	13:35:12	13:40:16
B	13:40:16	15:34:22
B	15:34:22	15:39:26
B	15:44:34	15:49:34
C	15:59:47	18:06:54
C	18:06:54	18:11:54
W	18:11:54	00:00:00

Grouped

Bld.	Start	End
W	00:00:00	09:30:00
A	09:30:00	09:35:07
B	09:35:07	12:13:51
W	12:13:51	13:30:03
A	13:30:03	13:35:12
B	13:35:12	15:49:34
C	15:59:47	18:11:54
W	18:11:54	00:00:00

Filtered (Passing by)

Bld.	Start	End
W	00:00:00	09:30:00
B	09:35:07	12:13:51
W	12:13:51	13:30:03
B	13:35:12	15:49:34
C	15:59:47	18:11:54
W	18:11:54	00:00:00

Figure 6.6: Preprocessing

6.2. Mobility analysis

As described in subsection 6.1.3 gaps in the data during which a device is not scanned are filled up with world states. It is however possible that the reason a device is not scanned at the campus is not because the device left the campus, but simply that the device is switched off or lost connection to the network. Especially in the case of laptops it is likely that several gaps in the data are present, due to the fact that the particular person closes its laptop for example to have lunch or go to a lecture. Furthermore it is likely that the laptop is only opened when the person starts studying and not when the person is actually entering the building. Mobile phones on the other hand are likely to have fewer gaps in the data as they are usually not switched of during the day. In terms of movement the results could be more accurate by filter out the laptops and only looking at mobile phones. Furthermore distinguishing between mobile phones and laptops enables comparison of the movement patterns of the different device types.
The main difference between laptops and mobile phones is that laptops are usually on switched on if a person is stationary at a certain location. Mobile phones on the other hand are usually also switched on when the person is moving over the campus or through the building. As described in subsection 6.1.4, records of moving people result in a session duration of 5 minutes. Therefore, the ratio between the number of records with a session duration of 5 minutes and the total number of records in the database gives an indication of the mobility of the device. This mobility ratio can be defined with the following formula.

$$Mobility\ ratio = \frac{number\ of\ records\ with\ a\ sesdur\ of\ 5\ min}{total\ number\ of\ records}$$

Figure 6.7 shows a histogram of the mobility ratio of all devices. Two distinctive peaks can easily identified, one around 0.1 and one around 0.5. The 0.1 peak relates to devices of which only approximately 1 out of 10 records has a session duration of approximately 5 minutes, these are likely to be the laptops. The 0.5 peak relates to devices of which 1 out of 2 records has a session duration of approximately 5 minutes, these are likely to be the mobile phones. A separate table is created in the database in which the mobility ratio for each device is stored, later this table is used to distinguish between the generally static devices (laptops) and the mobile devices (mobile phones).

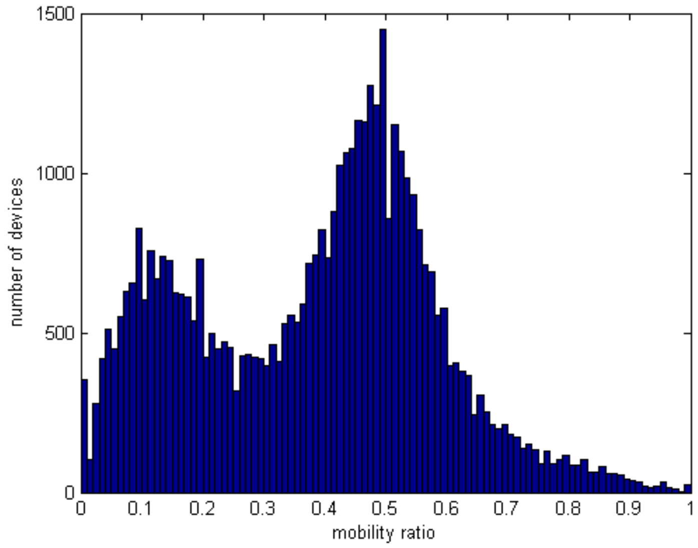


Figure 6.7: Histogram of mobility ratio of all devices in the wifilog

6.3. States to movements

The data resulting from preprocessing contains the states of where a particular device was located during a certain time period. Implicitly this also includes information on the movement of the device. If a device is first located in building A and subsequently in building B it must have moved from building A to B. However, in order to be able to retrieve the movement patterns of devices the movement should be stored explicitly. This means that each record should store the movement of one device from one building to another building or to world. Examples of movement patterns that can be retrieved from this data are: the number of devices moving from building A to B within a given time period, and the peak in movement from the canteen to all other building-parts. To create records for each individual movement first the preprocessed data is ordered on mac address and start time. By doing this all the subsequent states for every device are listed directly below each other (see Figure 6.9). As a movement is defined by the change of one state to another, movements records can be created from every two consecutive state records (see Figure 6.9. However, not every two consecutive states represent a movement. Only when the two states concern the same device and they are at different buildings they represent a movement. This means that movement records with different mac addresses or similar building ids are filtered out (see Figure 6.9. Figure 6.8 shows the creation of movement records graphically. The states are shown in black, the movements in red.

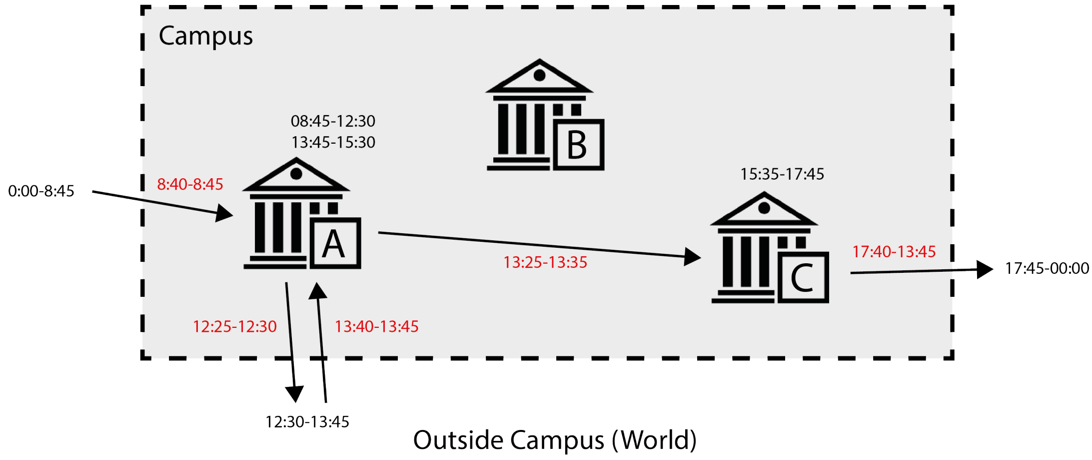


Figure 6.8: Graphic representation; retrieving movements from states

States

Mac	Bld.	Start	End
1	W	00:00:00	09:30:00
1	B	09:35:07	12:13:51
1	W	12:13:51	13:30:03
1	B	13:35:12	15:49:34
1	C	15:59:47	18:11:54
1	W	18:11:54	00:00:00
2	W	00:00:00	10:32:33
2	A	10:32:33	14:21:05
2	A	14:40:37	15:11:07
2	W	15:11:07	00:00:00

Movements

Mac	Mac2	Bld.	ToBld	Start	End
1	1	W	B	09:25:00	09:35:07
1	1	B	W	12:08:51	12:13:51
1	1	W	B	13:25:03	13:35:12
1	1	B	C	15:44:34	15:59:47
1	1	C	W	18:06:54	18:11:54
1	2	W	W	23:55:00	00:00:00
2	2	W	A	10:27:33	10:32:33
2	2	A	A	14:16:05	14:40:37
2	2	A	W	15:06:07	15:11:07

Movements (filtered)

Mac	Bld.	ToBld	Start	End
1	W	B	09:25:00	09:35:07
1	B	W	12:08:51	12:13:51
1	W	B	13:25:03	13:35:12
1	B	C	15:44:34	15:59:47
1	C	W	18:06:54	18:11:54
2	W	A	10:27:33	10:32:33
2	A	W	15:06:07	15:11:07

Figure 6.9: Database representation; retrieving movements from states

The start and end time of the movement are defined by the end time of the previous state minus 5 minutes, and the start time of the next state (see Figure 6.10). The reason that 5 minutes are subtracted from the end time of the previous state is that this is approximately the last moment in time the device was actually scanned at the location of the previous state. In the figure below the device is scanned 15:21 at building B. Approximately 5 minutes later (at 20:27) the device is scanned at building C. The state record of building B however continues all the way until 20:27, whilst the last time it was actually scanned at building B was 15:21. As a result it can be concluded that the movement from building B to C took place somewhere between 15:21 and 20:27. Therefore the start time of the movement between B and C can be approximated by subtracting 5 minutes from the end time of the state record at B. As can be observed in Figure 6.10 the movement from A to B is retrieved in the same way.

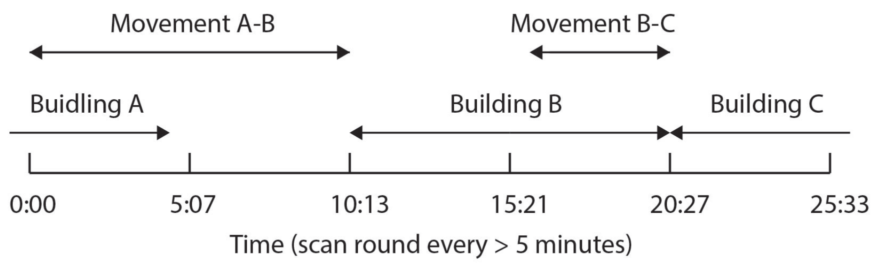


Figure 6.10: Defining the start and end time of a single movement

6.4. States to trajectories

An individuals trajectory is constructed as a sequence of locations in order of the scan time. Start and end time of a trajectory can be specified with a time interval. Two consecutive scans from the Wi-Fi log are considered in the same trajectory if and only if t_{s2} - t_{e1} <T_{split}, where T_{split} is the splitting threshold. The splitting threshold is important when dealing with people, who are not observed for a long duration of time, i.e. people moving home. For example, if a student leaves the campus at the end of the day, and returns the next morning, separate trajectories should be created. Because, T_{split} is larger than the threshold for identifying *'world'* (see section 6.1), the trajectory will always start and end with *'world'*. If p is a location, then a trajectory can be written as:

$$p_1 \rightarrow p_2 \rightarrow p_3 \rightarrow \ldots \rightarrow p_n$$

Given a time interval, there is a set of individual trajectories $S = \{t_1, t_2, t_3, ..., t_n\}$ where each t_i is the trajectory.

7 Movements

7.1. Introduction

Movement patterns are defined as how people regularly move on the campus. To answer this question, there are three sub questions to answer: when do people move, from where do they move and where do they move to. These patterns consists of spatial component and temporal component, thus they are called spatio-temporal movement patterns.

Movement pattern is actually a kind of behaviour pattern, which implicitly reveal how people use the campus and furthermore how they think and behave. In this chapter, several patterns related to time and space are discussed on building level. These movement patterns are about how and when people move between different buildings and according to these movement patterns, the reason why people move in this way at a certain time can be explored.

In this chapter, section 7.2 will discuss the methodology of exploring the movement patterns, mainly in four ways: all movements, mobile and static devices, week and weekend and from or to a building during a certain time period. Section 7.3 will discuss the details of the result in the following order: first in subsection 7.3.1, all movements in the database are described regardless of any time or spatial components, then in subsection 7.3.2, the difference of movements between mobile devices and static devices will be discussed. In subsection 7.3.3, temporal component is taken into consideration, the movement patterns in week days and weekends are going to be described. Finally in subsection 7.3.4, movements from or to a building in a certain time period will be explored.

7.2. Methods

A state is defined as a time interval during which a particular device is located in a certain area and movement is always from the location of one state to the location of another state, where two states can not be the same. On building level, movement is defined as a state between two successive scans in different buildings.In a database, one movement record contains the start time of the movement, the end time of the movement, the start building, the end building and whether a device is labelled as mobile device or static device. With time information, it is easy to distinguish week days and weekends, and look for movements only in a certain time interval. With building information, movement patterns from or to one specific building can be found. With the attribute 'type', static devices can be filtered out so that only mobile devices are kept in order to make results more reliable. Additionally, the difference between movements of static devices and mobile devices can also be explored in maps and graphs.

In order to find movement patterns, a GUI is made for automatic visual exploration of the data and movement patterns. For the following four topics, both graphs and maps are used to describe movement patterns. The graphs show the movement in time from 6:00 to 0:00 and the maps show the amount of movement between buildings. The movements are shown in straight lines between buildings because the outdoor space of the campus is not constrained. This makes it impossible to know how exactly people move from building to building, thus snapping the movements to the road network could give the readers the wrong impression.

The line width represents the amount of movements, the thicker the line is, the more movements there are. Besides, the color of the line also stresses the amount of movements, where red represents the most movements.

