MSc Thesis
Geomatics for the Built Environment

The identification of road modality and occupancy patterns by Wi-Fi monitoring sensors as a way to support the "Smart Cities" concept.

Application at the city centre of Dordrecht

Dimitrios Kyritsis

TU Delft Delft University of Technology

Challenge the future

The identification of road modality and occupancy patterns by Wi-Fi monitoring sensors as a way to support the "Smart Cities" concept.

Application at the city centre of Dordrecht

by

Dimitrios Kyritsis

Student number: 4412176
E-mail: dimkyritsis@gmail.com

in partial fulfillment of the requirements for the degree of

Master of Science in Geomatics

at the Delft University of Technology,
to be defended publicly on Thursday April 20, 2017 at 09:15 AM.

Main mentor: ir. Edward Verbree

Second mentor: dr. ing. Sisi Zlatanova

Co-reader: Robert Voûte

Delegate of the Board of Examiners: Tuuli Jylhä

TU Delft Delft University of Technology

On cover:

Drierivierenpunt Dordrecht

Photo by: “Watericonen - Biesbosch Dordrecht Kinderdijk”
http://www.watericonen.nl/de-iconen/dordrecht/

Ἐπεὶ δὲ μετρητική, ἀνάγκῃ δήπου τέχνη καὶ ἐπιστήμη.
Συμφήσουσιν.

Πλάτων (380 π.Χ.)

Being measurement, it necessarily must be an art and a science.
They will assent to this.

Plato (380 B.C.)

The text of Plato (427 – 347 B.C.) philosopher, as well as mathematician, in Classical Greece, is written in his work entitled “Protagoras” (357b).

The image of Plato originated from “The School of Athens”, one of the most famous frescoes by the Italian Renaissance artist Raphael, painted between 1509 -1510, in the Vatican.

Acknowledgements

In this section I would like to take the opportunity to thank a few people that had an impact on conducting this research. First of all, I would like to thank Edward Verbree for being an inspiring supervisor and for his great support and guidance that helped me overcome the many problems I encountered in the course of time. I am also sincerely grateful to Sisi Zlatanova for the helpful discussions, feedback and advice that she was always willing to give me throughout this period. Furthermore, I would like to thank my co-reader, Robert Voûte, and the delegate of the board of examiners, Tuuli Jylhä, for their critical feedback which was very important in the process of writing this thesis.

Moreover, a big thank to Professor Sokratis Basbas from the Aristotle University of Thessaloniki for the invaluable conversations, advice and suggestions he shared with me based on his very extensive experience in the field of transportation.

Also, I would like to express my gratitude to all the employees of the Municipality of Dordrecht and Libelium for their continuous support during this research and for providing all the technical equipment necessary for the data collection part. Additionally, I would like to thank the bureau RMC for the provision of data from its pedestrian counting camera in the city centre of Dordrecht.

Last but not least, as I complete this stage of my studies, it would be an oversight on my part not to thank all my distinguished professors at TU Delft who helped me strengthen my knowledge and broaden my horizons in this field of study.

Abstract

As world urbanization continues apace and total population increases, there is an immediate demand for better monitoring and exploitation of space. In view of the above, the “Smart Cities” concept has been developed and numerous efforts are made to deploy technology to this end. The main information needed for city development and planning is road modality and the relevant occupancy patterns. However, it is quite difficult to collect this information. There have been several different approaches towards providing this information and various methods have been used. However, each of them has weaknesses which do not allow it to be used on its own. On the other hand, thanks to the new technological developments and due to the growing needs of society over the last years, the system of Wi-Fi monitoring sensors has been increasingly used in outdoor environments. Many companies have already used this method to collect data and provide information about users’ behavior in places such as public areas, shopping centers and malls. Nonetheless, the contribution of this thesis is the study of the applicability of this method, the assessment of the reliability of its outcomes and the identification of crucial parameters which significantly affect the final accuracy.

Thus, the aim of this research is to investigate what kind of road modality and occupancy patterns can be recognized using Wi-Fi monitoring sensors in a city area as well as which setup parameters can influence the final outcome. The system is implemented in the city of Dordrecht, which constitutes the research area of this study. First of all, the design of the observation network is described and the relevant parameters are taken into account. Using the data collected by the system and the known distances between the sensors, the movement speed of each device is computed. Street-uses criteria of the research area are also used as input to the system, and in combination with the computed speed three categories of users are recognized and each device is categorized as “pedestrian”, “bicyclist”, or “vehicle”. Under this classification each street’s road modality is studied. The relationship between the categories throughout the day is investigated and preferred streets for each kind of users are recognized.

Based on the ability of the system to identify every device in the research area throughout the day, the movement behaviors of users are researched and similarities between them as well and the most frequent patterns are identified. Three sets of movement patterns are studied considering the number of sensors which scan the same device within a time period. Each set is investigated separately for every kind of users. Moreover, using the number of devices scanned at each sensor point, occupancy patterns are identified both for users as a whole and for each user category separately. It is argued that this constitutes an important advantage of the system. Rush hours, recession periods and movement trends are recognized for the different days of the week as well as the occupancy relationship between the research area and its surroundings. Finally, a questionnaire and random samplings with Bernoulli trial are used to validate the outcomes. A quite strong correlation between the system's results and reality is revealed, especially with regard to pedestrians and bicyclists. However, despite the quite promising findings, further implementation and testing of the system in different environments is needed in order to draw an indisputable conclusion about its effectiveness.

Key Words: Passive Wi-Fi monitoring, road modality, movement patterns, occupancy, Smart Cities.

Table of Contents

Acronyms

AP	Access Point
dB	decibel
dBm	decibels of the measured power referenced to one milliwatt
DDPA	Dutch Data Protection Authority
GPS	Global Positioning System
MAC	Media Access Control
mW	milliwatt
NP	Nondeterministic Polynomial
PNO	Preferred Network Offload
RMC	Retail Management Center
RSSI	Received Signal Strength Indication
SSID	Service Set Identifier
WLAN	Wireless Local Area Network

1. Introduction

1.1. Introduction - Motivation

Today, half of the world's population and 80% of European citizens live in cities, world urbanization continues apace and the total population is expected to double by 2050 (*Department of Economic and Social Affairs, 2015*). Furthermore, this increase has inevitably led to a simultaneous increase in the number of vehicles. The combination of such growth with the fact that the size of the Earth remains the same, results in a significant problem for humanity which must be dealt with (*Robinson, 2016*).

As an effort to solve this problem, there is the immediate need for better monitoring and exploitation of space, using intelligent and sustainable environments which will offer citizens a high quality of life. This need has led to the creation of the new term, coupled with a worldwide trend toward "Smart Cities" (*European Commission, 2013*). The significance of this effort can also be demonstrated by the fact that the European Union has already devised a strategy for supporting the concept of Smart Cities. It has developed a range of programs and is also subsidizing related research to 'smarten up' Europe's urban areas. It is estimated that the global market for smart urban services will be $400 billion per year by 2020 (*Department of Business for Innovation & Skills, 2013*).

Two main concerns underscore the turn toward smart cities: the search for sustainability as a way to support a more inclusive, diverse and sustainable environment, and create green cities with less energy consumption (*Deakin, 2014*) and the increasing use of new internet technologies such as mobile phones, smart devices, sensors and the Internet of Things (*Deakin, 2014*). The combination of the latter concern with theories and methodologies from other fields, such as knowledge and innovation management, means it is now possible to overturn the established methods of urban development and planning (*Deakin, 2014*). Three terms can be used to describe the main steps in the evolution of urban planning as part of the "Smart Cities" concept: 1) *interconnection*, as the ability to take advantage of technology and use of the internet in order to enable different parts of a system to be joined and communicate to each other *(Miller, 2015)*; 2) *instrumentation*, as the appropriate use of this system in a city to get data round the clock as key performance indicators; and 3) *intelligence*, as the ability to use the information gathered to develop behavior patterns and predictive models for urban flows (*Deakin, 2014*). Thus, existing new technologies in conjunction with the use of the internet can act as a way of collecting useful information for the urban planning procedure. This system can be used both before urban planning, as a pre-processing data provider, and after urban planning, as a means to evaluate the effect of changes (post-processing tool).

The information required for urban planning and development it consists of road modality data and the relevant occupancy patterns. Road modality can be defined as the combined and actual use of a road by people who use different kinds of means of transport. However, various distinctions between user categories can be made depending on the goal and the country of the research (*Hallenbeck et al., 1997*), (*VCA, 2017*), (*European Commission, 2017*). Thus, an initial identification of road modality can be the distinction between motor and non-motor means of transport. Nevertheless, these two categories can be further divided into subcategories. Hence, the non-motor category can be divided into pedestrians and bicyclists, while for the category of motor vehicles a distinction between two-wheeled and four-wheeled vehicles can be made. Motor vehicles can be further distinguished into motorcycles, passenger vehicles, and Heavy Goods Vehicles, such as buses, lorries and trailers (*European Commission, 2017*). Furthermore, occupancy patterns constitute the second piece of information required for urban planning and development. They represent the way all users, or each user category separately, occupy a region during the day, revealing in this way the existence of rush hours and recession periods as well as similarities in the way the region is used during the week. The above-mentioned knowledge is undoubtedly an essential starting point for urban design. However, apart from its use in urban planning, this information can be efficiently used for many different purposes, such as monitoring traffic networks, crowd control, facility usage as well as marketing purposes.

Despite the undeniably high importance of this information, the relevant collection procedure is the most difficult part of the effort. In previous years, random counting of vehicles and pedestrians by people located in specific parts of the research area was the most frequently used method of data collection. It is a method which is very accurate and the results represent reality but there are many disadvantages. Firstly, it is time-consuming and many employees are required to collect data just for a small area, as each of them can only count one category of road users. Furthermore, as this method is based on manpower, it is clear that it cannot be applied for long periods of time to a large area under, perhaps, difficult weather conditions. Finally, despite the accuracy of this method, it is not possible to have real-time results as well as monitoring which is required for the identification of occupancy patterns (*U.S. Department of Transportation, 2016*).

As a result of all the disadvantages of the classical method of counting, there has been a demand for alternative data collection methods, and the evolution of technology is a key factor for investigations of this kind. There are many different kinds of methods which can be used in including inductive loops, pressure sensors, infrared cameras, and video detection systems. However, each method has weaknesses which do not allow it to be used on its own. For instance, the inductive loops system cannot count pedestrians, especially when they are moving in groups; pressure sensors are quite expensive to install; infrared camera systems cannot distinguish between bicyclists and pedestrians; while algorithms for video detection systems are still being developed. Moreover, none of them offers the possibility to "follow" users, under the relevant privacy directions, in order to identify occupancy patterns, (*U.S. Department of Transportation, 2016*).

An alternative method, usage of which has been constantly rising over recent years, is Wi-Fi monitoring. Nowadays, most people carry one or more mobile devices around with them that have Wi-Fi and Bluetooth functionality. Furthermore, all contemporary vehicles are manufactured with the ability to link to a smartphone or already contain a kind of smartphone as a service. Thus, when users switch on the relevant Wi-Fi or Bluetooth functionality, the embedded devices start to constantly send out a signal in their search for a Wi-Fi access point. This signal contains an ID number; a MAC address, which is unique to each device. Taking advantage of the available technology, many companies have produced sensors which can detect and count these signal transmissions within a relative range as well as some more details, such as the received signal strength indication (RSSI) and the vendor of the device. Given that, it is possible to use this method as a way to count these devices but also to investigate their movement in space by installing a network of sensors in the research area (*Musa et al., 2012*).

An important advantage over other techniques is that Wi-Fi monitoring system does not require the active participation of users and it works without any modification, as it is not necessary to install any application on the relevant device. Furthermore, unlike other methods, such as GPS, it is not necessary to request the provider to give access to the relevant data since, thanks to the use of sensors, the owner can have real-time access to the information collected. Finally, apart from their monitoring ability, Wi-Fi sensors can also be used as a way to offer a (free) Wi-Fi network. In this way, the activation of Wi-Fi functionality by more users could be promoted, leading to a higher amount -and more representative- data.

Taking into account the above-mentioned advantages of the Wi-Fi system, it is clear that it can be used to develop real-time monitoring, if required, and a data collection system, which can be used as a way to support the main philosophy of the 'Smart cities' concept.

For this thesis, a part of the city of Dordrecht has been used as the research area, where the relevant methodology, tests, and results are applied and used to support both the scientific purposes of this research and "the Smart City of Dordrecht" concept. Chapter 1.3 describes the characteristics of the area and the reasons for selecting it in-depth.

1.2 Research objectives

1.2.1 Objectives

The main research question for this thesis is:

What kind of road modality and occupancy patterns can be recognized by Wi-Fi monitoring sensors in the city of Dordrecht in order to support the "Smart City" concept?

To be able to answer this main question, secondary questions have been formulated:

- What influence does the Wi-Fi monitoring setup have?
- What are the Wi-Fi monitoring performance parameters and how can we measure them?

- What kind of movement patterns can be recognized by the Wi-Fi monitoring system?

- What is the road modality in the research area of Dordrecht during different times of the day and month?
- What kind of road modality can be recognized by the Wi-Fi monitoring system?

- What is the occupancy pattern in the research area of Dordrecht during different times of the day and month?
- Which occupancy patterns can be recognized by the Wi-Fi monitoring system?
- Is it possible to identify the effect of the weather on road modality?

As is clear, the abovementioned sub-questions can be grouped into four categories based on their focus. Thus, the first one is related to the technical parameters of the method used, the next category focuses on the movement patterns while the last two sets of sub-questions are associated with identifying road modality and occupancy patterns respectively.

1.2.2 Scope of research

This thesis focuses on the use of Wi-Fi monitoring sensors data from the research area in order identify the relevant road modality and occupancy patterns. Identification of the most appropriate Wi-Fi network configuration is not investigated in this study. However, the influence of some parameters on the final result, such as the total number of sensors, is studied.

1.3 Research Area

As mentioned in the previous chapter, this study was carried out in part of the city of Dordrecht, and more specifically in the area between the city center and the central railway station. There are many reasons to justify the choice of this area. First of all, Dordrecht is a city with a great deal of interest in and enthusiasm for the "Smart Cities" idea. Many research projects have already begun on the development of the city and the exploitation of technology, like the "Smart City Dordrecht" project.

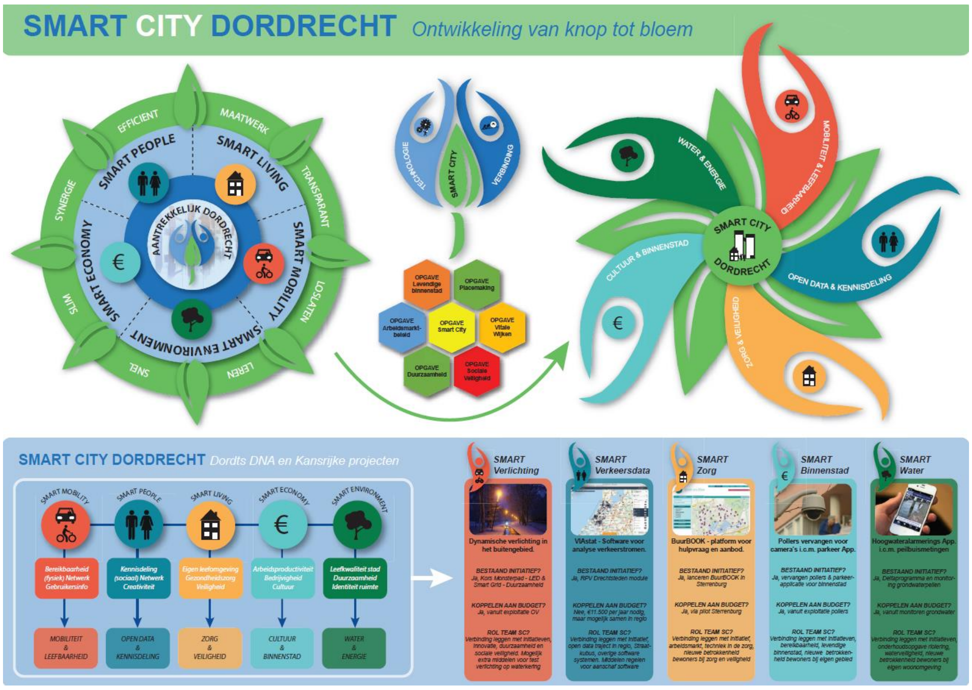


Figure 1: Poster about the Smart City project from the Municipality of Dordrecht

The selected part of the city constitutes the interface between the city center, where the majority of shops and offices are situated, and the central station, which comprises the basic means of transport to and from neighbouring cities like Rotterdam. This close proximity to Rotterdam is also one of the main reasons that the city is primarily a residential area. Thus, significant changes in the movement flows and occupancy patterns are expected during the day when citizens go to and return from their jobs. Furthermore, due to the significance of this region, local authorities want to rebuild it in order to change the land uses and increase the level of public services. There is no preliminary information about road modality and occupancy patterns for the region and thus the outcome of this research will be very useful for urban planning in the area (as a pre-processing tool). Finally, the additional ability of Wi-Fi sensors to act as Wi-Fi routers is directly related to the willingness of the Municipality of Dordrecht to set up a free WLAN network, while a possible future re-application of this system could be used to evaluate changes, i.e. as a post-processing tool.

Figure 2: Part of the city of Dordrecht. Green line represents the boundary of the research area while the orange one the city center

1.4 Methodology

In order to achieve the ultimate target of this study, the research process was split into four main phases, as outlined below.

1. **Literature Review**: by reviewing the literature various already-used methods for similar purposes were examined in order to gain an insight into their characteristics and their drawbacks, which led to the need for alternative research techniques. Moreover, the literature review elaborated on the Wi-Fi monitoring method, in order to obtain a deep understanding of the system, the technical specifications of the devices, its benefits over other methods as well as the privacy issues which have to be considered. Finally, relevant applications of the Wi-Fi monitoring method for researching human behavior at the spatiotemporal level and the computation of movement and occupancy patterns were investigated. The knowledge acquired

from the literature review constituted a basis for understanding the basic concepts and parameters which had to be taken into account throughout this research.

2. **Design of the observation network, zero-level test, and data collection**: the whole set of rules about the observation network and its characteristics were defined. Based on the knowledge from the literature review, the decisions about the locations of sensors, the duration of the data collection period, the recording frequency, and the kind of recording antennas were taken. As a pre-collection procedure, a zero-level test, using various devices from different manufacturers, was designed in order to check and verify the appropriateness of the locations chosen, especially from the perspective that each sensor covers the initially designed area. Furthermore, other potentially useful data sources were investigated and stored, simultaneously with Wi-Fi data collection, such as weather information and data on pedestrian numbers from relevant cameras in the city centre of Dordrecht.

3. **Data preparation and analysis**: before the main part of data analysis, a data preparation procedure was required. Due to privacy issues, all the MAC addresses of recorded devices were hashed while static devices which work with the use of Wi-Fi or devices which were detected for only one sensor during a fixed time period were filtered. Taking into account the distances between sensors, the time differences between the records, and the land uses of the research area, the relevant speed of movement for each device was computed enabling us to characterise it as “pedestrian”, “bicyclist”, or “vehicle”. Moreover, having the spatiotemporal behaviour of devices, movement and occupancy patterns were identified at different timeslots and days of the research period. Finally, the influence of the weather and the total number of sensors on road modality and final outcomes respectively were investigated.

4. **Data validation**: Simultaneously with the Wi-Fi data collection procedure, a questionnaire-based survey and the sample counting method were used as a way to ensure data validation and test the accuracy of the final outcomes. Based on the questionnaire, the percentage of people who had Wi-Fi functionality enabled on their devices was calculated and in combination with the results of the counting method, the expected numbers of pedestrians and bicyclists were estimated. In this way and performing a simple random sampling with Bernoulli trial the accuracy of the Wi-Fi monitoring method could be examined to obtain a relevant confidence level. Moreover, street-use criteria, such as the existence of one-way streets and roads under construction, tests with known devices as well as the use of other data sources, like pedestrian counting cameras and inductance loops, were investigated in order to assess the system's reliability.

1.5 Structure

The remainder of the report is structured according to the discrete stages of the abovementioned methodology. The second chapter provides a thorough literature review on the methods which are used for the computation of road modality, the Wi-Fi monitoring method, its attributes and relevant applications in the identification of movement and occupancy patterns. In Chapter 3, the design of the observation network as well as the zero-level test and data collection procedure are described. Chapter 4 contains the data filtering and all the analysis steps right up to computation of the final results, which are analyzed, while later on, in Chapter 5, the validation process and its outcomes are presented. Finally, the last two sections discuss the findings with respect to the initial research questions, provide the conclusions and the contribution this work makes and recommend options which have not yet been investigated, which could be examined in future research.

2. Related Work

2.1 Existing techniques and their use for the computation of road modality

Location monitoring techniques can be classified into two categories:

- Systems which require the “active” participation of people (active systems). This means that the person is carrying an electronic device which sends information to the system.

- Systems using passive localization (passive systems). This means that the position is estimated based on the variance of a measured signal (*Deak et al., 2011*).

Over the past years, taking advantage of the evolution of technology, many “active” and “passive” techniques have been invented in order to be used for the identification of road modality and occupancy patterns. Each of them has unique benefits over the others but also significant drawbacks which make their separate application impossible (*U.S. Department of Transportation, 2016*).

According to U.S. Department of Transportation (*2016*), the most frequently used and known techniques are the following:

- *Inductance Loop*: Inductance loop detectors operate by circulating a low alternating electrical current through a formed wire coil embedded in the pavement. An electromagnetic field is created by the alternating current above the formed wire coil and a conductive object, like a car or a bike, disrupts the field by a measurable amount when passing through it. The detection accuracy of motorcycles or bicycles can be increased by changing the sensitivity of the inductance loop. However, an overcounting of cars is often observed in case of increased sensitivity. Furthermore, despite the applicability of this method on bicycles and cars, it cannot be used for the research of pedestrian movement while accuracy problems may also arise with regard to groups.

-*Magnetometer:* In this method, changes in the normal magnetic field of the Earth caused by a ferrous metal object are used as a way of magnetometers operation. This method, just as the first one, cannot be applied to pedestrian movement while there are very few commercially available magnetometers designed for bicycle detection and counting.

-*Pressure, Seismic sensor:* Pressure sensors identify force changes, such as weight, in order to record a movement, while seismic sensors detect movement using the passage of energy waves through the ground caused by feet, bicycle tires, or other non-motorized wheels. The high cost of their underground placement, the limited number of commercially available sensors as well as the low accuracy when it comes to distinguishing pedestrians from bicyclists are considered to be the main disadvantages of these methods.

-*Video imaging system:* Another technique is the use of video image processing. It uses visual pattern recognition in order to identify and count pedestrian or bicyclist movement through a specific camera range. Despite the significant improvement of the pattern recognition algorithm, problems with regard to distinguishing individuals in cases of group travelling and the influence of weather and lighting conditions on the accuracy of the outcomes have yet to be solved (*Zervos, 2013*).

Apart from the video system's automatic detection, manual detection is also possible by viewing recorded video from intersection control or surveillance cameras. This manual approach is practical and low-cost for periodic short-term counts, but is not sustainable for continuous monitoring purposes. Despite being characterized as a very accurate method, its long-term application significantly increases costs while, in cases of groups or crowds, a reduction in accuracy is observed.

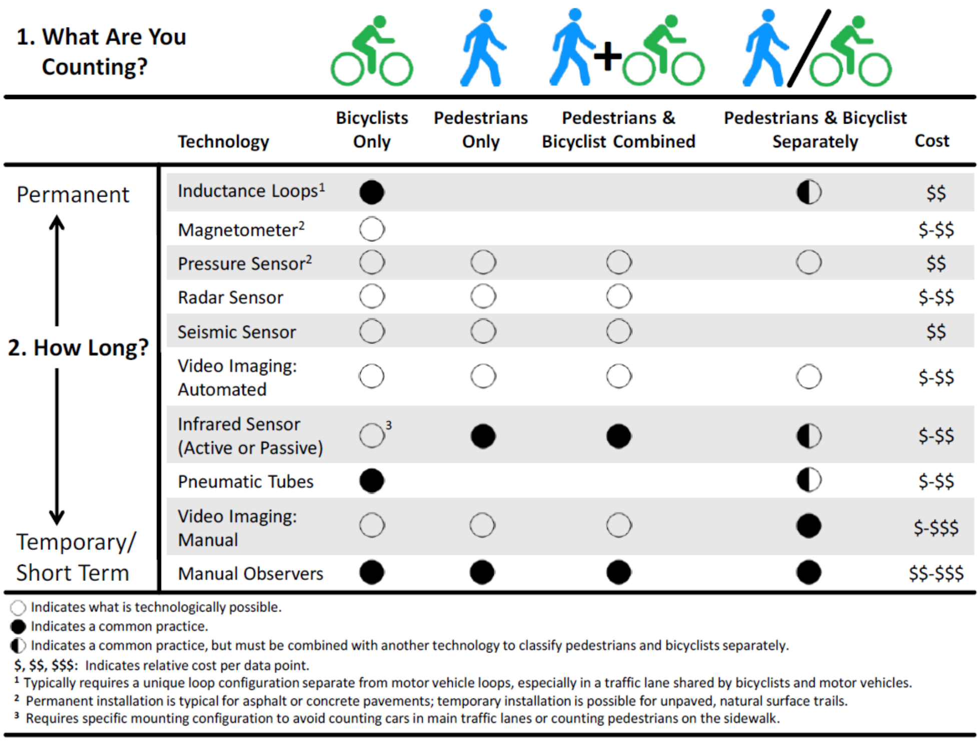
-*Infrared cameras:* As the name implies, in this technique a specific light sensor is used to detect a selected light wavelength in the Infrared spectrum. Thus, difficulties arise in distinguishing bicyclists from pedestrians as well as in cases of multiple people moving.

-*Pneumatic Tube:* In this method, an air switch is used to detect short bursts of air from a passing motorized or non-motorized vehicle. Despite the fact that it is considered a low-cost and portable approach, it can be utilized only to count bicyclists.

Technology	Typical Applications	Strengths	Weaknesses
Inductance Loop	Permanent counts Bicyclists only	Accurate when properly installed and configured Uses traditional motor vehicle counting technology	Capable of counting bicyclists only Requires saw cuts in existing pavement or pre-formed loops in new pavement construction May have higher error with groups
Magnetometer	Permanent counts Bicyclists only	May be possible to use existing motor vehicle sensors	Commercially-available, off-the-shelf products for counting bicyclists are limited May have higher error with groups
Pressure sensor/pressure mats	Permanent counts Typically unpaved trails or paths	Some equipment may be able to distinguish bicyclists and pedestrians	Expensive/disruptive for installation under asphalt or concrete pavement
Seismic sensor	Short-term counts on unpaved trails	Equipment is hidden from view	Commercially-available, off-the-shelf products for counting are limited
Radar sensor	Short-term or permanent counts Bicyclists and pedestrians combined	Capable of counting bicyclists in dedicated bike lanes or bikeways	Commercially-available, off-the-shelf products for counting are limited
Video Imaging – Automated	Short-term or permanent counts Bicyclists and pedestrians separately	Potential accuracy in dense, high-traffic areas	Typically more expensive for exclusive installations Algorithm development still maturing
Infrared – Active	Short-term or permanent counts Bicyclists and pedestrians combined	Relatively portable Low profile, unobtrusive appearance	Cannot distinguish between bicyclists and pedestrians unless combined with another bicycle detection technology Very difficult to use for bike lanes and shared lanes May have higher error with groups
Infrared – Passive	Short-term or permanent counts Bicyclists and pedestrians combined	Very portable with easy setup Low profile, unobtrusive appearance	Cannot distinguish between bicyclists and pedestrians unless combined with another bicycle detector Difficult to use for bike lanes and shared lanes, requires careful site selection and configuration May have higher error when ambient air temperature approaches body temperature range May have higher error with groups Direct sunlight on sensor may create false counts
Pneumatic Tube	Short-term counts Bicyclists only	Relatively portable, low-cost May be possible to use existing motor vehicle counting technology and equipment	Capable of counting bicyclists only Tubes may pose hazard to trail users Greater risk of vandalism
Video Imaging – Manual Reduction	Short-term counts Bicyclists and pedestrians separately	Can be lower cost when existing video cameras are already installed	Limited to short-term use Manual video reduction is labor-intensive
Manual Observer	Short-term counts Bicyclists and pedestrians separately	Very portable Can be used for automated equipment validation	Expensive and possibly inaccurate for longer duration counts

Table 1: Commercially-available bicyclist and pedestrian counting technologies (*U.S. Department of Transportation, 2016*)

Table 1 summarizes the characteristics of each method including their benefits and drawbacks, while Figure 3 shows the same techniques grouped by their application and the duration of the data collection.



Technology	Bicyclists Only	Pedestrians Only	Pedestrians & Bicyclist Combined	Pedestrians & Bicyclist Separately	Cost
Inductance Loops[1]	●			◐	$$
Magnetometer[2]	○				$-$$
Pressure Sensor[2]	○	○	○	○	$$
Radar Sensor	○	○	○		$-$$
Seismic Sensor	○	○	○		$$
Video Imaging: Automated	○	○	○	○	$-$$
Infrared Sensor (Active or Passive)	○[3]	●	●	◐	$-$$
Pneumatic Tubes	●			◐	$-$$
Video Imaging: Manual	○	○	○	●	$-$$$
Manual Observers	●	●	●	●	$$-$$$

○ Indicates what is technologically possible.
● Indicates a common practice.
◐ Indicates a common practice, but must be combined with another technology to classify pedestrians and bicyclists separately.
$, $$, $$$: Indicates relative cost per data point.
[1] Typically requires a unique loop configuration separate from motor vehicle loops, especially in a traffic lane shared by bicyclists and motor vehicles.
[2] Permanent installation is typical for asphalt or concrete pavements; temporary installation is possible for unpaved, natural surface trails.
[3] Requires specific mounting configuration to avoid counting cars in main traffic lanes or counting pedestrians on the sidewalk.

Figure 3: Simplified flowchart for selecting non-motorized count equipment (*U.S. Department of Transportation, 2016*)

As it becomes clear from the short description of the techniques, none of them can be applied independently for the identification of road modality. Each method focuses more on one category; hence there is no unique technique which can simultaneously count vehicles, bicyclists and pedestrians. Furthermore, none of the above-mentioned methods can be used for the investigation of movement patterns, as they have the ability to count but not to monitor devices as users change locations.

One more method which is frequently used for detecting, analyzing, and modeling human movement behavior is the Global Positioning System (GPS). The position of the device is calculated based on the information it receives from several satellites and the computation of the relevant distances to them (*Hofmann et al., 2008*). Using GPS technology, researchers can collect data on the routes followed by individuals in order to analyze how the latter interact with the geographical features in the area. However, despite the fact that this method makes it possible to "follow" the devices in space, having access to these data is very difficult, as it is necessary to request them from the provider, and potentially pay an additional charge. In addition to real-time access and analysis not being possible, filtering of data based on the research area may also be an issue. Finally, in high-density areas, such as city centers, signal reflections due to surrounding buildings can considerably reduce accuracy (*Van der Spek et al., 2013*).

2.2 The Wi-Fi method and its use for the computation of movement and occupancy patterns

In order to overcome the problems posed by the methods mentioned above, Wi-Fi signal can be used as an alternative, namely via access points, APs, or by the use of scanning devices made for this purpose (*Henniges, 2012*). Over the past years, smartphone sales and their use have seen explosive growth, while the majority of them come with a Wi-Fi network interface. When searching for available access points, devices periodically emit Wi-Fi packages. The type of these emitted packages is strongly related to the Wi-Fi connectivity of the device. When the device is not connected to a Wi-Fi network, it only sends probe requests, whereas when connected, it sends out all type of packets, supporting the basic idea of Wi-Fi roaming (*Bakker, 2016*). Based on that, it is possible to detect these emissions by making use of Wi-Fi monitoring equipment. In this case, the Wi-Fi monitoring procedure is carried out passively, as detecting the Wi-Fi signal emitted by devices does not require any settings to be changed or additional software to be installed. The time difference between two consecutive received packages is strongly related to the type of the smartphone, its battery level as well as the installed apps. The rate of received probe requests ranges from a minimum of 9 to a maximum of 513 requests per hour (*Bakker, 2016*).

Furthermore, when a device sends out a Wi-Fi package, it also sends its MAC address which can be used as a unique identifier for each device. Hence, with this indicator, the Wi-Fi monitoring system can be used not only as a counting method, but also as a means to research movement patterns in a specific area of interest (*Musa et al., 2012*). However, over the last years the use of randomized MAC addresses is promoted, with Apple and its iOS 8 being the first to prevent the research of customers' movements thus reducing the applicability of the system (*Bakker, 2016*). iOS 8 uses a randomly generated MAC address instead of the device's real address, in the cases of PNO and ePNO scans. According to Apple Inc. (*2016*), Preferred Network Offload (PNO) scans are conducted when a device is not associated with a Wi-Fi network and its processor is idle, determining in this way if a user can connect to a preferred Wi-Fi network. On the other hand, ePNO scans are conducted when a device is not associated with a Wi-Fi network or its processor is idle and are used when a device uses Location Services for apps which use geofences.

Due to privacy issues, however, MAC addresses collected by these sensors must be hidden or hashed in order to avoid possible misuse of the information. In Europe, according to the European Personal Data Protection Directive, a MAC address is considered private personal data. In the Article 29 Working Party (WP29) Opinion 9/2014 on device fingerprinting, access of a Wi-Fi device to a MAC address is considered to be covered by Article 5 of the ePrivacy Directive (*Duynstee et al., 2016*). This is supported by the opinion that, although the MAC address itself does not exist as a means of direct access to personal information, it is permanently linked to a unique device and by the use of a Wi-Fi network adapter, router, or a simple sensor, this address could be intercepted. Therefore, a MAC address can be used to monitor human movement behavior as the MAC addresses of different individuals are detected by sensors at different locations (*European Digital Rights, 2015*). Besides the need for MAC address hiding, as mentioned in Article 6, personal data must be collected for specified, explicit and legitimate purposes, including historical, statistical or scientific purposes, and in such cases the appropriate safeguards should be defined. Thus, all measures mandated by the Personal Data Protection Directive of the European Union, as well as by Directive 2002/58/EC, which protects personal data in electronic communications, should apply to the Wi-Fi system sensors' data collection procedure both for this thesis and for the use of its outcomes. Nonetheless, it is noteworthy that the use of random MAC address, on its own, does not guarantee user privacy. Wi-Fi probe requests contain data which still keep enough information to perform tracking, even after MAC address randomization. Parameters such as frame sequence numbers, the amount of information elements in probe requests as well as Wi-Fi radios have also to be taken into account and changed (*Vanhoef et al., 2016*).

As a way to fill the legal vacuum created from the increased computer use and the advancing data mining techniques, there is a trend to replace the existing personal data protection legislation with new Directives (*van Loenen et al., 2016*). Based on the new legislation and due to the number of citizen complaints the Data Protection Authorities received, the right of each citizen to not be subject to tracking of their movements has been enhanced. Specific rules and instructions have been put in place regarding the implementation of the Wi-Fi monitoring method and for each data collection method related to personal data. In the Netherlands, the Dutch Data Protection Authority, DDPA, set the legal framework governing the use of the Wi-Fi monitoring method in shops or in public spaces (*Authoriteit Persoonsgegevens, 2015*). A fundamental principle of privacy is that the collection of personal data should not take place covertly (*Datatilsynet, 2016*). This is why the Personal Data Act requires data collectors to notify users whenever Wi-Fi or Bluetooth tracking is taking place. Moreover, people should know exactly what the purposes of data collection are, who the collector and controller is and how they can contact them in order to get further information. With regard to recording MAC addresses, their deletion or anonymization is required to be done as soon as possible and, in any case, before data analysis. According to DDPA, collected information can be retained for 24 hours only, and then it must be destroyed or irreversibly anonymized, while in the public domain this procedure has to be done immediately upon collection. Finally, apart from notification signs, the collector has to offer users and local residents the opportunity to apply and exempt their devices from data storage and analysis. In general, it can be mentioned that despite the above described instructions and the possible limitations they entail, the Wi-Fi monitoring system can continue to be used in outdoor environments following the relevant rules and with the support of the Data Protection Authority.

Wi-Fi signal can be used either in an active or in a passive way. In the former case, Wi-Fi localization requires the modification of each device in order to receive data from the un-instrumented stationary access points. In the latter case, the passive Wi-Fi monitoring system consists of a number of sensors and a central server without requiring the active participation of users; this being one of the main advantages of the method (*Musa et al., 2012*). Taking into account the capabilities of this method as well as the fact that the range of the Wi-Fi monitoring sensors is larger than the ordinary width of city streets, it is clear that this system can be used as a permanent way of data collection. Furthermore, apart from the monitoring capability, Wi-Fi sensors can also act as conventional Wi-Fi routers; they can be used for the installation of a WLAN, thus supporting the ever increasing trend of WLANs becoming available in public spaces, under the main philosophy of the "Smart Cities" concept (*Komninos et al., 2012*). Finally, unlike the above-mentioned techniques, it is feasible to detect the total number of devices with Wi-Fi functionality within the sensors' operating range, without any limitations regarding the category in which the owner of the device belongs to (pedestrian, bicyclist, and vehicle). However, due to this massive data collection, further research is required for the identification of road modality.

During the last years, many efforts to use the Wi-Fi system of sensors have been made. In the research of Rose et al. (*2010*), Wi-Fi detections were used to predict bus and train arrival times based on Wi-Fi access points installed in these vehicles, while Musa and Eriksson (*2012*) use Wi-Fi monitors in order to track unmodified smartphones. Moreover, many efforts focus on the research of human behavior under different circumstances. Duynstee et al. (*2016*) installed a network of four Wi-Fi monitoring sensors in the city center of Dordrecht and recorded their findings for a period of two weeks. Using this system, details about all devices with their Wi-Fi functionality enabled were collected and, after a relevant analysis, pedestrian movement patterns were identified. Differences in street usage frequency as well as rush hours and changes during the week were investigated. As a way of validation, data from the camera counting system were used, designating a close linear relation between the total amount of pedestrians and the total number of detected devices, this way verifying the efficiency of this method. Meneses and Moreira (*2012*) used a Wi-Fi infrastructure of more than 550 access points in order to perform a large scale movement analysis in a University campus, to detect human motion, its relationship to the characteristics of the space as well as the connectivity between different attraction places. A Wi-Fi monitoring system was also used by Verbree et al. (*2015*) in the campus area of TU Delft, in order to detect shared facilities of different buildings through the amount of people who visit multiple buildings in a certain timespan. Finally, in the same campus area of TU Delft, taking advantage of the Wi-Fi monitoring data and the Markov model, Van der Spek et al. (*2016*) identified the activity of different users.

The evolution of technology over the last fifteen years has significantly improved monitoring technology, providing the opportunity to have a wide range of georeferenced disaggregate spatial behavior data (*Jung et al., 2012*) and to investigate the moving point of objects over time (*Kwan 2004; Andrienko et al., 2008; Orellana et al., 2010*). Unlike the potential diversity of movement, people usually follow simple and predictable movement patterns which can be very much used to explain the interactions between moving entities and between those entities and the environment (*Orellana et al., 2012*). For instance, Figure 2 illustrates the trajectories of four moving entities over twenty steps. From these trajectories we can identify: a flock of three entities over five time-steps, a periodic pattern in which an entity shows the same spatio-temporal pattern with periodicity, a meeting place where three entities meet for four steps, and finally, a frequently visited location which is a region where a single entity spends a lot of time (*Gudmundsson et al., 2008*).

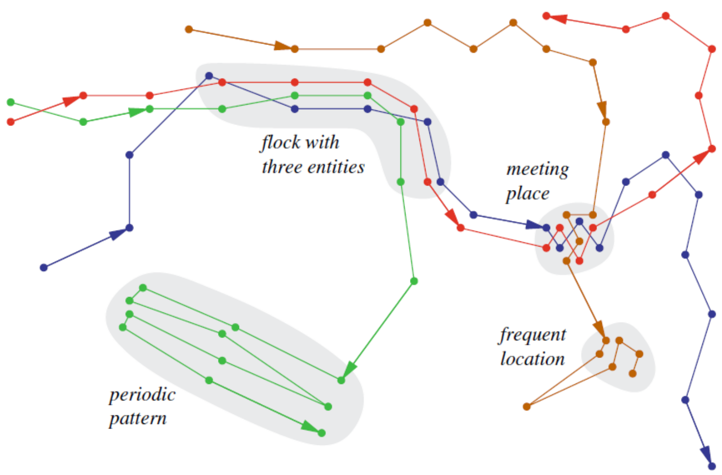
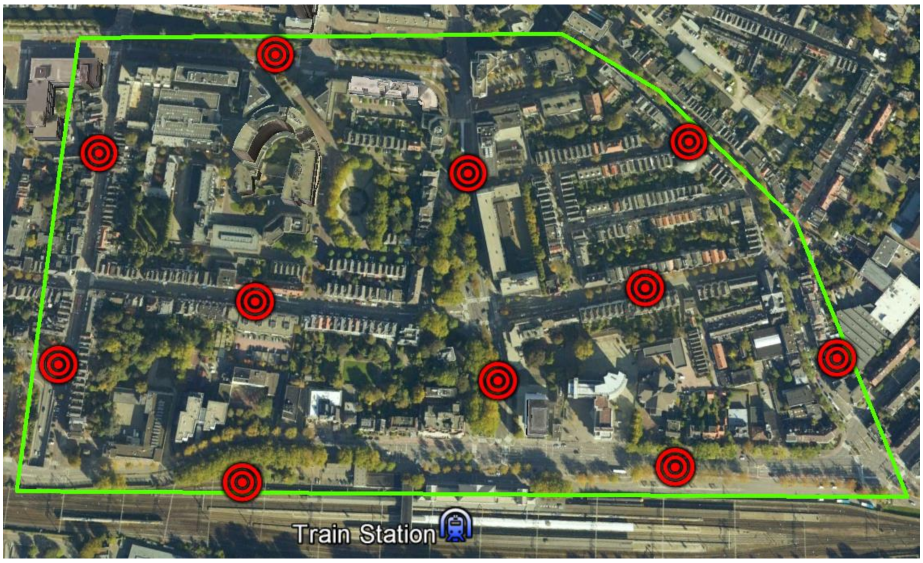


Figure 4: Patterns of Trajectory Movements
(*Gudmundsson et al., 2008*)

3. Design of the observation network, zero-level test & data acquisition

3.1 Observation network design

In designing the observation network, many parameters had to be taken into account. First of all, the eight available Meshlium sensors had to be evenly distributed over the research area so as to provide the same level of cover and allow for representative data collection. Considering that parameter and after experimenting with many combinations there were two final options, one of which had to be chosen. In the first scenario, sensors would be placed in the middle of the main streets, while in the second scenario, they would be placed in the cross sections of the streets.



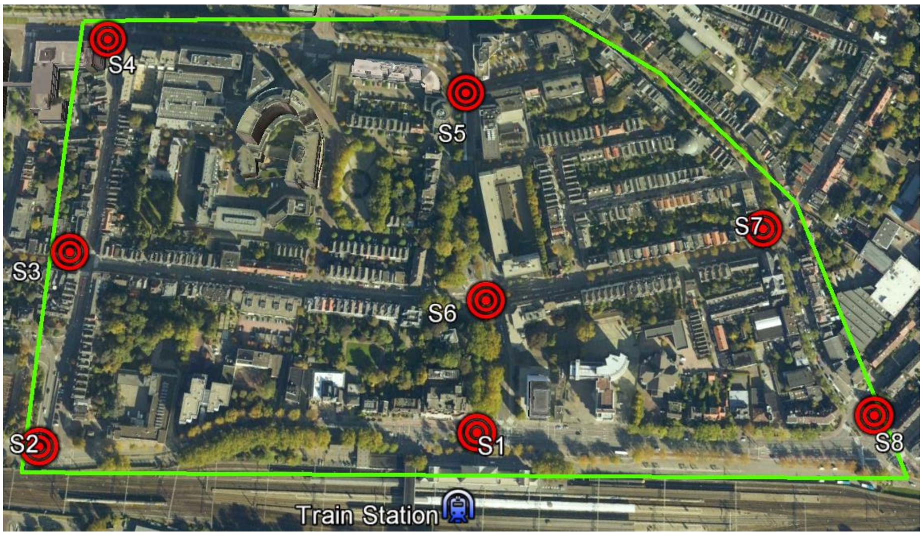


Figure 5: Visualization of the two scenarios for the observation network in the research area

Each scenario has its advantages and disadvantages. In the first case, the existence of one device in a specific street is more certain than in the second one. Especially in cases in which the sensors' scanning range is quite big, the device location at the time of the record cannot be ensured when sensors are placed at street intersections. Furthermore, in cases in which the most crowded streets are known and the main goal is to study the total movements and not the movement in each street separately, in the first scenario fewer sensors are required to be placed in each intersection. For example, in the area between sensors S2 and S4 or S1 and S5 or S2 and S8, as they are illustrated in the second scenario, two sensors are needed instead of three under this scenario.

On the other hand, as shown in Figure 5, the number of sensors required to cover the whole area in the first scenario is much bigger than in the second one. Moreover, an advantage of the second option is that it allows for movement research to be conducted for each street separately. In this way, the correlation between the streets can also be investigated.

After deciding the approximate placement of the sensors, both the technical specifications of the devices, such as the selected antenna, its range, waterproofing and the need for continuous access to electric power, were taken into account in order to decide on the exact locations of the devices. Two types of antennas can be used with these sensors. The first one records signals from the area around it, while the second one can be used as a directional antenna due to the fact that it scans only at an angle of 180 degrees.

Figure 6: The two types of sensor antennas for Meshlium scanners (the directional antenna is illustrated on the right side)

Thus, when the sensors were placed in the middle of the local research area, in traffic lights for example, the first type of sensor was preferred to cover the whole regional perimeter. However, in cases like balconies or windows in buildings, a directional antenna was used in order to record signal from the street area exclusively.

Figure 7: The choice of antenna based on the location of the sensor

Taking all the above-mentioned parameters into account, the observation network was designed and implemented in cooperation with the technical department of the Municipality of Dordrecht by placing the sensors in the selected locations. Finally, due to privacy issues, relevant notification signs were placed throughout the research area in order to inform people about the existence of Wi-Fi sensors.

Figure 8: Wi-Fi sensor placement in the research area in cooperation with the technical department of the Municipality (left) and notification signs informing people about their existence (right)

3.2 Zero-level test

According to the technical specifications of the sensors, their scanning range (50-200 meters) is larger than the ordinary city streets width. Hence, each sensor should cover the whole area of the observation network.

However, before collecting all the data, a zero-level test is required to verify them. To conduct this test, different kinds of devices from various manufacturers were used. More specifically, three mobile devices (Samsung, Apple, LG), one tablet (Lenovo) and one laptop (Toshiba) were placed in various locations around each sensor, with their Wi-Fi functionality enabled, in order to check if the relevant sensor can record devices there or if the region is not included in the scanning range of the device.

A factor used during the zero-level test was the received signal strength indicator (RSSI). Knowing their MAC addresses and using the sensor devices software, the relevant RSSI was computed. There is not a simple equation to calculate the distance between a device and a scanning point using the RSSI, as there are many factors which affect the outcome, including the type of the device, its inveteracy, its manufacturer as well as where users keep it (holding it in their hand, having it in their pocket, backpack, etc.). However, it is considered an indicator of transmission quality and average distance. Thus, the higher the signal strength the closer the device is to the sensor.

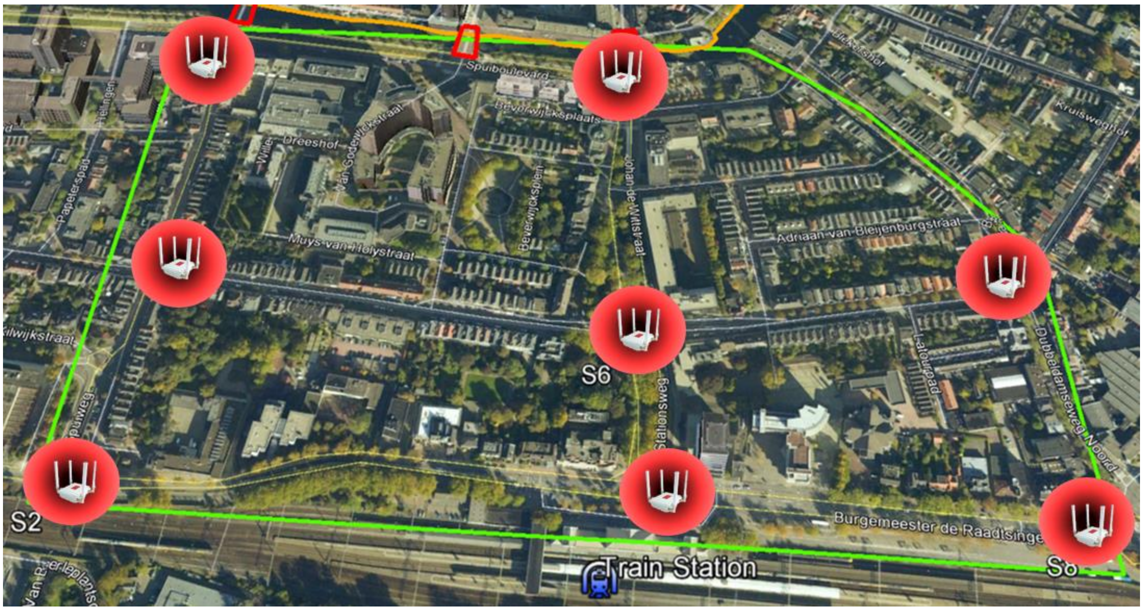


Figure 9: Visualization of the final observation network

Following the results of the zero-level test, the actual range of each sensor was computed and compared to the range specified by the manufacturer. It was confirmed that the network had been properly designed, as not a single location was found where the sensors failed to record the signal of the devices used. In general, all RSSI values in the tested points around the cross sections of the streets were quite high, indicating that the transmission quality was adequate both on the streets and the pavements. Figures 10 and 11 show the average RSSI values of the tested devices in different locations around sensors 1 and 8, as well as the shape of their respective range. Based on them, it is worth noting the different application of the two antenna types. In sensor 1, Figure 10, the standard antenna was used, covering the whole surrounding area (spherical shape). On the contrary, in sensor8, Figure 11, the directional antenna was installed, thus scanning only the one-side area indicated by the semicircular shape. More details and photos on the implementation of the observation network and the zero-level test can be found in the appendices.

Figure 10: Average RSSI values around sensor1 (left) and its range shape (right)

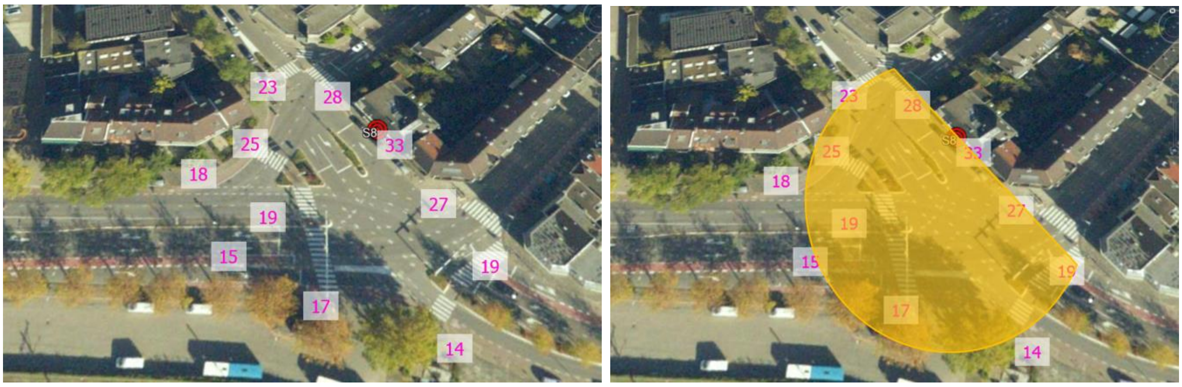


Figure 11: Average RSSI values around sensor8 (left) and its range shape (right)

3.3 Data acquisition-limitations

3.3.1 Wi-Fi data

Following the zero-level test, data were collected for a period of one month, from mid-September till mid-October. A short overview of the plan is shown in Figure 12. Due to the fact that no internet connection was available in the places where the sensors were installed, manual download of the data was periodically required. Thus, the recorded data were removed from the local memory of the device, while time settings were checked in order to avoid any problems that could arise due to sensor desynchronization.

Date	S1	S2	S3	S4	S5	S6	S7	S8	Weather
13/9/2016			Installation				Installation		Download
14/9/2016	Installation	Installation		Installation	Installation	Installation		Installation	Download
15/9/2016			Download-Update	Download-Update			Download-Update	Download-Update	Download
16/9/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download
17/9/2016									Download
18/9/2016									Download
19/9/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download
20/9/2016									Download
21/9/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download
22/9/2016									Download
23/9/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download
24/9/2016									Download
25/9/2016									Download
26/9/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download
27/9/2016									Download
28/9/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download
29/9/2016									Download
30/9/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download
1/10/2016									Download
2/10/2016									Download
3/10/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update		Download
4/10/2016								Download-Update	Download
5/10/2016									Download
6/10/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download
7/10/2016									Download
8/10/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download
9/10/2016									Download
10/10/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download
11/10/2016									Download
12/10/2016									Download
13/10/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download
14/10/2016	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download-Update	Download

Figure 12: Overview of the data collection procedure

During the data collection procedure, every time a sensor received a signal from a device with enabled Wi-Fi functionality, it stored the following parameters:

- The timestamp: the relevant date and time of signal reception.

- The MAC address of the device's wireless interface, which acts as a unique identifier.

- The service set identifier, namely SSID, which is the name identifier of the Wi-Fi network in which the device is connected.

- The strength of the signal (RSSI), which indicates the average distance between the device and the scanning point. It is used to determine the quality of a received signal to a client device in arbitrary units. According to the standards of Institute of Electrical and Electronics Engineers (*IEEE 802.11*), RSSI can range from 0 to 255. However, each chipset manufacturer can define their own “RSSI_Max” value and RSSI range (*Lui et al., 2013*). The RSSI range is entirely up to the manufacturer, but the general idea is that the higher the RSSI value, the better the signal. Unlike the different RSSI Max values, all chipset manufacturers usually use zero as the lowest value. Nevertheless, there are cases in which the value of -1 is used, which indicates a very low-quality signal and unusable records.

However, it is important to mention that due to this fluctuation in the RSSI range, signal strength is often preferred as an indicator of quality. The unit of measurement used to determine the strength of the signal is the dBm which is an abbreviation for the power ratio in decibels (dB) of the measured power referenced to one milliwatt (mW) . According to the IEEE 802.11 standard, signal strength for the wireless network ranges from -100 to -10 dBm, which represent the minimum and maximum received signal power respectively. However, it does not define any relationship between RSSI values and power level in dBm.

Data Output | Explain | Messages | History

	id_frame integer	timestamp timestamp without time zone	mac character varying(17)	ssid character varying(32)	rssi character varying(3)
1	18498390	2016-10-12 23:59:42	A0:EC:80:47:6D:FC	H368N476DFC	3
2	18498380	2016-10-12 23:59:42	00:1D:68:70:EC:47	SpeedTouchvanBram	7
3	18498476	2016-10-12 23:59:42	88:03:55:C2:6C:01	KPN Fon	33
4	18498471	2016-10-12 23:59:42	00:1D:AA:E2:8F:68	VFNL-E28F68	25

Figure 13: Example of stored details at each sensor’s record

All the above-mentioned indicators will be used in the main part of data analysis in Chapter 4.

3.3.2 Other data sources

It was decided that, in addition to Wi-Fi data, further data sources would be used as well. In the city center of Dordrecht, which is very close to the research area, there are four pedestrian counting cameras installed by the Municipality. Furthermore, in a different street of the same area, the Retail Management Center consultancy (RMC) had also installed a similar camera.

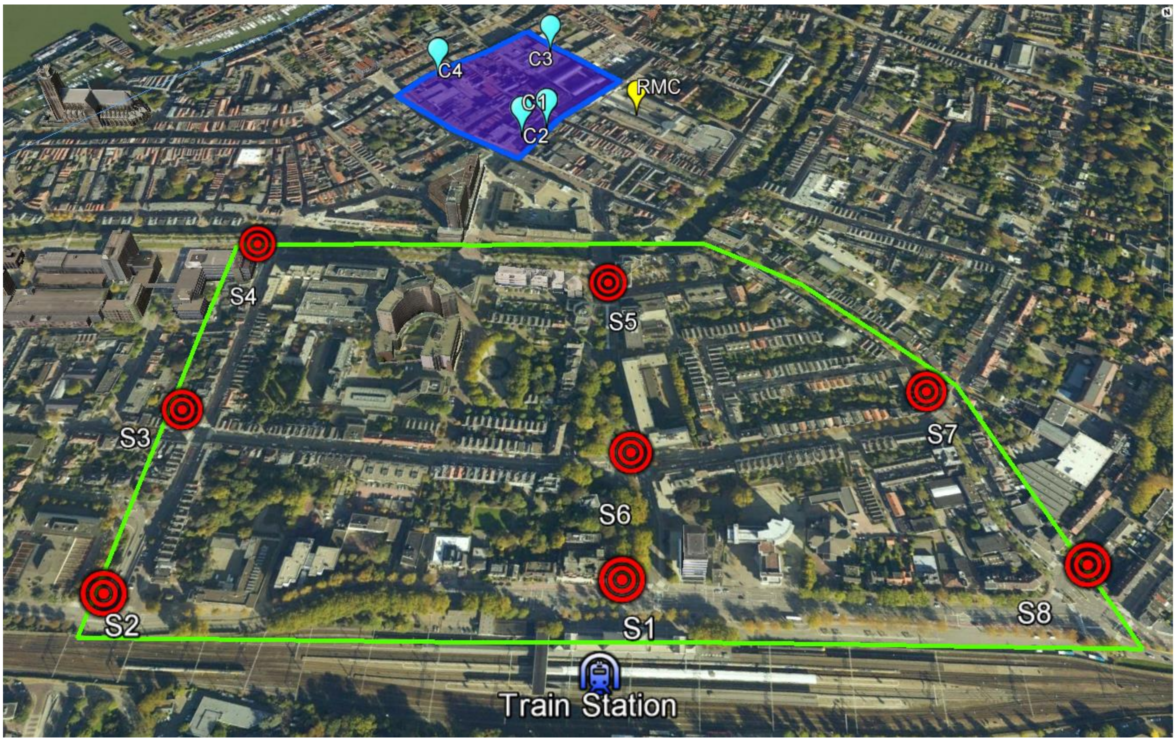


Figure 14: Map showing part of Dordrecht. The light green, blue, light blue, and yellow colors represent the research area borders, the main part of the city center, the four locations of the Municipality cameras and the location of the RMC camera respectively.

Based on that and after a relevant application to both organizations, hourly datasets from each camera were collected throughout the observation period. As it was expected and as illustrated in Figure 15, the relationship between the Municipality and RMC datasets is linear and very close, as the high R-squared value of the trendline also proves. The main purpose for the use of these datasets is the identification of occupancy patterns and specifically the relationship between the occupancy patterns in the city center and the research area, which is described in Chapter 4.

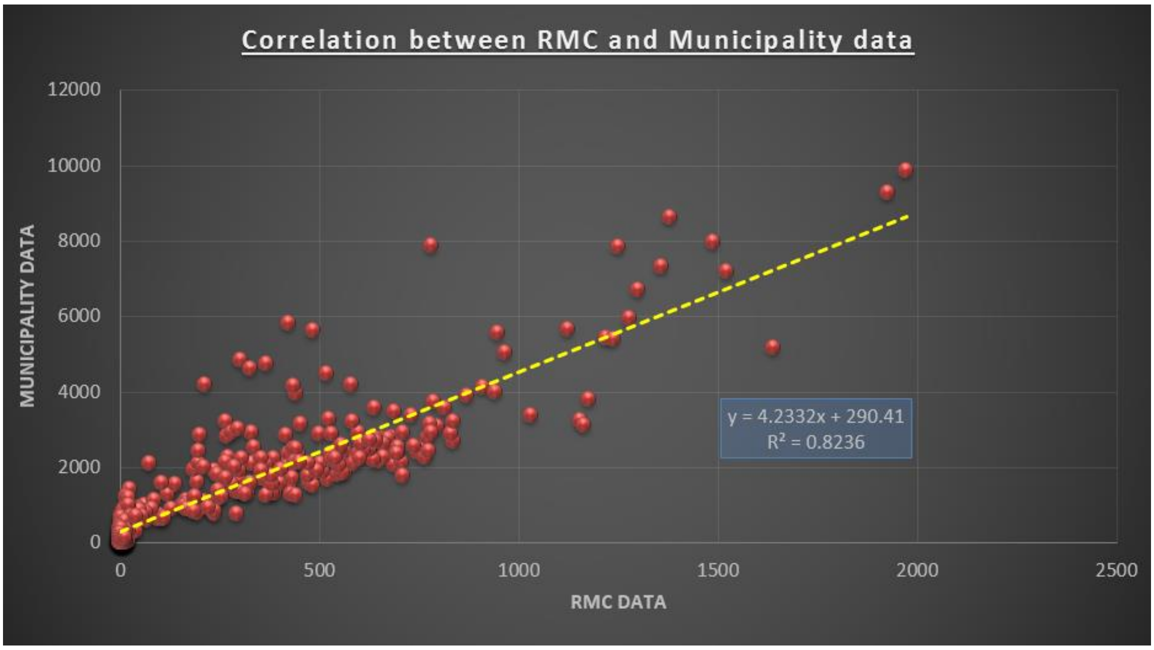


Figure 15: Correlation plot between RMC and Municipality pedestrian counting datasets

An additional data source is the questionnaire-based survey. During the observation period, pedestrians and bicyclists passing through different streets, on different days and at various times throughout the day, were asked to fill out a survey. In this way, general information on the percentage of people who have their device's Wi-Fi functionality enabled will be collected. Thus, the percentage of the representativity of research outcomes can be assessed while the survey results will also be used in the validation procedure, which is described in Chapter 5.

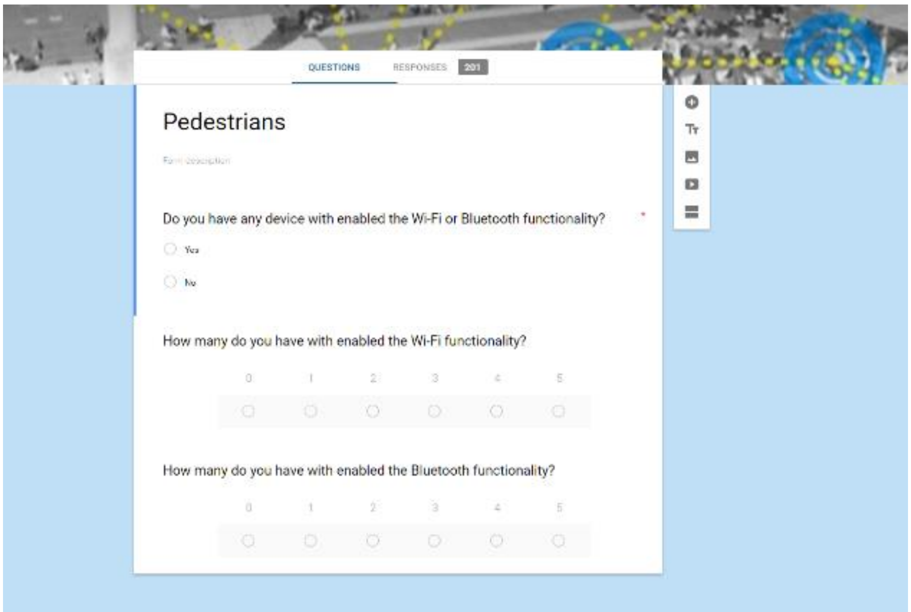


Figure 16: Overview of the pedestrians' questionnaire

Last but not least, hourly weather information was recorded in the research area. Hence, the effect of the weather to road modality and occupancy of the area can be investigated, as described in Chapter 4.

SATURDAY	2pm	3pm	4pm	5pm	6pm	7pm	8pm	9pm
Forecast	Showers	Partly Cloudy	Showers	Partly Cloudy	Partly Cloudy	Partly Cloudy	Cloudy	Cloudy
Temp (°C)	14°	14°	14°	14°	13°	12°	11°	10°
RealFeel®	12°	13°	12°	12°	11°	10°	10°	10°
Wind (km/h)	13 NE	13 NE	13 NE	13 NNE	11 NNE	11 NNE	7 NNE	7 NNE

Figure 17: Example of the weather information stored

3.3.3 Data limitations

There are some limitations to the various datasets which were collected. First of all, the Wi-Fi dataset represents only part of the actual number of users. However, the steadily increasing trend of this percentage constitutes one of the main purposes of this research and, at the same time, increases the representativeness of the outcomes of this method. Moreover, although the data source of pedestrian counting in the city center can be characterized as very useful for researching occupancy, there are no relevant data sources in the research area which could be used directly for the validation of the method outcomes. Thus, the validation procedure significantly depends on the questionnaire, which is quite slow to process, especially in cases where the personnel is limited.

4. Data preparation & analysis

4.1. Hashing of MAC address

As mentioned in literature review, there are some privacy issues which have to be taken into account when using the Wi-Fi monitoring method. Despite the fact that the scientific research of this work is one of the purposes protected under Article 6 of Directive 95/46/EC, it was decided to hash the MAC address of the recorded devices.

During the data collection procedure with the use of Meshlium sensors, the user can opt to anonymize the MAC addresses. However, a significant disadvantage of this cipher is that the hashed address of each device changes every day. Additionally, one more problem is that each sensor encrypts the same device in a different way and, thus, it is not possible to use the final data list to identify movement patterns (*Duynstee et al., 2016*). Due to these problems, automatic encryption was not used during data collection; a script code written in Python (see appendices) was used instead at the end of the data collection period in order to create a one-way hash-function and thus replace the original MAC address of the devices before the following steps of analysis.

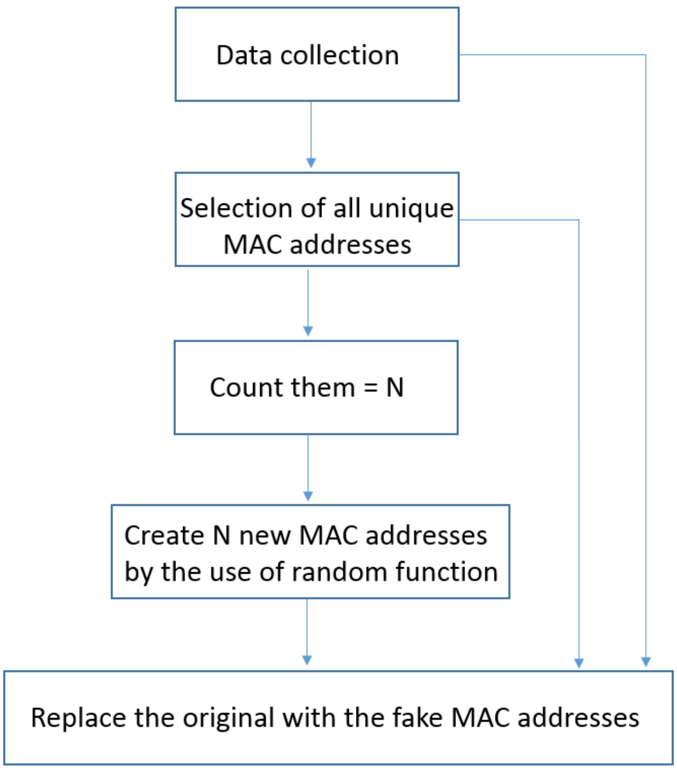


Figure 18: Flowchart of the main steps of the hashing procedure

Figure 18 illustrates the main steps included in the hashing procedure. First of all, when data collection was completed, the records of all sensors were combined and their unique MAC addresses were separated, counted, and represented as N. Then, using a random number generator function, N unique and new MAC addresses were created and replaced the initial ones in the records file before proceeding with the next steps of the analysis procedure. In this way, MAC addresses can be used as a unique ID to research user behavior without any legal or privacy issues.