7.3. Results

7.3.1. All movement

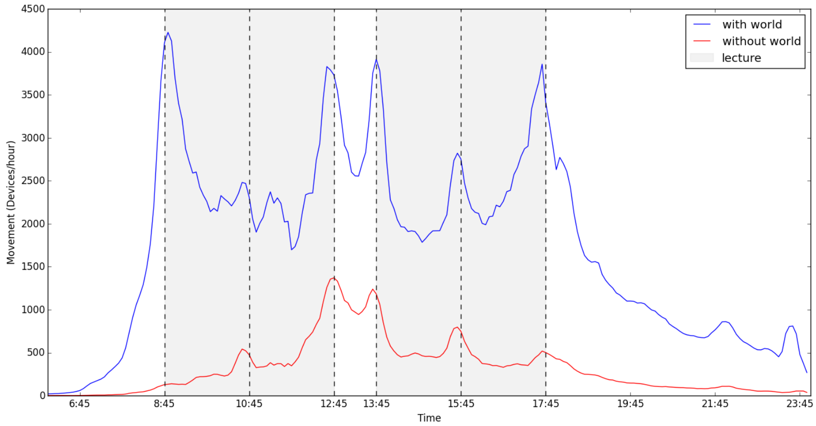


Figure 7.1: Graph of all movements

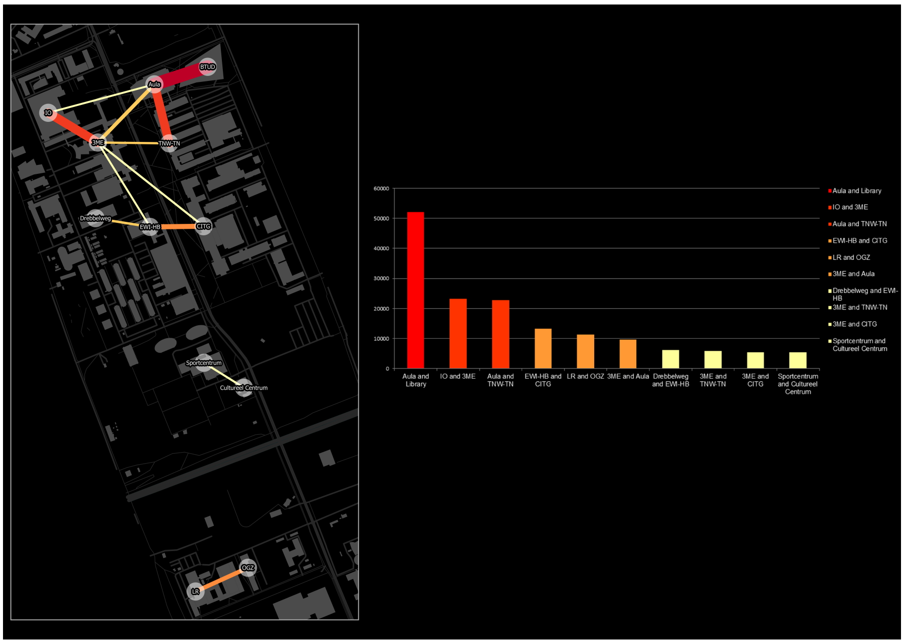


Figure 7.2: Maps of all movements

All movements from April 1st to May 27th without filtering static devices is shown in the graph (Figure 7.1) and map (Figure 7.2). As shown in the graph, there are several peaks at 8:45, 12:45, 13:45, 15:45, 17:45, which are the beginning and the end of the lecture. With 'world', it is clear to see the morning rush hour at around 8:45 when students all come to the campus and around 17:45 when they leave the campus. Both the curve 'with world' and 'without world' show peaks at which movement increases. The amount of movement with 'world' included is higher, because there are lots of users moving from outside to the campus, more than user that travel only between buildings on the campus. The map in Figure 7.2 shows the top 10 movements between buildings on the campus. It is clear that between Aula and library, there are over 50000 movements, which is much more than the others. The other buildings are building pairs which are more connected and close to each other. People most often travel between these buildings. So the map actually shows the connectivity of the buildings. It is clear to see that the amount of movements are determined by the locations of the buildings. Normally many movements will happen between two buildings near to each other. Moreover, it also depends on the functionality of the buildings. For example, there are many students of EWI having lectures also in CITG, that could explain why there are many connections between these two buildings.

7.3.2. Mobile vs static

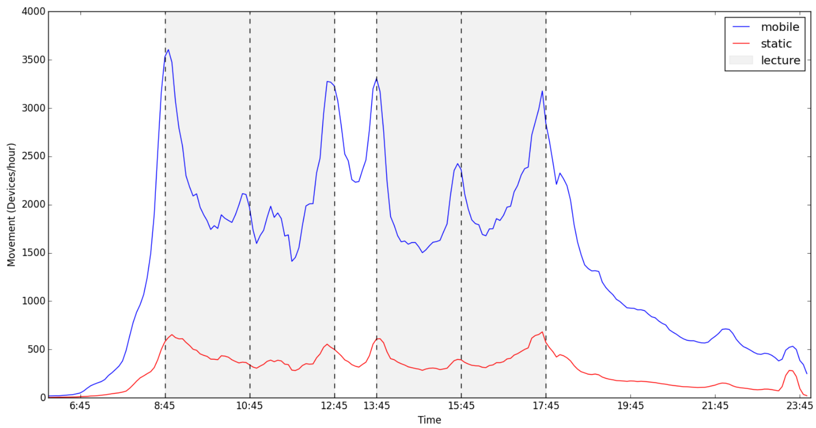


Figure 7.3: Graph of movement of static and dynamic devices

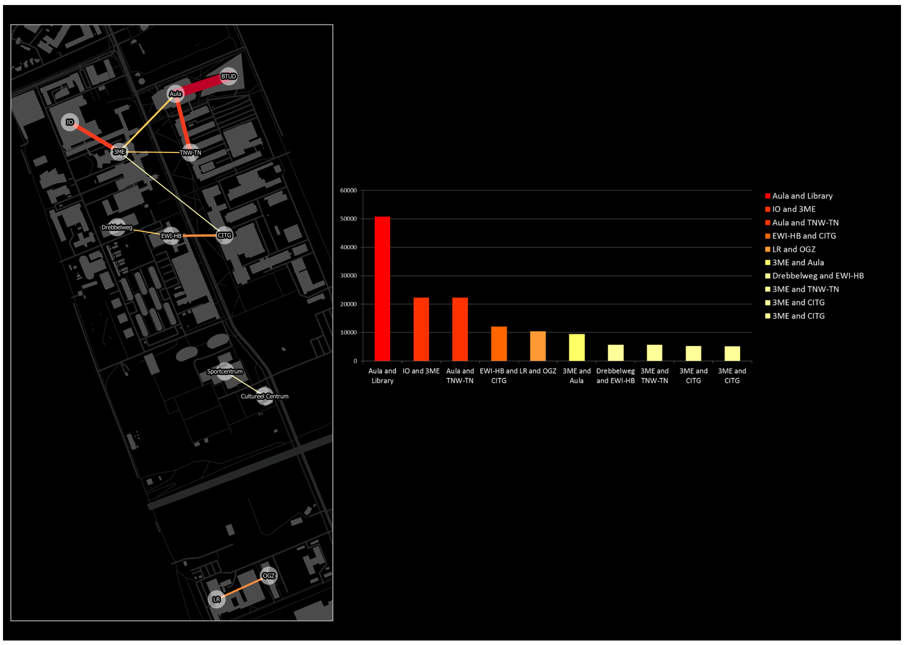


Figure 7.4: Map of mobile devices

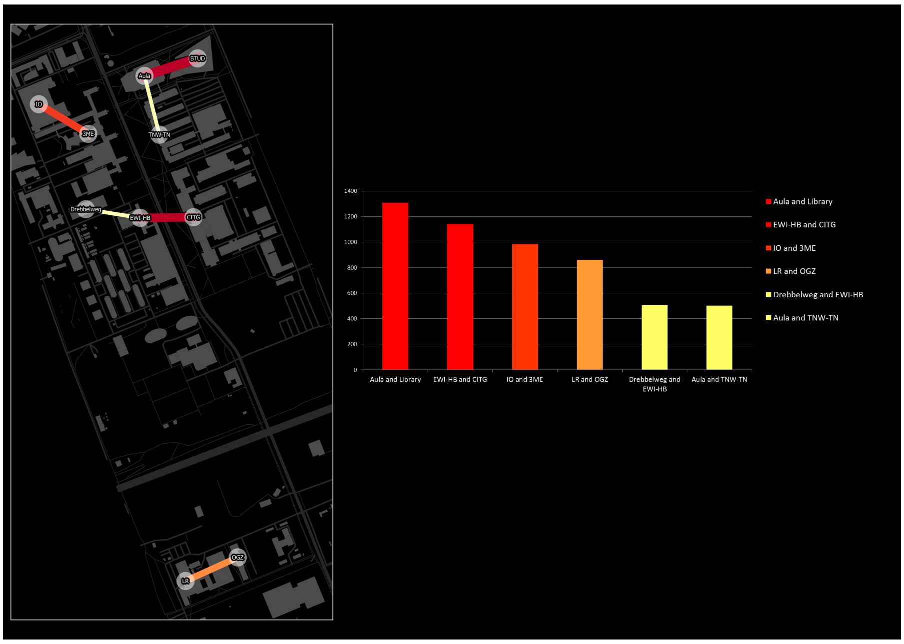


Figure 7.5: Map of static devices

The graph (Figure 7.3) shows the difference between mobile and static devices. It is obvious that compared with peaks of mobile devices, the peaks of static device are more flat, which indicates that mobile devices are more mobile than static devices. It also implicitly proves the correctness of the classification of mobile device and static device. The two maps, Figure 7.4 and Figure 7.5, shows how mobile devices move between Library and Aula much more than static devices. Additionally, there is more movement between the Library and the Aula than between other buildings.

7.3.3. Week vs weekend

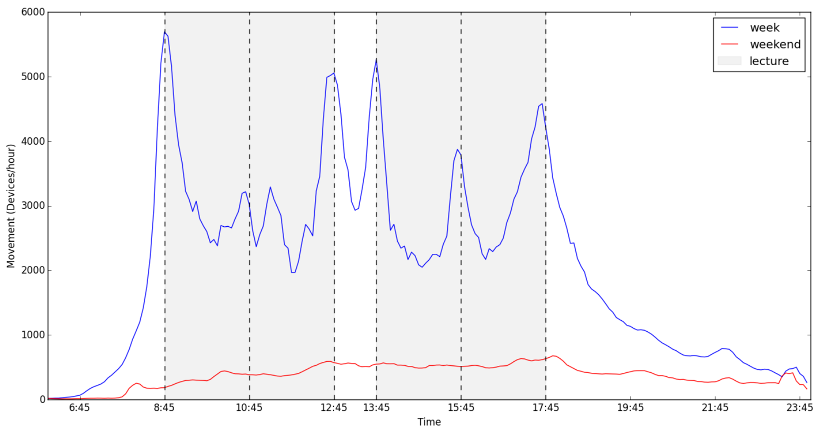


Figure 7.6: Graph of weekdays and weekends

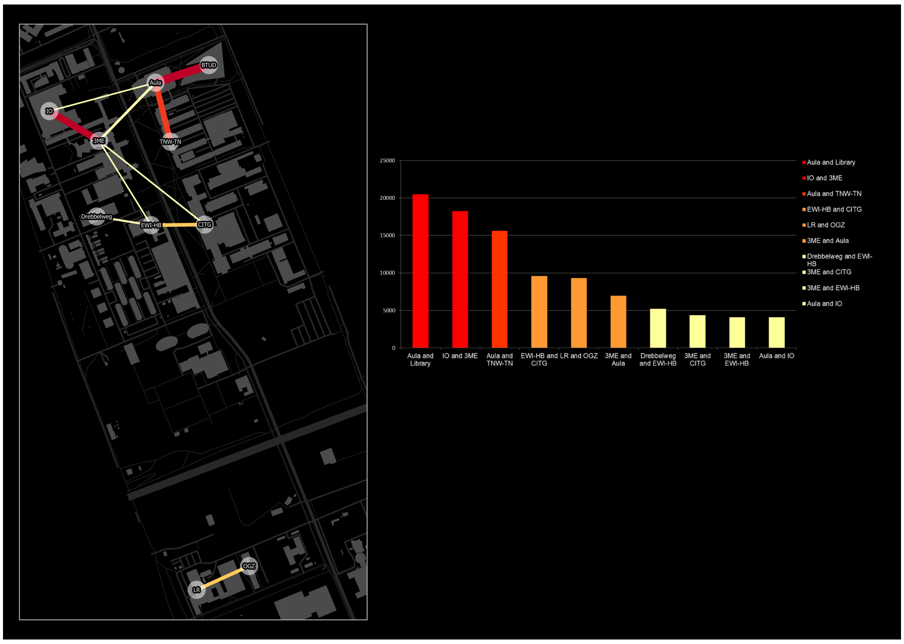


Figure 7.7: Map of weekdays

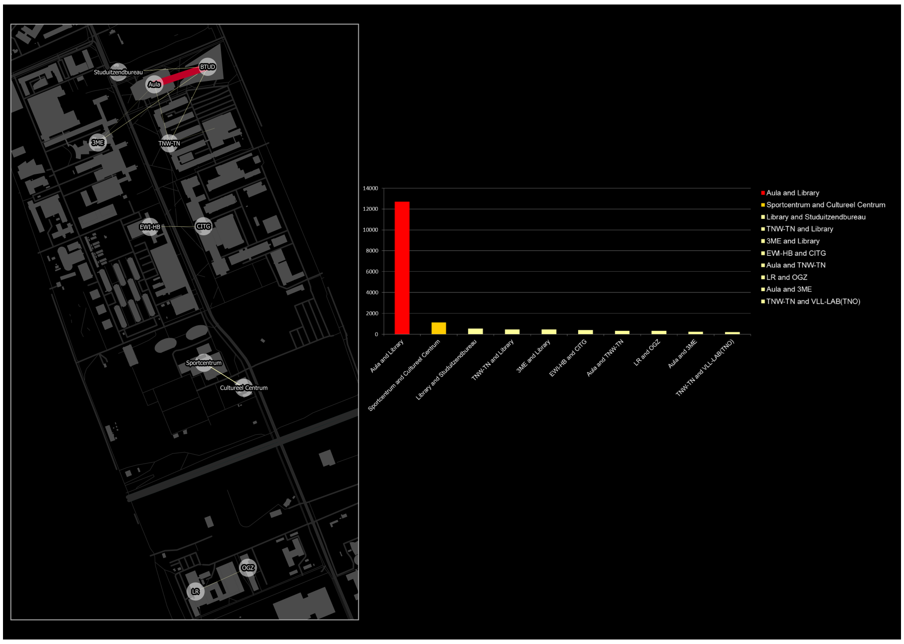


Figure 7.8: Map of weekends

The Figure 7.6 shows the different movement patterns in weekdays and on weekends. It is clear that in weekdays, the movements are more regular, the peaks of movements match the lecture time very well. However on weekends the movements is much less and also less regular than weekdays. On weekends. the curve is flat, indicating that the amount of movement is more or less the same throughout the day . The map of weekdays (Figure 7.7) is similar to Figure 7.2, but the movement between sportcentrum and cultureel centrum is no longer in the top 10 movements. However in the map of weekends(Figure 7.8), there are far more movements between Aula and Library, and sportcentrum and cultureel centrum is top 2. Comparing figure Figure 7.7 and Figure 7.8, it is clear to see that more people visit the sportcentrum and cultureel centrum during the weekends.