4.2 Correction of record time

The second step of data preparation is the correction of record times. As mentioned in Chapter 3.3.1, during the placement of the sensors in the research area as well as after downloading the records, the time settings of the scanners were checked and synchronized with the time of the same device, thus achieving a synchronization accuracy with a deviation of less than one second. However, there were many instances in which significant differences were observed between the actual time and the time of the sensors, with time differences even longer than two hours. This would significantly affect the computation of movement patterns and lead to wrong outcomes. For instance, if there is an error in sensor 2 involving a delay in the record times, movement from sensor 1 to sensor 3 and then to sensor 2 will be computed, instead of the actual movement from sensor 1 to sensor 2 and then to sensor 3.

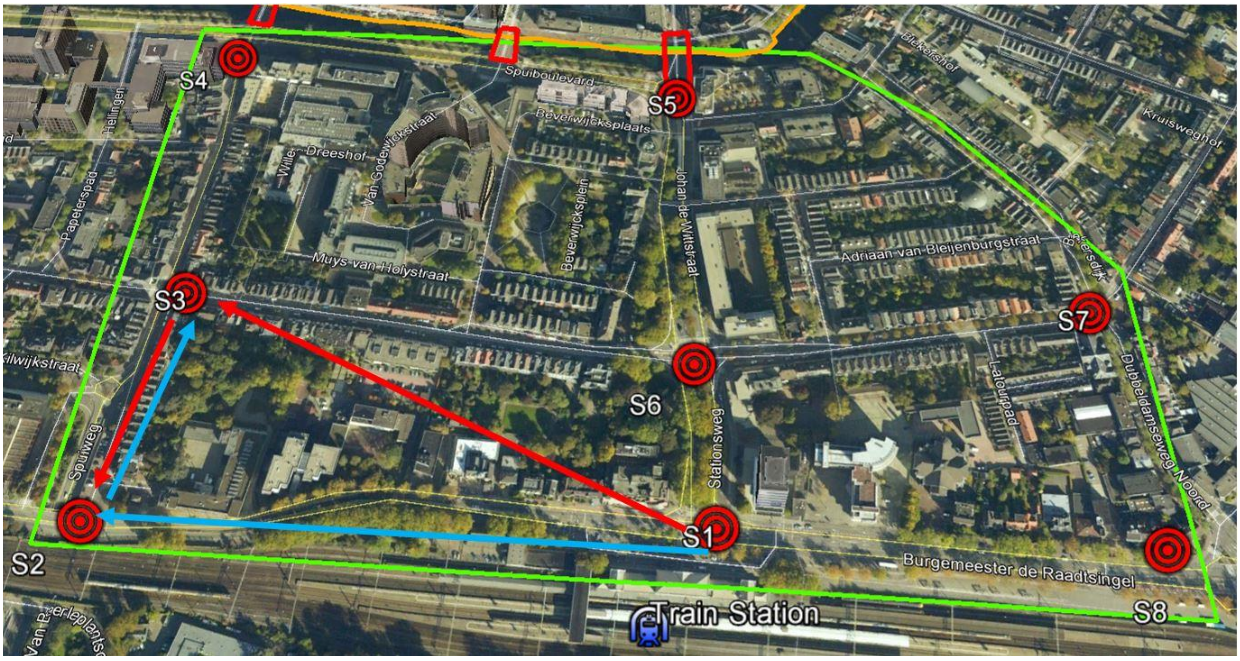


Figure 19: Example of the effect of time synchronization on the computation of movement patterns. The light blue line represents the actual movement while the red one represents the computed movement in case of a sensor 2 desynchronization.

Since this problem is already mentioned in other similar studies, precautions were taken before implementing the system. According to the available literature, there are two types of time in the system of the Meshlium sensors: the system time which is the problematic one and is used as the record time of the scanned devices, and the hardware time which remains always synchronized. Thus, the Putty software was used in order to create a log file for each sensor in the system and set the automatic storing of two different types of time for each minute. With the use of this log file, all record times were checked and corrected before proceeding with the analysis procedure, in order to avoid the wrong computation of movements.

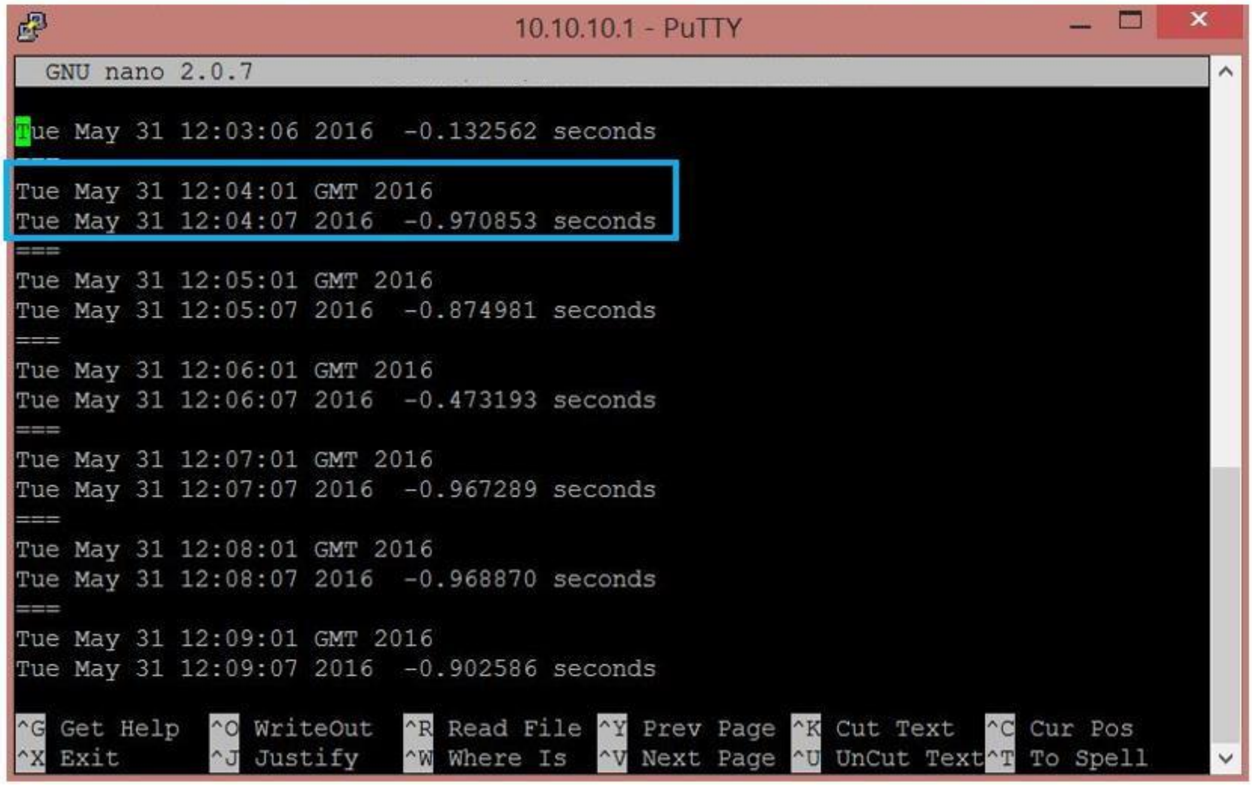


Figure 20: Example of the log file, the storing of the system (first raw) and hardware time (second raw) as well as the relevant time difference.

4.3 Filtering

Another necessary step is filtering. Among the 33 million records throughout the whole research period, there are many which are useless for the purposes of this work, and can be characterized as outliers and be removed, thus reducing the size of the dataset and increasing the speed of the procedure. The outlier records can be appertained to one of the following categories:

- **Devices which were scanned by only one sensor**. With the evolution of technology over the years, there are now a lot of devices which support Wi-Fi functionality. Many of them are indoor devices, such as printers and televisions, but there are also a lot of outdoor ones, such as automatic food dispensers, pay parking devices, etc. All these devices do not move into space and they can be characterized as “static”. However, each of them was continuously scanned by one sensor and thus the relevant records had to be removed.

Figure 21: Example of static devices which can be scanned by Wi-Fi sensors

Moreover, this category can also include devices which were not really static but were scanned at only one scanning point. For example, devices with enabled Wi-Fi functionality, which belonged to people who moved in the borders of the research area and afterwards moved away from it can also be characterized as “static” and removed from the dataset.

- **Devices which were continuously scanned for a period longer than twelve hours**. This category includes devices which were continuously scanned, every thirty minutes, by one or more sensors for a period longer than twelve hours. When it comes to such devices, it can be assumed that they did not belong to humans and thus they were characterized as “unrepresentative” and were removed from the dataset.

- **Records whose time difference is longer than two hours**. Records for the same device with a time difference longer than two hours between them constitute the third category. If, for instance, a device is scanned in the morning by sensor 1, in the afternoon by sensor 2 and at night by sensor 3, it would be useless to investigate its movement inside the research area. It was considered that no more hours would be required for a user to move in the research and consequently be scanned by more than one sensors. This phenomenon could be justified by the fact that the user activated and deactivated Wi-Fi functionality on a regular basis, or in the period between the records they only moved in the area within the range of the sensors or outside the research area. For this reason, only records with a time difference shorter than two hours are kept in the dataset. Moreover, this can also be explained by the fact that users may not move continuously within the research area, they may move-stay-move-stay. In this case, the user should be scanned by at least two different sensors when they move within a two-hour period, and thus only the movement timeslots are kept for the analysis procedure.

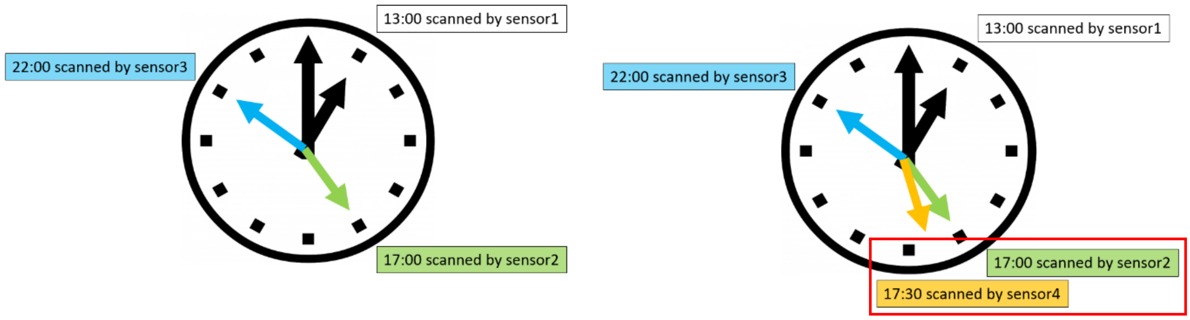


Figure 22: Two examples of records which belong to this category. In the first one (left) all records are removed, while in the second one (right) only the middle records (at 17:00 and 17:30) are kept in the dataset.

- Records with negative received signal strength indicator (RSSI).

As mentioned in Chapter 3, RSSI constitutes an indicator of transmission quality based on the general assumption according to which the higher the signal strength, the closer the device is to the sensor. As the relevant distance between the sensors in the observation network was not that large, there is a possibility that a device signal was scanned with delay and was debilitated by a nearby sensor. Given that and taking into account that the RSSI value of all tested devices in the covered area around the sensors was higher than ten, it was decided to remove all the records with negative RSSI values. Hence, it is expected that filtering and deleting false records and movements will be much more beneficial than filtering devices which were scanned by the correct sensor and had negative RSSI values only.

After the implementation of the above-mentioned filters, it is possible and interesting to investigate the relationship between the unfiltered and filtered devices throughout the research period. In Figure 23 the hourly numbers of these two parameters are visualized, proving the close relationship between them by the simultaneous increases and reductions.

- Devices which are connected to the Wi-Fi network of trains

As the research area is quite close to the railway, it is possible that, despite the existence of a tall wall which acts as a barrier with regard to signal transmission, the system recorded devices of users who were aboard the train. For this reason, all devices which were connected to the Wi-Fi network of trains were removed from the dataset. Despite the fact this filter was used, there is also the risk that some of the scanned devices were connected to the Wi-Fi network and thus could be considered as belonging to users in the street. However, as described in detail in Chapter 5, there is no difference in the validation procedure between the streets next to the railway and those away from it.

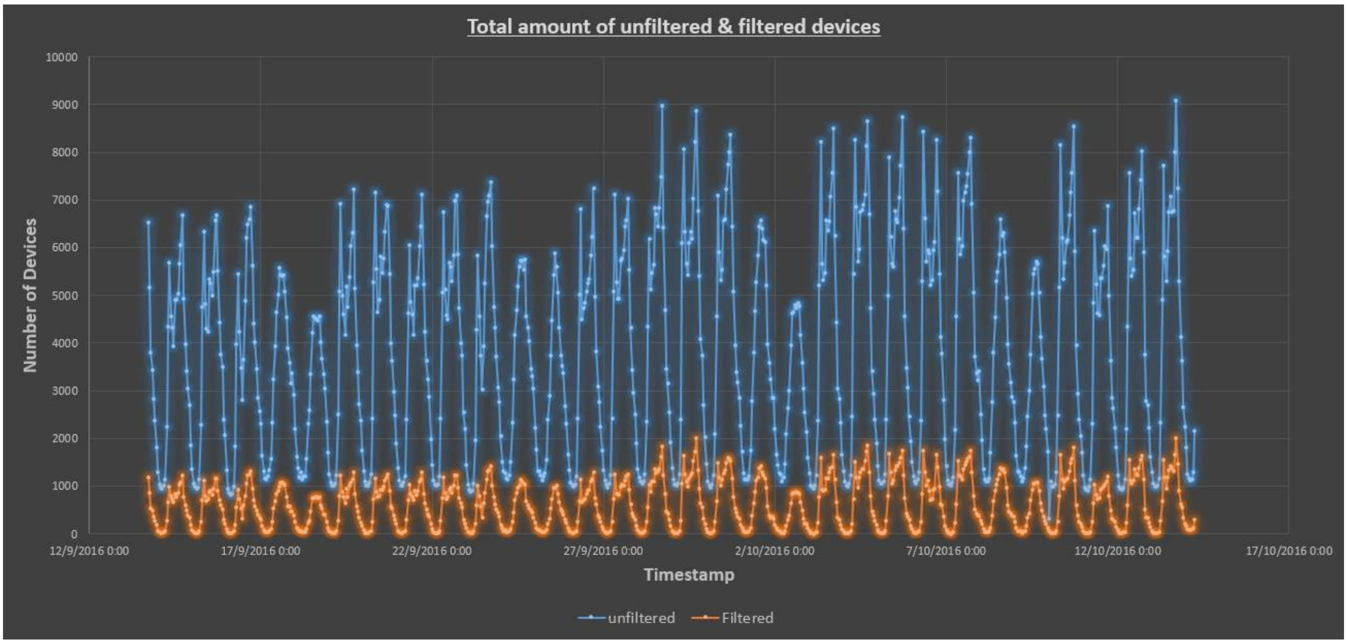


Figure 23: Hourly number of unfiltered and filtered devices throughout the research period

Figure 24: Linear regression between unfiltered and filtered devices (Intercept=0)

Figure 24 illustrates the hourly combinations of unfiltered and filtered devices throughout the research period as well as the best fitted trendline of the form y = ax and its equation. The unfiltered devices represent the initial number of devices scanned by the system while the filtered category includes those devices that remained in the database after applying the whole set of filtering rules. As it is clear, there is a very close and linear relationship between the two variables, which shows that filtering is systematic and is done in the same way. Furthermore, based on the equation of the trendline and its slope, it can be pointed out that overall about 18% of the initial devices is kept into the dataset after the filtering procedure. As the above Figure shows, the percentage is not constantly stable, as it depends on the number of unfiltered devices. Thus, in cases in which the number of scanned devices is quite small, the percentage is smaller than 18% while it gradually increases as the number of unfiltered devices goes up. Finally, there is an average number of 1150 permanent static devices in the area and, as shown in Figure 25, if the best fitted trendline of the form y=ax+b is researched, the average proportion between filtered and unfiltered devices is 1:4.5, equals to a percentage of around 22%. At this point, it is important to explain this linear regression as well as the high value of the R2 index. Regardless of the number of unfiltered devices throughout the day, the percentage of them, which is used for the analysis procedure, remains almost the same. This percentage can be expected to be higher. However, the number of unfiltered devices increases significantly not only by the number of static devices filtered but also by all the above-mentioned parameters as well as the location and the size of the research area.

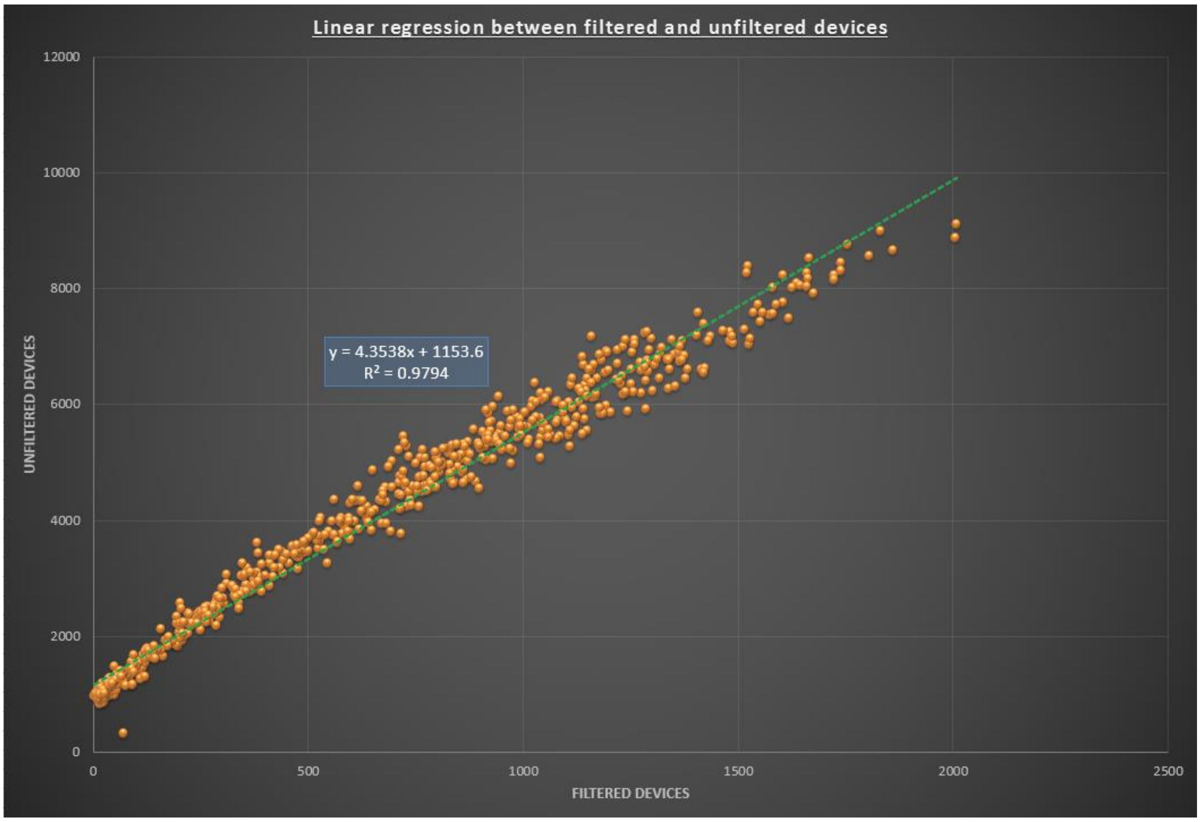


Figure 25: Linear regression between filtered and unfiltered devices

4.4 Computation of road modality

Following the filtering procedure, the dataset of Wi-Fi devices can be used for the computation of road modality. This procedure consists of the following main steps, which are visualized in the flowchart of Figure 26:

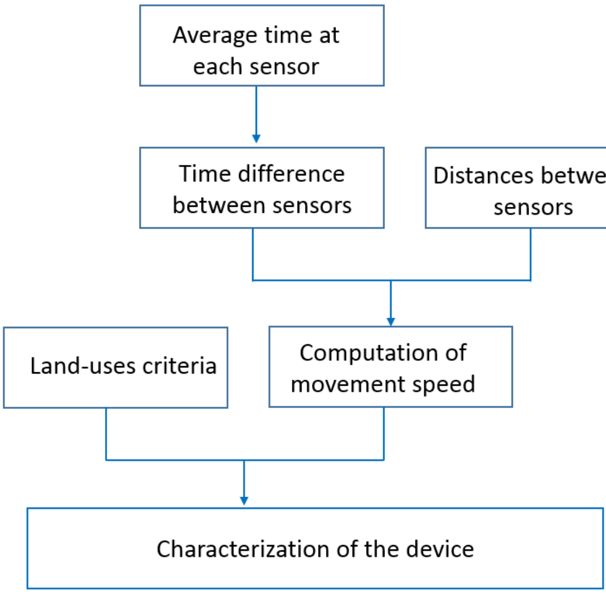


Figure 26: Flowchart of the main steps of road modality computation

- Computation of average time at each sensor.

As people move in the research area and as the range of the sensors is quite big, most of the devices were scanned multiple times by each sensor. Thus, a problem which has to be solved is the need to define the time each device was in the area of the sensor. The first option is to use the record time of one of the relevant records. However, this scenario's disadvantage is that the exact location at that time is unknown, which renders the computation of the relevant distances between the sensors problematic. For this reason, it was preferred to use the average time of the records for those cases in which the time difference between the first and the last record is less than one minute. Otherwise, two record times are computed; one for the computation of the time up to this sensor and one for the computation of time from this sensor to the next one, which are computed as the average time of the first and the last records respectively.

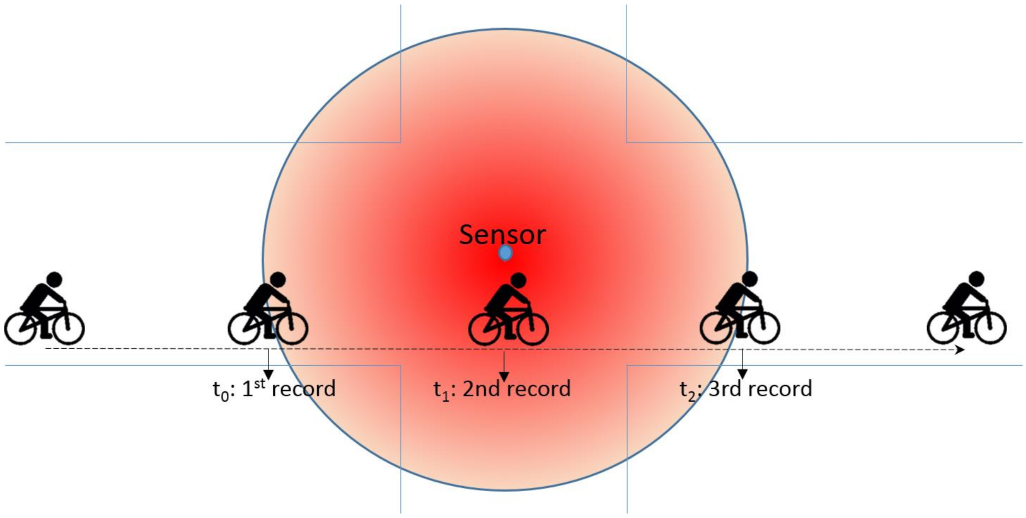


Figure 27: Visualization of the main idea for the average record time

- Computation of the time difference between sensors.

For the computation of a user's movement duration between two sensors, the average time at each sensor is used. Thus, if for instance a bicyclist goes from sensor 1 to sensor 2 then the duration of this movement (Δt) is calculated as:

$$\Delta t = \text{(average time at sensor 2)} - \text{(average time at sensor 1)}$$

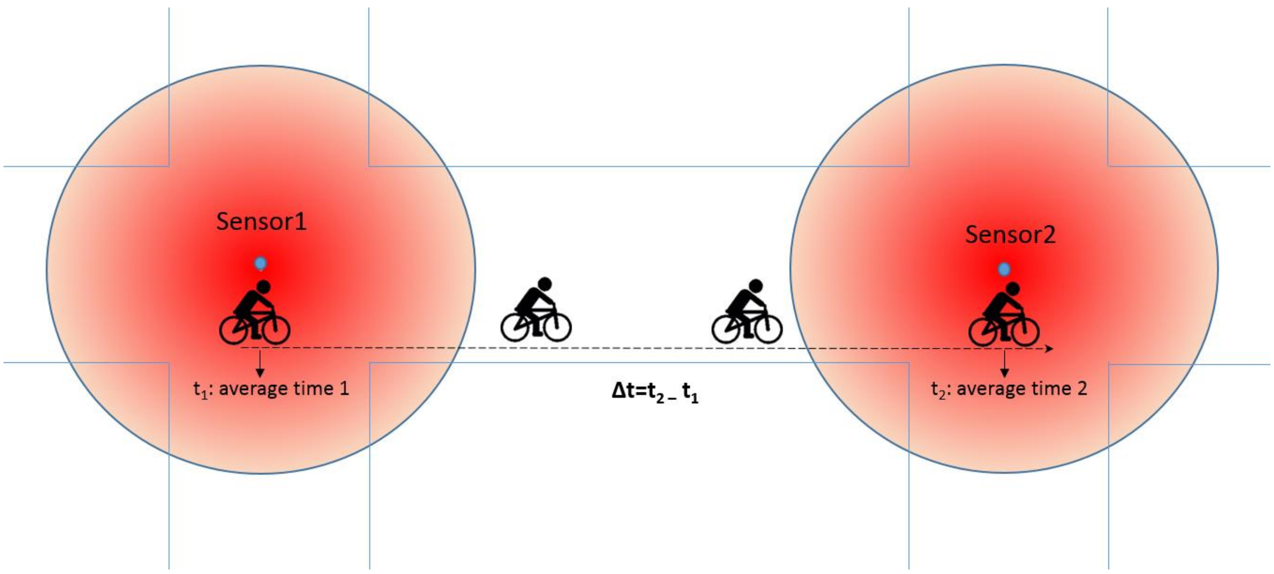


Figure 28: Visualization of the main idea about the computation of time difference between sensors

- Computation of movement speed.

The movement speed of each device between two consecutive sensors was computed based on the following equation:

$$u = \frac{d}{\Delta t} \quad (1)$$

where

u – the movement speed of a user between two sensors,

d – the relative distance between two sensors,

Δt – the time difference of the record times at each sensor.

Taking advantage of the fact that the average value of the records at each sensor was used for the computation of the movement duration, the exact distances between the sensors, and not the distances between the borders of the sensors range, were integrated into the formula.

Figure 29: Distances between the sensors of the research area

- Combination of movement speed and street-uses criteria for the characterization of devices.

Using the computed movement speeds and combining them with speed and street-uses criteria, each device was characterized as "pedestrian", "bicycle", or "vehicle" - the three categories being investigated in this research. Different speed ranges were set as speed criteria for each category based on the outcomes of related works (*Rastogi et al., 2013*) (*Bussche, 2015*) while attributes of the research area, such as the existence of one-way streets and roads under construction, were also taken into account. For instance, a car was not allowed to move from sensor 3 to sensor 2 or from sensor 4 to sensor 3. Furthermore, due to construction works being in progress, vehicles were prohibited in the street which connects sensors 5 and 7 during the data collection period.

User category	Speed criterion
Pedestrians	< 7 km/h
Bicycles	7 - 20 km/h
Vehicles	> 20 km/h

Table 2: Speed criteria for the computation of road modality

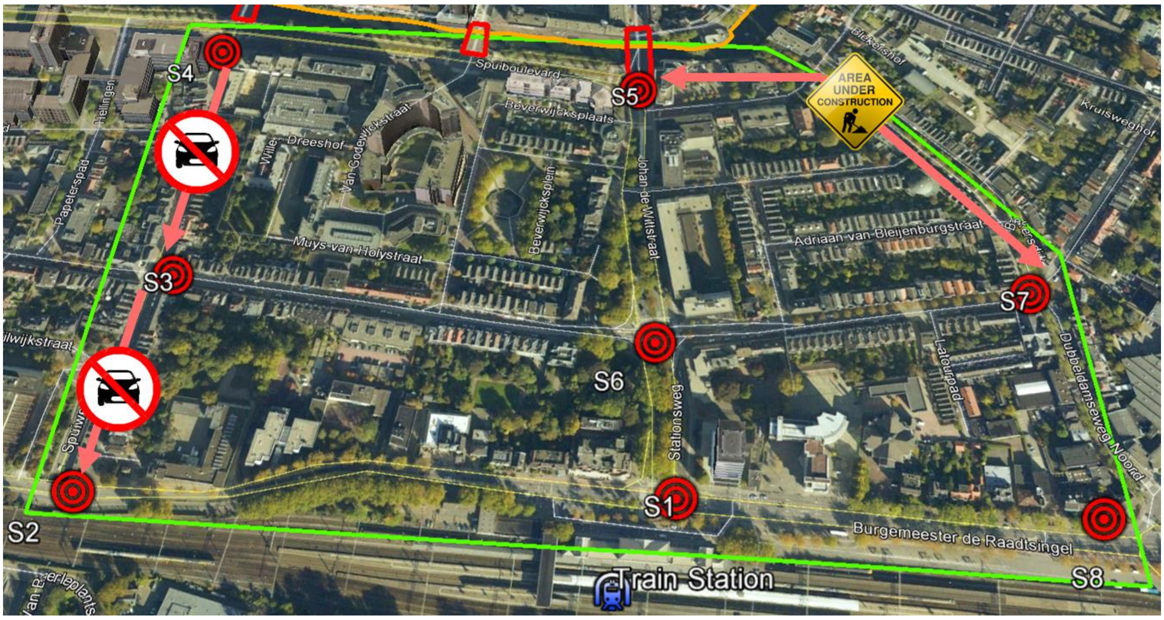


Figure 30: Street-uses criteria in the research area

Having completed the steps described above and the classification of devices, it is interesting to investigate the distribution of the road modality throughout the research period. However, before that, the relationship between the total number of identified movements and the number of unfiltered and filtered devices can be studied. Figure 31 illustrates the total amount of identified movements for the whole data collection period, while Figure 32 also contains the relevant numbers of unfiltered and filtered devices. Based on the latter diagram, simultaneous changes of the three variables can be observed, indicating a very close relationship between them.

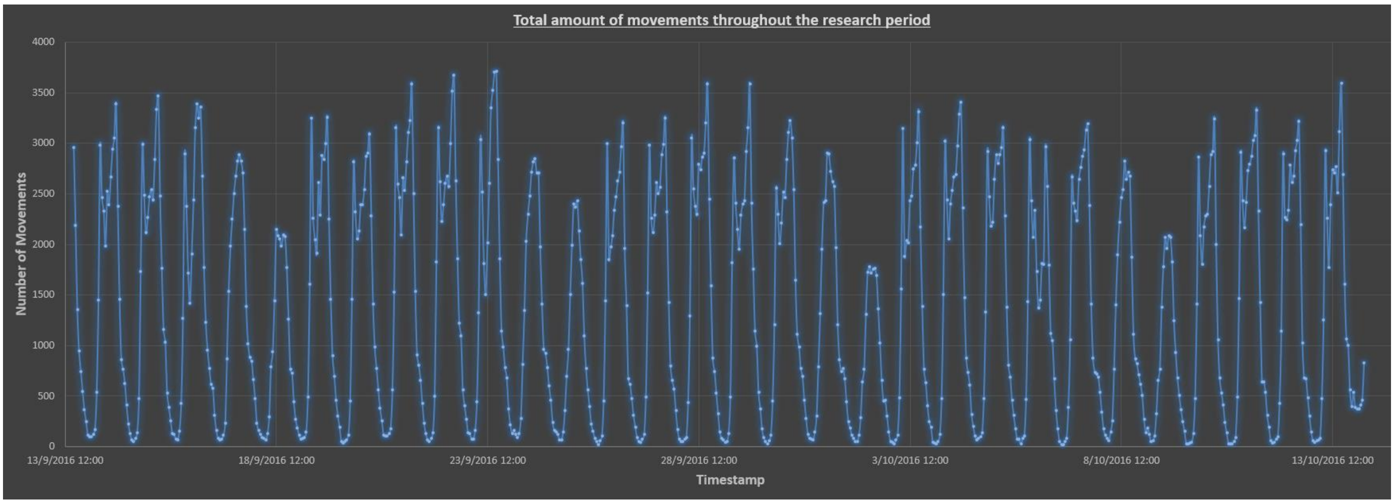


Figure 31: Total (number) of movements throughout the research period

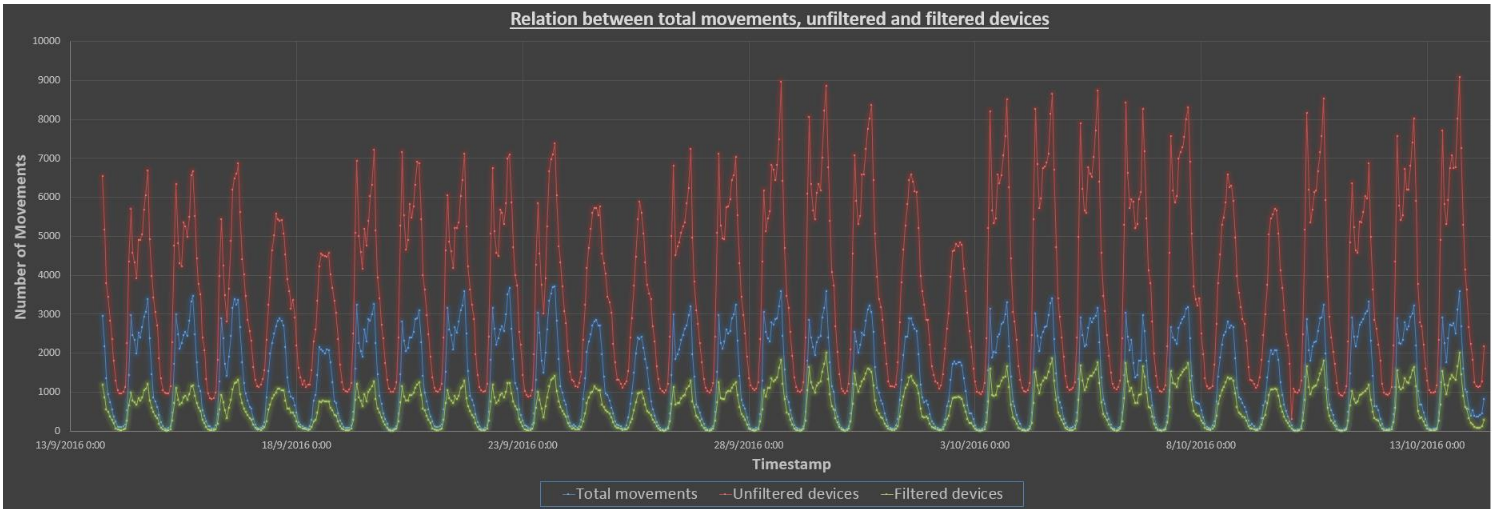


Figure 32: Total number of identified movements (blue color), unfiltered (red color), and filtered devices (light green color) throughout the research period

This expected close relationship is verified in Figures 33 and 34, which show the outcome of a linear regression between the total amount of identified movements and the unfiltered and filtered devices respectively. As mentioned in the previous section, there are about 1200 static devices in the research area while approximately two movements for each filtered device are observed. Furthermore, the systematic and similar analysis procedure is verified by the fact that the computed linear equation fits very well to the related datasets and it is confirmed by the very high values of the R-squared indicator.

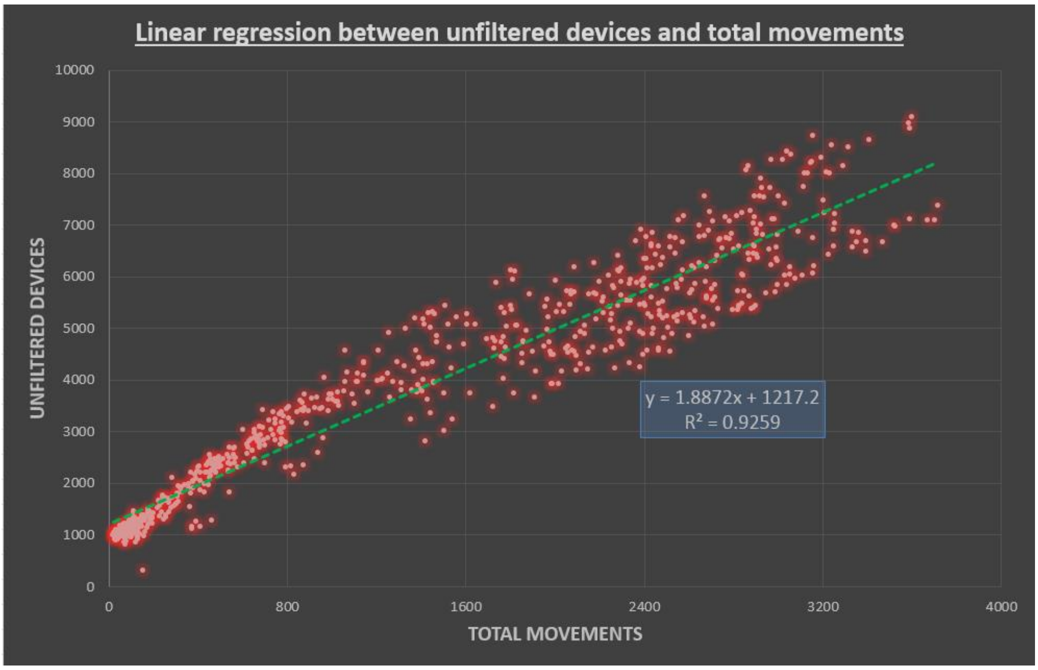


Figure 33: Linear regression between unfiltered devices and total number of movements

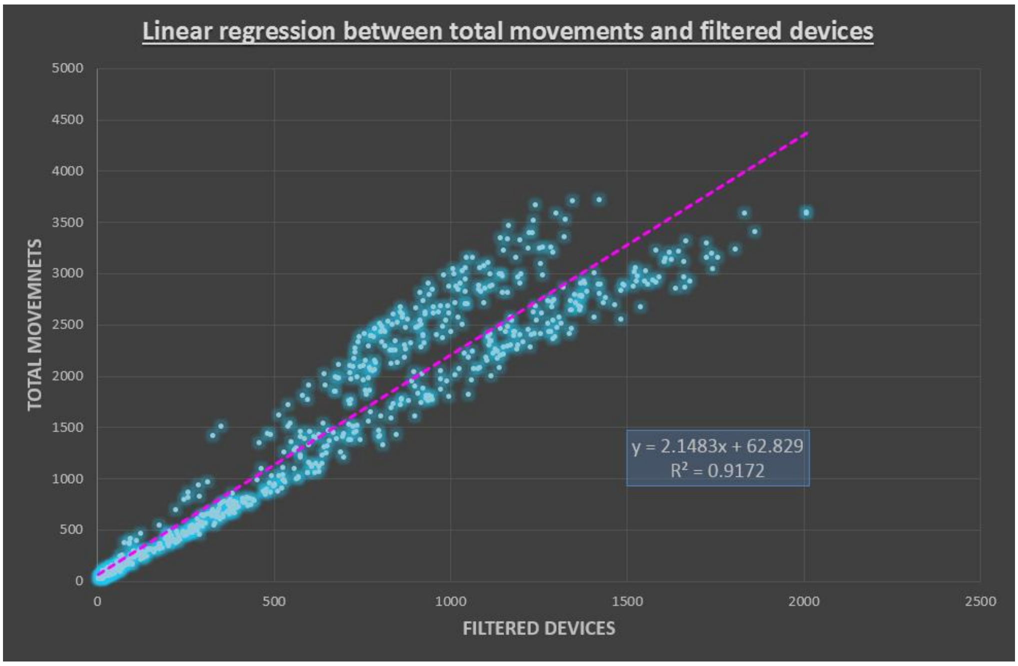


Figure 34: Linear regression between filtered devices and total number of movements

Now focusing on road modality throughout the research period, Figure 35 shows the relevant hourly number for each user category in the whole area while Figure 36 also includes the total number of movements. Despite the scale of the diagrams being quite small, a similar behavior of the contained variables can be observed. Due to the different scale of values, the positive monotonic transformation of logarithm is used in Figure 37 as a way to reduce the scale of the values without changing the relationship between the variable. Thus, it is easier to understand that all user categories have peaks and recessions at the same time, which is also proved from the very high and positive correlation values between them, as these are detailed in Table 3. Finally, Figure 38 presents, for example, the distribution of the total daily movements of each user category on Wednesday 14/09/2016 verifying that, despite the different total number of movements for each category, they are distributed in a similar way.

Correlations	Values
Pedestrians -Vehicles	0.965
Pedestrians - Bicyclists	0.981
Bicyclists - Vehicles	0.980

Table 3: Correlation values for the different sets of user categories

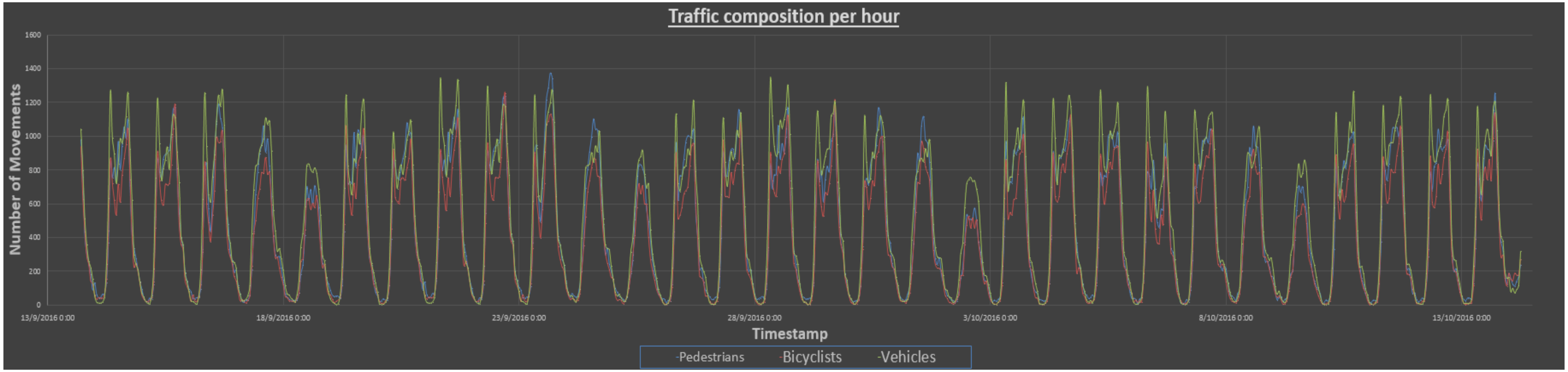


Figure 35: Hourly number of pedestrians (blue), bicyclists (red), and vehicles (light green) in the area throughtout the research period

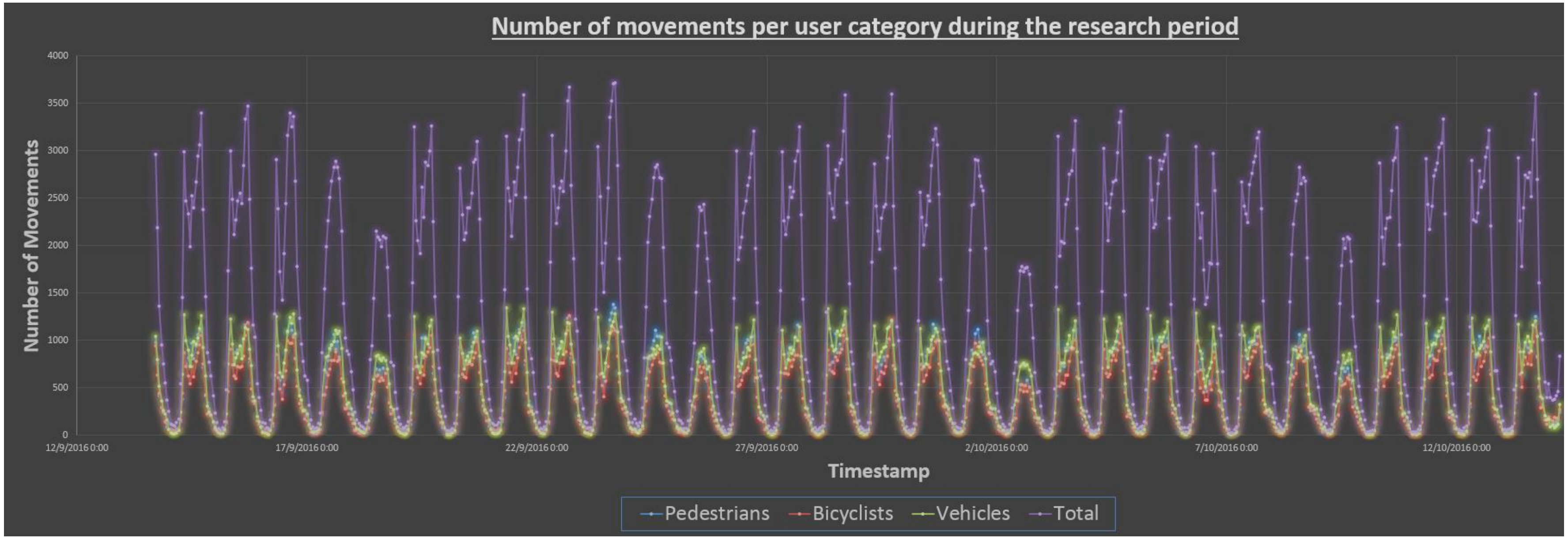


Figure 36: Hourly number of pedestrians (blue), bicyclists (red), vehicles (light green) and total movements (purple) in the area throughtout the research period

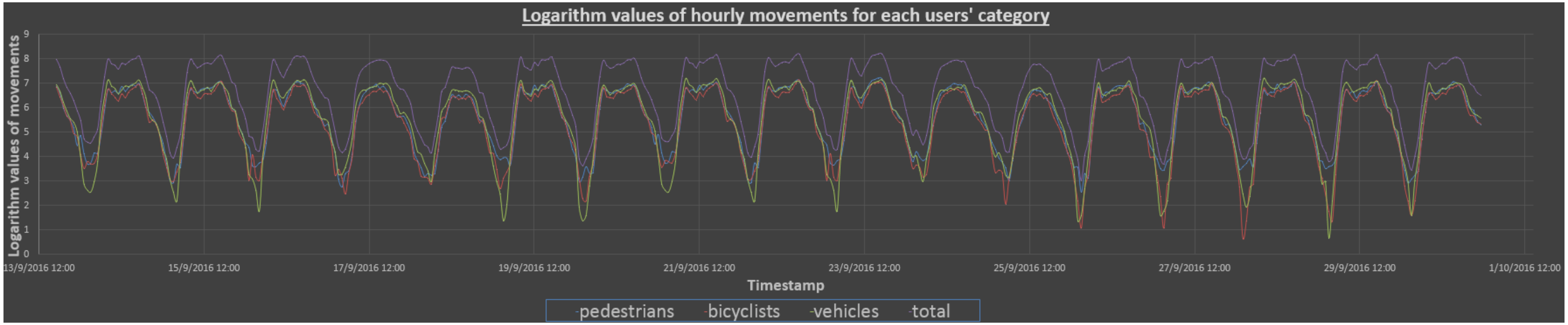


Figure 37: Logarithm values of hourly pedestrian movement (blue), bicyclists (red), vehicles (light green) and total movements (purple) in the area for the first half of the research period

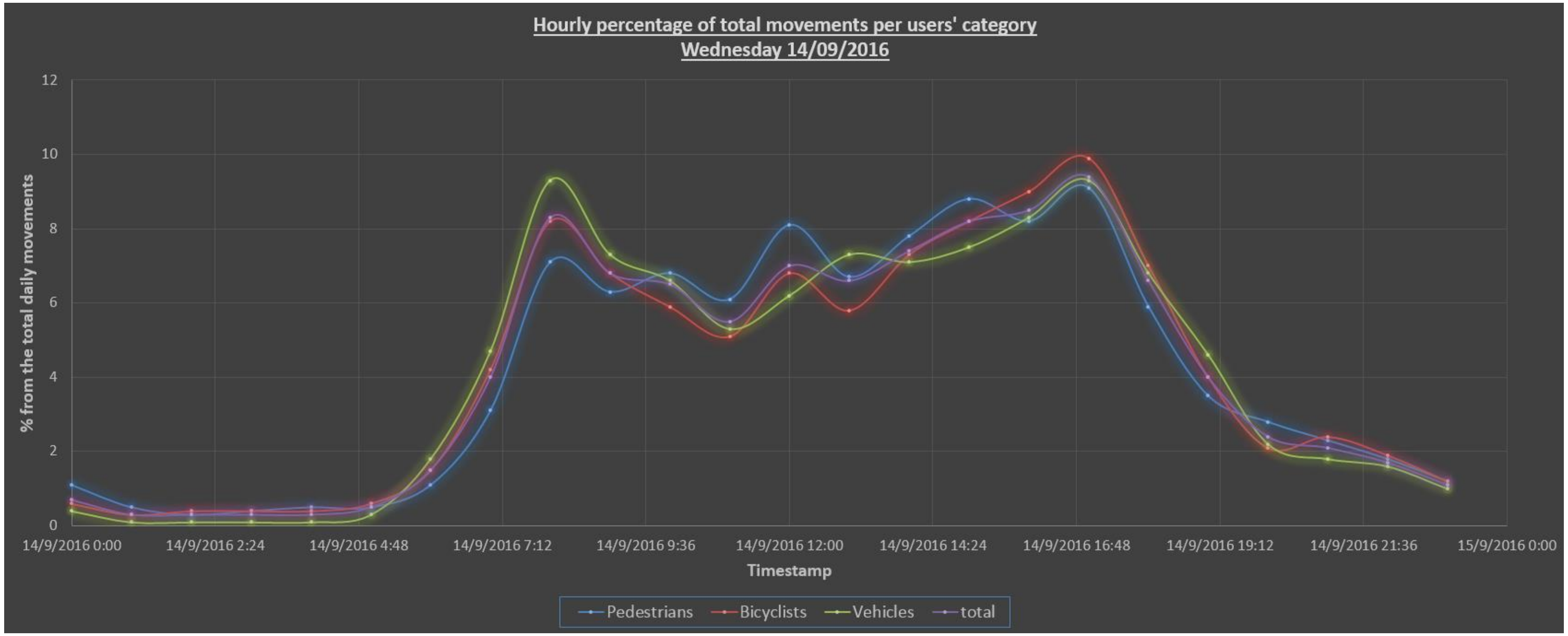


Figure 38: Distribution of total daily movements of each user category for Wednesday 14/09/2016

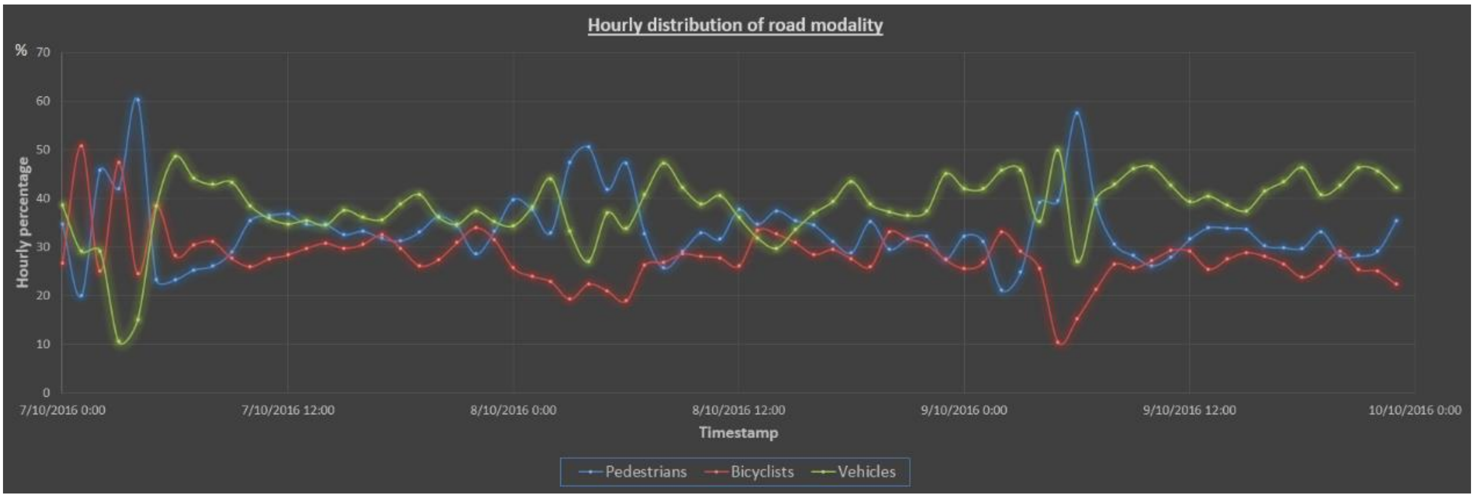


Figure 39: Hourly distribution of road modality for a period of three days

In addition to comparing the total numbers of the three categories, their percentages per hour can also be computed. Figure 39 visualizes the distribution of hourly road modality for a period of three days. Based on this Figure, it can be pointed out that in some timeslots there is a signifant difference of the relevant percentages and one category represents more than the half of the total movements in the research area. However, in Figure 40 it is obvious that this phenomenon occurs only in periods where the total number of movements is quite small. Thus, despite the fact that the percentage of pedestrians, for instance, is sometimes almost equal to 80% due to the very low total number of movements, the real difference between pedestrian and bicyclist movement is minimal. Contrariwise, smaller percentage differences are observed during rush hours. In general, it can be mentioned that the higher the total number of movements in the area the smaller the dispersion of the percentages.

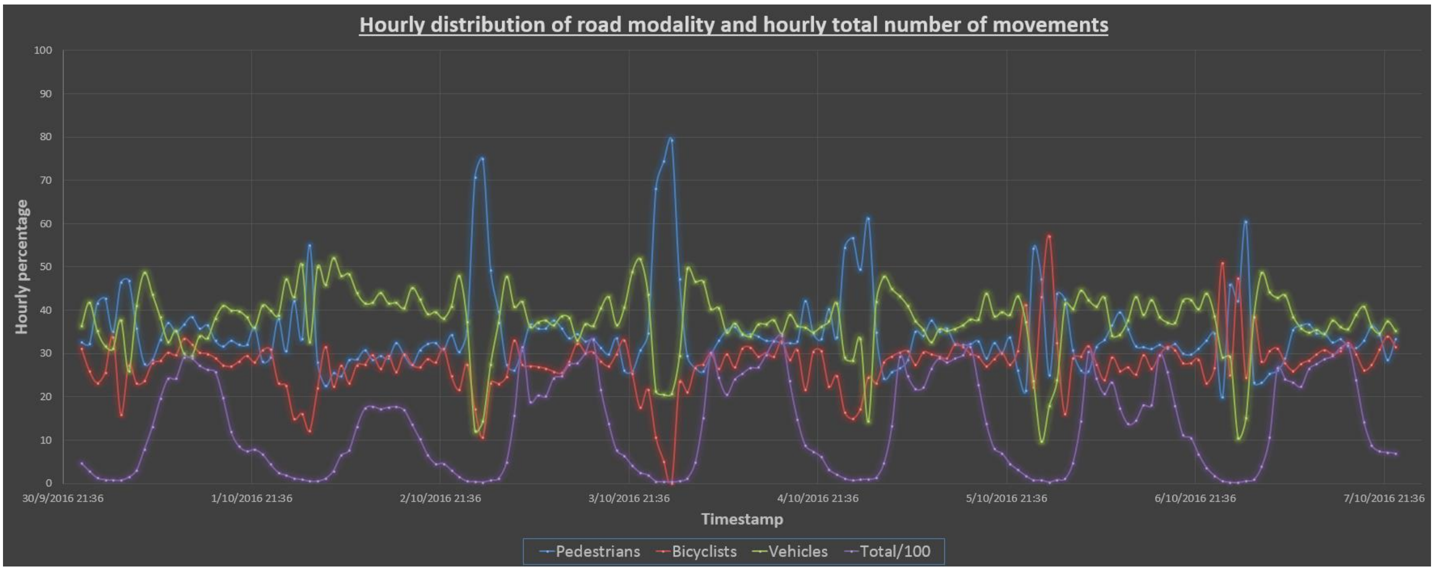


Figure 40: Hourly distribution of road modality and the relevant total number of movements

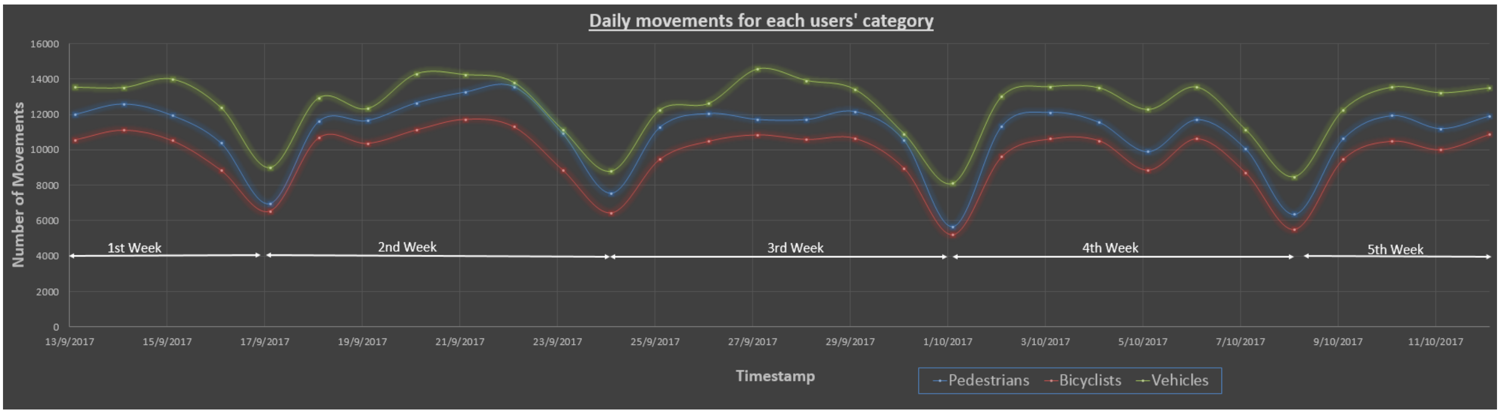


Figure 41: Daily number of movements for each investigated category

If road modality is computed in total for each day, it is easier to come to an overall conclusion about the relationship between the three investigated categories. Figure 41 presents the number of movements for each user category on a daily basis throughout the data collection period. The quite similar behavior in the course of time can be easily observed in this diagram as well, despite the differences in the scale of values on diffent days. Moreover, the order of the categories based on their values remains the same, with "vehicles" being in the first place followed by "pedestrians" and "bicyclists". The same hierarchy is also presented in Figure 42 which includes the percentages of the total daily number of movements. Finally, in case an overall outcome for road modality is required, 37%, 33%, and 30% percentages can be used for "vehicles", "pedestrians", and "bicyclists" respectively. However, it is important to mention that the above-mentioned percentages are directly related to the relevant percentage of the users of each category which have enabled the Wi-Fi functionality; a topic which is described in detail in Chapter 5: Data validation.

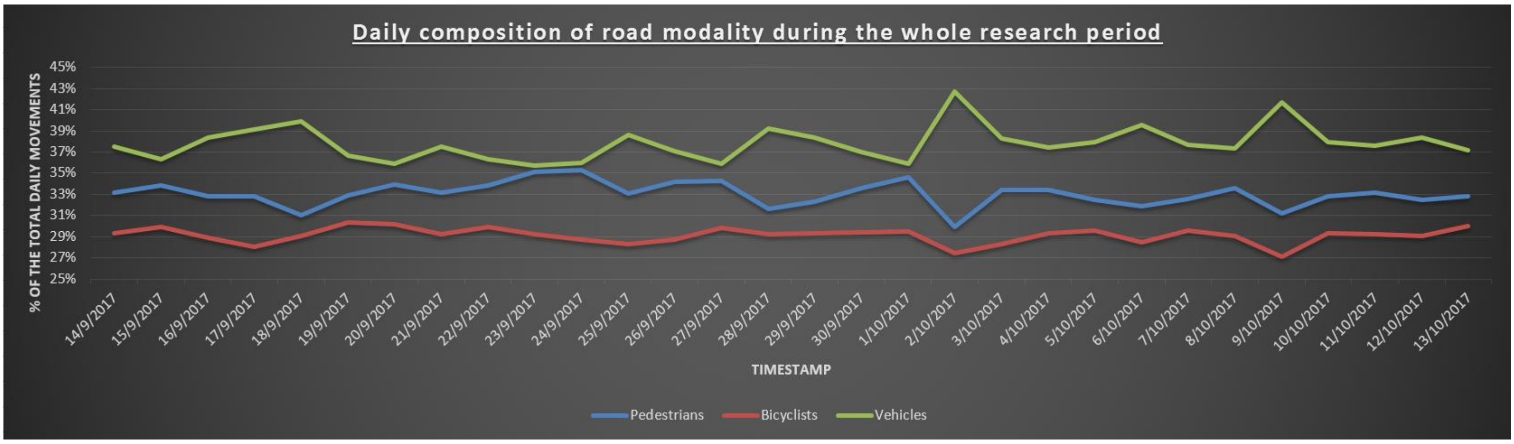


Figure 42: Daily percentage for each investigated category throughout the research period

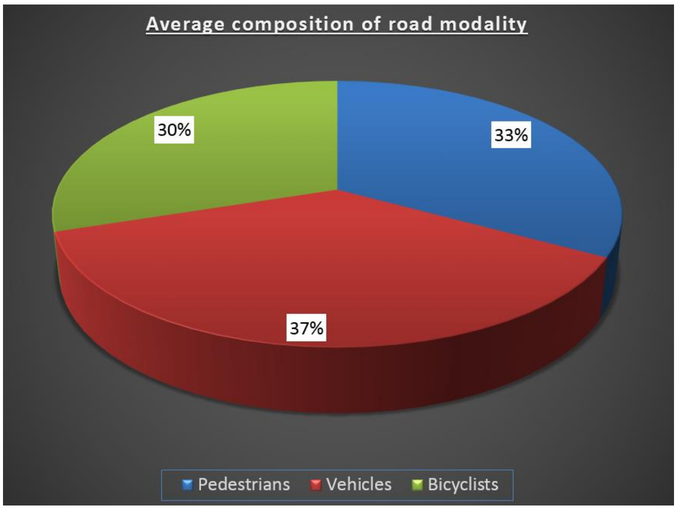


Figure 43: Overall road modality for the research period

4.5 Computation of movement patterns

In this section, the identification of movement patterns in the research area will be investigated. One of the main advantages of the Wi-Fi monitoring system is that, apart from counting the devices which pass through the area of each sensor and characterizing them based on moving speed and street-uses criteria, it is possible to look at the bigger picture, that is people's movement behavior in that area. Using the each device's unique indicator, their total movement in space can be researched and thus movement patterns can be identified.

In this work, three cases of movement patterns are investigated. First of all, as the sensors were placed in the intersections of streets, the movements of the three user categories in each street, namely the area between two sensors, can be analyzed. Moreover, the movement of users which were scanned by three and four sensors in a period of one hour are researched separately, leading to the identification of longer patterns and common areas throughout the day.

	Pedestrians	Bicyclists	Vehicles
Monday-Thursday	11729	10418	13280
Friday	12348	10782	13685
Saturday	10474	8853	11393
Sunday	6641	5929	8595

Table 4: Average number of movements for each user category on different days of week

Tables 5, 6, and 7 illustrate the percentages of the three above-mentioned categories of patterns for pedestrians, bicyclists, and vehicles respectively. As it is clear, the percentage of patterns for each user category is almost the same on all days of the week, with 70% of users having been scanned by two sensors, about 25% by three, and only 5% of them by four sensors. As Table 4 shows, unlike the similarity of the pattern percentages during the week, significant differences between the absolute numbers of movements are observed. Given that, it was decided to use daily percentages instead of daily numbers, in order to facilitate the comparison of the outcomes from different days with regard to the relationship between different streets as well as the identification of the most crowded of them in the research area. However, the charts with absolute values instead of percentages can be found in the appendices.

Patterns	Monday-Thursday %	Friday %	Saturday %	Sunday %	Overall %
2	74.2	74.6	73.4	73.2	73.9
3	20.8	20.5	21.4	21.6	21.1
4	4	3.9	4	4.1	4.0
Other	1	1	1.2	1.1	1.1

Table 5: Percentage of each movement pattern for pedestrians

Patterns	Monday-Thursday %	Friday %	Saturday %	Sunday %	Overall %
2	70.5	70.4	69.8	69.4	70.0
3	23.7	23.8	24.4	24.1	24.0
4	4.8	5.1	5.1	5.5	5.1
Other	1	0.7	0.7	1	0.9

Table 6: Percentage of each movement pattern for bicyclists

Patterns	Monday-Thursday %	Friday %	Saturday %	Sunday %	Overall %
2	67.4	67.2	67.3	66.5	67.1
3	25.8	25.9	26.1	26.3	26.0
4	5.7	5.9	5.5	6.3	5.9
Other	1.1	1	0.9	0.9	1.0

Table 7: Percentage of each movement pattern for vehicles

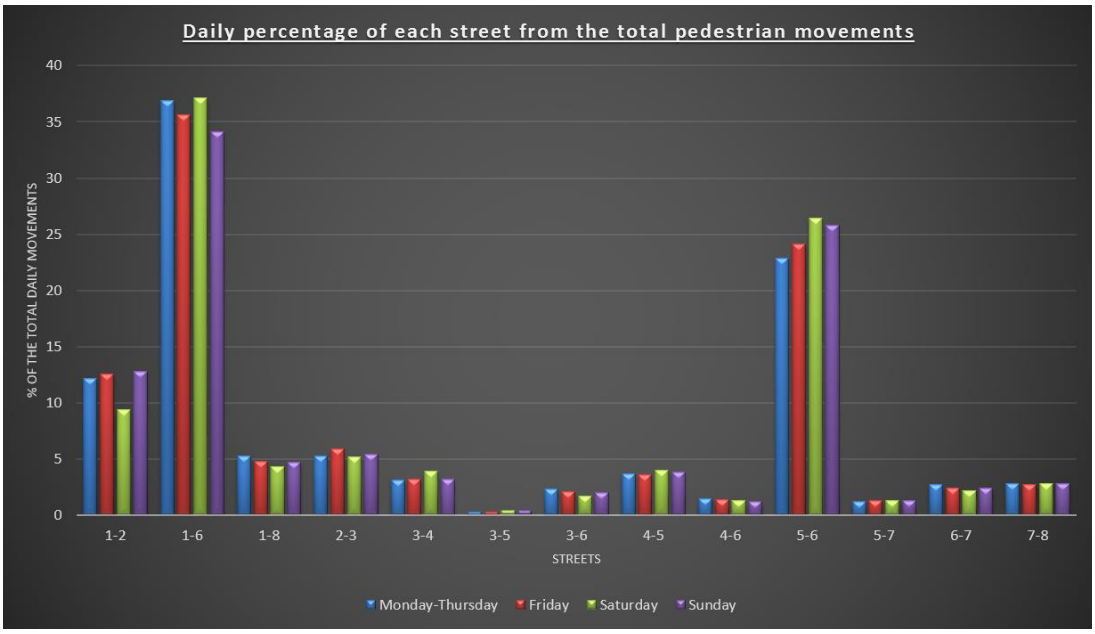


Figure 44: Daily percentage of pedestrian movements for each street

Figures 44, 46, and 48 present the daily percentage of pedestrians, bicycles, and vehicles respectively in each street throughout the week. Based on these diagrams and in combination with Table 4, it is observed that despite the significant changes in the total number of movements during the week, the relevant percentages remain almost the same. For instance, the total number of pedestrian movements on Fridays is almost double that on Sundays. However, the relevant percentages for each street are nearly the same. Hence, it can be stated that the behavior of people does not fluctuate throughout the week with regard to the way they move in space and the streets they "prefer". Also, it should be mentioned that the investigation about the changes of the absolute number of movements is described in detail in Section 4.5 of this Chapter: Occupancy patterns.