7.3.4. From and to

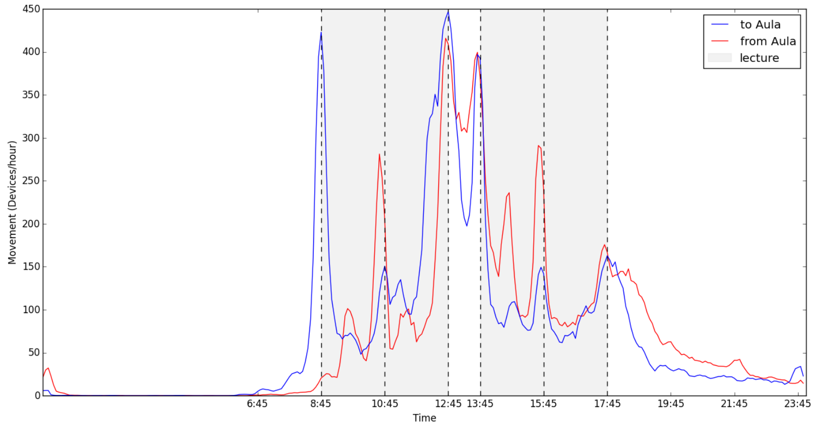


Figure 7.9: Graph of movement from and to aula

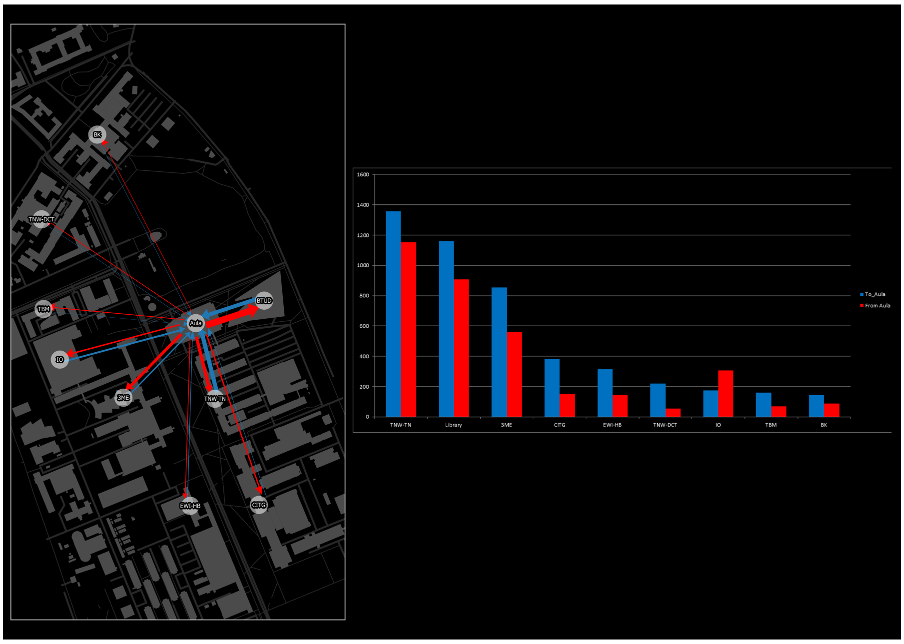


Figure 7.10: Map of from or to Aula between 13:15 and 14:00

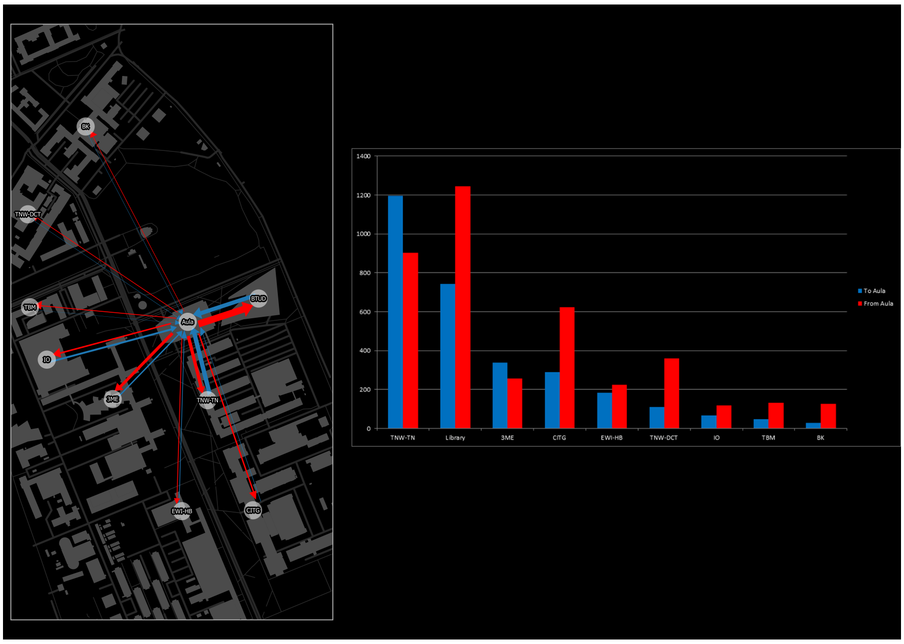


Figure 7.11: Map of from or to Aula between 13:15 and 14:00

The graph in Figure 7.9 shows the movements from and to Aula in time. At 8:45, there are many movements to Aula, the reason might be that there are also lectures in Aula at this time. And many people leave at 10:45 when the lectures end. During lunch time between 12:15 to 14:00, there are almost the same amount of people moving to and from Aula. The two maps Figure 7.10 and Figure 7.11 respectively show the movements from and to Aula from 12:15 to 13:00 and from 13:15 to 14:00. The figures show which other faculties use the Aula most during lunchtime. Most people go from TNW-TN to Aula, from Library to Aula and from 3ME to Aula. In general, more people move to Aula than from Aula between 12:15 and 13:00. From 13:15 to 14:00, more people move from Aula to other buildings, the Library in particular. The buildings in Figure 7.10 match the buildings in Figure 7.11, which to some degree proves that people from these faculties do use Aula as canteen during lunch time.

8

Trajectory patterns

8.1. Introduction

This GSP attempts to identify people's movement patterns from anonymized Wi-Fi logs. Chapter 7 described movement patterns including spatial and temporal aspects of single movements of a crowd of people. Another way of looking at movements, is by tracking individual movement for a longer time interval. A large set of individual trajectories can be used for the identification of typical movements among users of the campus. The method uses concepts from sequential pattern mining.

This chapter presents a method for identifying movement patterns using individual trajectories. As described in chapter 5, if moving individuals share some locations in their trajectory, you can speak of co-location in space. When the order of the shared locations are similar for multiple trajectories, you can speak of typical movement. This concept is explored for the identification of movement patterns, and thus the usage of the campus. This approach can answer different questions than looking at single movements, as is done in chapter 7. For example, 'how many places the user frequently visits', 'at what order the user visits places', 'how often a trajectory happens', 'how many places contained in a frequent trajectory'.

First, this chapter will describe the problem description, including the extraction of locations of a user, the mining of individual trajectories from an anonymized Wi-Fi scan list, and finally the mining of movement patterns from a set of trajectories using the PrefixSpan algorithm.

8.2. Problem description

The data provided by the eduroam network enables a detailed view of people's movement on campus. The large coverage of the eduroam network allows to track users for a large part of the day when they enter the campus. However, the observation space is limited to the extent of the size of the campus, making it not possible to track people outside the eduroam network. A second disadvantage is the spatial resolution of the positioning method. The range a mobile device can be connected to an AP, influences the accuracy of the estimated location of a mobile device. For indoor environments of the TU Delft campus, this is just a few tens of meters wide. This resolution allows tracking movement at a building level by re-locating mobile devices to the closest AP. Data between two re-locations is not available. Therefore, an individual's trajectory is depicted by connecting the re-locations as a sequence of APs. These individual trajectories are used to identify patterns.

A location represents a geographic position where a user stays, i.e. a user is in state. For identifying movement patterns from Wi-Fi monitoring, we are interested in movement between two locations where an individual stays for a longer time period. Such a location, or stay place, can be detected when a user is connected to the same AP for a longer time. To detect buildings as a location (i.e. contains multiple APs), two consecutive Wi-Fi scans must contain APs of the same building. With a data collecting interval of 5 minutes, it means that people will be filtered out if their stay duration is less than 10 minutes. Based on this assumption, people with a shorter stay duration are considered passing by, as explained in section 6.1.

8.2.1. Trajectory Pattern

An individual's trajectory is constructed as a sequence of locations in order of the scan time.

$$p_1 \rightarrow p_2 \rightarrow p_3 \rightarrow \ldots \rightarrow p_n$$

From a set S of trajectories, different patterns can be identified using sequential pattern mining algorithms. Frequency of a trajectory by all users of the campus can be detected. This can be represented as a trajectory T with a support s. Support means how many times the same sequence, or sub-sequence, is shared in the set of trajectories. This gives valuable information on the order common buildings are used and what order of buildings occurs the most. Using a minimum support threshold, sequential mining returns all movement patterns that satisfy $n > 2$ and support $T > S_{min}$ Furthermore, the length of common trajectories can be discovered. This allows for identification of movement patterns of a specific length n. Also, when location is not considered, but only the length of a trajectory, the mobility pattern of an individual can be described in terms of how many times he/she re-locates. Figure 8.1 illustrates a trajectory pattern of length 3, and has a support of 3.

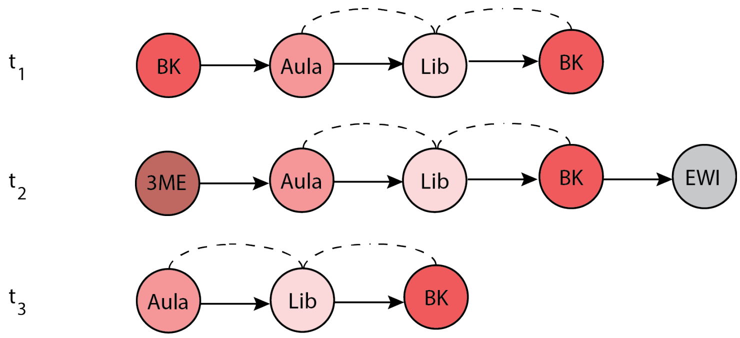


Figure 8.1: trajectories of three days, and the trajectory pattern

For this study, a trajectory pattern is a sequence of states with $n > 2$ and support $> S_{min}$. We are only considering trajectory patterns with $n > 2$, because chapter 7 already looked at two consecutive states.

There exists many developed sequential pattern mining algorithms. For this study PrefixSpan Pei et al. 2004 is used to identify common shared trajectories or sub-trajectories. This sequential pattern mining algorithm can find re-occurring sequences or sub-sequences from a set of trajectories. For every common sequence, a support value is computed.

8.3. Implementation

For this analyses the same data is used as for the single movement analysis. As described in section 6.1, from the raw Wi-Fi log states are extracted. More than 2.8 million states are identified from the dataset. This information is stored including a unique mac address, a number representing a building, the start time of the state and the end time of the state. The states are used to construct individual sequences ordered by date and time. The T_{split} is used to create separate trajectories for different days, for each individual. For this study, a new trajectory is created when there has not been a connection for 5.5 hours, i.e. a state of outside campus ('world') > 5.5 hours. This threshold is suitable for identifying people moving home at the end of the day and coming back the next morning. After splitting the sequences, over 950.000 trajectories are created, with temporal granularity of one day. Every trajectory starts and ends with 'world', i.e. people start and end there trajectory outside the campus. A sample of a constructed trajectory can be seen in Figure 8.2

World ⟶ Bk ⟶ World ⟶ Aula ⟶ World
World ⟶ 3ME ⟶ Lib ⟶ Aula ⟶ Lib ⟶ World
World ⟶ EWI ⟶ World

Figure 8.2: sample of individual trajectories

Based on the created trajectories, trajectory patterns with a support value are detected by applying the PrefixSpan algorithm. Figure 8.3 shows an example of the detection of patterns with a support value given by the sequential pattern algorithm. Logically, the pattern with the highest support is a length-1 sequence. The longer patterns get, the lower the support will be.

```
0 23 0                    ([0, 21], 4)
0 23 0                    ([21, 0], 4)
0 32 0                    ([21, 21], 4)
0 22 20 0                 ([0, 0, 21], 4)
0 22 0                    ([0, 21, 0], 4)
0 21 0 21 0       ──→     ([0, 21, 21], 4)
0 21 0 21 36 0            ([21, 0, 21], 4)
0 21 0 21 0               ([0, 21, 0, 21], 4)
0 36 0 36 0               ([0, 0], 5)
0 21 0 21 0               ([0], 10)
```

Figure 8.3: sequential pattern mining sample

8.4. Results of trajectory pattern identification

This section describes the result of the trajectory pattern mining. We used trajectories of mobile devices only, see section 6.2.

8.4.1. Length of pattern

From the individual trajectories, the length can be retrieved. This trajectory length is plotted in a histogram, showing how many different places users visit during a day. Figure 8.4 shows that most trajectories consist of three states, i.e. entering the campus, visit one building and leave the campus. The average number of states in a single trajectory is 3.95, this provides information about movement behaviour of users on the campus.