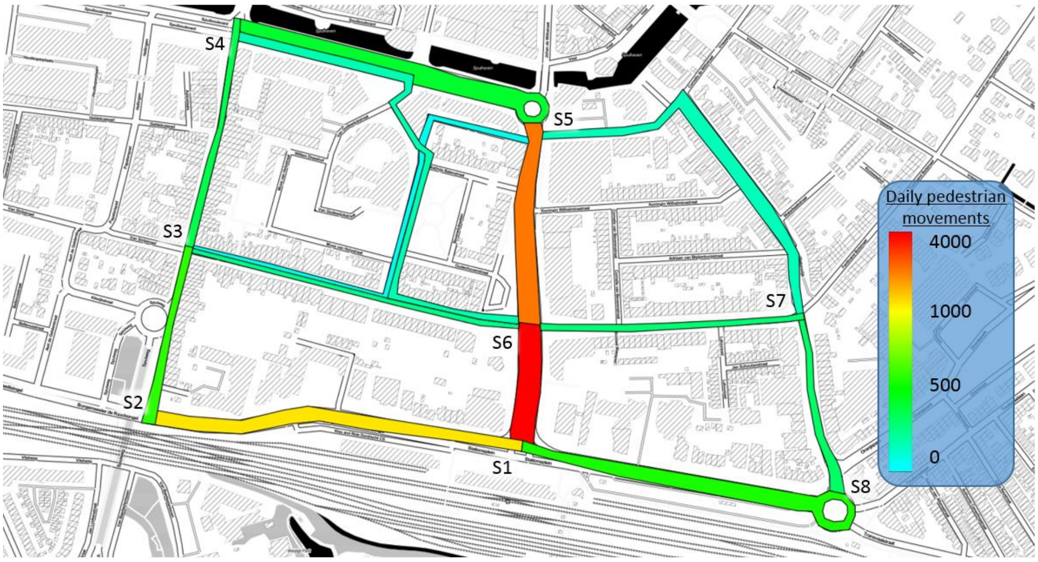


Figure 45:Visualization of daily number of pedestrian movements for each street

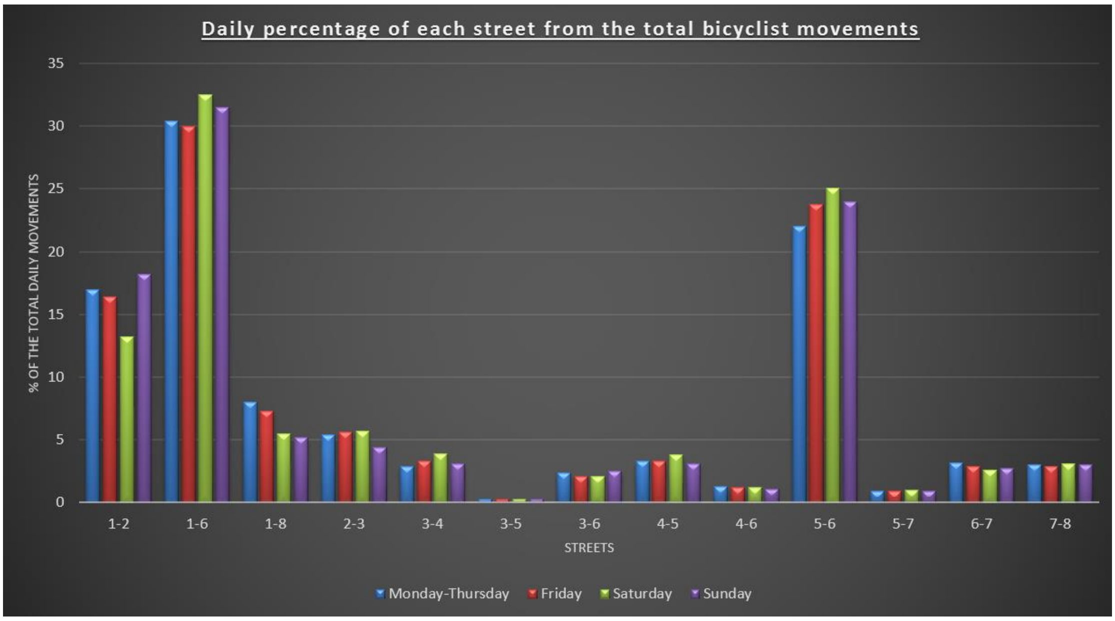


Figure 46: Daily percentage of bicyclist movements for each street

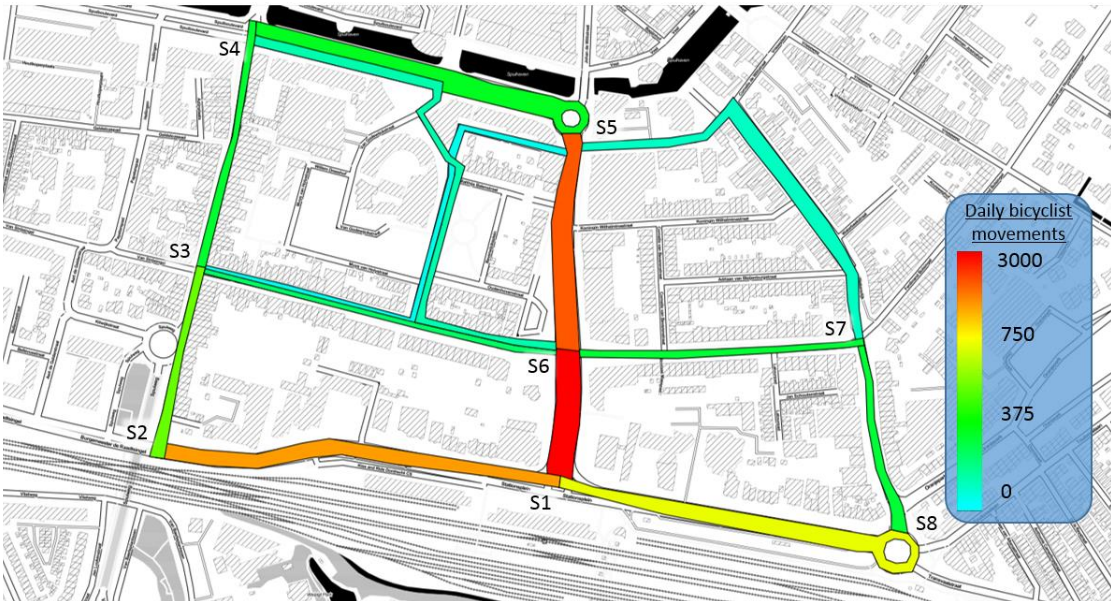


Figure 47:Visualization of daily number of bicyclist movements for each street

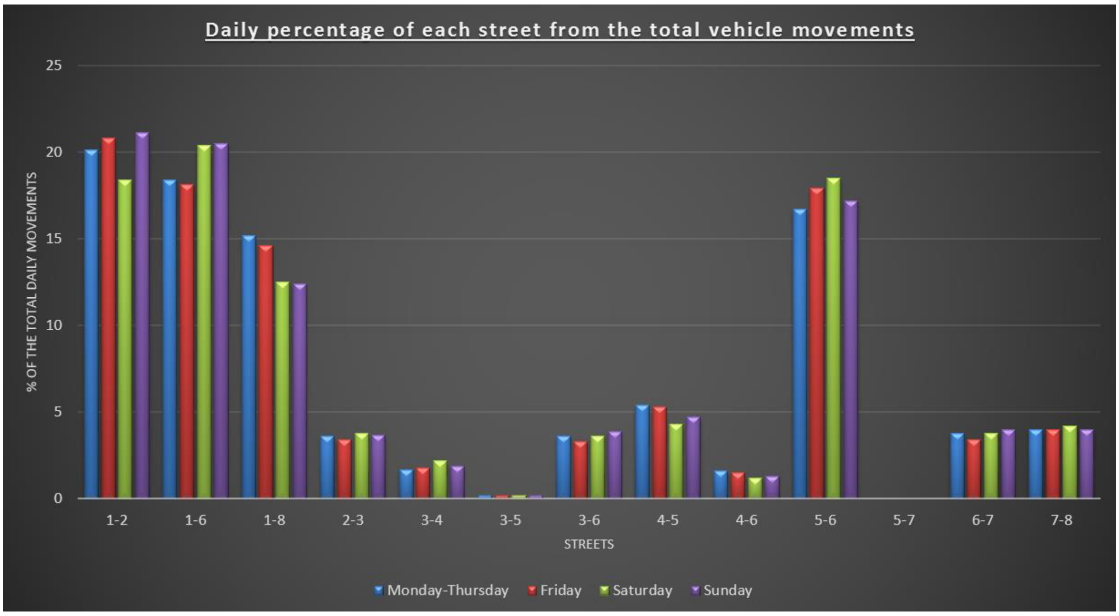


Figure 48: Daily percentage of vehicle movements for each street

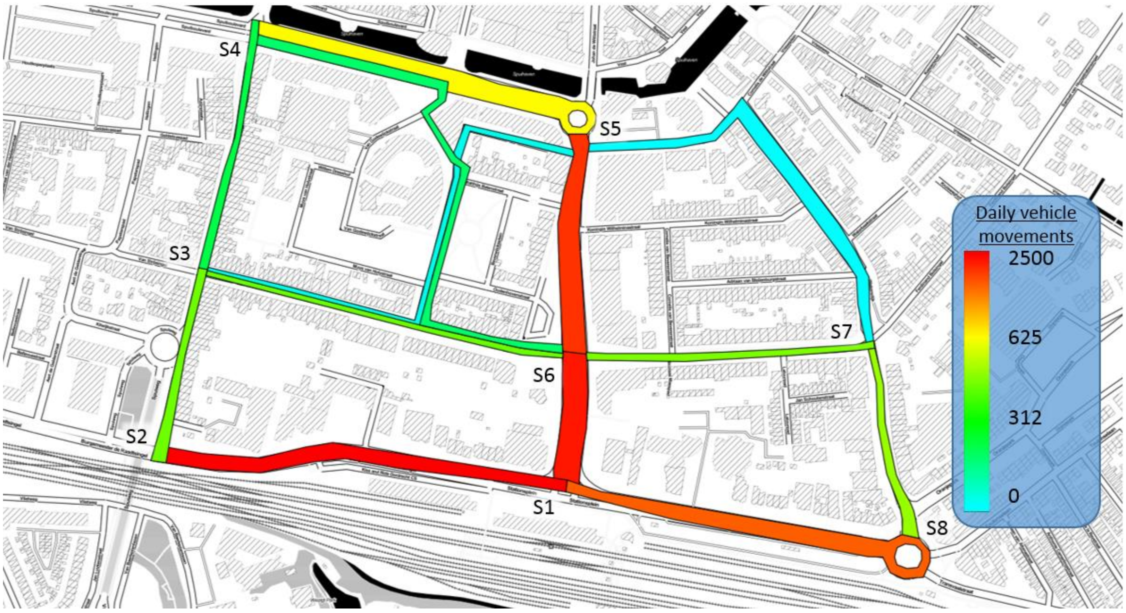


Figure 49:Visualization of daily number of vehicle movements for each street

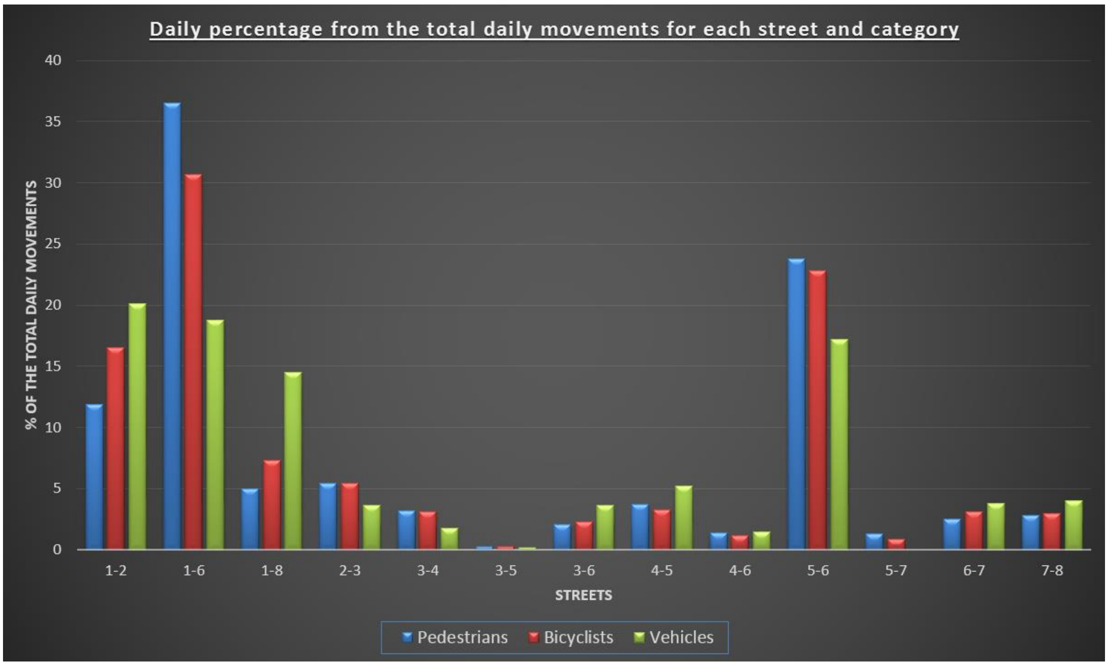


Figure 50: Overall percentage of the daily movements of each user category in each street

Taking into account the stable percentage of each street for each user category during the week, it is possible to summarize the outcomes on a weekly basis, as visualized in Figure 50. Based on this chart, it is clear that the relationship between the streets of the research area remains the same. For example, the percentages of all the categories of streets 1-6 are higher than those of streets 1-8. However, there are notable differences in the road modality of each road separately and in the main category of users who use each street. Thus, in some cases, such as streets 1-6 and 5-6, the majority of users are pedestrians followed by bicyclists and vehicles. On the other hand, streets 1-2 and 1-8 are mainly occupied by vehicles unlike streets 2-3 and 3-4, in which bicyclists prevail. At this point, it is important to mention that the absolute number of movements which are presented in the diagrams and maps in this study, such as those on Figures 51 and 49 respectively, does not represent the total set of the real number of movements but only those identified by the Wi-Fi monitoring system.

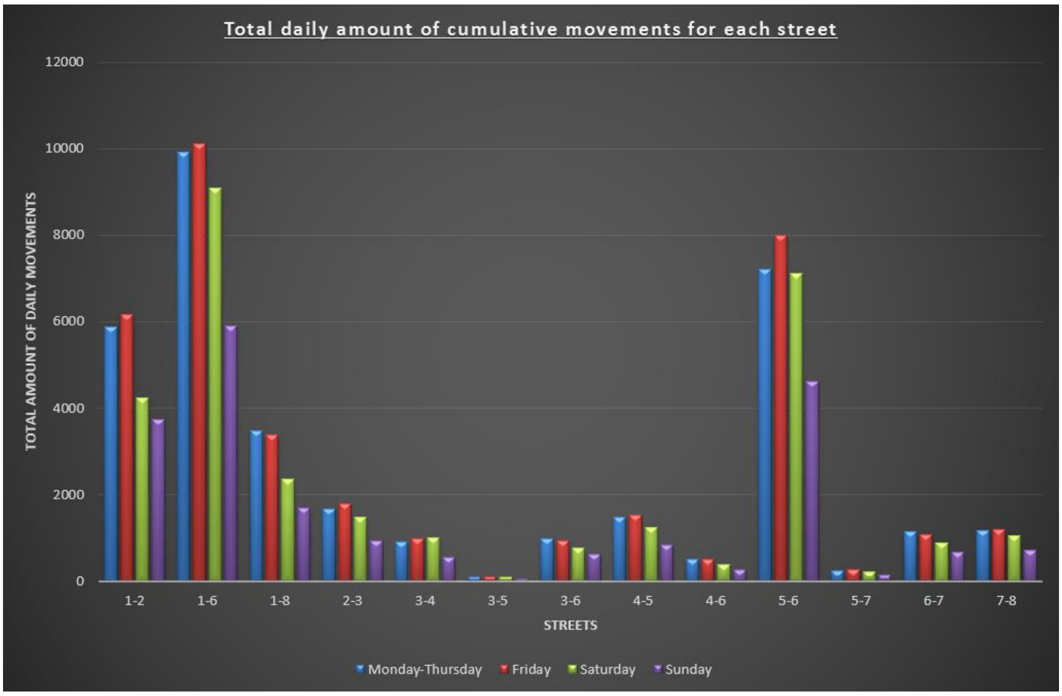


Figure 51: Daily amount of cumulative movements for each street

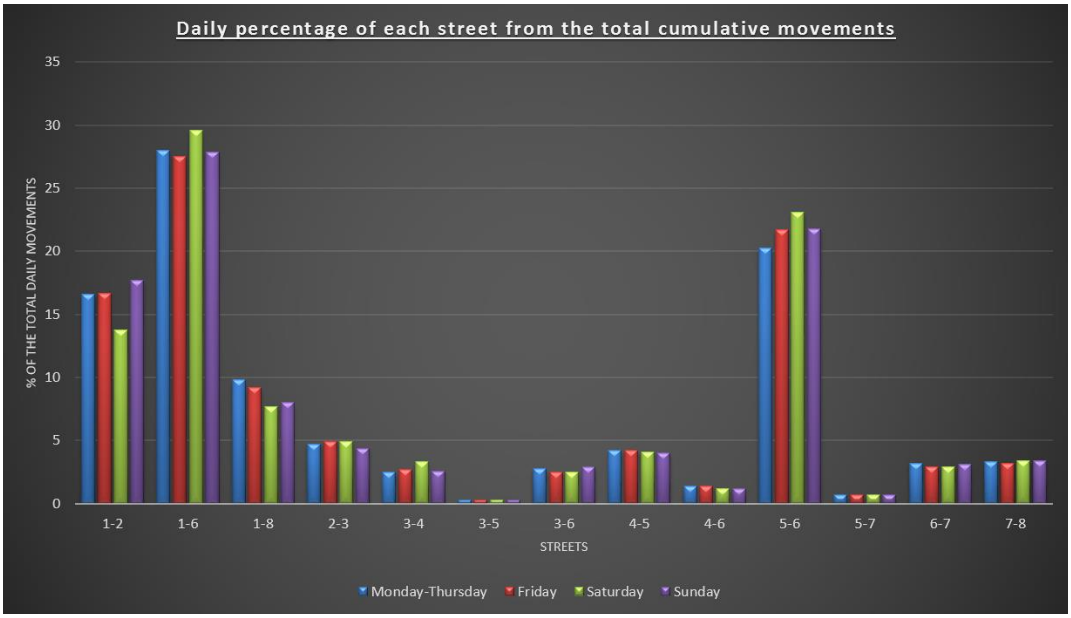


Figure 52: Daily percentage of cumulative movements for each street

As a means to identify the order of the most crowded streets in the area, the datasets were aggregated so as to get a cumulative indicator about the total number of movements. Figures 51 and 52 present the corresponding diagrams about the absolute values and percentages respectively. Despite the significant differences in the absolute values during the week, the percentages of user accumulation in each street are similar for different days. Thus, based on the total number of movements an overall conclusion about the hierarchy in streets can be drawn. Almost half of the movements were detected in streets 1-6 and 5-6 which connect the city's train station with its downtown. Moreover, 25% of the daily movements occurred in streets 1-2 and 1-8, which again connect the train station with the East and West part of the city respectively, while the remaining percentage is distributed in other streets of the area.

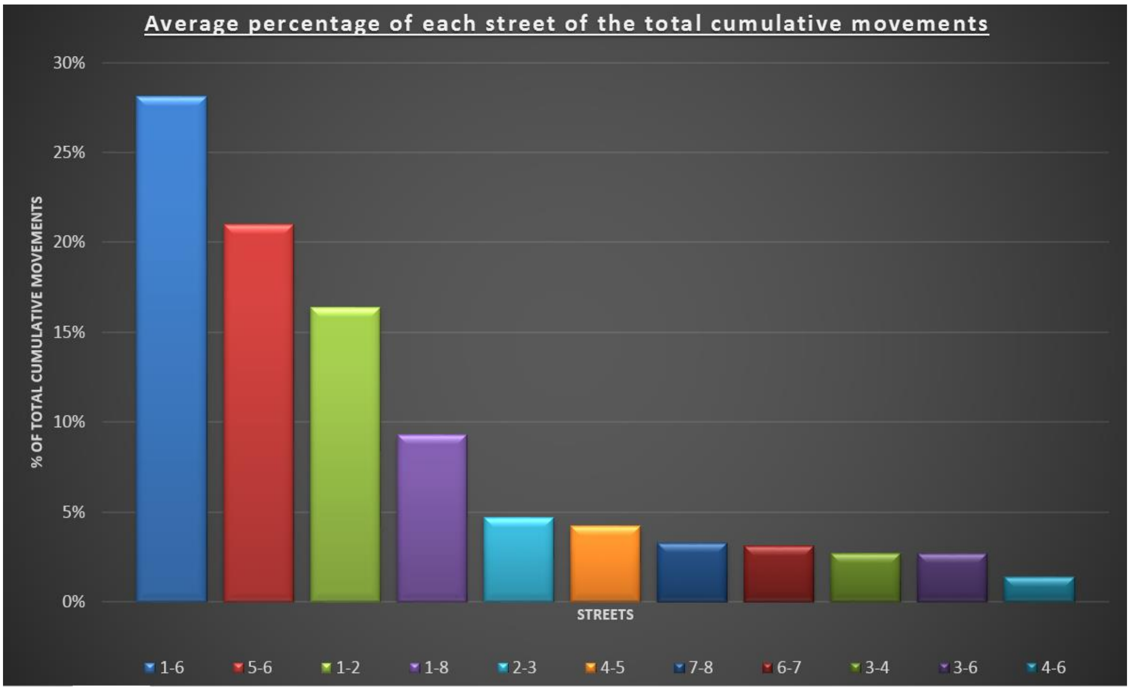


Figure 53: Average percentage of each street about the corresponding number of cumulative movements

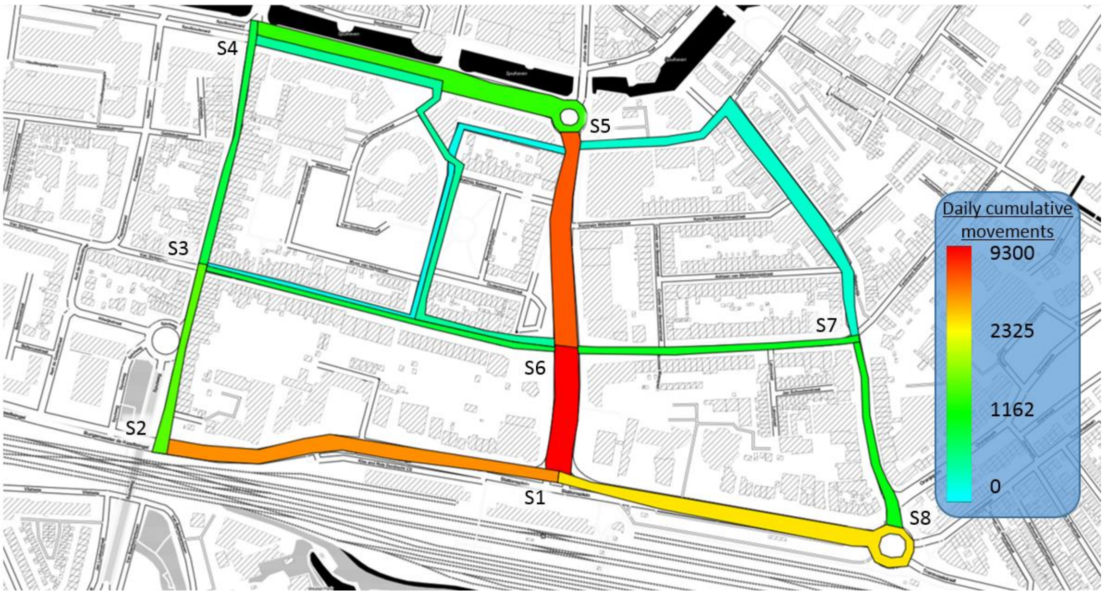


Figure 54: Visualization of average number of each street about the corresponding number of cumulative movements

Following the movement analysis ofeach street in the research area, the same procedure can be applied for the patterns between three sensors. Table 8 shows the most frequent movement patterns for each user category from Monday to Thursday. For each pattern there is an average number of movements per day, the percentage of this pattern from the set of this kind of pattern as well as the percentage from the total set of daily movements. For instance, the patterns from sensor 1 to sensor 6 and then to sensor 5 and vice versa constitute almost 39% and 30% of the pedestrian movements for this kind of pattern (relative percentage). Thus, it is clear that they are the two main patterns of the combinations of three locations. However, these movements corrенspond to the percentages of 8% and 6% of total movements (absolute percentage).

Figures 55, 57, and 59 present the most frequent patterns for each user category separately, showing the average number of movements per day as well as their relative and absolute percentages. Tables for the other days of the week, similar to Table 8, can be found in the appendices.

As it becomes obvious, in both figures, the most frequently used patterns are those which connect the train station and the city centre of Dordrecht, regardless of the user category. Differences between user categories can only be identified for the rest of the movements. Streets between sensors 2, 3, and 4 are used as the third pattern by pedestrians and bicyclists, unlike vehicles, which move more in the streets between sensors 2 and 8 that connect the East and West part of the city going past the train station.

Pedestrians Patterns	Daily Amount	Relative% (absolute%)	Bicyclists Patterns	Daily Amount	Relative% (absolute%)	Vehicles Patterns	Daily Amount	Relative% (absolute%)
165	942	38.6 (8.0)	**165**	771	31.2 (7.4)	**561**	675	19.7 (5.1)
561	726	29.8 (6.2)	**561**	683	27.7 (6.6)	**165**	542	15.8 (4.1)
612	64	2.6 (0.5)	**234**	107	4.3 (1.0)	**218**	373	10.9 (2.8)
456	57	2.3 (0.5)	**218**	89	3.6 (0.9)	**812**	358	10.4 (2.7)
234	57	2.3 (0.5)	**812**	78	3.1 (0.7)	**456**	148	4.3 (1.1)
216	49	2.0 (0.4)	**216**	74	3.0 (0.7)	**236**	143	4.2 (1.1)
761	49	2.0 (0.4)	**461**	63	2.5 (0.6)	**216**	133	3.9 (1.0)
432	49	2.0 (0.4)	**612**	52	2.1 (0.5)	**816**	112	3.3 (0.9)
461	34	1.4 (0.3)	**618**	44	1.8 (0.4)	**167**	112	3.3 (0.9)
654	30	1.2 (0.2)	**432**	41	1.6 (0.4)	**781**	97	2.8 (0.7)

Table 8: The most frequently used movement patterns of 3 sensors for each user category for Monday to Thursday

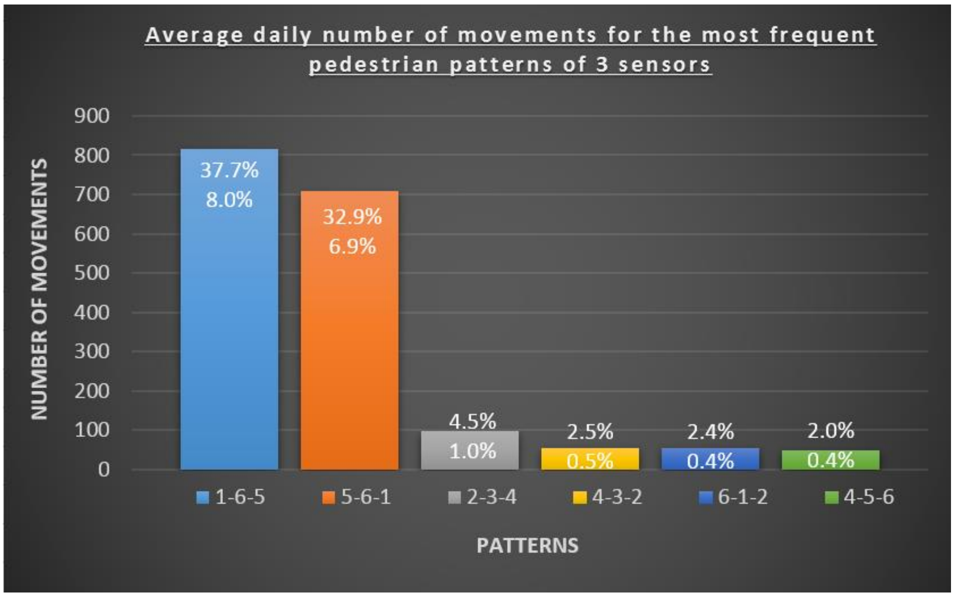


Figure 55: Average daily number of pedestrian movements for the most frequently used patterns of 3 sensors. In each column, the corresponding relative (above) and absolute percentage (below) of this number is also included.

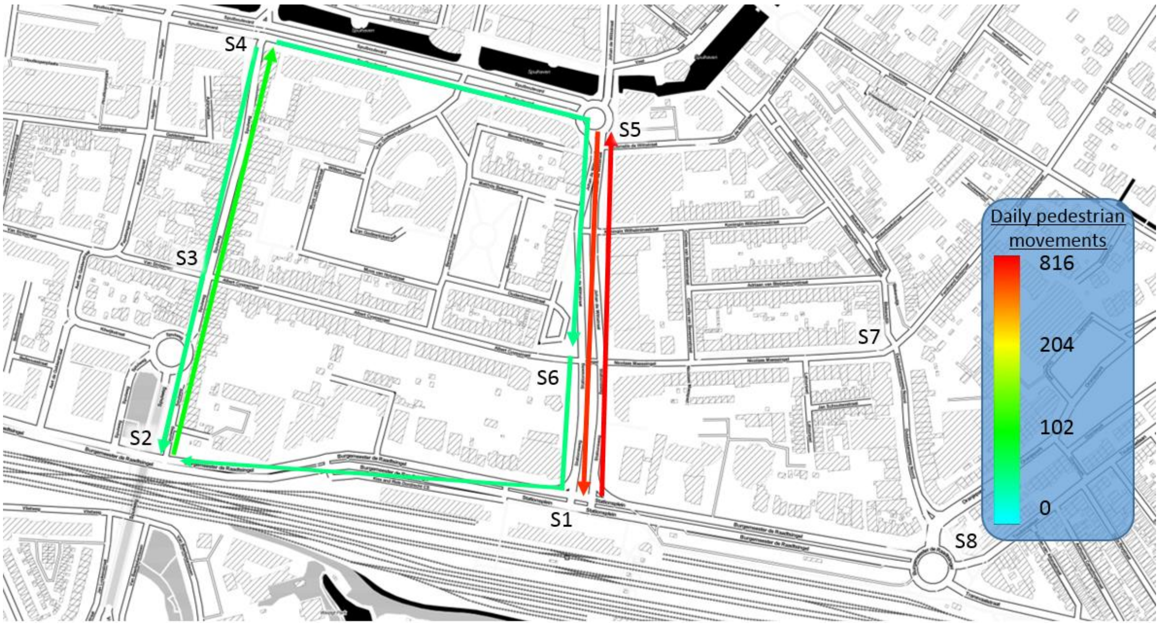


Figure 56: Visualization of average daily number of pedestrian movements for the most frequently used patterns of 3 sensors.