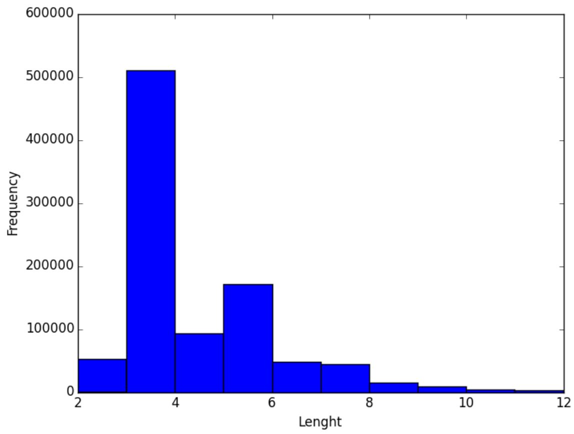


Figure 8.4: length of individual trajectories as a sequence of states at buildings

In the same way, trajectories at the spatial level of inside a building can be analysed. In chapter 9 more on indoor movement is discussed. We analysed the trajectories of individual users inside the Faculty of Architecture and the built environment. The length of more than 150.000 trajectories is plotted in Figure 8.5. This figure shows, compared to Figure 8.4, that people visit more different places indoor, than the number of different buildings, i.e. people are more mobile inside the Faculty of Architecture and the built environment compared to movement between buildings. The average number of states in a single trajectory inside the Faculty of Architecture and the built environment is 4.66

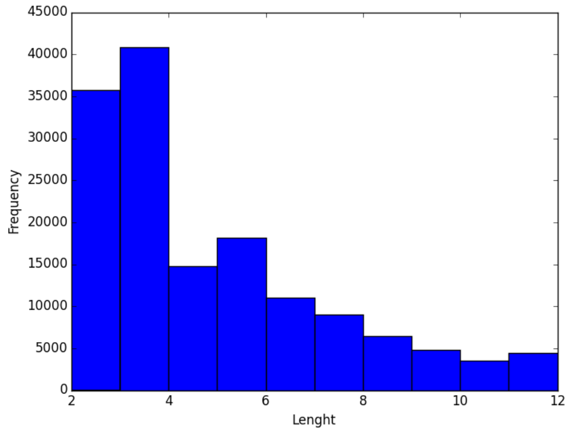


Figure 8.5: length of individual trajectories as a sequence of states at indoor building-parts

8.4.2. trajectory length-4 pattern

For the identification of trajectory patterns, the trajectories from the building spatial level are used. To retrieve information about the most common order of at least four distinct locations, we only considered trajectory patterns with a length of four and longer, including 'world'. The three most frequent used order of four distinct locations with support > 1000 is shown in Figure 8.6.

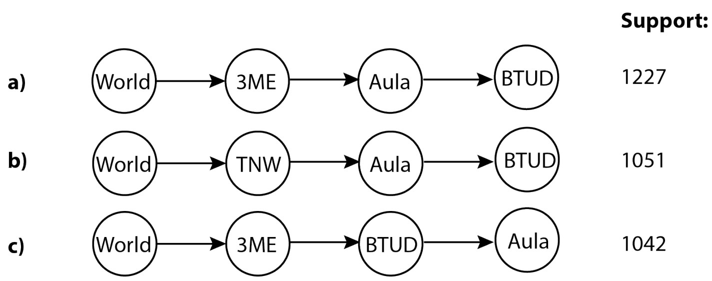


Figure 8.6: trajectory pattern with support > 1000 and four distinct locations

These three movement patterns (A, B and C) from users of the TU Delft Campus are visualized on the map in Figure 8.7, Figure 8.8 and Figure 8.9.

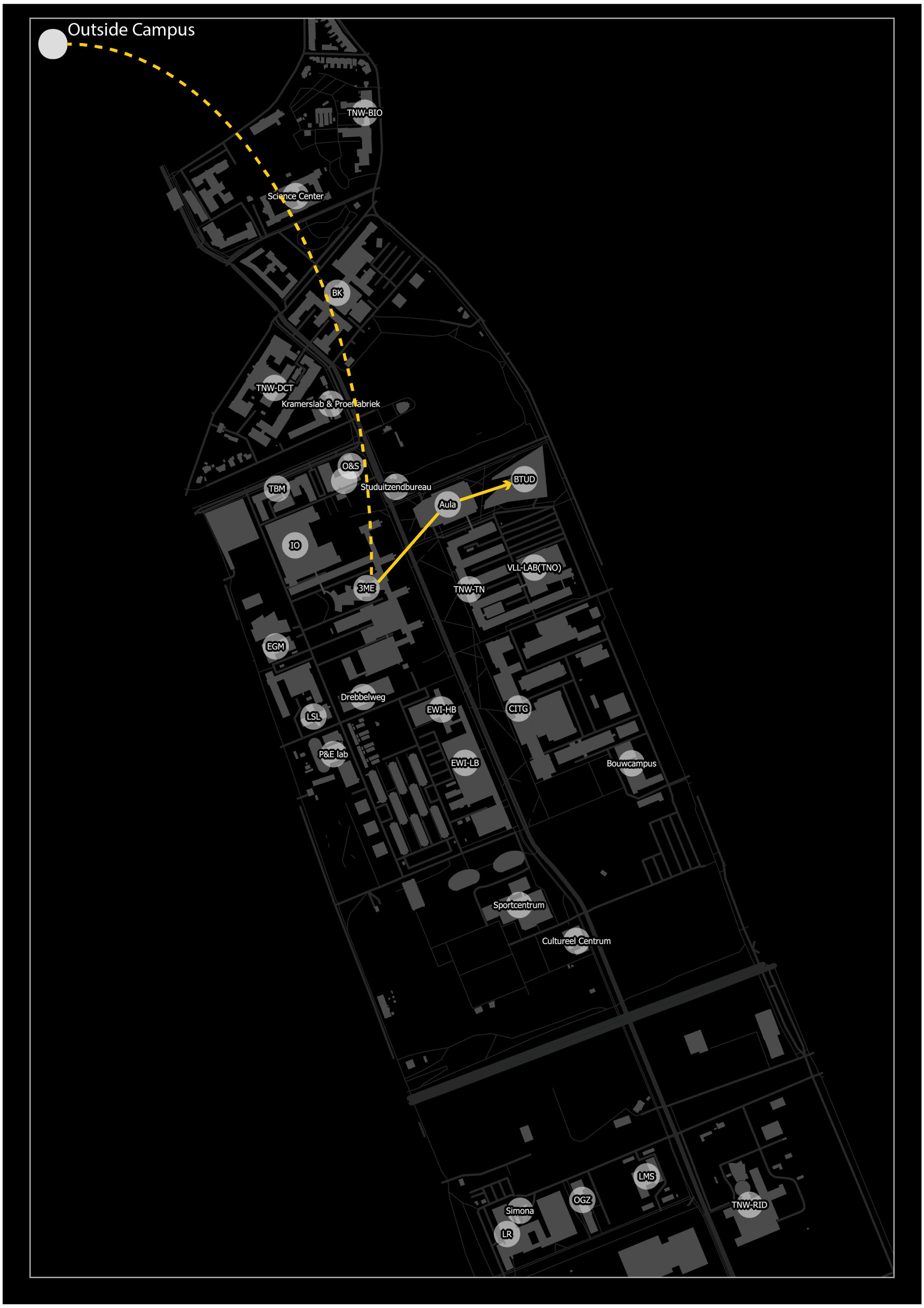


Figure 8.7: Trajectory pattern A,

$world \rightarrow 3me \rightarrow aula \rightarrow btud$

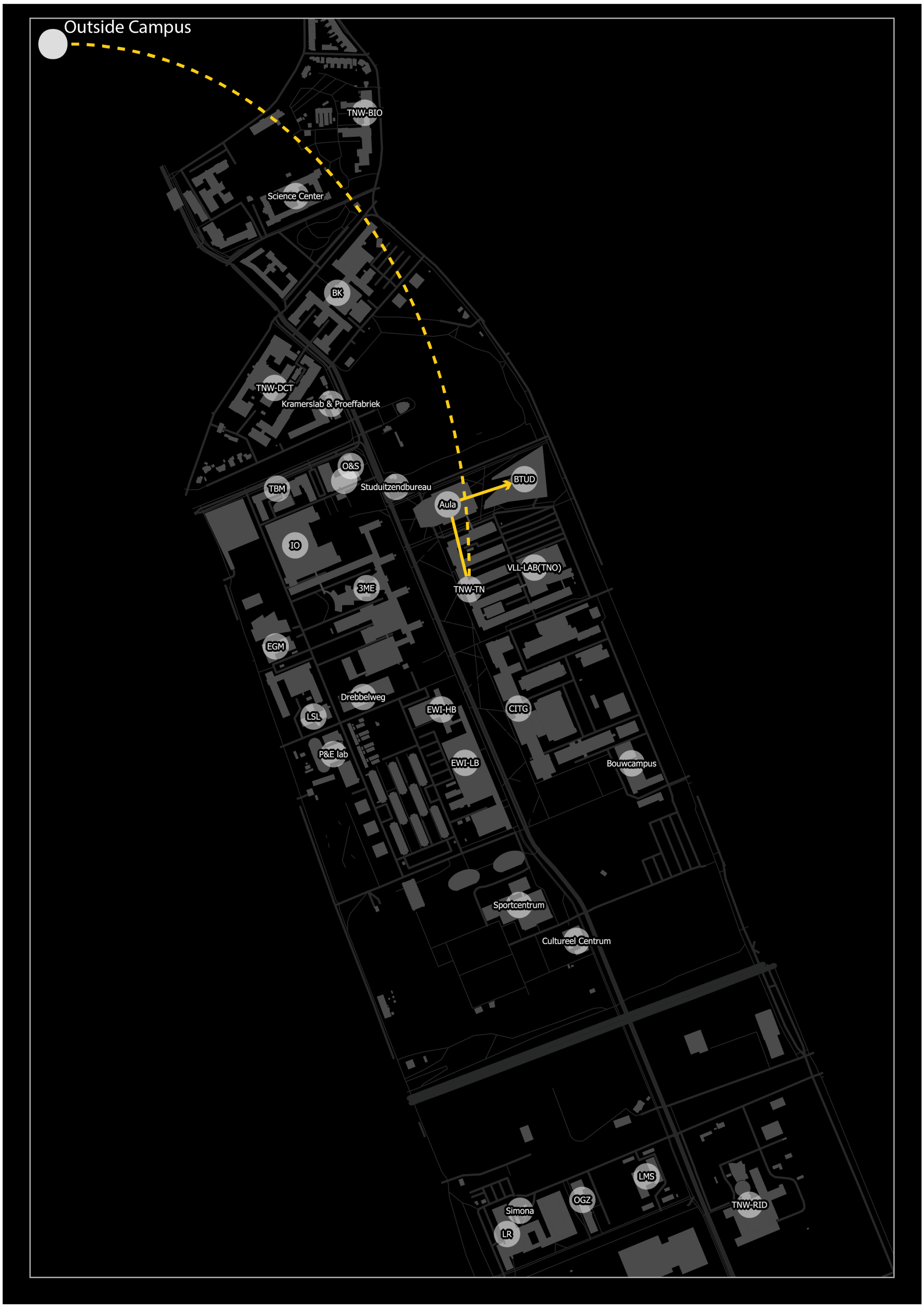


Figure 8.8: Trajectory pattern B,

$world \rightarrow tnw \rightarrow aula \rightarrow btud$

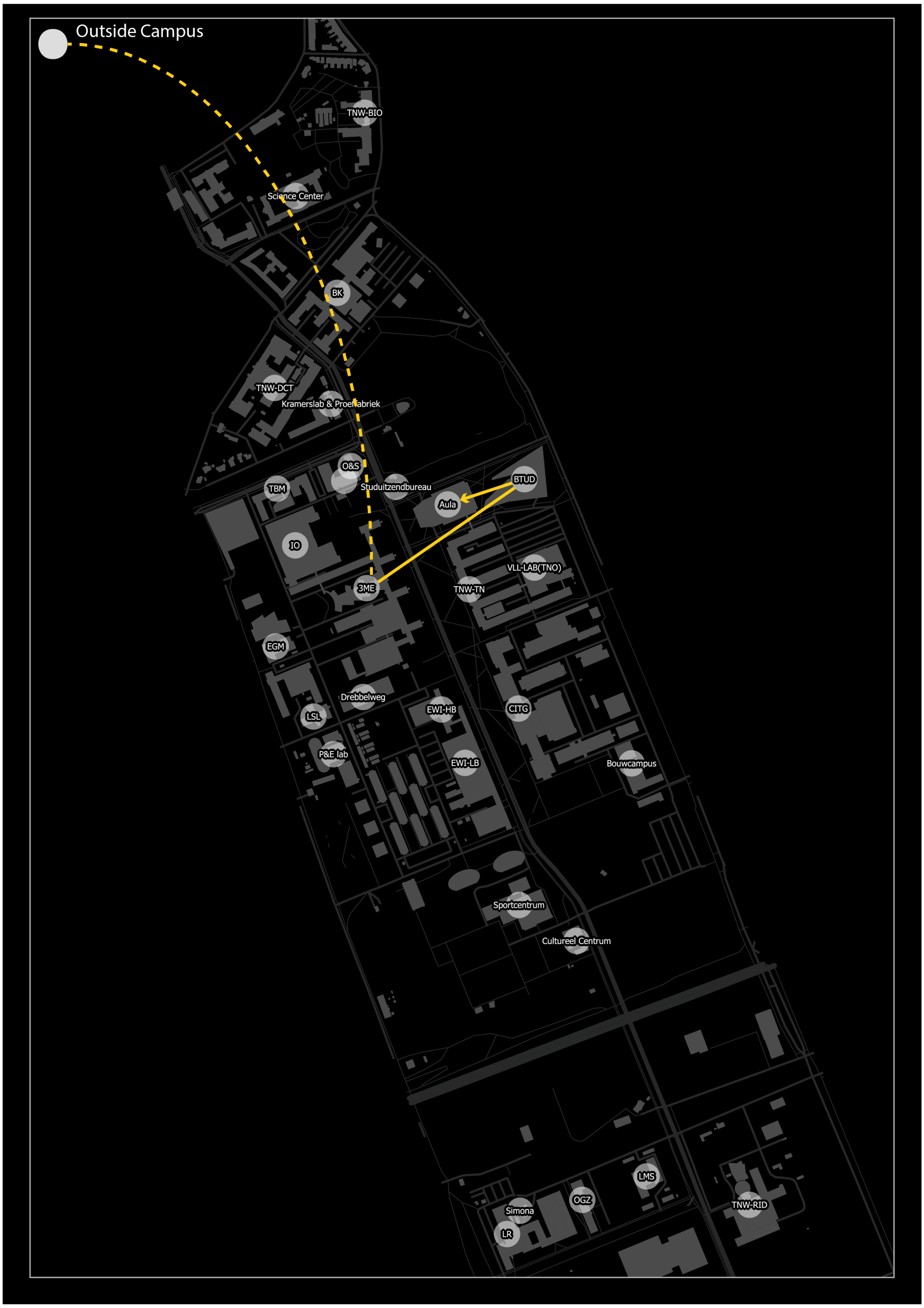


Figure 8.9: Trajectory pattern C,

$world \rightarrow 3me \rightarrow btud \rightarrow aula$

8.4.3. Trajectory filtering

Trajectories can serve another purpose. Although a splitting threshold of 5.5 hours was used, we noticed numerous trajectories with a temporal granularity of more than one day. Considering human life patterns, and mobile devices are for this research related to people, trajectories should present human behaviour. This means that people usually come to work in the morning, leave in the evening to go home and sleep during the night. Trajectories that are stretched over more than one day do not illustrate human behaviour, as the mobile devices do not leave the 'working environment' to go home and sleep. Besides a longer time interval of several trajectories, they also show a systematic pattern in there states. Only by creating trajectories and based on knowledge of human movement behaviour, can these trajectories be filtered out.

9

Indoor movement

9.1. Introduction

As described in the firs part of this report, Wi-Fi tracking data can be used to identify movement between buildings. Given that indoor areas are usually better covered with Wi-Fi access points than outdoor areas, it is natural to also look at movement inside buildings. The following section describes the method of identifying and visualizing indoor movement in the Faculty of Architecture and the built environment of the TU Delft.

The process of indoor movement analysis is conducted along the steps below, thus the section also follows this structure:

1. Delineate building parts based on the layout of access points and the division of the building (e.g. department, canteen, building wing), and group the access point into building parts.

2. Identify movements in the data between building parts.

3. Create a route network that connects the building parts and is constrained to the corridors of the building.

4. Assign the movements to the route network.

5. Visualize the amount of movement along the indoor network.

9.2. Theory / methods

After identifying movement between different buildings, the next level is to do so between different parts inside a building. These parts represent functional or spatial divisions inside a building, e.g. departments, community areas, building wings and are referred to as building-part.

A prerequisite of the method is to know the at least room level location of the access points in the respective building. At the time when the project was carried out, the detailed access point locations were available only for the Faculty of Architecture and the built environment. Thus the focus on this particular building.

As opposed to outdoor pedestrian movement which is not necessarily constrained on a fixed network, indoor movement is constrained by the layout of the respective building. The building parts of the Faculty of Architecture and the built environment can be represented by its underlying graph, having the building parts as nodes and the corridors as edges Figure 9.1. Then indoor movement is necessarily constrained on this underlying graph.

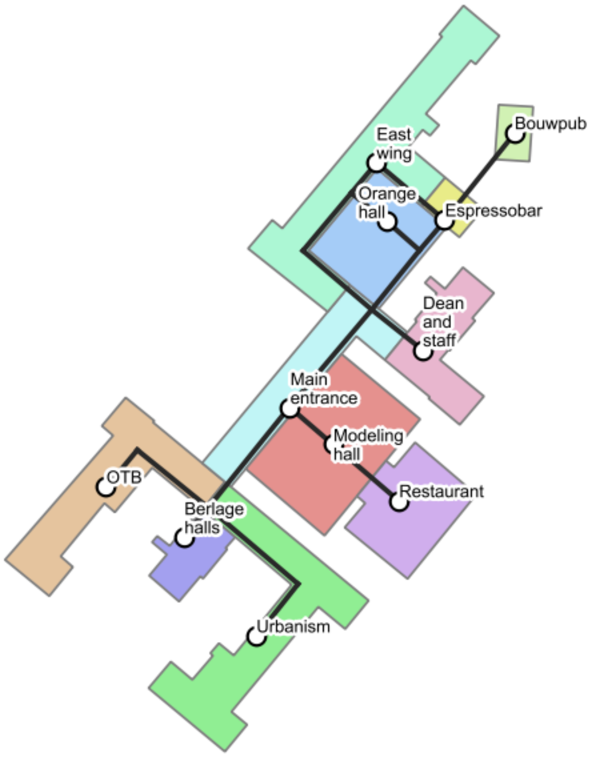


Figure 9.1: Building parts on the ground floor of the Faculty of Architecture and the built environment and its underlying graph.

The Wi-Fi system of the TU Delft campus has a five minute scan interval, which is too coarse to catch detailed movement indoor. As five minutes is sufficient to reach any two locations in the building taking any route. Therefore not the movement trajectory itself is identified from the data, but the fact of relocation from origin to destination. Then the path of the movement can also be identified by analysing the layout of the building. For example if a person stayed at the canteen, then soon after he stayed at the orange hall, he necessarily had to traverse the corridors in-between these two locations. Our method is based on this assumption.

Due to the building layout, in most of the cases there is only one possible direct route between two building parts. However, in case of multiple route options, the exact route of a movement is assumed to be the shortest route between origin and destination.

9.3. Implementation

The identification and visualization of indoor movement in a procedure requires various tools and steps. While some steps can be automated, others need to be done manually. The detailed description of these steps are as follows.

9.3.1. Delineation of building parts

There are two factors that define what is considered a building part, the layout of the building and the layout of the access points. The layout of the building defines the functional divisions, e.g. departments or common areas. Additionally, it is necessary to have at least one access point in each of these divisions, or preferably more access points equally distributed in the division. Considering the signal range of an access point, it is not desirable to have access points close to the boarder of two neighbouring divisions, as in that case the user could be falsely located in the neighbour division if he is picked up by the respective access point. The combination of a functional division and the access points within define a building part.

In case of the Faculty of Architecture and the built environment Figure 9.2 displays the provided access point map and the manually overlaid functional divisions, thus defining the building parts.

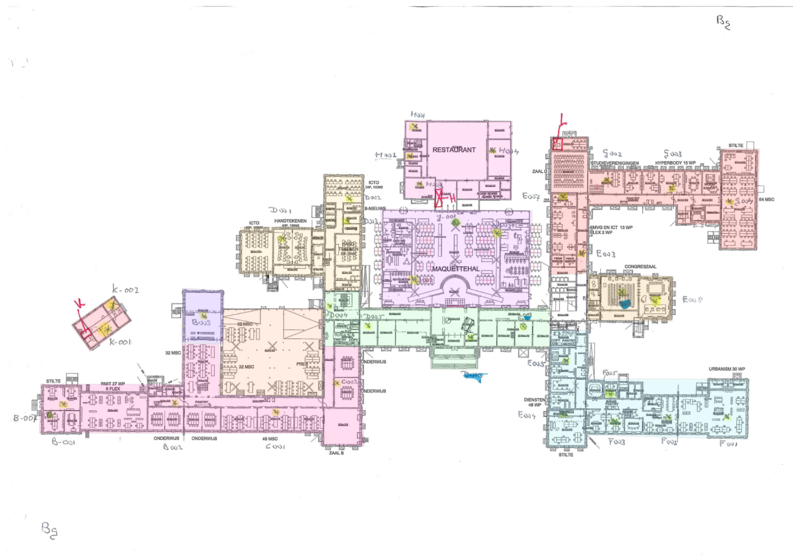


Figure 9.2: Access point map where yellow dots mark the access points, and the functional divisions (coloured areas) on the ground floor at the Faculty of Architecture and the built environment.

9.3.2. Movement between building parts

The method how indoor movements are identified is described in section 9.2.

9.3.3. Indoor route network

The route network of the Faculty of Architecture and the built environment where nodes represent building parts and edges represent corridors was drawn manually in *QGIS*, following the floor plan of the building. However, the resulting *spaghetti network* does not contain the topological relations that are required to calculate a shortest route. Therefore the topological relationships were created with the PostGIS extension *pgRouting*. Using a database-based solution for storing the data, creating topology and calculate shortest routes allowed us to easily match the movements, which were calculated in the database, to the route network.

9.3.4. Mapping traffic to the route network

In the *movements table* every record represent a single move of a person from origin to destination. In order to display these movements, identical moves that have the same origin-destination pair are aggregated, resulting in a table of unique origin-destination pairs with the amount of related moves (Table 9.1).

Origin	Destination	Count
OTB	Restaurant	126
Main entrance	Espressobar	543

Table 9.1: Aggregated moves between building parts

Then the shortest route between each origin-destination pair is calculated and the movement counts are added to each edge that is traversed in the network. Thus if the shortest route of two distinct movements share edges, the movement count is summed up on the common edges, resulting in the traffic load of a given edge (Table 9.2).

Edge ID	Traffic	Line width
45	6151	1.10
46	1994	0.64

Table 9.2: Traffic load on the indoor route network

9.3.5. Visualization of the movement

The visualization method, as well as the route network, is two-dimensional. However, three-dimensionality is imitated by using an *exploded view* common in architectural visualizations, that shifts overlapping elements (e.g.floors) by a certain angle.

In this graphic the *nodes* that represent the building parts are the approximate centroids of the polygonal area of the building part. The nodes were manually adjusted to better match the route network.

The route network is represented with straight lines, where the *line width* is proportional to the traffic load of a given edge. However, line widths cannot be compared across graphics, as in order to facilitate consistent scale the line width variable is normalized to the range of 0.5-5 units. The range of 0.5-5 units is chosen to provide a visually appealing and clear graphic. Colours mark the four separate floors and the staircases (grey) in the building (Figure 9.4).

9.4. Results

Considering the movement from any origin to any destination at the Faculty of Architecture and the built environment, throughout the whole measured period, our results clearly indicate the peak hours in the morning, before and after lunch Figure 9.3.

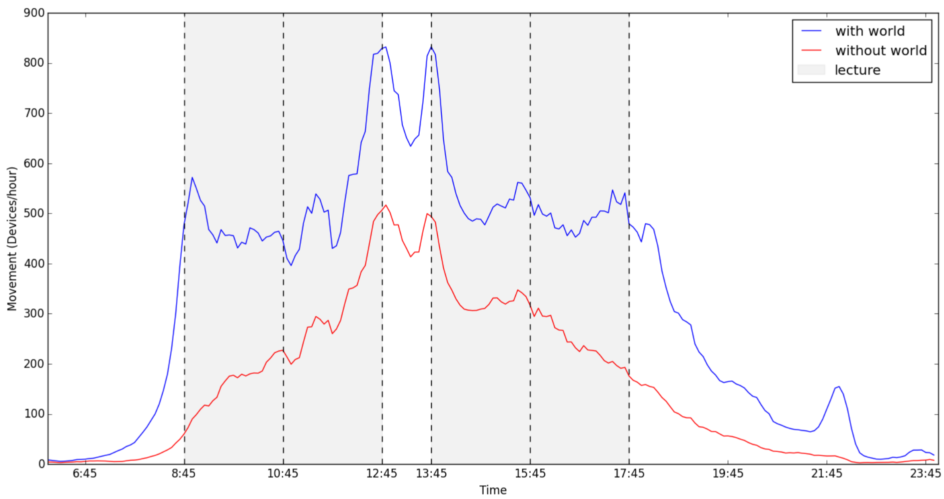


Figure 9.3: Total movement at the BK

Furthermore, if the same data is visualized using the previously described method, we can observe the occupation of corridors in the building. The advantage of this method that it provides insight into the usage of those spaces where data is not directly available. See the limitations of the eduroam system to track detailed indoor movement in section 9.2.

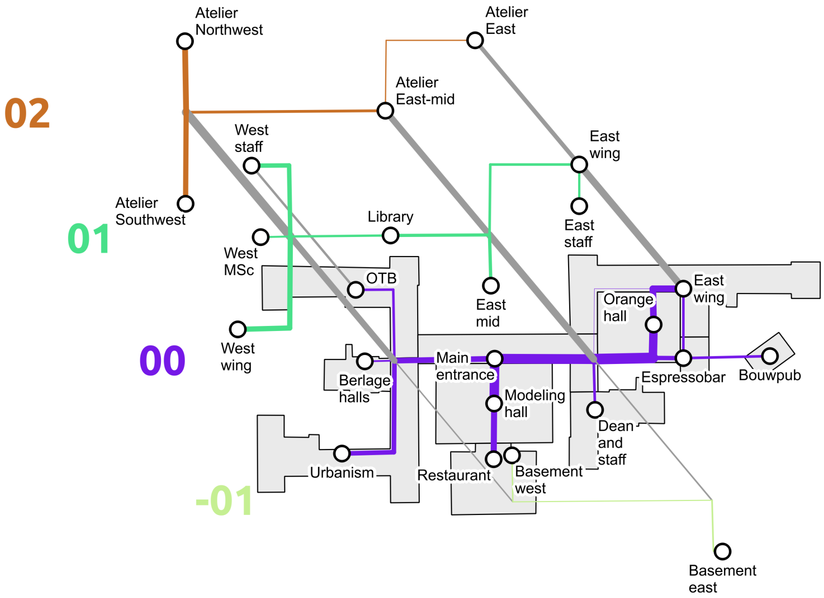


Figure 9.4: Occupation of corridors at the BK

Additionally, we analysed movements of mobile and static devices, weekdays and weekends, to the Bouwpub and to the Restaurant at the Faculty. These results are displayed in Appendix A.

10 Conclusions

To understand human motion behaviour for better decision making, many studies have been conducted based on location data collection. Wi-Fi tracking technology is increasingly used due its cost effectiveness and ability to track people at a large scale. For this study, we used the eduroam network of the TU Delft Campus to identify movement patterns. Firstly, states are extracted from the raw Wi-Fi logs. Subsequently, the event of going from one state to another can be detected as movement. Finally, by counting the number of movement for an observation period, movement patterns can be identified. This project tried to illustrate to what extend movement patterns in and between buildings can be identified from large scale Wi-Fi based location data of the eduroam network. In order to answer this question, there are three applied sub questions:

- What patterns can be identified moving from and to the TU Delft campus?
- What movement patterns can be identified between buildings on TU Delft campus?
- What movement patterns can be identified between large indoor regions of the Faculty of Architecture and the built environment?

First of all it can be concluded from the results, that the Wi-Fi network data is to some extent suitable for retrieving movement patterns of people. Expected patterns such as a movement peak between buildings during lunch time, and a morning and afternoon peak of people entering and leaving the campus can be clearly distinguished in the data. Additionally, the data shows actually high movement peaks just before the lectures start. More specific patterns between particular buildings and/or during certain time intervals can easily be derived due to the automated workflow.

At building level, the rhythm of the campus is illustrated by time profiles showing the amount of movement for different observation periods. Peaks of high movement could be distinguished just before a lecture starts (at times 8:45, 10:45, 13:45 and 15:45). Additionally, aggregated movement on the map show the expected result that Aula-Library is the most frequently travelled path.