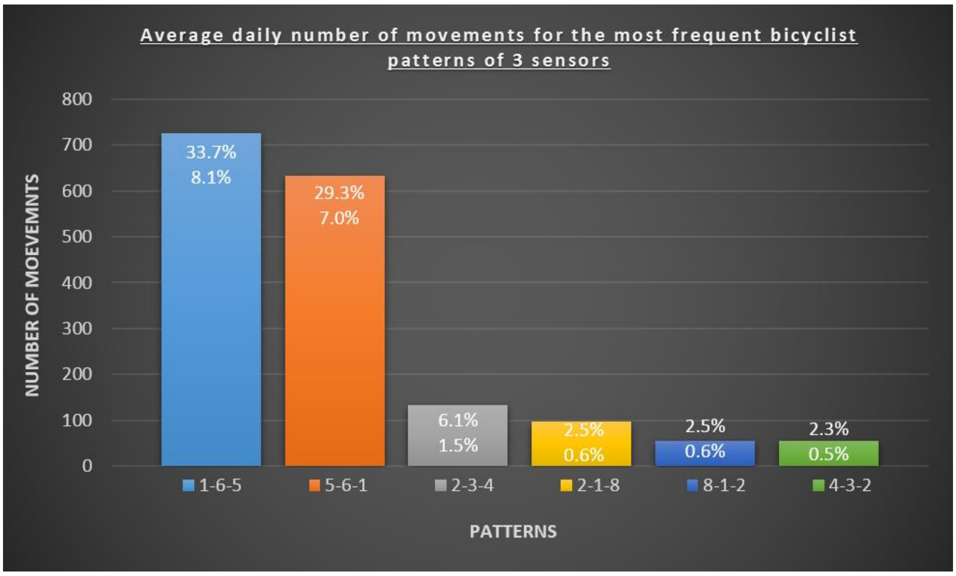


Figure 57: Average daily number of bicyclist movements for the most frequently used patterns of 3 sensors. In each column, the corresponding relative (above) and absolute percentage (below) of this number is also included.

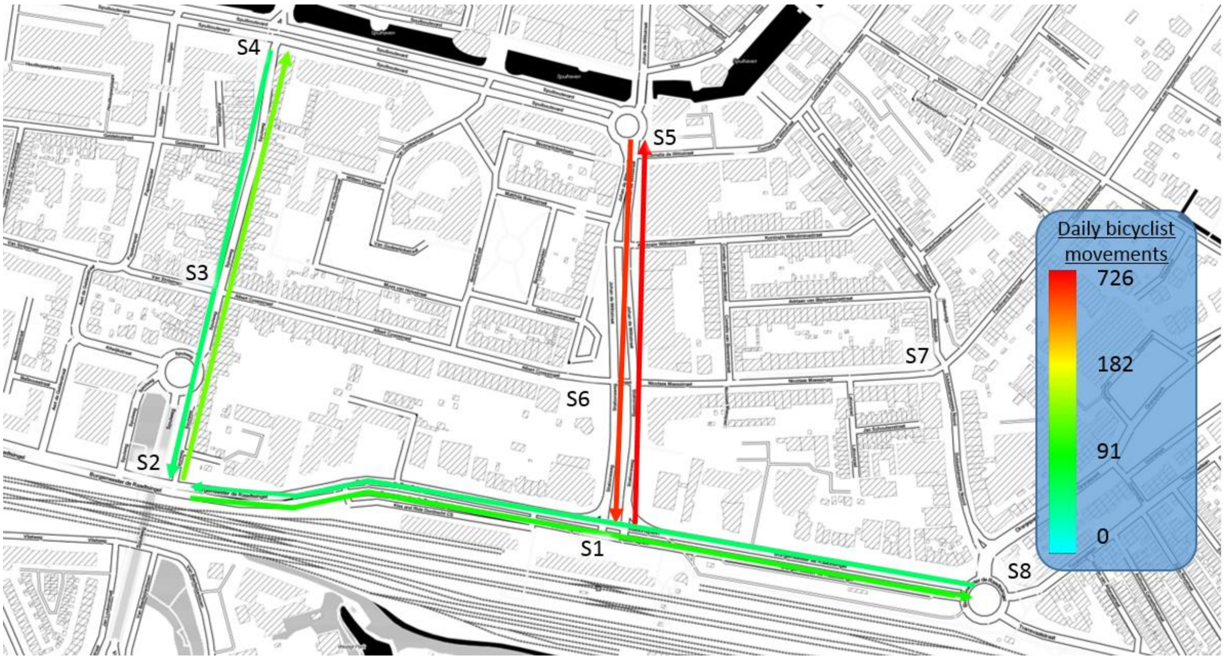


Figure 58: Average daily number of bicyclist movements for the most frequently used patterns of 3 sensors.

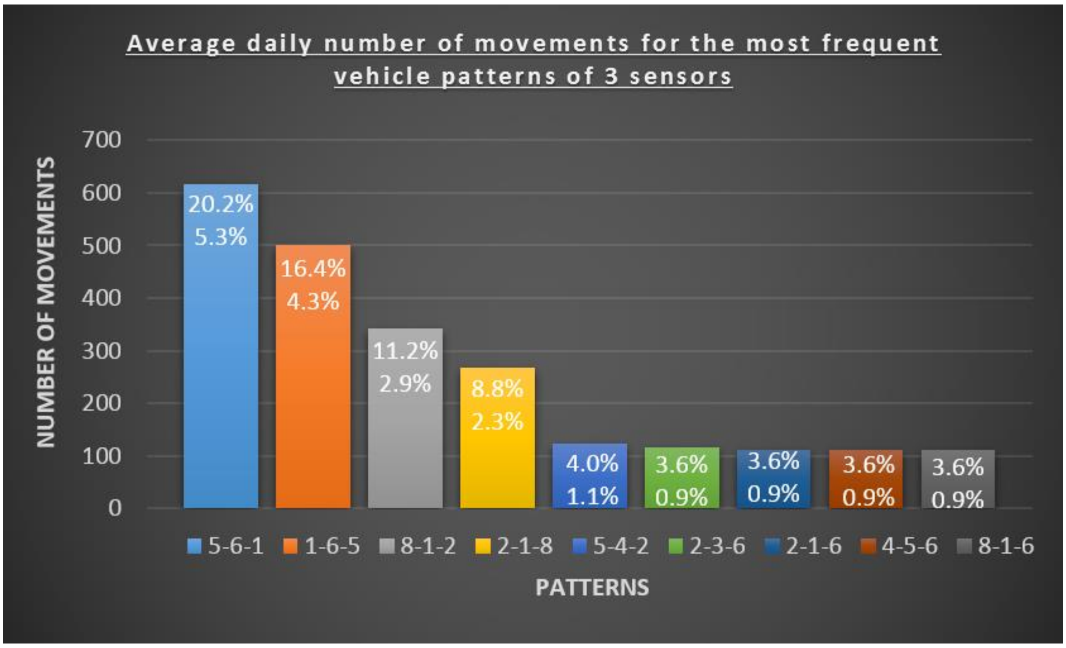


Figure 59: Average daily number of vehicle movements for the most frequently used patterns of 3 sensors. In each column, the corresponding relative (above) and absolute percentage (below) of this number is also included.

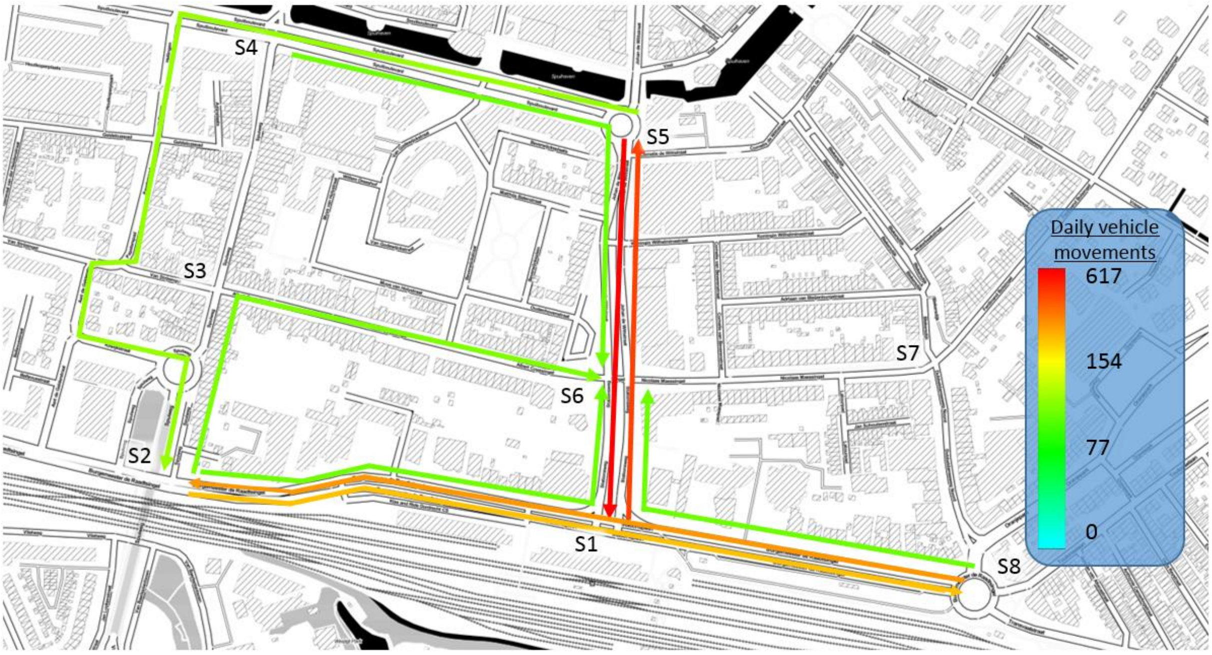


Figure 60: Visualization of average daily number of vehicle movements for the most frequently used patterns of 3 sensors.

The outcomes of the investigation of the last category of movement patterns, with combinations of four sensors, are similar to the ones described above. As Figures 61 to 66 and the following Table show, it is obvious that the most frequent patterns are again related to sensors 1, 6, and 5, which connect the train station with the city centre, regardless of the user category. These sensors are always included in all movement patterns while the last sensor is alternated with sensor 4, whose patterns constitute the major ones for all the user categories, sensor 8, and sensor 2.

Pedestrians Patterns	Daily Amount	Relative% (absolute%)	Bicyclists Patterns	Daily Amount	Relative% (absolute%)	Vehicles Patterns	Daily Amount	Relative% (absolute%)
4561	141	30.0 (1.2)	**1654**	167	33.3 (1.6)	**4561**	208	27.4 (1.6)
1654	94	20.0 (0.8)	**4561**	90	18.1 (0.9)	**1654**	173	22.8 (1.3)
5618	66	14.0 (0.6)	**5618**	76	15.3 (0.7)	**5618**	88	11.7 (0.7)
8165	38	8.0 (0.3)	**8165**	63	12.5 (0.6)	**8165**	61	8.1 (0.5)
2165	28	6.0 (0.2)	**2165**	42	8.3 (0.4)	**2345**	58	7.6 (0.4)
2367	28	6.0 (0.2)	**5612**	28	5.6 (0.3)	**2167**	35	4.6 (0.3)
2167	19	4.0 (0.2)	**2345**	21	4.2 (0.2)	**2367**	35	4.6 (0.3)
8754	19	4.0 (0.2)	**2167**	14	2.8 (0.1)	**5612**	27	3.6 (0.2)
5612	19	4.0 (0.2)				**7812**	27	3.6 (0.2)
2345	19	4.0 (0.2)				**7612**	15	2.0 (0.1)

Table 9: The most frequently used movement patterns of 4 sensors for each user category for Monday to Thursday

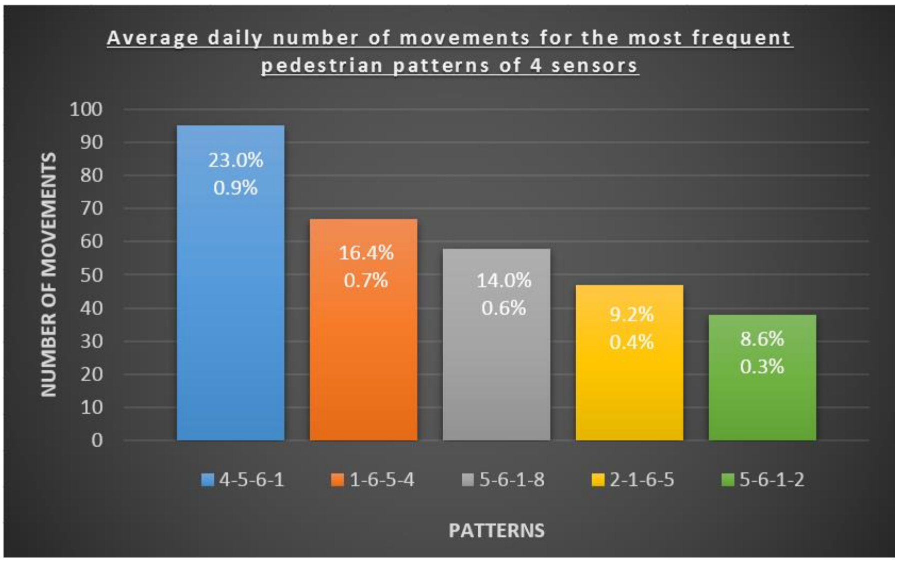


Figure 61: Average daily number of pedestrian movements for the most frequently used patterns of 4 sensors. In each column the corresponding relative (above) and absolute percentage (below) of this number is also included.

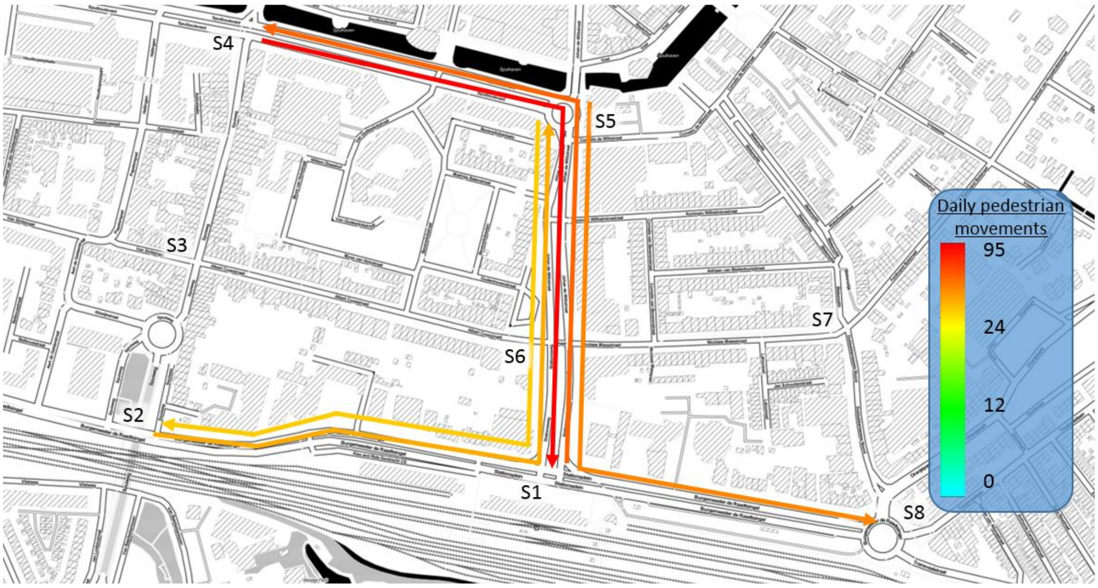


Figure 62: Average daily number of pedestrian movements for the most frequently used patterns of 4 sensors.

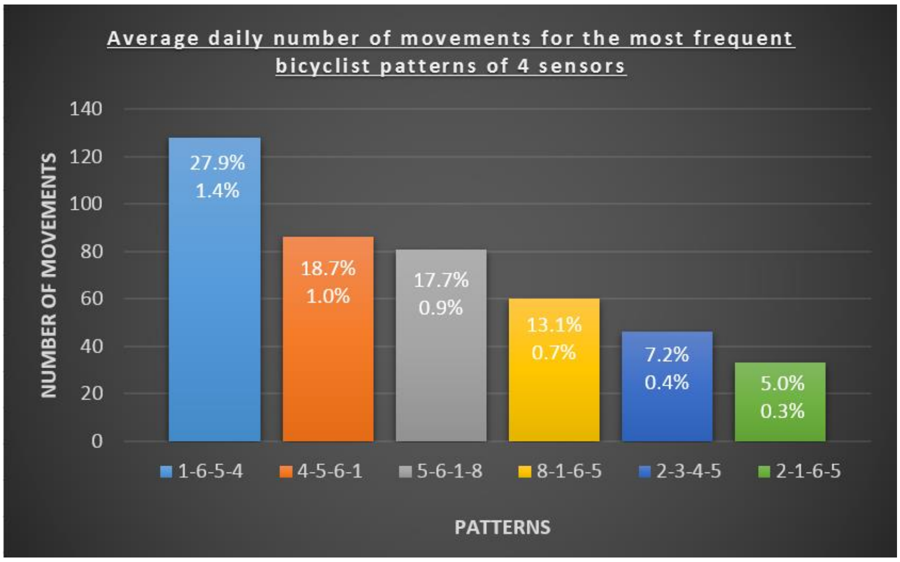


Figure 63: Average daily number of bicyclist movements for the most frequently used patterns of 4 sensors. In each column the corresponding relative (above) and absolute percentage (below) of this number is also included.

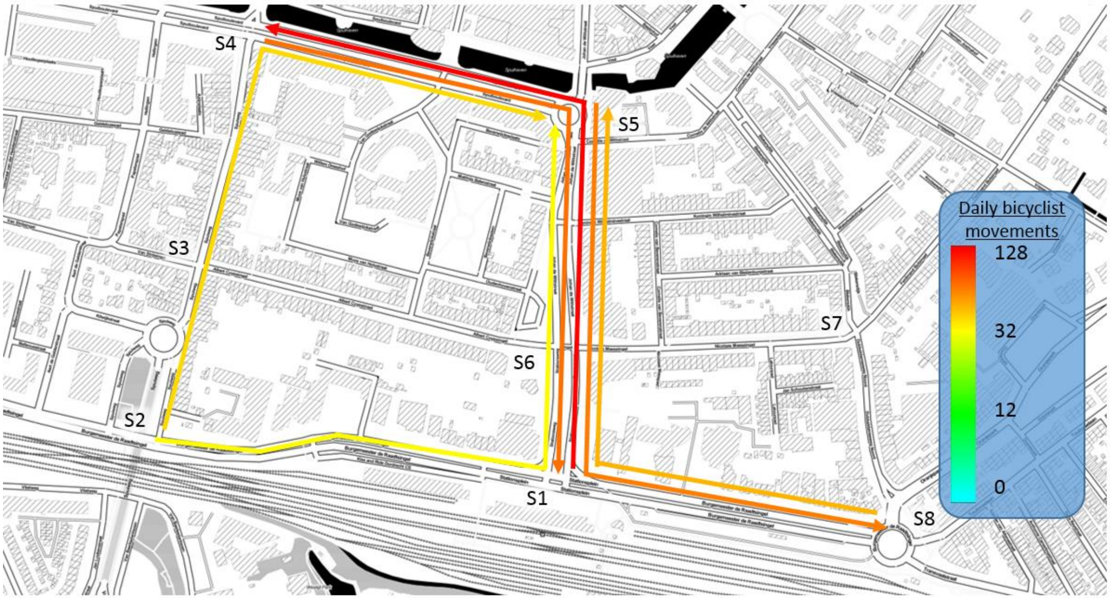


Figure 64: Visualization of average number of each street about the corresponding number of cumulative movements

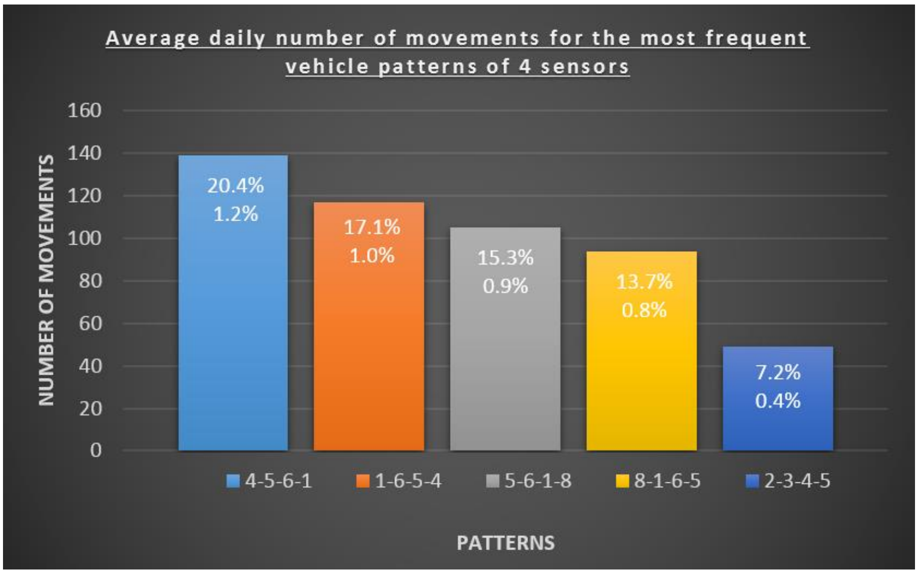


Figure 65: Average daily number of vehicle movements for the most frequently used patterns of 4 sensors. In each column the corresponding relative (above) and absolute percentage (below) of this number is also included.

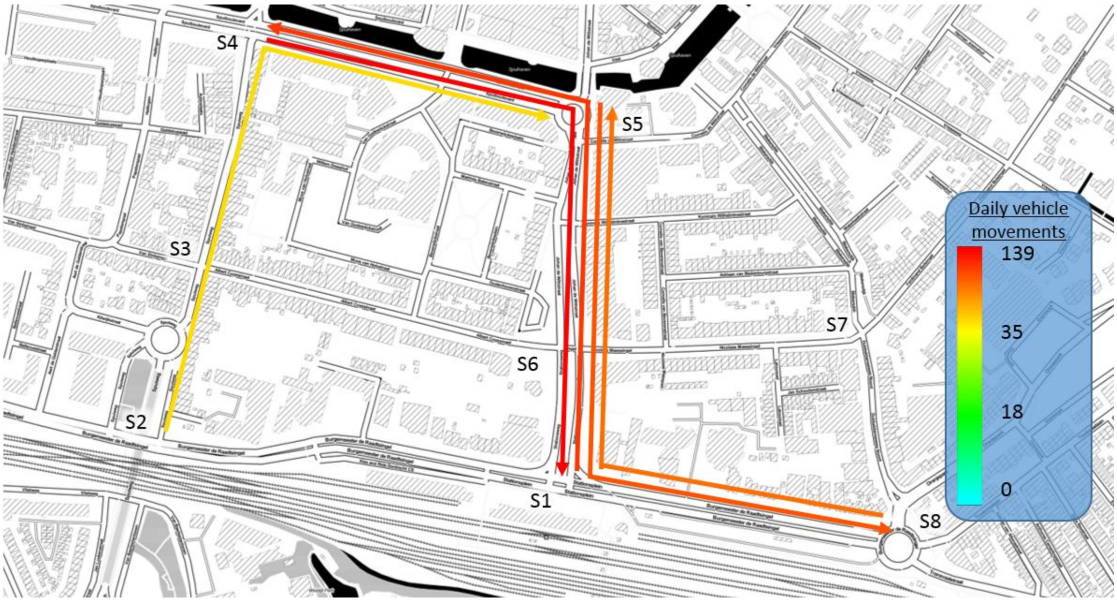


Figure 66: Visualization of average daily number of vehicle movements for the most frequently used patterns of 4 sensors.

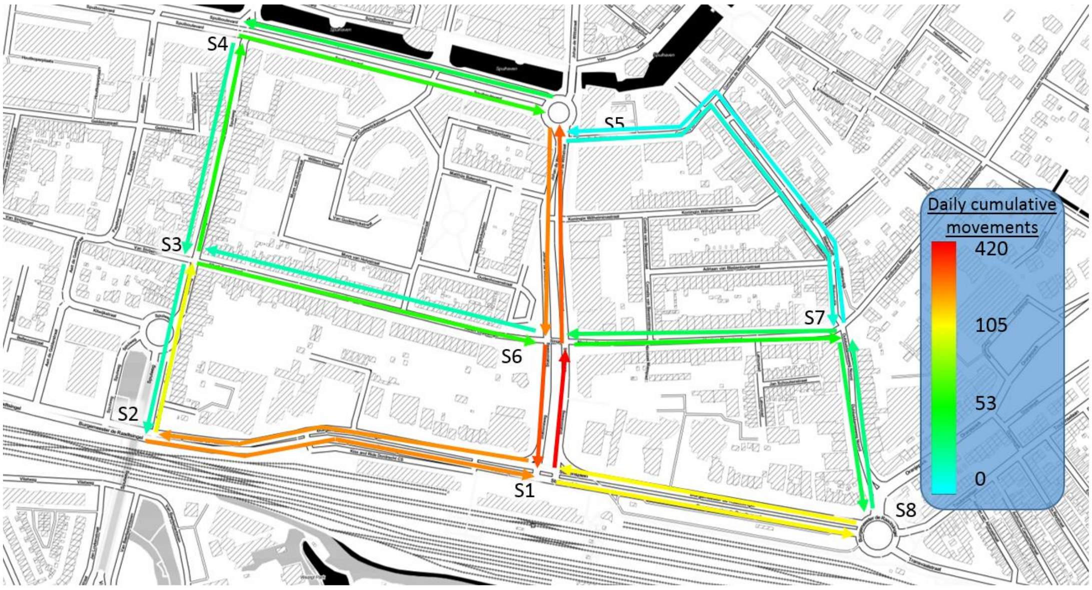


Figure 67: Visualization of cumulative movement flows for Thursday 22/09/2017 and the time period between 17:00 – 18:00

4.6 Computation of occupancy patterns

Apart from the movement patterns analyzed in the previous section, the investigation of occupancy patterns is also possible taking advantage of the fact that there are continuous data for a four-week period. Thus, similarities, differences and repeatabilities between days can be identified both in the research area and its surroundings.

4.6.1 Occupancy patterns in the research area

Figures 68 illustrates the hourly cumulative number of movements for each day of the research period. As it becomes apparent, the representation lines of the different weeks share significant similarities, which indicates a recurrence of the same condition and verifies the existence of occupancy patterns in the area. Based on that and in order to collect the information available, the average number of movements for each day of the week was computed, and is shown in Figures 69 and 70. Looking at these figures, three different movement patterns can be identified:

- An occupancy pattern for the days from Monday through Thursday: During these days, an initial peak is identified in the morning, between 08:00 and 09:00, followed by a steep decline in the number of movements until 11:00, when recession is observed. After that time, more and more people move in the area until the timeslot between 17:00 and 18:00, when a second (afternoon) peak is observed. Finally, a progressive reduction until the end of the day is observed again.

- An occupancy pattern for the days of the weekend: Unlike the other days of the week, only one peak exists during the weekend days - at around 14:00 to 15:00. Generally speaking, it can be stated that the occupancy pattern of these days can be represented by the form of the normal distribution with the mean at 14:00 in both days, but with higher value of standard deviation on Saturday. Thus, people move for longer timeslots on Saturday unlike Sunday, when the fluctuation of values is quite faster.

- An occupancy pattern for Friday: The occupancy pattern of Friday can be characterized as a transitional state between the above-mentioned patterns. Despite the fact that, as it has again two peaks, the pattern is almost similar to that of the other working days, a longer duration of the afternoon peak is observed which shows that the weekend approaches.

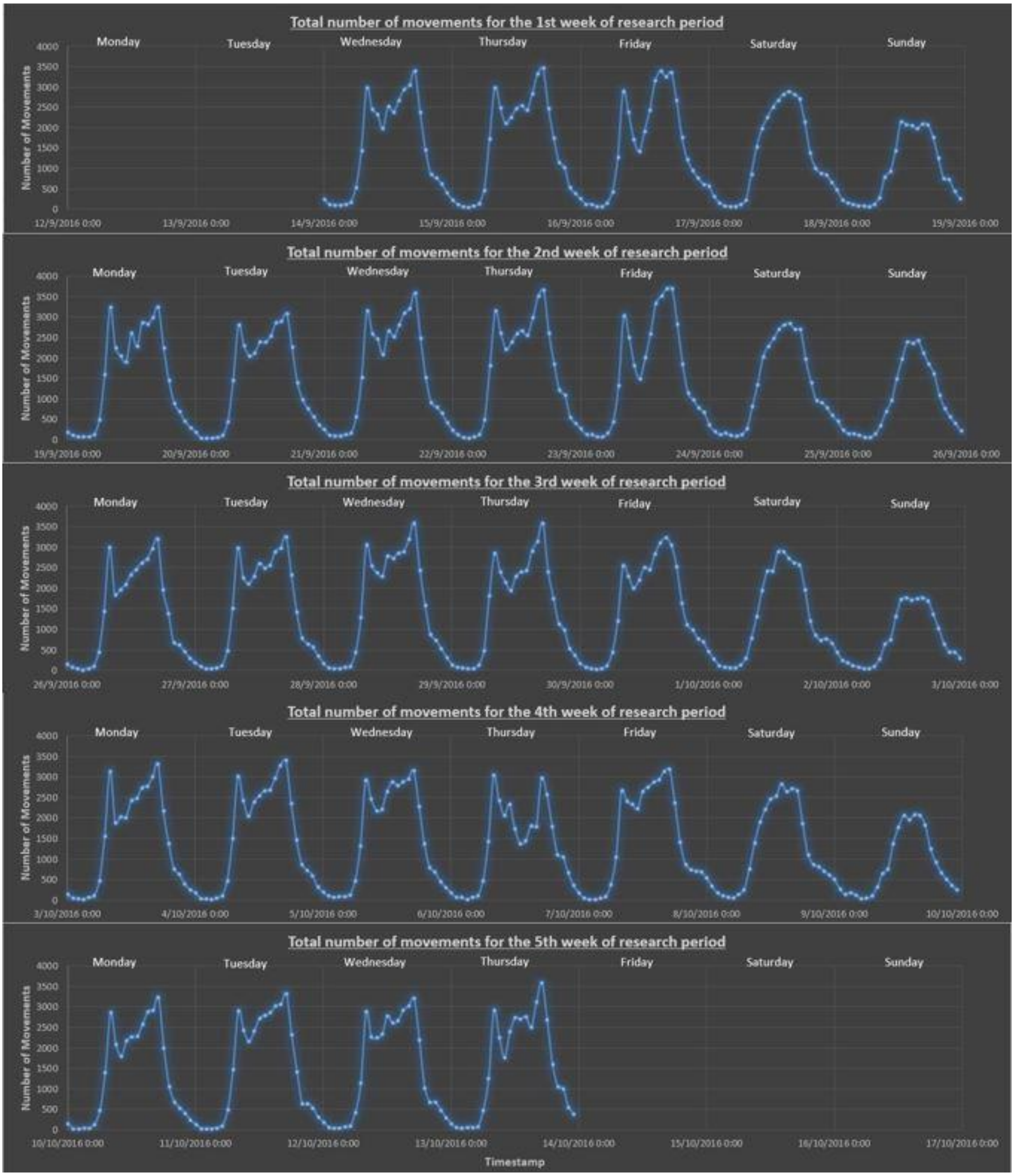


Figure 68: Hourly cumulative number of movements for each day of the research period

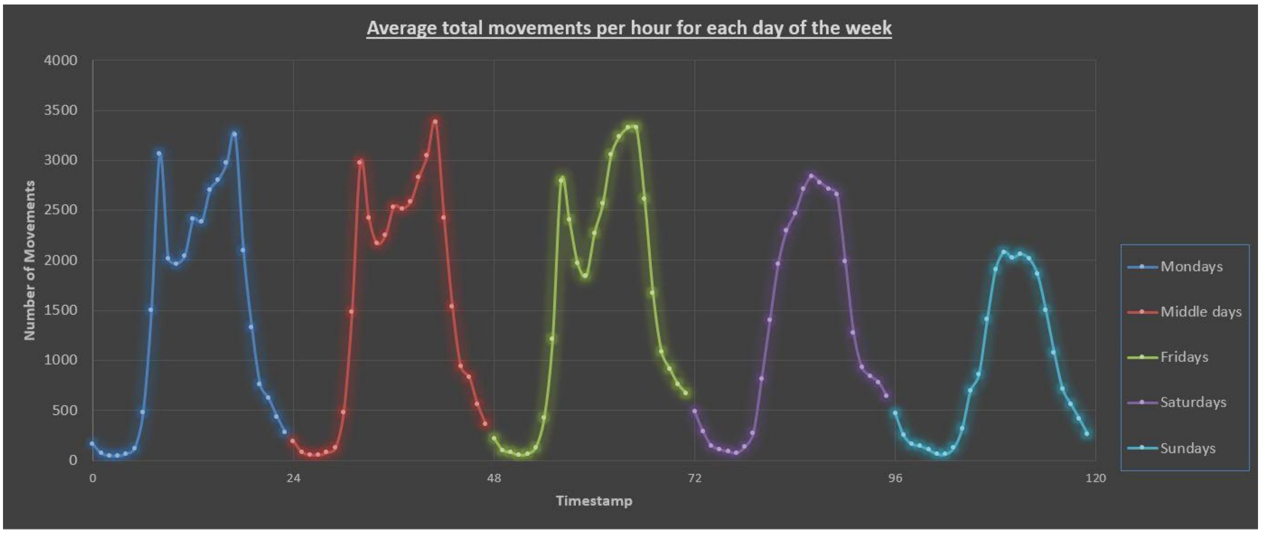


Figure 69: Average number of cumulative movements per hour for each day of the week

If we compare each day's total number of movements, it becomes obvious that, excluding weekends, all other days share almost the same peak and recession values, with the afternoon peak of the day being always a bit higher than the morning one. On the contrary, fewer movements and lower peak values are observed on Saturday and even fewer on Sunday. Finally, notable differences are identified in users' behavior on Friday and Saturday night. In contrast to other days, the number of movements during these two nights and the early hours of the next morning is significantly higher, which can be explained by the fact that many people do not have to go to their jobs on the following day, and can thus hang out during the night.

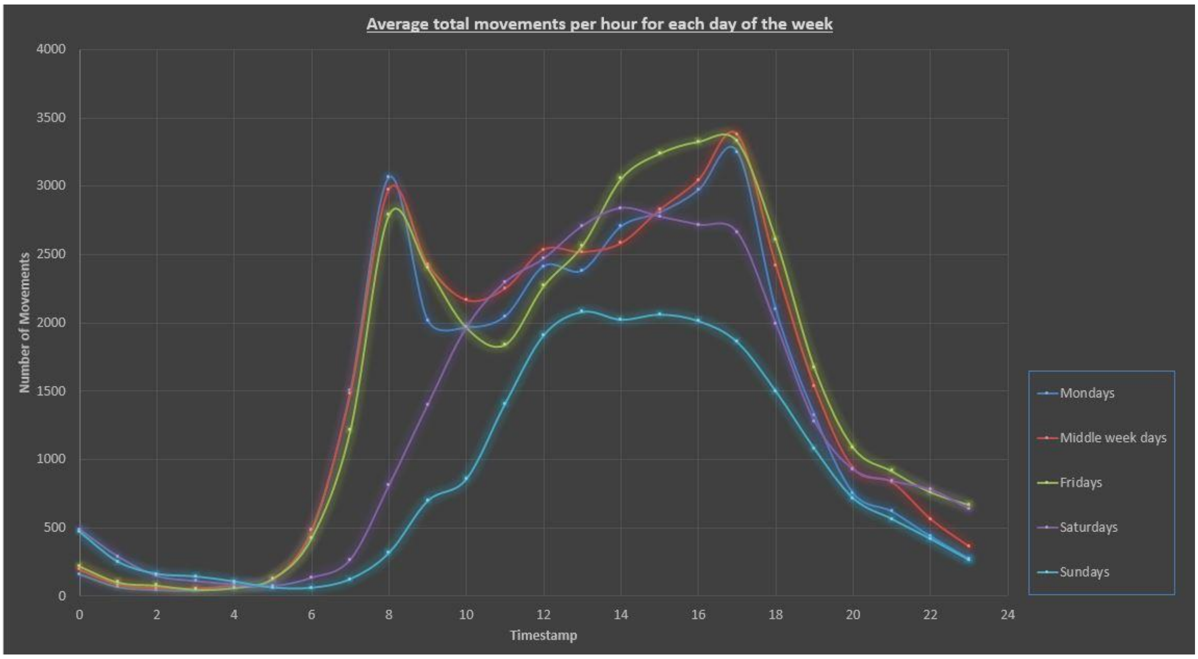


Figure 70: Average number of cumulative movements per hour for each day of the week

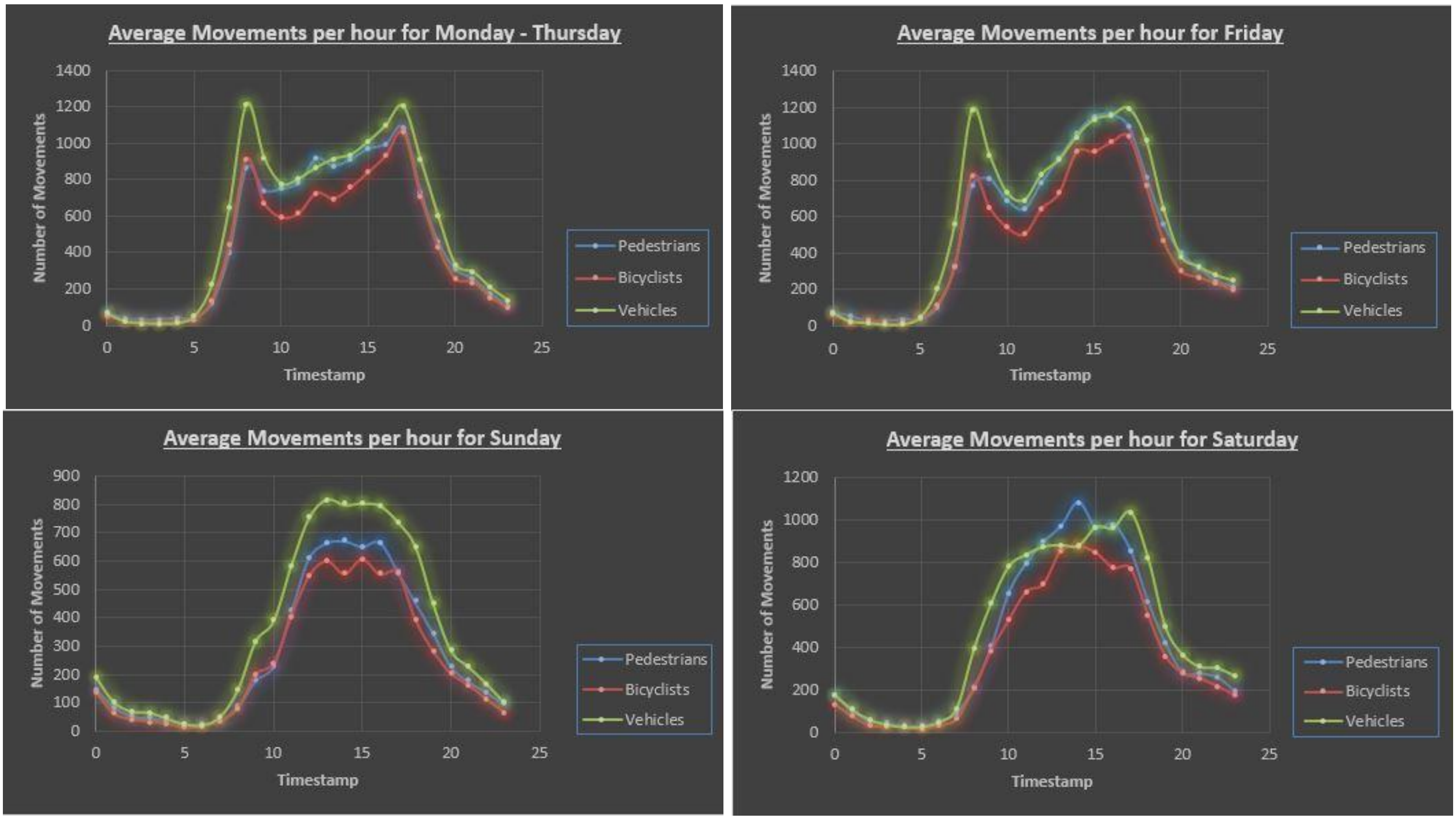


Figure 71: Average number of movements for each user category on different days of the week

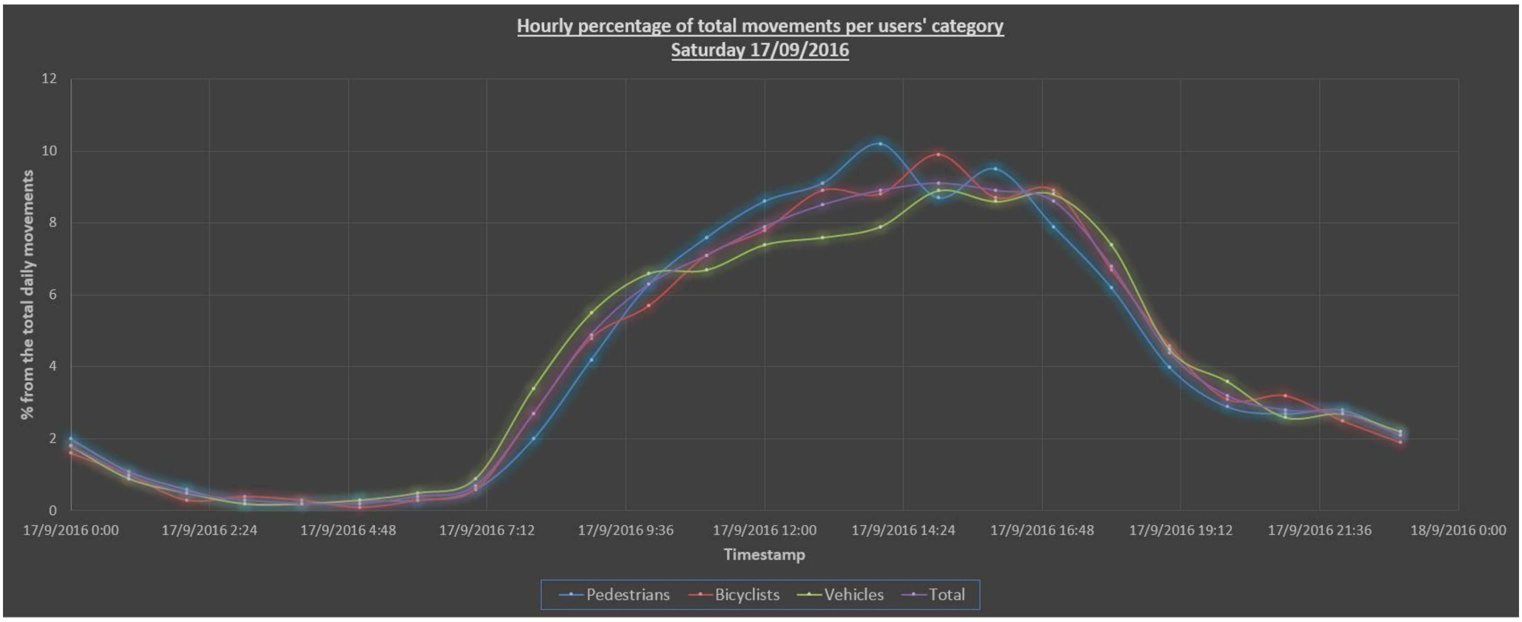


Figure 72: Distribution of the daily number of movements of each user category as well as of their sum on Saturday 17/09/2016

If each user category is investigated separately, it is observed that each of them can also be represented by the occupancy patterns described above. Despite the fact that they are part of the total number of movements, it is possible that the “behavior”, and thus pattern, of some categories differs from that of others and the cumulative pattern. However, as shown in Figures 71 and 72, a simultaneous fluctuation of the number of movements of each category is observed, with very few negligible differences. Hence, the peaks and recessions for all categories observed are the same.

Finally, significant resemblances at the level of streets are also observed. If the occupancy patterns from Figure 69 are compared with the representation lines of Figures 73 and 74, it becomes clear that they share many similarities. Especially with regard to the most crowded streets, the distribution of the relevant movements is almost the same as the distribution of the total number of movements, presenting two peaks at the same time on Friday as well as the shape of normal distribution on Saturday.

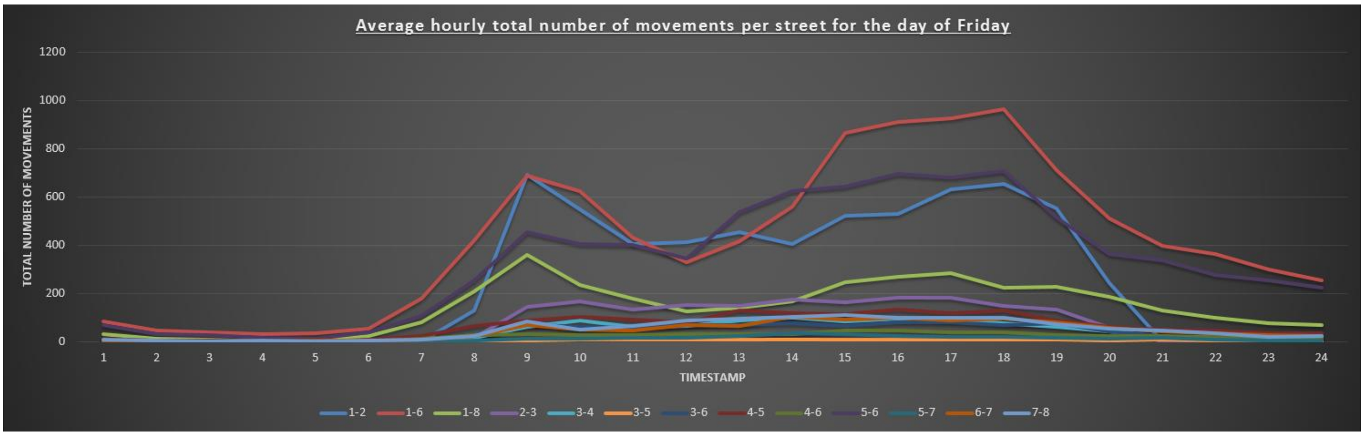


Figure 73: Distribution of the total number of movements of each street over the day of Friday

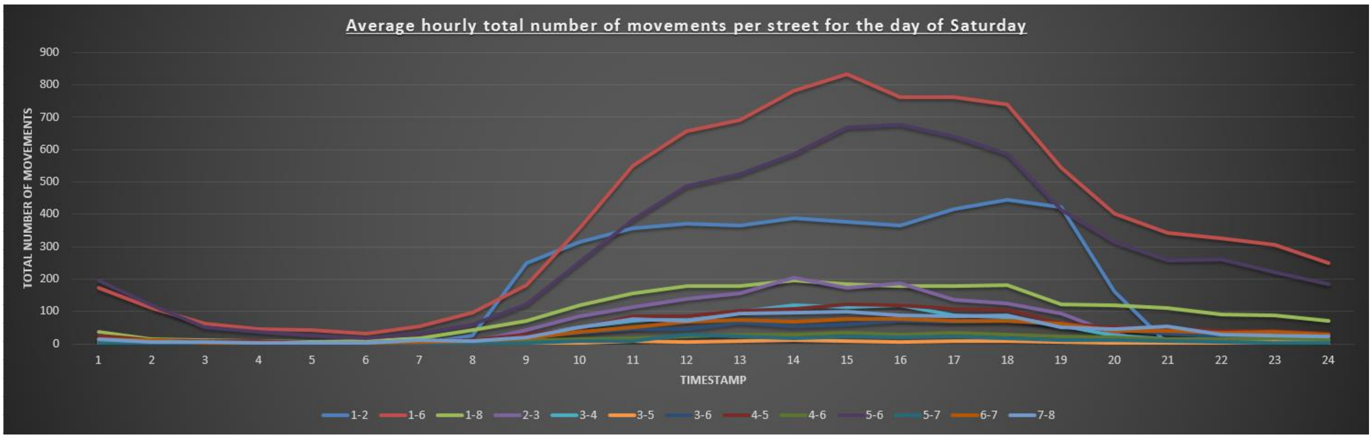


Figure 74: Distribution of the total number of movements of each street over the day of Saturday

4.6.2 Occupancy patterns in the surrounding area

Apart from the research area, it is possible to investigate the identification of occupancy patterns in the surrounding area as well. This can be done by deploying the data sources from the city center of Dordrecht, and more specifically those recorded by the counting cameras of the Municipality and the RMC company. Figure 75 illustrates the total number of movements in the research area as well as downtown Dordrecht. In this Figure, the same relationship between the two areas can be observed throughout the research period. To be more specific, throughout the whole week, there is only one peak in the city centre around 14:00, which coincides with the recession of the movements in the research area during working days. This can possibly be explained by the fact that, on working days, many people are moving towards the city centre for their lunch break. During the weekend though, both regions feature similar occupancy patterns, with a common peak at midday but with a much higher number of movements in the city centre.

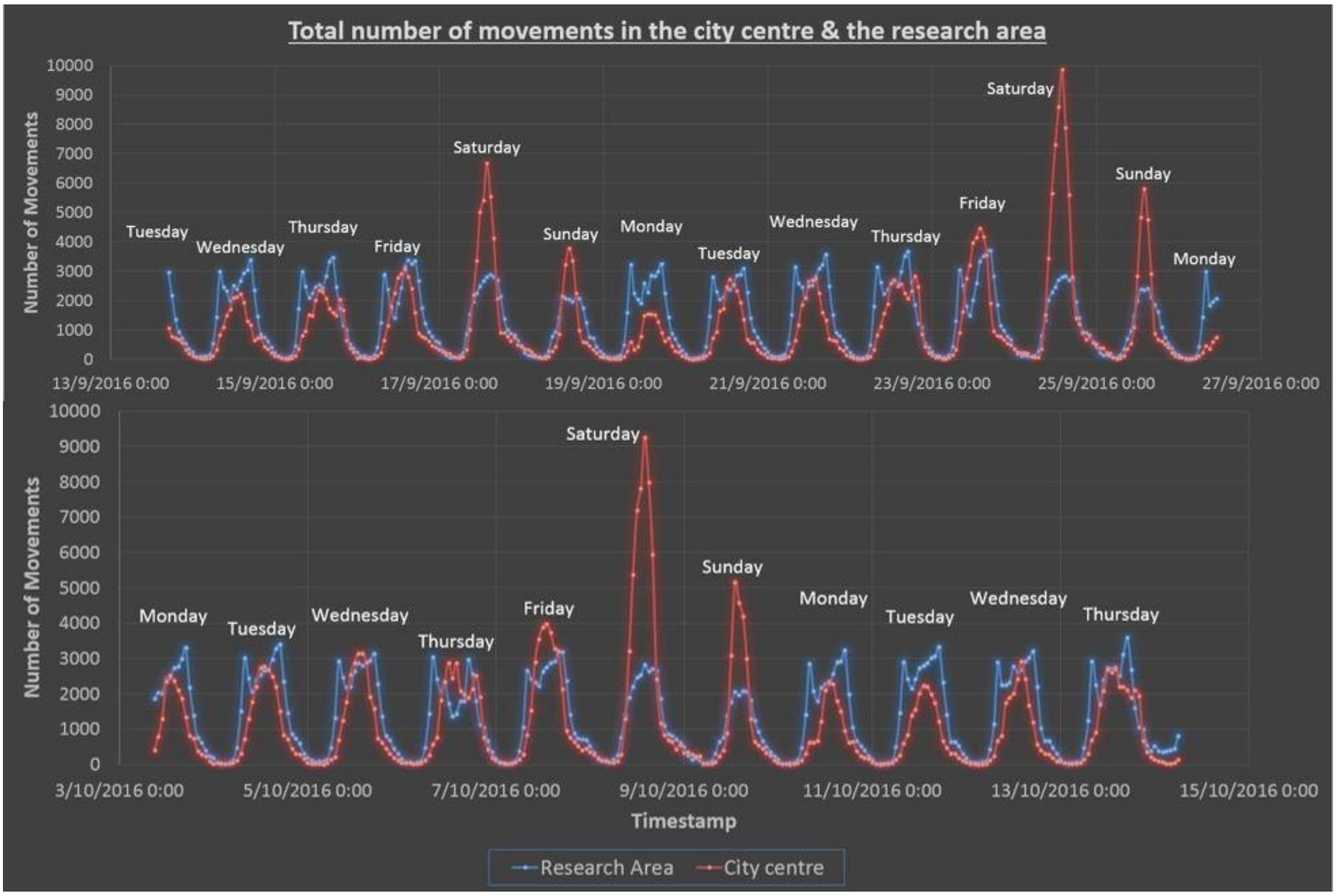


Figure 75: Total number of movements in the research area and in the city centre

4.6.3 Influence of weather conditions

Another parameter which was decided to be investigated in this research is the influence of weather conditions on the total number of movements as well as on road modality. Based on related works (*Thurau, 2013*), temperature has a higher influence on user movement behavior than rain. As mentioned in the previous chapter, during the Wi-Fi data collection period weather conditions, including temperature, wind speed, and sunshine were also stored. Thus, the same days of the week with the highest weather conditions difference were selected in order to compare the relevant traffic attributes.

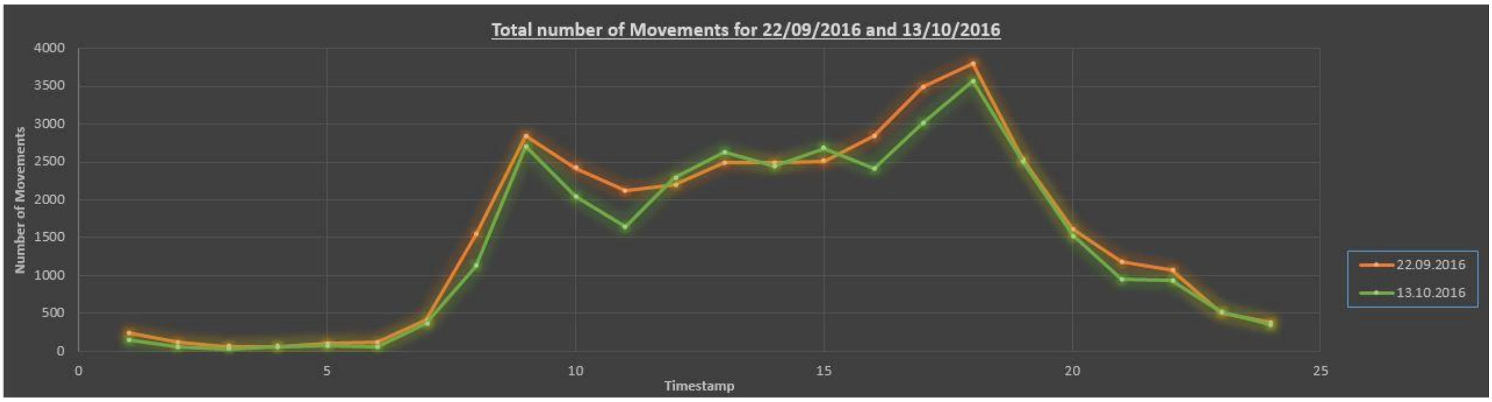


Figure 76: Cumulative number of movements in the research area for Thursday 22/09/2016 (orange) and Thursday 13/10/2016 (light green)

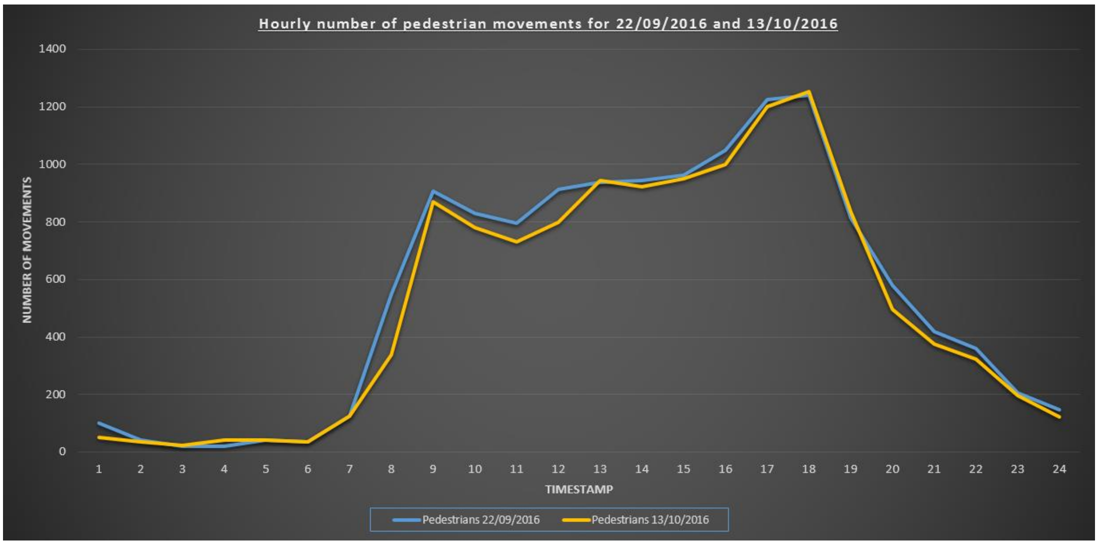


Figure 77: Hourly number of pedestrian movements in the research area for Thursday 22/09/2016 (light blue) and Thursday 13/10/2016 (yellow)

Thursday 22/09/2016	12am	1am	2am	3am	4am	5am	6am	7am	8am	9am	10am	11am
Forecast	Partly Cloudy	Partly Cloudy	Partly Cloudy	Cloudy	Cloudy	Cloudy	Mostly Cloudy	Mostly Cloudy	Partly Cloudy	Partly Cloudy	Showers	Partly Cloudy
Temp (°C)	13°	13°	13°	12°	12°	12°	12°	12°	13°	14°	16°	18°
RealFeel®	12°	13°	13°	13°	12°	12°	12°	12°	13°	15°	16°	19°
Wind (km/h)	9 E	7 E	7 E	6 ESE	6 ESE	6 SSE	6 SSE	6 SSE	6 SSE	7 SSE	7 SSE	9 SSE

Thursday 22/09/2016	12pm	1pm	2pm	3pm	4pm	5pm	6pm	7pm	8pm	9pm	10pm	11pm
Forecast	Partly Cloudy	Showers	Partly Clou	Partly Cloudy	Partly Cloudy	Showers	Showers	Partly Cloudy	Cloudy	Cloudy	Partly Cloudy	Partly Cloudy
Temp (°C)	19°	20°	21°	22°	21°	20°	20°	19°	17°	16°	15°	15°
RealFeel®	20°	20°	21°	22°	20°	19°	17°	17°	17°	16°	15°	15°
Wind (km/h)	9 S	11 S	11 S	11 SSW	11 SSW	9 SW	11 SSW	9 S	9 S	7 S	7 SSW	7 SW

Thursday 13/10/2016	12am	1am	2am	3am	4am	5am	6am	7am	8am	9am	10am	11am
Forecast	Cloudy	Mostly Cloudy	Mostly Cloudy	Cloudy	Partly Cloudy	Mostly Clear	Mostly Clear	Mostly Clear	Mostly Clear	Mostly Sunny	Mostly Sunny	Partly Sunny
Temp (°C)	7°	7°	6°	6°	6°	6°	5°	5°	6°	7°	8°	9°
RealFeel®	4°	3°	3°	3°	2°	2°	2°	1°	1°	3°	3°	4°
Wind (km/h)	13 ENE	15 ENE	15 ENE	15 ENE	15 ENE	15 ENE	17 E	17 E	17 E	18 E	22 E	28 E

Thursday 13/10/2016	12pm	1pm	2pm	3pm	4pm	5pm	6pm	7pm	8pm	9pm	10pm	11pm
Forecast	Partly Sunny	Partly Sunny	Partly Sunny	Partly Sunny	Partly Sunny	Partly Sunny	Partly Sunny	Partly Cloudy	Cloudy	Cloudy	Cloudy	Cloudy
Temp (°C)	10°	11°	11°	12°	12°	11°	11°	10°	10°	9°	9°	9°
RealFeel®	5°	6°	7°	8°	8°	7°	7°	6°	6°	5°	5°	5°
Wind (km/h)	30 E	30 E	28 E	28 E	28 E	26 E	26 E	24 E	20 E	18 E	18 E	20 E

Figure 78: Weather conditions for Thursday 22/09/2016 and Thursday 13/10/2016

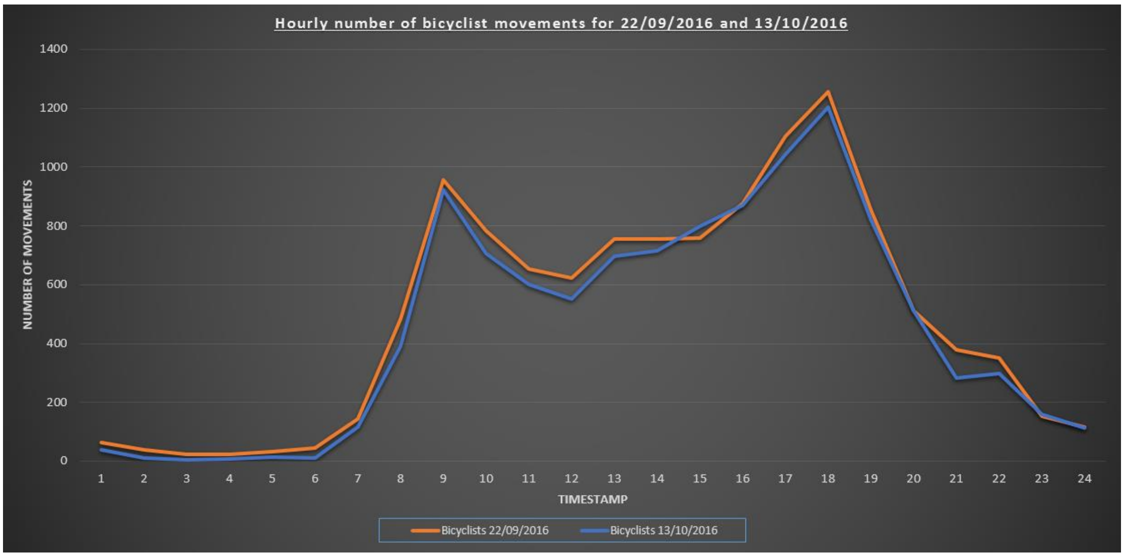


Figure 79: Hourly number of bicyclist movements in the research area for Thursday 22/09/2016 (orange) and Thursday 13/10/2016 (blue)

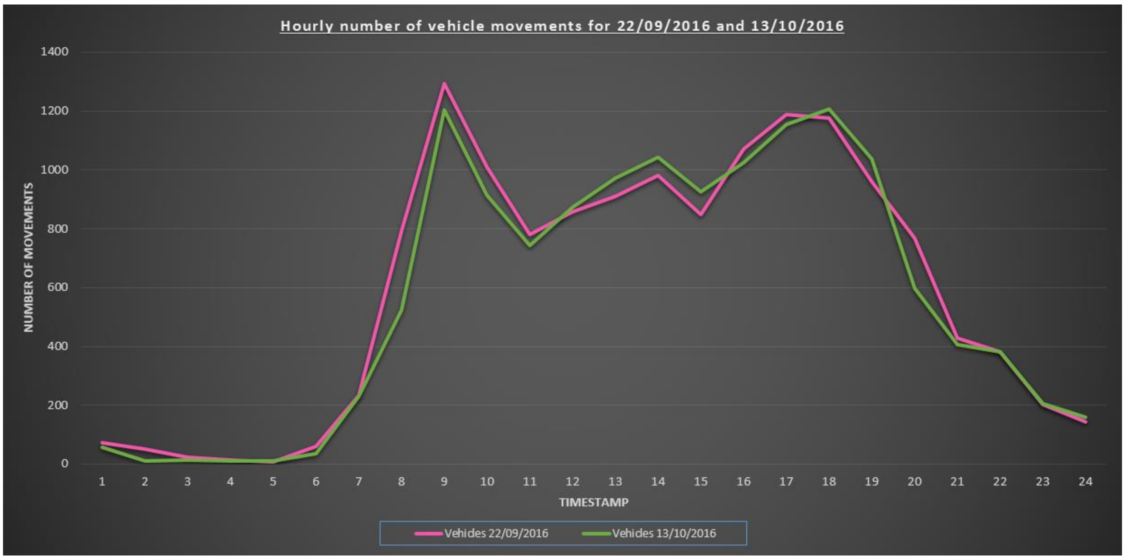


Figure 80: Hourly number of vehicle movements in the research area for Thursday 22/09/2016 (pink) and Thursday 13/10/2016 (light green)

Figure 78 illustrates the weather conditions on Thursday 22/09/2016 and on Thursday 13/10/2016 while Figures 76, 77, 79 and 80 present both the total hourly number of movements and the hourly number of movements for each user category separately. As it can be observed, despite the fact that there is a notable difference in weather conditions between these two days - with a temperature difference of about ten degrees and a change from sunny to rainy weather - no other significant change can be identified neither in the number of movements nor on road modality. However, it is important to mention that the duration of the data collection period of this work is too short to draw reliable conclusions. During the period between mid-September and mid-October, getting a temperature deviation of more than ±10 C was quite difficult. The further investigation of the influence of weather conditions might be possible, with the use of longer datasets providing information about the same region in different seasons.

4.7 Influence of Wi-Fi monitoring setup

Besides describing research methodology, its application and relevant results, it would also be interesting to investigate the influence of the Wi-Fi monitoring setup on the final outcome. As it was mentioned in Chapter 1, due to technical issues and the inability to change the location of the sensors throughout the data collection period, the influence of the way the observation network was designed cannot be studied in this work. However, the effect of the total number of sensors used on the final results can be investigated.

Related works focused on the theory of Network Analysis. Also, the Graph Theory was studied in relation to the common problem of n-nodes. Unfortunately, based on this theory, the influence of the number of sensors falls within a broader problem which is known as NP-complete problem* and maybe constitutes the most important open problem in the field (*Goldreich, 2010*).

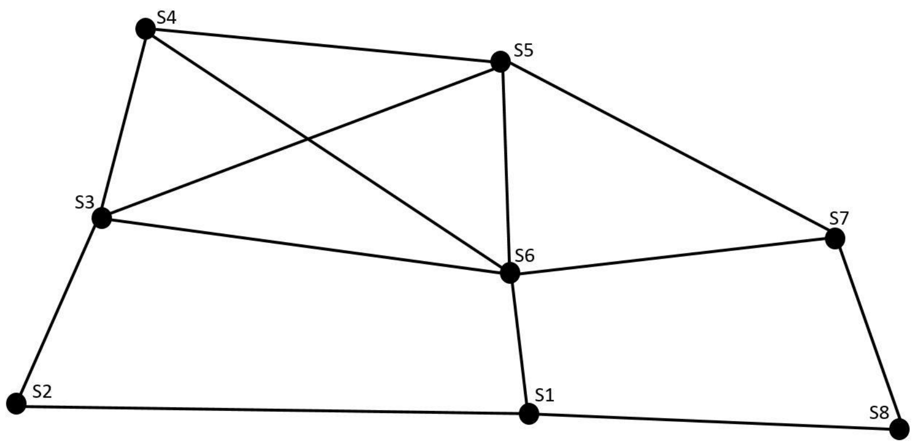


Figure 81: Node network of the observation area

However, the theory suggests the breakdown of the network into smaller subsets with less nodes. Based on that, it is possible for individual cases to be studied with combinations of n=7,6,5,4 sensors, etc., instead of n=8, choosing different sensors to be removed each time.

* NP- nondeterministic polynomial time complete problems are a set of problems to each of which any other NP-problem can be reduced in polynomial time, and whose solution may still be verified in polynomial time (*Goldreich, 2010*).