Moreover, the indoor movement analysis shows that users of the Faculty of Archicture and the built environment deviate from the trend. This could be explained by the open form of education, because designing in a studio is not limited to strict lecture times. Additionally, the the faculty is also functioning as a meeting place for project groups. An indoor network graph was created of the underlying building floor plan. This successfully illustrates the occupied space for movement. However, the range of APs can extent between building-parts and floors and limits the accuracy of the analysis.

11

Recommendations

11.1. Entrances

Introduction
This section will describe the work that is done to find out what, when and how frequent entrances of the Faculty of Architecture and the built environment are used. This is an interesting and challenging use case at the same time. The Faculty of Architecture is a building having multiple entrances; five to be precise. Knowing what, when and how frequent these entrances are used, will give insight into the use of a building, the spatial context and the relation between these two.

Methodology
In order to find what entrance someone uses to enter or exit a building, we will look in the part of a sequence in which the device is recorded by an AP in a building and subsequently recorded by an AP in another building. More specific, we will look at which AP(first or last) is used in a movement from one to another building. For this two different approaches can be distinguished. The first approach does not take in account the devices that might get recorded when passing by the building. In the second approach we will make use of the pre-processed data which excluded the passing by events.

Hypothesis
Our hypothesis is that finding clear answers to the question whether it is possible to identify what entrances are most frequently used, is going to be hard. Firstly, because the existing layout of APs is not designed for the purpose of tracking people. For this reason there is not always an AP located near an entrance. Secondly, because the logging frequency of the system is a little more than 5 minutes. Ideally the system records the connected device at the very first AP it connects with. The chance the device is recorded at the moment it is connected with the very first access point is small. However we still expect to see some results. Although the time interval in which the system logs the connected devices is relatively large, an AP located near an entrance would still pop up as one of the most frequently used AP as first connection (assuming people disseminate over the building after entering).

First approach: including passing by events
The first approach makes use of the raw wifilog data, by finding the part in a sequence in which a device is recorded by an AP in a building and is subsequently recorded in another building. The states in which a device is scanned once are not filter out. These single records imply that a device only passed by the building, and thus was not located in the building.

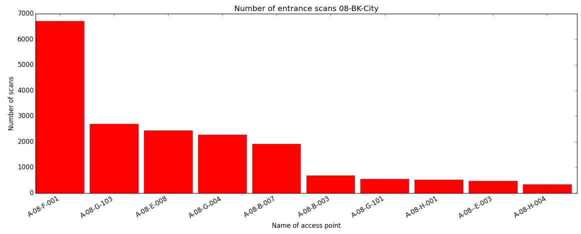


Figure 11.1: Most frequently recorded APs in a movement to the Faculty of Architecture and the built environment

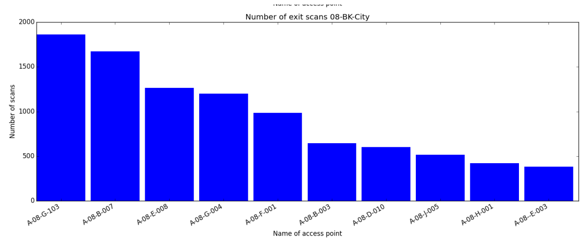


Figure 11.2: Most frequently recorded APs in a movement from the Faculty of Architecture and the built environment

The floor plans of BK, enriched with the location of APs, are used to locate the most frequently used APs on the map (see Figure 11.3). The result is interesting, since most APs are not located near an entrance but are located at one of the corners of the building. Most of the them are located at the western part of the building. Knowing that lots of people are passing in the street next to this part of the building, we can conclude the result of this analysis is distorted due not filtering out the devices that are recorded when passing by the building.

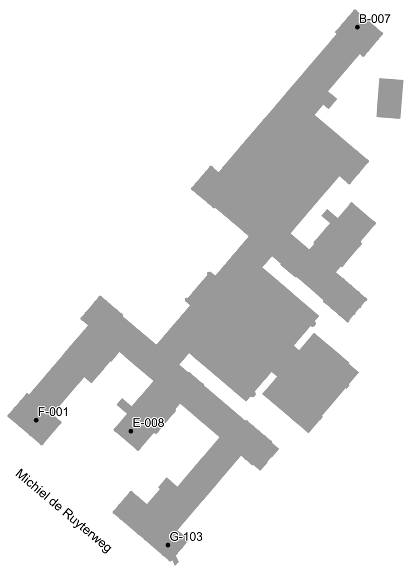


Figure 11.3: The location of the most frequently used APs that are used to record the first and/or last connection of a device in the Faculty of Architecture and the built environment

Second approach: excluding passing by events

Table Table 11.1 shows the individual states as a result of the pre-processing (see chapter pre-processing).

The records represent the states for each mac, including the first and last recorded AP (ap_start, ap_end).

mac	**building**	**ts**	**te**	**ap_start**	**ap_end**
000c+YfkIi..	0	30-3-2016 23:34	6-4-2016 22:39	NULL	NULL
000c+YfkIi..	21	6-4-2016 22:39	6-4-2016 23:30	A-21-0-005	A-21-0-045
000c+YfkIi..	0	6-4-2016 23:40	10-4-2016 19:53	NULL	NULL
000c+YfkIi..	0	10-4-2016 20:03	10-4-2016 21:13	NULL	NULL
000c+YfkIi..	21	10-4-2016 21:13	10-4-2016 21:34	A-21-0-046	A-21-0-046
000c+YfkIi..	21	10-4-2016 22:04	10-4-2016 22:19	A-21-0-045	A-21-0-046
000c+YfkIi..	0	10-4-2016 22:19	10-4-2016 23:14	NULL	NULL
000c+YfkIi..	0	10-4-2016 23:24	11-4-2016 12:27	NULL	NULL
000c+YfkIi..	21	11-4-2016 12:27	11-4-2016 13:25	A-21-0-043	A-21-0-043
000c+YfkIi..	20	11-4-2016 13:25	11-4-2016 13:56	A-20-0-008	A-20-0-045

Table 11.1: Individual states as a result of the pre-processing

The table also includes 'world' (in Table 11.1 represented by NULL) which implies the device is not located on the campus. A simple SQL query is used for plotting the most frequently used first and last recorded APs in a stay (Figure 11.5)

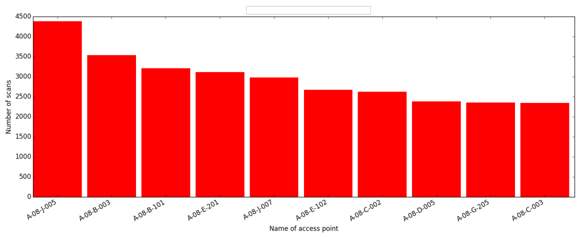


Figure 11.4: Most frequently recorded APs in a movement to the Faculty of Architecture and the built environment

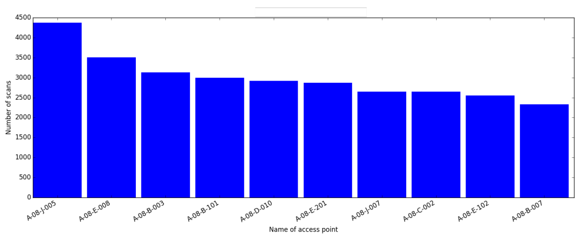


Figure 11.5: Most frequently recorded APs in a movement from the Faculty of Architecture and the built environment

The most frequently used access point, A-08-J-005, is located high up in the modelling hall and thus not near an entrance (see Figure 11.6). Although this location is different than expected there might be a reason for it. The access point is placed in an open space in which no objects could seriously block the Wi-Fi signal.

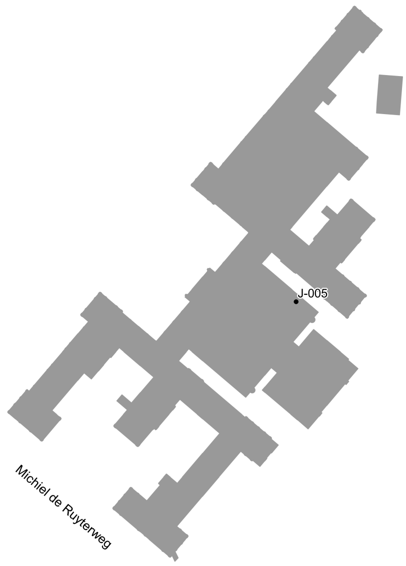


Figure 11.6: The location of the most frequently used APs that are used to record the first and/or last connection of a device in the Faculty of Architecture and the built environment

In order to know with what APs a device connects when entering a building, some experiments are conducted. By looking at the MAC address of the access point the device is connects with, it would be possible to identify the location of that AP. This experiment is conducted for entering the Faculty of Architecture and the built environment via the East, West and main entrance. Table 11.2 shows the results of the experiment.

entrance	**MAC address**	**apname**	**maploc**
east entrance	00-15-C7-80-9A-60	not found	not found
west entrance	00-22-90-5E-66-F0	A-09-E-102	1st floor West MSc
main entrance	00-22-90-38-7F-D0	not found	not found

Table 11.2: The AP a device connects with when entering the Faculty of Architecture and the built environment

The AP a device connects with when entering the building via the West entrance, E-102', can also be found in Figure 11.5. Though it does not stand out compared to other APs. The MAC addresses of the APs the device connects with when entering the building via the East or Main entrance are not found, meaning the APs are not located on the map or listed in the table of APs. This implies it is not possible to relate the results to the data.

Recommendation

The results and conducted experiments has shown it is not possible to clearly find what, when and how frequent entrances of the Faculty of Architecture and the built environment are used. The first and most important reason for that, is the time interval of approximately 5 minutes in which the system is recording. A person could be anywhere in the building at the moment of recording. A smaller time interval between the moments of recording would help in finding answers to the questions regarding the use of the entrances. Also, the existing layout of APs in the Faculty of Architecture and the built environment is currently not designed for any other purpose than allowing a Wi-Fi compliant device to connect with the wireless eduroam network. Locating APs near the entrances of a building might help. Moreover, the fact the Faculty of Architecture and the built environment has multiple entrances, in combination with the large time interval of recording, is what makes identification of the entrances difficult.

11.2. Association rules

The following section describes how movement patterns can be derived on building level, without considering the direction or order of the movement. An association rule mining algorithm (Agrawal, Imieliński, and Swami 1993) was used to identify groups of buildings that frequently visited in combination with each other. Firstly the algorithm is described briefly, then the results are presented and a recommendation is given.

Association rules mining

Association rule mining is a technique to analyse what variables or items are commonly associated with each other in large databases. Probably the one of the main application is to analyse which items are commonly bought together by customers of a supermarket. As an example for this use case is an association rule of an item set {bread, butter}, tells that in 80% of those transactions including {bread, butter}, also {milk} was present. In other words, 80% of the people who buy bread and butter also buy milk (Agrawal, Imieliński, and Swami 1993). Compared to sequence mining, association rule mining does not consider the order of items neither within, nor across transactions.

Thus every rule is composed by two item sets, the *antecedent* {bread,butter} on the left-hand side, and the *consequent* {milk} on the right-hand side. The rule is denoted as {bread, butter} => {milk}.

Association rules of buildings

When a trajectory is simplified into a set of distinct buildings that the person visited, association rules for buildings can be derived. In this case the rule describes the set of buildings, or building set, that are commonly visited in combination. For example the rule {BK_City, Aula} => {Library} tells that a group of people who visited the buildings BK_City and Aula also visited the Library.

As association rule mining does not consider the order of buildings, nor the time spent in a building, it is important that these variables are appropriately handled and noise is filtered out prior running the algorithm.

In the first version the building sets were stored in a table as below, where the field *mac* contains the mac-address of a device and each remaining field represents a building. Value 1 is given if the device was recorded in a building, otherwise no value is given. This binary encoding is rather simplistic as it does not consider the amount of time spent in a building and therefore it does not allow to differentiate between occasional or regular visits.

mac	aula	bk_city	bouwcampus	btud	ctig	...
A	1	1			1	
B			1	1		
C		1			1	
D	1					
E	1		1			

Table 11.3: uncategorized buildingset table

Therefore in the second version a distinction between *occasional, regular* and *frequent* stays was added to the building sets. The division between the categories is based on the 40 hour work week and 1.5 hour lecture durations (see Table 11.4).

Category	hours/week	ID
occasional	≤ 0.5	1
regular	$> 0.5, \leq 5$	2
frequent	> 5	3

Table 11.4: Stay duration categories

The trajectories of approximately 14,000 devices were used to create the first set of association rules with categorized stay duration. At this stage only the noise was filtered from the data but not the stationary devices, and people carrying two devices were not accounted for. The time range of trajectories spanned from 31.03.2016 to 02.05.2016, approximately one month.

Although there are several measures to evaluate the interestingness of an association rule (Zhang et al. 2009), only *support* and *confidence* were used for testing purposes.

Support
“The support for a rule is defined to be the fraction of transaction in the dataset that satisfy the union of items in the consequent and antecedent of the rule.” (Agrawal, Imieliński, and Swami 1993). In case of the rule {BK_City, Aula} => {Library}, the support is the percentage of the total dataset that includes BK_City, Aula and Library.

Confidence
Confidence measures the strength of the rule, and is considered as a conditional probability. In case of the rule {BK_ City, Aula} => {Library}, the confidence is the probability that Library is in the trajectory if both BK_ City and Aula are in the trajectory (Agrawal, Imieliński, and Swami 1993; Anbukkarasy and Sairam 2013).

The most interesting rules are displayed in Figure 11.7:

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.02	0.86	0.02	6.92	6.51	0.01	drebbelweg=2, ewi_lb=2	→	ewi_hb=2
0.01	0.74	0.01	24.80	3.38	0.00	btud=2, drebbelweg=2, tbm=2	→	citg=2
0.01	0.70	0.01	30.21	3.21	0.00	aula=2, lr=2, ocp_io=2	→	citg=2
0.01	0.70	0.01	29.22	2.44	0.00	aula=2, lr=2, ogz=2	→	btud=2
0.01	0.72	0.02	6.60	5.49	0.01	btud=2, ewi_lb=2	→	ewi_hb=2
0.01	0.73	0.01	15.17	5.52	0.01	aula=2, btud=2, ewi_lb=2	→	ewi_hb=2
0.01	0.72	0.01	16.24	5.44	0.00	btud=2, citg=2, ewi_lb=2	→	ewi_hb=2
0.01	0.90	0.01	17.31	6.81	0.01	btud=2, drebbelweg=2, ewi_lb=2	→	ewi_hb=2
0.01	0.85	0.01	20.16	6.43	0.00	citg=2, drebbelweg=2, ewi_lb=2	→	ewi_hb=2
0.01	0.74	0.01	10.08	5.59	0.01	ewi_lb=2, tnw_tn=2	→	ewi_hb=2
0.01	0.86	0.01	21.85	6.51	0.00	drebbelweg=2, ewi_lb=2, tnw_tn=2	→	ewi_hb=2
0.01	0.72	0.01	25.66	2.52	0.00	aula=1, tbm=2	→	btud=2

Figure 11.7: Building set

In the building set of approx. 14,000 devices 2% was recorded in all of the buildings *Drebbelweg, EWI-LB, EWI-HB* (Support = 0.02). There is an 86% chance that if a device is recorded in the buildings *Drebbelweg, EWI-LB*, then it is also recorded in *EWI-HB* (Confidence = 0.86). And they spent on average between half hour to five hours a week in each building (drebbelweg=2, ewi_ lb=2, ewi_ hb=2).

Recommendation
Association rule mining is a suitable technique to analyse the occupancy of a group of buildings, but it is less suitable for analysing movement patterns. Therefore it is not handled more intensively in this project. However, this technique can potentially answer questions such as,

- *Which are the most visited buildings?*
- *Which buildings are islands?*
- *If a group of people visit building A, how likely that they will also visit building B?*
- *All the people who visit building A, what other buildings do they visit as well?*

11.3. Distinguishing user groups

Individual trajectories contain detailed information about the movement patterns of people. As is discussed in chapter 8 can trajectories from Wi-Fi scans be used to identify co-location in space. However, this pattern mining approach only considers location and the order of locations. When also time is considered and stored for each state in the trajectory, new patterns can be identified. When multiple trajectories share more than one location at the same time and order, moving groups can be identified. Detecting co-location in space and time is not considered for this research, but will be an interesting topic for further analysis.

11.4. Occupancy

As discussed in section 6.2, all devices can be classified as either static or dynamic device, such as laptops and smart phones. For this project, mostly the dynamic devices are used to find movement patterns, because these devices are most probably carried around the campus with the user and thus gives the best representation of the actual movement. But this assumption leaves out all the devices that can provide other valuable information. The information that a static device is carrying can help finding patterns in occupation of rooms. For example, when an individual in the Library is leaving his or her laptop at a work space, but he is going to the Aula for lunch, the work space is still occupied. This information can be of great value when assessing the use of the Library.

For future research, it would be useful to consider both static and dynamic devices and depending on the question that is asked use either one of them. For occupational research, static devices would be used, for identifying movement, dynamic devices would be used.

11.5. AP system

The set-up of the system that logs the devices connected to access points is directly connected to the accuracy of the processed data. Currently, the APs register every device that is connected to it and the logging system receives all connected devices approximately every five minutes. Additionally, all access points are located indoors, logging every device carried by people using that building. These two aspects of the AP system limit the accuracy of the processed data and thus the movement patterns that can be derived.

Because the system logs every connected devices once every five minutes, a device will only be registered if the device is connected at the time of logging. This will result in discrepancies in the processed data. Devices and thus people walking by an AP will probably not be registered, for they are most likely not connected to that AP at the time of logging. This is unfortunate, because a person can travel a rather long distance in five minutes, e.g. making it hard to track people indoors. If the system would be logging every device all the time, irrespective of the time the device is connected, the tracking data would contain every AP that a device would connect to and thus provide much more accurate tracking data. Understandably, logging every user every second would result in huge amounts of data, which would most definitely result in performance issues.

Secondly, because all scanners are located inside buildings, there is little to no information on people when they move from one building to another. Surely something can be told from the time it takes a device from the last scan in one building, to the first scan in the second building. But for outdoor tracking purposes, this system is limited. From some experiments that were conducted on the TU Delft Campus it can be concluded that a device located outdoors near a building can be detected by APs inside the building, but this depends on the antenna in the devices and the exact location of the device in respect to the AP. If more detailed information about movement outdoors is desired, it would be wise to also include outdoor APs in the system. If adding outdoor APs would become too expensive, another improvement can be made by publishing a map with the locations of every AP. That way, at least the APs near the outer walls can be distinguished from the APs in the center of the building.

To improve further research, it is recommended to take the system of APs into account before actually conducting the research. If outdoor movement tracking is desired, outdoor APs are required. And if tracking indoors is one of the goals, the frequency of logging should be set to an interval that is in the order of magnitude of 10 seconds to one minute, taking data size in consideration.

11.6. Data reasoning

During this project a lot of data is handled. With all the data available and the processing to derive movement patterns one could ask: 'How reliable is the data?' and 'How accurately can we derive these movement patterns?'. Determining the working of the system of APs as described in subsection 4.6.2, was a great step towards a reliable outcome. Knowing how the system works helped improve the processing steps that were taken, because the systems flaws could be taken into account and avoided. Additionally, when the first movement patterns were derived, common knowledge and knowledge about the TU Delft campus and its layout helped in validating these patterns.

Because the working of the system of APs is known, the dataset can be improved by filtering out people that are only registered for less than five minutes at one AP, indicating that they only were only passing by. This means that the states derived from the data are actually stay places of an individual. Another perspective could be that exactly those people that are only passing by are valuable for the dataset. When an individual is registered at four consecutive APs and each scan was less than five minutes, it can be concluded that this person is moving between those four APs. However, the current set-up of the AP system is not suitable enough to use only devices that have a session duration of less than five minutes. This would only work when the frequency of logging is increased.

Moreover, the knowledge acquired from previous courses in the Geomatics programme and common knowledge about buildings and the TU Delft campus can be used to validate certain outcomes of the data processing. For example, it would seems very illogical that an individual could travel from Architecture to Aerospace Engineering and then to Industrial Design in five minutes. Such requirements could improve the final outcomes. This kind of reasoning became even more useful when zooming in to spatial level building-part. Using the knowledge of the building layout of Architecture, it could be concluded that moving from one floor to another is impossible without using one of the staircases. Such a conclusion could then be included in the processing, e.g. validating only movement between floors if one of the staircases is used.

For future research, it is desirable to use a higher frequency for logging the connected devices. This will ensure that a device is always registered and that its movement can be easily identified.

11.7. Visual exploration

In the course of this project, a lot of visualizations have been used to represent movement between buildings. In any data visualization question, it is important to consider the right type of visualization. The types of visualization that have been used in this project include ***1)*** Static maps; ***2)*** dynamic maps and ***3)*** interactive maps.

In static visualization, all movement between buildings are aggregated. The map shows the amount of movements between buildings. All buildings are represented as points and movements as straight lines between these points. Line width is used to represent the amount of movements. The thicker the line is, the more movements there are. However, in this visualization, all lines have same color and since the basemap is openstreetmap, these lines are not highlighted. All the possible movements between these buildings are shown on the map, which makes the map chaotic. The users cannot comprehend what the map wants to emphasize. Besides, the static map disregards the temporal component which is also important for analyzing movement patterns. Considering movement is dynamic, so a dynamic visualization is adopted in the next stage.

Instead of aggregating movements, dynamic visualization focuses on individual movement that each line represents one movement (one movement record in database). Each line has two timestamps, the start time and the end time of the movement. All the lines overlap and the lines will be darker if there are more movements. The line will appear and disappear according to the start time and end time, so the dynamic visualization shows how movements change in time during one day. However the animation runs fast and with many lines popping up at the same time, the users can only get an overview when the campus is the busiest, but dynamic visualization is not suitable to find movement patterns, because it doesn't provide detailed information.

In order to find movement patterns. An automatic visualization including GUI is developed. Users can choose dates and buildings, a graph and a map will be generated automatically. The graph and the map together combine static visualization and automatic visualization. The graph shows the change of movements in time and the map shows the aggregated movements so that the users are able to know at which time there are the most movements with the graph and between which buildings are there the most movements according to the map. The line width still represents the amount of movements and the color of the lines is gradient from red to green. Green line means the movement is symmetric that there are similar amount of movements for both directions and red line means the movement is not symmetric. The basemap of automatic visualization is openstreetmap and the color of the lines doesn't make a strong contrast, which makes the map not readable enough.

Acknowledgements

We would like to take the opportunity to express our gratitude and regards to everyone that contributed to this project.

First, we would like to thank Edward Verbree, our supervisor, for the feedback provided during the course of the project. Additionally, we would like to thank Wilko Quack for his patience and support with all the issues we encountered when using the database.

We also would like to thank Bart Valks and Iljoesja Berdowski from the department of Facility Management and Real Estate for their feedback and guidance throughout the project and the challenges they proposed.

Additionally we would like to thank Jorge Gill for his feedback on our visualizations and his help for creating even better visualizations.

Appendix A

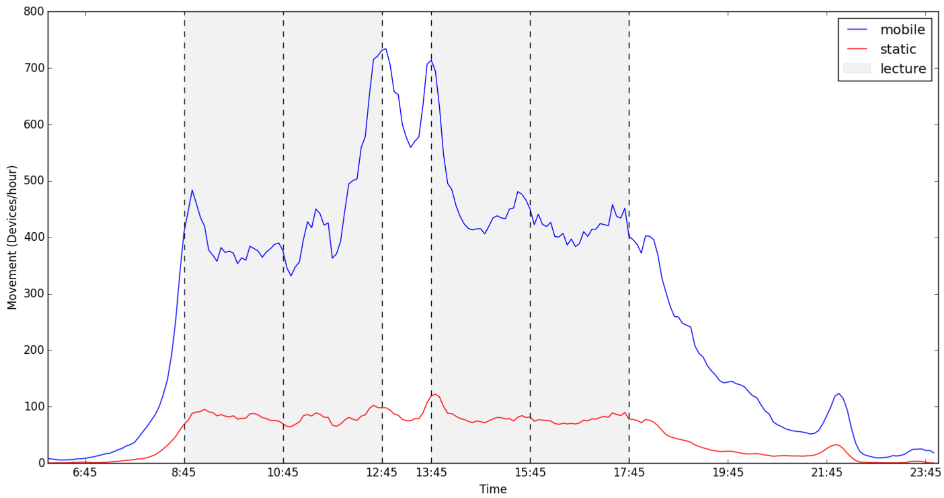


Figure 8: Movement on during weekdays and weekends at the BK

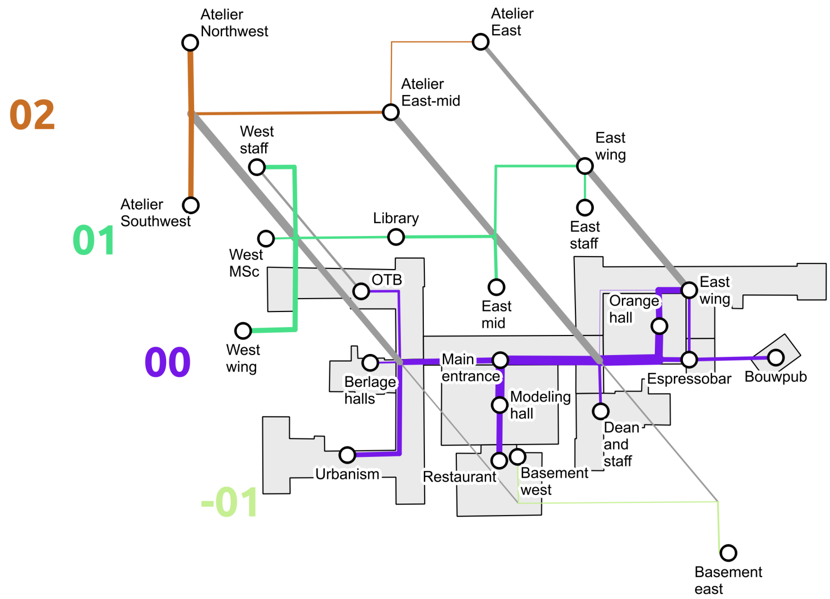


Figure 9: Movement on during weekdays at the BK

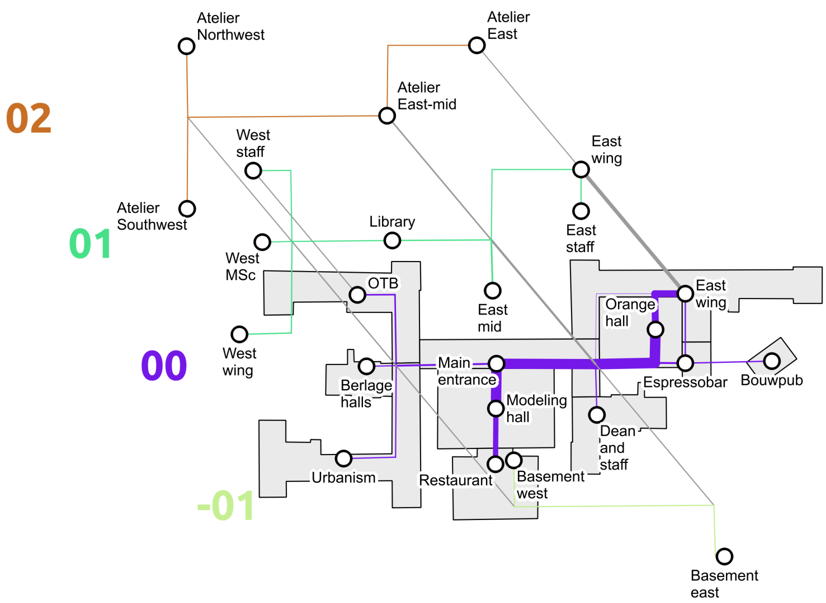


Figure 10: Movement on during weekends at the BK

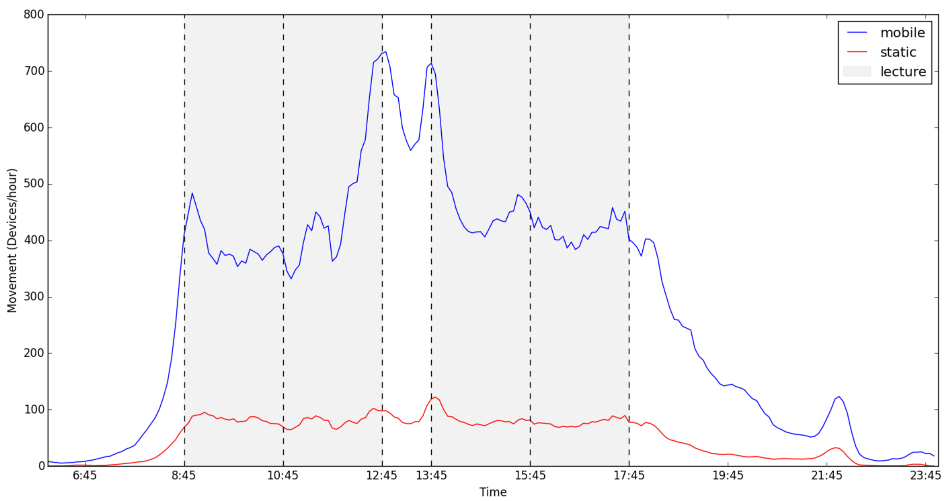


Figure 11: Movement of mobile and static devices at the BK

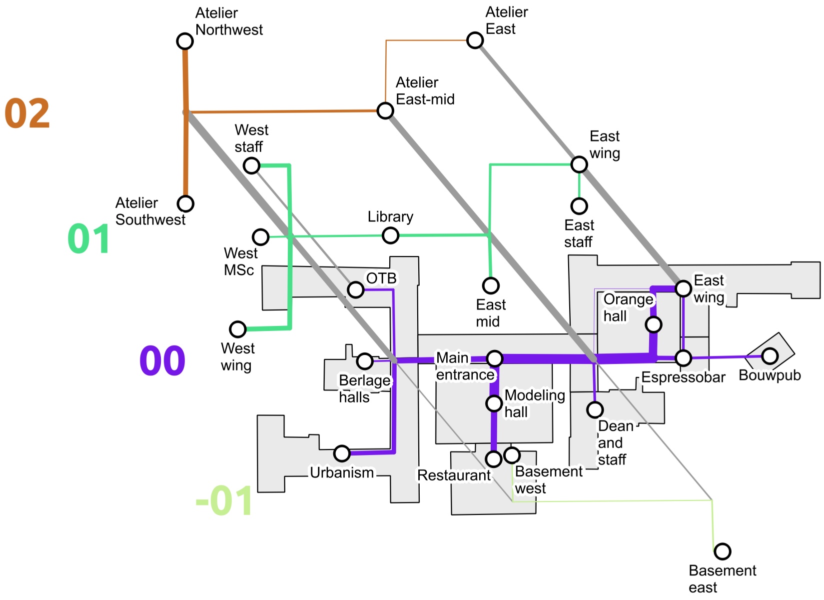


Figure 12: Movement of mobile devices at the BK

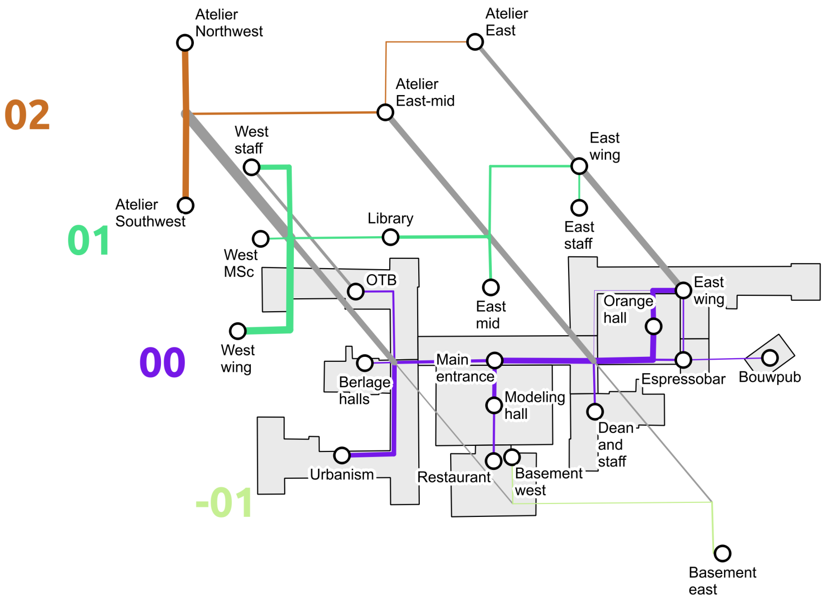


Figure 13: Movement of static devices at the BK

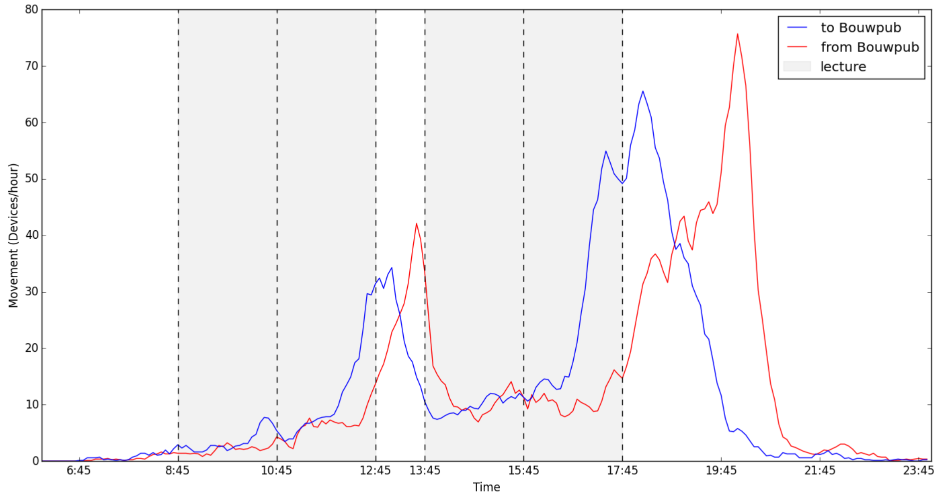


Figure 14: Movement to and from the Bouwpub at the BK

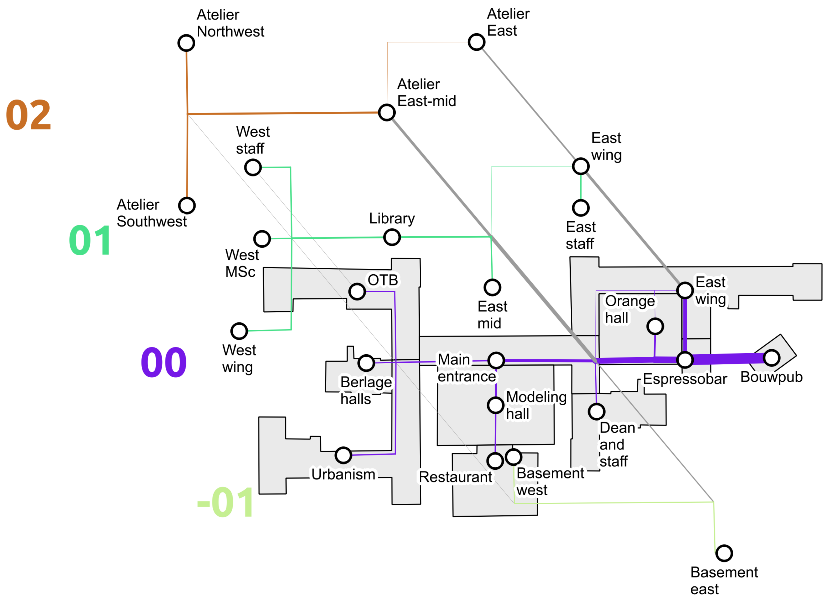


Figure 15: Movement to the Bouwpub at the BK

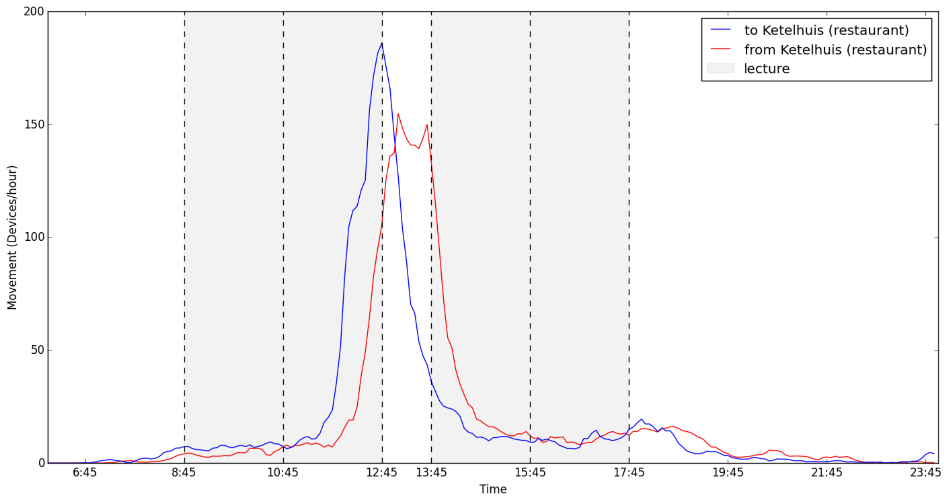


Figure 16: Movement to and from the Restaurant at the BK

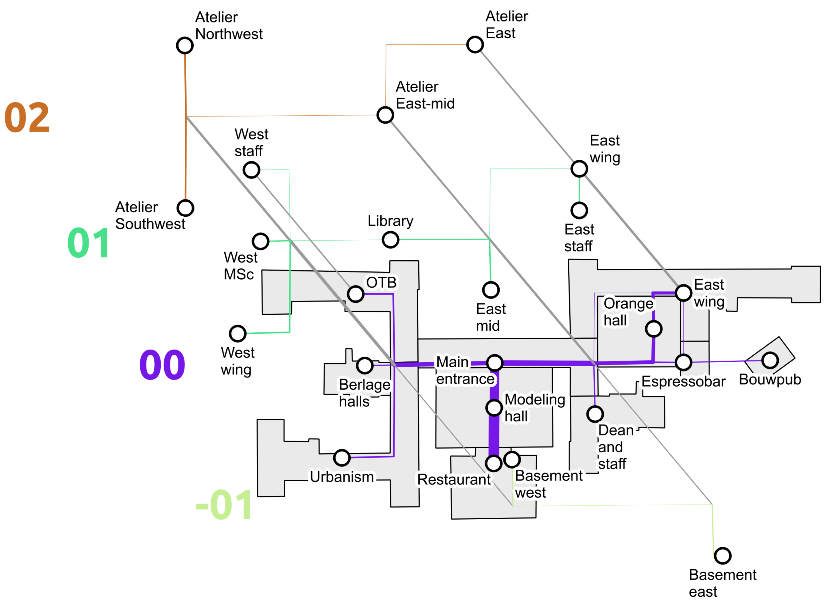


Figure 17: Movement to the Restaurant at the BK

Bibliography

Agrawal, Rakesh, Tomasz Imieliński, and Arun Swami (1993). "Mining association rules between sets of items in large databases". In: *ACM SIGMOD Record*. Vol. 22. ACM, pp. 207–216. (Visited on 04/09/2015).

Agrawal, Rakesh and Ramakrishnan Srikant (1994). "Fast algorithms for mining association rules". In: *Proc. 20th int. conf. very large data bases, VLDB*. Vol. 1215, pp. 487–499. URL: `https://www.it.uu.se/edu/course/homepage/infoutv/ht08/vldb94_rj.pdf` (visited on 04/09/2015).

Anbukkarasy, G. and N. Sairam (2013). "Interesting Metrics Based Adaptive Prediction Technique for Knowledge Discovery". In: *International Journal of Engineering and Technology* 5.3, pp. 2069–2076. (Visited on 05/16/2016).

Dodge, Somayeh, Robert Weibel, and Anna-Katharina Lautenschütz (2008). "Towards a taxonomy of movement patterns". In: *Information visualization* 7.3-4, pp. 240–252.

Heijer den, Alexandra (2012). "Managing the University Campus: Exploring models for the future and supporting today's decisions". In:

Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment". In: *Computing In Science & Engineering* 9.3, pp. 90–95.

Kalogianni, E et al. (2015). "Passive WiFi Monitoring of the Rhythm of the Campus". In: *Proceedings of The 18th AGILE International Conference on Geographic Information Science*. AGILE, pp. 1–4.

Mautz, Rainer (2012). "Indoor positioning technologies". PhD thesis. Habilitationsschrift ETH Zürich, 2012.

Meneses, Filipe and Alberto Moreira (2012). "Large scale movement analysis from WiFi based location data". In: *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*. IEEE, pp. 1–9.

Parliament and Council of European Union (1995). *Directive (EC) 95/46/EC*. `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML`.

Pei, Jian et al. (2004). "Mining sequential patterns by pattern-growth: The prefixspan approach". In: *Knowledge and Data Engineering, IEEE Transactions on* 16.11, pp. 1424–1440.

Radaelli, Laura et al. (2013). "Identifying Typical Movements among Indoor Objects–Concepts and Empirical Study". In: *Mobile Data Management (MDM), 2013 IEEE 14th International Conference on*. Vol. 1. IEEE, pp. 197–206.

Spek, SC van der (2008). "Mapping Pedestrian Movement: Using Tracking Technologies in Koblenz". In: *Lecture Notes in Geoinformation and Cartography*, pp. 95–118.

Zhang, Yuejin et al. (2009). "A Survey of Interestingness Measures for Association Rules". In: *International Conference on Business Intelligence and Financial Engineering, 2009. BIFE '09*. International Conference on Business Intelligence and Financial Engineering, 2009. BIFE '09. IEEE, pp. 460–463. DOI: `10.1109/BIFE.2009.110`.

Zhao, Shao et al. (2014). "Discovering People's Life Patterns from Anonymized WiFi Scanlists". In: *Ubiquitous Intelligence and Computing, 2014 IEEE 11th Intl Conf on and IEEE 11th Intl Conf on and Autonomic and Trusted Computing, and IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UTC-ATC-ScalCom)*. IEEE, pp. 276–283.