The following equation gives the total number of possible combinations of *r* objects from a set of *n* (*Grimaldi, 2003*):

$$\binom{n}{r} = {}_{n}C_{r} = \frac{n!}{r!(n-r)!}$$

Based on this equation, in cases, for instance, of seven, six, five, or four sensors there are respectively ${}_{8}C_{1}=8$, ${}_{8}C_{2}=28$, ${}_{8}C_{3}=56$, and ${}_{8}C_{4}=70$ combinations which have to be checked separately. Thus, there is a significant increase in the complexity of the problem when the number of removed sensors also increases, as Figure 82 illustrates.

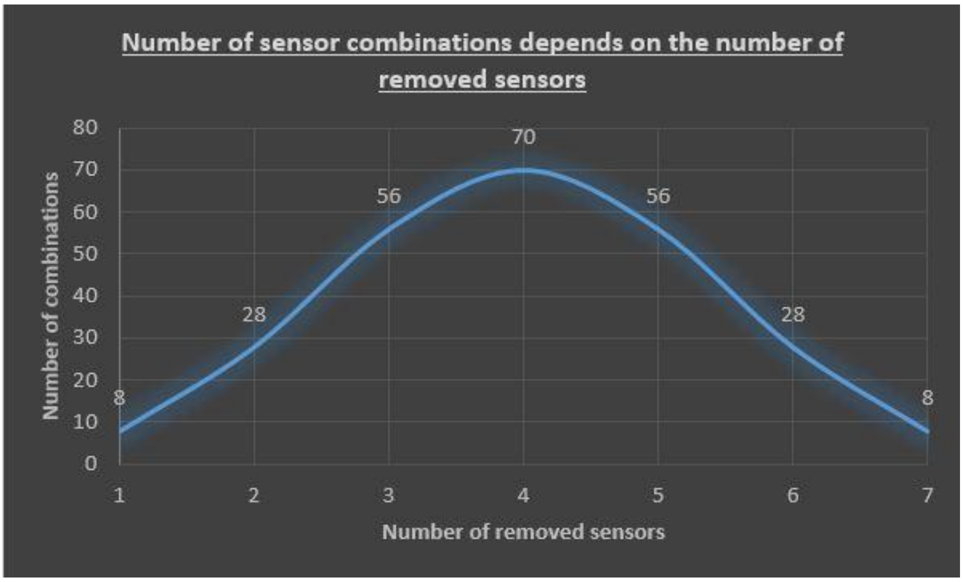


Figure 82: Relationship between the number of removed sensors and the number of combinations which have to be checked

Due to this complexity of the problem, the in-depth study of all possible combinations is quite difficult to carry out. Table 10 presents the outcome of the analysis of all combinations, with each row representing a different scenario in which one sensor from the observation network is removed. The first column contains the number of the sensor which can be removed while the second contains the number of streets connected to this sensor.

Sensor	Number of connected streets	Average number of related movements	Affected % of the total daily movements
1	3	16487	54.3
2	2	6479	21.4
3	4	3249	10.7
4	3	2553	8.4
5	4	8311	27.4
6	5	17669	58.2
7	3	2188	7.2
8	2	3750	12.4

Table 10: Overall influence of each sensor on the final outcome

During Wi-Fi data analysis in the previous chapter, the average number of movements was computed for each street of the research area. Based on this information, it is possible to calculate the average number of movements associated with each sensor (as shown in the third column of table 10). For instance, sensor 2 affects the computation of movements as well as the relevant road modality for streets 2-1 and 2-3. Finally, it is possible to calculate the percentage of the daily movements affected by each sensor, by comparing the above-mentioned movements with the average total number of movements in the whole area. Looking at the fourth column of Table 10, it is clear that the removal of each sensor does not affect the final outcome in the same way. Hence, the final outcome depends not only on the number of sensors, but also on their location. For instance, removing sensors 7 and 4 would have fewer repercussions than removing sensor 6, based on the assumption that the main goal of the research remains the same. Another parameter which, of course, has to be taken into account in the general investigation of the influence of the monitoring setup is the use of assumptions. For instance, if sensor 6 is removed and a device was initially scanned by sensor 1 and, a few minutes later, by sensor 5, an assumption can be made that the user passed from the location of sensor 6. At this point, however, it is worth mentioning that, based on the outcome of the data analysis and as illustrated in Tables 5, 6, and 7, just a percentage of approximately 30% of users were scanned by more than two sensors in the area; a finding which can also be used in case two or more sensors are removed from the observation network.

5. Data validation

As always, when observations are made based on a certain methodology, it is important that the accuracy of the undoubtedly useful outcomes be validated. For instance, it is meaningless to know a distance between two points if this information is not associated with details about how well the relevant observation was made, and how close this value is to reality. Thus, apart from describing the Wi-Fi data collection procedure, the analysis procedure and its final outcomes, it considered necessary to find a way to evaluate the accuracy of this method and the extent to which its results are representative of the real conditions in the research area.

Having reviewed all related works, it was decided to perform random sampling using the Bernoulli trial (Yes/No) to validate the outcomes. Specifically, the area between two sensors, for instance S1 and S2, was inspected each time. An observer counted the total number of passing pedestrians *N* per direction, i.e. from sensor 1 to sensor 2, over a period of one hour, while someone else performed a simple random sampling using the Bernoulli trial (Yes/No), by asking an *n* number of pedestrians whether they have a device with its Wi-Fi functionality enabled or not.

Using the proportion sample $\hat{p}_x$ of pedestrians with Wi-Fi enabled devices, the population proportion *p* that had the Wi-Fi functionality enabled was estimated. The proportion sample $\hat{p}_x$ follows the binomial distribution. If the sample size is large then the binomial distribution-, and, therefore, the sampling distribution of the proportion-is approximately normal. The confidence interval, for the population proportion *p*, at level of confidence 1-a is:

$$\hat{p}_x - z_{1-\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}} \leq p \leq \hat{p}_x + z_{1-\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}}$$

For 1-a = 0.90, meaning that there is a 90% probability, the confidence interval is:

$$\hat{p}_x - 1.645\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}} \leq p \leq \hat{p}_x + 1.645\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}}$$

Where α - the significance level.

For example, if 500 pedestrians moved from sensor 1 to sensor 2 within a period of one hour, 80 of them would be asked and 40% of them would have their devices' Wi-Fi functionality enabled, then for a confidence level of 90%:

- N = 500

- n = 80

- $\hat{p}_x$ = 0.40

$$0.40 - 1.645\sqrt{\frac{0.40 \cdot 0.60}{80}} \leq p \leq 0.40 + 1.645\sqrt{\frac{0.40 \cdot 0.60}{80}}$$

$$0.31 \leq p \leq 0.49$$

Thus, the confidence interval limits for the total number of pedestrians who walked from sensor 1 to sensor 2 during this timeslot are equal to:

$$(500 * 0.31 \,,\, 500 * 0.49)$$

$$(155 \,,\, 245)$$

Finally, the number resulting from data processing, which was provided by the sensors and associated with the devices characterized as "pedestrian", was checked as to whether it falls within the confidence interval limits. The above-mentioned process was repeated multiple times on different days, hours and streets for the categories of "pedestrians" and "bicyclists". Table 11 contains the relevant results in total and, as it is evident in all sampling tests, the number resulting from the Wi-Fi monitoring method always fell within the relevant confidence interval limits. Furthermore, it is important to mention that the questionnaire was also used in various timeslots during the data collection period in the research area both for pedestrians and bicyclists, in order to get a more representative indicator of the percentage of people who had devices with enabled Wi-Fi functionality. Thus, based on the questionnaire the relevant percentages for the category of pedestrians and bicyclists were equal to 43% and 40% respectively.

Day	Date		Time	Direction	Tested category	Counted (N)	Questionnaire Sampling (n)	Enabled Wi-Fi (%)	Confidence Interval 90%	Sensors Outcome
Sunday	18-Sep-16		16-17	5→6	Pedestrian	343	62	42%	(109 , 179)	129
"	"		17-18	"	Bicyclist	254	32	39%	(63 , 135)	107
Monday	19-Sep-16		08-09	1→2	Pedestrian	381	66	42%	(122 , 198)	146
"	"		09-10	"	Bicyclist	226	31	38%	(53 , 118)	95
Tuesday		27-Sep-16	10-11	6→1	Pedestrian	348	59	41%	(106 , 179)	129
"		"	11-12	"	Bicyclist	281	33	40%	(73 , 152)	100
Wednesday	21-Sep-16		12-13	1→6	Pedestrian	406	64	41%	(125 , 208)	182
"	"		13-14	"	Bicyclist	153	32	43%	(44 , 88)	64
Thursday		29-Sep-16	14-15	2→1	Pedestrian	98	48	43%	(31 , 54)	39
"		"	15-16	"	Bicyclist	143	31	38%	(34 , 75)	56
Friday	23-Sep-16		16-17	6→5	Pedestrian	388	61	45%	(134 , 215)	169
"	"		17-18	"	Bicyclist	227	34	41%	(62 , 125)	97
Saturday		01-Oct-16	13-14	4→5	Pedestrian	102	49	47%	(36 , 60)	39
"		"	14-15	"	Bicyclist	97	32	44%	(29 , 57)	38

Table 11: Overall outcome of the data validation for the categories of "pedestrians" and "bicyclists"

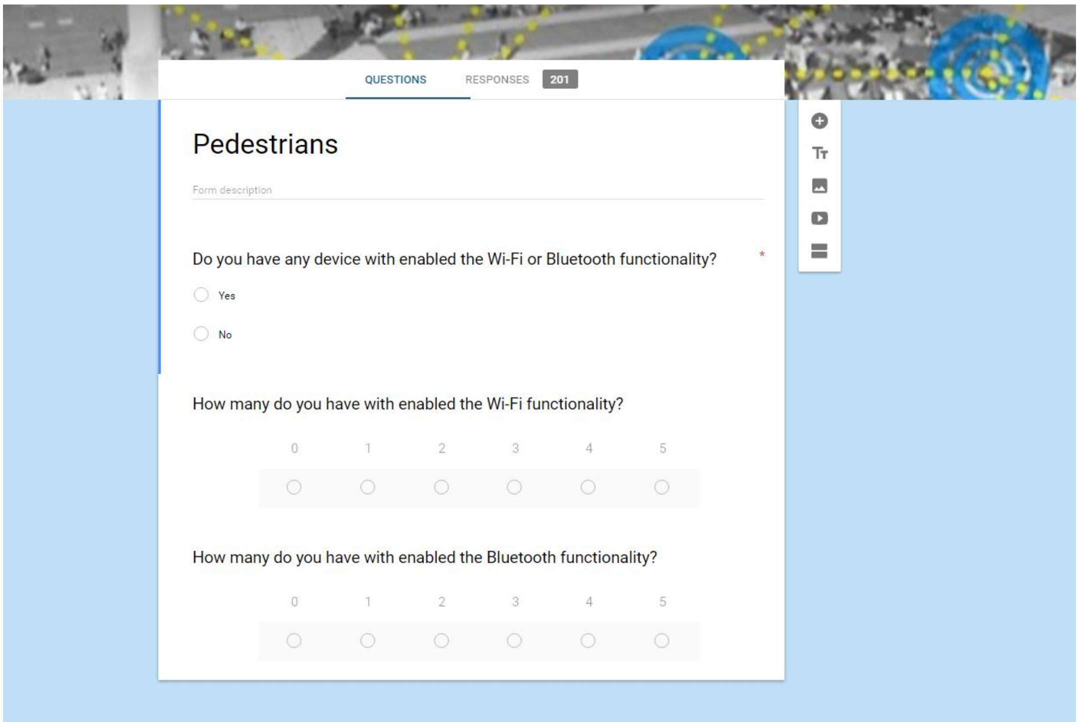


Figure 83: The questionnaire used for the category of pedestrians

Unlike the categories of pedestrians and bicyclists, the questionnaire could not be used for the category of vehicles. For this reason, the total number of passing vehicles was counted for different timeslots and days, as shown in Table 12. Based on these values and under the assumption that there is one user per each vehicle and two users per every two vehicles, that is three users in every two vehicles, an estimation about the total number of users can be made. Finally, based on this number of users the relevant percentage of the devices with enabled Wi-Fi functionality can be estimated, as presented in the last column of Table 12.

Apart from the use of the questionnaire and the random sampling using the Bernoulli trial, in order to draw a conclusion about the accuracy and the applicability of the Wi-Fi monitoring system, it was important to mention and take into account some other things as well. First of all, despite the efforts made, it was impossible to identify devices, and thus users, who moved in the research area by public transport. However, some devices which were connected to the Wi-Fi network of the flixbus were scanned almost simultaneously by different sensors. Based on that, it can be considered that these devices belong to users who travelled on the flixbus; a conclusion which is enhanced by the fact that the order and the time of movement was verified by the bus schedule. Furthermore, despite the research for devices connected to the Wi-Fi network of buses or the train and their filtering or separate study, it is possible that some devices with their Wi-Fi functionality enabled were not connected to the network but were scanned by the sensors. In this case, there is a risk that these devices might be considered to belong to users in the streets and be characterized as pedestrian, bicyclist, or vehicle after the analysis. However, based on the random sampling and the relevant results of the validation, there is not any notable difference between the streets close to the railway or the bus route and those away from them. Finally, based on the outcomes of the Wi-Fi monitoring system there was not any "vehicle" which moved between sensors S5 and S7, an area that was under construction and vehicles were not permitted, as well as from S4 to S3 and from S3 to S2 which are one-way streets. This enhances the reliability of the system.

Taking into account all the above-mentioned validation procedures as well as the relevant outcomes, the following conclusion can be drawn: the accuracy of the Wi-Fi monitoring method can be characterized as particularly high and the relevant results closely reflect the real street conditions in the research area.

Day	Date		Time	Direction	Counted (N)	Sensors Outcome	Estimated number of users	Estimated % of Enabled Wi-Fi (based on assumption)
Sunday	18-Sep-16		13-14	1→6	110	76	165	46.1
"	"		14-15	6→5	88	63	132	47.7
Monday	19-Sep-16		16-17	1→2	193	130	290	44.9
Tuesday		27-Sep-16	09-10	6→1	132	87	198	43.9
Wednesday	21-Sep-16		08-09	4→5	53	33	80	41.5
Thursday		29-Sep-16	12-13	3→6	22	16	33	48.5
"		"	17-18	1→8	141	99	212	46.8
Friday	23-Sep-16		15-16	5→6	165	111	248	44.8
"	"		18-19	6→7	44	28	66	42.4
Saturday		01-Oct-16	10-11	3→4	40	27	60	45.0
"		"	11-12	2→1	101	71	152	46.9

Table 12: Overall outcome of the data validation for the category of "vehicles"

6. Discussion

The main aim of this work was to investigate what kind of traffic composition and occupancy models can be identified using a Wi-Fi monitoring system of sensors in the urban area of Dordrecht. Thus, this work includes a detailed description of the findings, the accuracy of the computation as well as their representativeness.

At the beginning of the research, separate objectives were set and expressed in the form of research questions and sub-questions. The outcome was tested in order to verify that the objectives had been met. The findings of the research will be further discussed with respect to the research questions and the four categories they were divided into. Additionally, the limitations of this method will be pointed out to allow for future improvements to the system and its implementation.

What kind of road modality and occupancy patterns can be recognized by Wi-Fi monitoring sensors in the city area of Dordrecht in order to support the "Smart City" concept?

In order to achieve the main aim of the research, a process of four main steps was followed. The first step was to thoroughly understand the Wi-Fi monitoring system and study its indoor and outdoor applications. The technical specifications of the system as well as the limitations and suggestions from related works were used as a starting point to design the observation network and choose the locations of sensors during the second step. In the third step, all necessary data were collected by the Wi-Fi system and were analyzed in order to create the metadata set, which was the basis on which the main research question was answered. Three user categories were recognized and, thus, road modality was split into the categories of pedestrians, bicyclists and vehicles. Movement behavior was investigated both on an overall and a per category basis leading to the identification of certain rush hours on different days of the week in the research area, the occupancy relationship between the research area and the city centre as well as differences in region occupancy throughout the week. Last but not least, during the fourth step, the outcome of this work was validated to confirm the accuracy of the above-mentioned findings.

1. What influence does the Wi-Fi monitoring setup have?

Due to technical issues and time limitations, the identification of the most appropriate Wi-Fi network configuration is not part of this research. However, the influence of some parameters to the final result was studied. First of all, the choice of the kind of antenna is one parameter which affects the range of sensors and more specifically its shape. Unlike standard antennas, the directional antenna allows for the scanning range to be limited from 360 to 180 degrees. In this way, sensors can record devices only within a specific region. This could be very useful in case different streets are connected or passing through in the area around the sensor. Furthermore, it significantly speeds up the data filtering procedure, as there are no records from indoor devices such as printers, laptops, etc. Furthermore, another integral parameter is time synchronization between the sensors. A potential time error of mere seconds in the setting time of a scanner could significantly change the final outcomes. The moving patterns of users would be wrongly computed while the gradual change of this time error would also significantly influence the computation of road modality, especially in cases in which the distances between the sensors are relatively small.

Another setup parameter, which can affect the final result, is the relative distances between the network sensors. In case sensors are placed very close to each other there will be an overlap between their ranges. Thus, in some places devices will be simultaneously scanned by more than one sensor or with a quite short time difference. In the second option there is a risk that the number of total movements will be overestimated and wrong movement patterns will be identified as there will be many consecutive records between the two sensors. For instance, if a user moves from sensor 1 to sensor 2 and the ranges of the two sensors overlap, the order of the records could be “s1-s1-s1…-s1-s2-s1-s2-s1-s2-s2-s2…-s2” instead of “s1-s1-s1…-s1-s2- s2-

s2-s2…-s2”. One possible solution for this problem is the use of directional antennas as a way to avoid overlapping.

Finally, the influence of the number of sensors to the final outcome was investigated. Different scenarios for the subtraction of one or more sensors were studied. However, the complexity of this research increases as more and more sensors are removed. Having studied all possible cases, it was found that the number of the sensors used, or the number of the ones removed, does not itself influence the final outcome which primarily depends on the locations sensors will be placed at. Thus, the result is not only affected by “how many” sensors will be placed but also by “where” these sensors will be placed. Undoubtedly, it is impossible to get the same number of identified movements using fewer sensors as each of them has a greater or lesser influence. However, it was identified that the influence of the location (where) is sometimes more decisive than the influence of the total number (how many). For instance, the repercussions from taking away one sensor from the main street can be much higher than removing two or three sensors from streets with less traffic. However, it is important to mention that the influence of the setup is strongly related to the main objective. Hence the influence of the setup is different depending on whether the aim of the research is to investigate as many traffic movements as possible or as many streets as possible.

2. What are the Wi-Fi monitoring performance parameters and how can we measure them?

Based on the experience of this work and literature review, as regards the Wi-Fi monitoring performance parameters, the following can be pointed out:

• Accuracy: One of the main parameters used in the sensor Wi-Fi monitoring method is the received signal strength indicator, RSSI. Despite the fact that RSSI is an indicator of the distance between the device and the sensor, this distance cannot be calculated accurately enough as there are many parameters which influence signal strength. Furthermore, the system can identify and store the RSSI but not the incoming signal direction. Thus, the exact location of the device around the sensors, for instance the side of the road the user is on, cannot be accurately defined. Only an estimation of the distance can be made without any information about the direction. On the other hand, the system is pretty accurate when it comes to scanning and storing the signal strength of all devices around it. Undoubtedly, as in all observation methods, how accurate the identification is depends on how far from the sensor a device is; hence the closer the device to the sensor the higher the possibility to be recorded. Given that, a parameter which influences the accuracy of the system with regard to device identification, and as a result of the computation of road modality, is the width of the street. If a street is pretty wide there is a risk that the sensor will not be able to record devices which are close to the boundary of its range leading to unreliable results. Thus, the significance of the zero-level test becomes clear. Finally, based on the validation results of this work, the reliability of the system results, that is the relative percentages of each user category, for example, can be characterized as very satisfactory. However, additional data sources, questionnaires or counting cameras data are always required to evaluate this accuracy.

• Availability: The system's function and performance within the specified coverage area at the start of an intended operation can be characterized as very good. After activating the system, only few minutes are required in order for the sensors to start scanning the research area. This indicator can be measured with the zero-level test as the time difference between the activation of the system and the timestamp of the first records.

• Continuity: Apart from a one-second pause at the end of each scanning period the system is able to scan and store data continuously for a long time without the need for any interruption. The only parameter which can affect the continuity of the system is the storing location of the records. If sensors are connected to a WLAN, the stored data can be uploaded to a database. Otherwise, interrupting the function of the system is inevitable in order to avoid possible problems which may be cause due to local memory overloading.

• Integrity: The integrity of the system itself is quite poor as additional data sources are required in order to evaluate the accuracy of the Wi-Fi monitoring system with regard to road modality computation. However, in case a sensor stops scanning, the system produces a warning message notifying its operator.

• Yield and consistency: To define these performance parameters, outcomes from multiple implementations of the system in different environments are required. However, based on this research they can be considered as quite good and fair respectively. As the range of sensors is significantly larger than the average city street width and a large number of devices can be scanned and stored simultaneously, the system can be applied in different environments and in various region scales. With regard to consistency, it is estimated that the accuracy will be lower in case of traffic jams or in cases in which the movement speed criterion cannot be used so much for the categorization of the device.

• Overhead: The signaling and computational overhead of the system can be characterized as fair. This is a parameter which, in this method, is strongly associated with the size of the network as well as the main aim. For instance, in cases in which road modality in crowded areas is investigated or the aim of the research is not only to create an almost real-time system but also to store metadata for future research and analysis, a very large database and processor are required.

• Power consumption: Sensors do not come with a local battery and need to be continuously supplied with power. Thus, this is a very power demanding system.

• Latency: The latency of the system can be characterized as very low. Just a few seconds after scanning a device, the relevant record is available to be used and analyzed. This means that if the sensor is connected to a WLAN and uploads the stored data to a database within a few seconds after scanning a device, the script for data analysis can be used to identify this device. Thus, the Wi-Fi monitoring system can also be used as an almost real-time system for the computation of road modality.

• Roll-out and operating costs: The cost of the system can considered as moderate. The initial installation of the system, the creation of the necessary script as well as the collection of data for the validation procedure account for the main part of the costs. However, this performance is not an absolute indicator as it always depends on the cost of the alternative methods available.

---

3. What kind of movement patterns can be recognized by the Wi-Fi monitoring system?

Each time a user passes by a scanning point and is scanned, the relevant record is created in the memory of the sensor and thus in the system. Taking into account the timestamp and the region of the records for each device, it is possible to assess user movement behavior in space. By applying this method to the whole set of per hour records, common movement patterns as well as the most frequent ones were identified both for the total set of users and for each user category separately. The identified movement patterns were split into three main categories: patterns between two, three, and four sensors, each of which constitutes approximately 70%, 25%, and 5% of the total set of movements respectively.

---

4. What is the road modality in the research area of Dordrecht during different times of the day and month?

Road modality can be defined as the combined, actual use of a street by different categories of users, including cars, bicycles, pedestrians, etc. It provides information about the relationship between the different categories and each user category percentage during a specific time unit.

In the research area, no significant changes in the overall road modality throughout the days were found. The user categories percentages were pretty much equal apart from those timeslots in which very few users moved in the area leading to notable differences between the relevant percentages. However, when each street was studied separately, significant differences were observed among their modalities. Thus, in some streets the proportion of users was almost the same, unlike other ones which were notably occupied more by one user category throughout the research period.

5. What kind of road modality can be recognized by the Wi-Fi monitoring system?

Based on the use of the Wi-Fi monitoring system, road modality can be divided into three categories of users: those who walk, those who ride bicycles, those who use motorized vehicles such as cars, motorcycles, buses, etc. Thus, the traffic composition of each street can be computed on a per hour basis and useful findings regarding those streets and the whole research area can be generated. There is undoubtedly a risk that some movements identified by the system may be assigned to a wrong user category, false positive error, or may not be captured by the sensors despite their occurrence, false negative error. Moreover, the number of users who move in the area but do not carry a device with its Wi-Fi functionality enabled can either be included into the latter category or can be considered as a general limitation of the method. However, based on the validation results of this research a quite strong relationship between the system’s results and reality is revealed, especially with regard to pedestrians and bicyclists.

---

6. What is the occupancy pattern in the research area of Dordrecht during different times of the day and month?

To plan and design improvement works in the research area, it is essential to know the distribution of people on a spatiotemporal level. The way all users, or each user category, occupy the region between the city centre and the train station throughout the day could provide information on rush hours and recession periods. Furthermore, occupancy patterns observed on different days can be compared in order to identify similarities between them and possibly repeatability of patterns during the week and month.

In the research area of Dordrecht, three main occupancy patterns were recognized. The first one has to do with the period from Monday to Thursday, during which two rush-hour periods were identified; one in the morning and one in the afternoon. Also, a significant recession was observed in-between them. On the other hand, during the weekend there was only one traffic peak around lunchtime with a gradual increase and decline in occupancy before and afterwards.

The third pattern represents the occupancy pattern on Friday, which is quite similar to the first one but the second rush period was longer, as the weekend approached. Furthermore, a high level of repetition was observed in the above-mentioned patterns throughout the four weeks of the researched period. Apart from the identification of rush hours, useful outcomes were obtained about the occupancy level between the different days of the week, showing a significant reduction in the total number of movements during the weekend.

7. Which occupancy patterns can be recognized by the Wi-Fi monitoring system?

Using a Wi-Fi monitoring system, various information about the occupancy of a region during a timescale can be collected. Based on the system's ability to scan those who have enabled the Wi-Fi functionality of their devices at a certain points, the occupancy level of the research area can be studied. Apart from congestions, recessions, and trends throughout the day, differences between the relevant occupancy levels can be noted. Moreover, an important advantage of the Wi-Fi monitoring system is the fact that it allows for the characterization of each device as “pedestrian”, “bicyclist”, or “vehicle” and the identification of occupancy patterns for each user category separately.

8. Is it possible to identify the effect of the weather on road modality?

Taking into account the outcomes of this work, it can be stated that no notable differences in road modality due to weather changes were observed in the research area. Cases with the highest weather condition differences were studied. Thus, same days of the week with a temperature difference higher than ten degrees Celsius or sunny and rainy days were used to compare road modality. The duration of the data collection period may be considered as not being long enough for this kind of investigation and data from different seasons are required.

Limitations

Overall, the effectiveness of the investigation of road modality and the identification of occupancy patterns using a Wi-Fi monitoring system is validated based on its results. However, there are some aspects which affected the final outcome of this research and have to be considered. First of all, due to technical issues and lack of internet access, data from the sensors could not be “uploaded” to a server database. Thus, manual downloading was required every two days in order to clear the scanners' local memory and avoid any problems due to overloading. Apart from the practical difficulties, there was a time period of some hours during which not all sensors worked simultaneously. During data downloading, scanning of the local area had to be interrupted and thus every half hour one different sensor was paused. As a consequence, the amount of data available for the computation of road modality of surrounding streets decreased.

Moreover, despite the validation of the results and their increased correlation with reality, the sampling tests size and the number of those who filled out the questionnaire were quite small; so reliable conclusions about the accurate application of the system for this purpose could not be drawn. The main reason for this limitation is that it the initial plan was not followed. According to this plan, students of the Da Vinci College of Dordrecht would support this research by applying the questionnaire and the random samplings for a few hours every day throughout the data collection period. Thus, the whole procedure of validation was done only by one person. Thus, the amount of data collected was significantly smaller than initially expected. Furthermore, apart from the use of the questionnaire and the random sampling tests, no other source of data could be used to enhance the validation procedure.

Finally, one of the conclusions drawn from this work is that the duration of the data collection period was not long enough to allow for more detailed investigation of the weather influence on road modality. During the one-month observation period there were not any significant changes in the weather conditions of the research area; on most days the weather was sunny with no significant changes in temperature. Based on that, datasets from a longer period could be required to fulfill the aim of such an investigation.

7. Conclusions and future research

This research has contributed to the ongoing investigation concerning the computation of road modality and occupancy patterns with the use of an outdoor Wi-Fi monitoring system, as a way to support the main rationale of the ‘Smart cities’ concept.

In designing the observation network, both findings from literature review about the method’s technical attributes and relevant applications were taken into account. Parameters such as the range of the sensors, the width of the streets, the need for electric power supply, the number of available scanners and the street network in the research area were the main factors which determined the way the system was designed as well as the exact position each sensor was placed at. The ability to detect all devices with enabled Wi-Fi functionality within the range of each sensor is the main advantage of this method. Based on that, the movement behavior of each device in the region covered by the installed sensors can be recorded, leading to the identification of similarities between movements and of the most frequent movement behaviors. Knowing the distance between sensors and using the time difference between the records of a device on consecutive sensors, the relevant movement speed can be computed. This computed speed can be combined with street-uses criteria for the recognition of different user categories and, thus device characterization. Based on the classification of devices into different categories, each street's road modality can be computed. The relationship between these categories throughout the day can be studied and preferred streets as well as movement patterns for different kinds of users can be recognized.

An important advantage of the Wi-Fi monitoring system is its ability to measure the number of users in a certain place at a specific time, which allows for the identification of occupancy patterns both for users as a whole and for each user category separately. Rush hours, recession periods and movement trends can be recognized for the different days of the week as well as the occupancy relationship between different regions in large research areas.

The system itself cannot provide information about the computation accuracy and the reliability of the outcomes. Additional data sources, such as pedestrian counting cameras and inductance loop systems, are required to validate the system. However, a quite strong correlation between the system’s results and reality has been found. Based on the validation procedures, the accuracy of the Wi-Fi monitoring method can be characterized as particularly high.

Finally, based on literature review as well as on the experience from this research, there were some useful findings about the setup parameters and their influence on the outcomes, and the reliability of the system. The choice of the appropriate kind of antenna, a directional or a non-directional antenna, determines how the scanned area (research area) is defined in order to ensure the appropriateness of the recorded data and at the same time minimize the filtering procedure. Moreover, a significant parameter which influences the outcomes of the system is the total number of the available sensors in relation to the size of the research area and the number of streets. However, based on the findings of this research, the number of available sensors cannot be considered as the only contributing factor, since each sensor's exact position is also critical and must be taken into account when designing the observation network. Thus, parameters such as the need for waterproofing and continuous power supply also affect the final outcome, but indirectly, as they are taken into account in selecting the exact position of sensors.

The research's outcomes are discussed here and some recommendations are provided for future research.

Recommendations

As already mentioned, although the influence of some parameters, such as the total number of sensors, on the final result is investigated in this research, the identification of the optimal Wi-Fi network configuration is not. However, future studies could lead to useful findings about the outdoor application of the Wi-Fi monitoring system for similar purposes; thus some recommendations can be provided based on the literature review and the experience of this thesis. First of all, cases with overlapping in sensor ranges can be examined taking advantage of the system's applications in indoor environments and using it as foundation. Additionally, the system can be used in larger research areas and different environments, such as traffic jams, in order to evaluate the range of its applicability and the reliability of its outcomes. Despite the use of fixed speed criteria for the computation of road modality, it would be interesting to study the use of fluctuated speed limits depending on time. For example, in cases of traffic jams the average speed in streets reduced markedly and, thus, a device which belongs to a vehicle could be characterized by the system as bicyclist. Moreover, the total number of scanned devices can also be taken into account and used as an indicator for choosing speed limitations.

Apart from the speed and the street-use criteria, another parameter that could be investigated is the use of the total number of records at each sensor for the different user categories. As the movement speed is different, pedestrians need more time than bicyclists to cross the scanned area of a sensor and even more time than a vehicle. Based on that, it is expected that the total number of consecutive records at each sensor would be different for devices that belong in different categories and, thus, this can be taken into account as an extra parameter for the computation of road modality. Furthermore, the use of longer timeslots instead of hourly sets can be examined, in particular with regard to identifying meeting points and frequently visited places.

With regard to the system validation, the use of additional data sources, such as pedestrian counting cameras and inductance loop systems, is strongly recommended. Moreover, comparing at the same time the system's results with real data for all user categories is also suggested. Either a video camera system or the method of counting can be used for this kind of validation, leading to the simultaneous validation of the overall road modality. Additionally, a research on the average number of people per each vehicle as well as the number of devices per user is strongly suggested. They are two parameters which can significantly improve the final reliability of the Wi-Fi monitoring system and eliminate overestimation phenomena.

From a technical perspective, the systematic uploading of records to an online database as well as the use of external batteries in combination with solar panels are both strongly recommended. In this way, the local device memory will not be affected by any overloading problems and the degree of freedom in the choice of sensors' locations will increase significantly. Technical suggestions are further described in Appendix A.

Finally, some recommendations can be provided also for the part of the data analysis. The identification of a higher number of user categories can be investigated, especially in order to divide the category of vehicles into subcategories. Furthermore, as each device is represented by a unique ID in the system, it can be quite interesting to study the repeatability in regard to the use of means of transport. For example, each day can be split into four time periods - morning, noon, afternoon, and night- and whether there is any record for each device and its relevant characterization can be studied for each time period. However, the outcomes will be associated with devices and not users, as it is impossible to know whether each device is always used by the same person or not. Moreover, the calculation of the average speed of each user category during different time periods throughout the day is suggested. Comparing the average speed with the occupancy level based on this calculation could lead to useful findings about the relationship between these two parameters. Also, the relevant capacity level for each street can be computed, and act as an indicator of the quality of service provided to users and the potential need for improvement works.

References

Andrienko G., Andrienko N., Kopanakis I., Ligtenberg A., and Wrobel S. (2008) "Visual analytics methods for movement data" in Giannotti F. and Pedreschi D. (eds) "*Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*" (pp. 375–410) Berlin, Springer

Apple Inc. (2016) "iOS Security Guide - White Paper"
Downloaded from (https://www.apple.com/business/docs/iOS_Security_Guide.pdf)

Autoriteit Persoonsgegevens (2015) "Dutch DPA investigates WiFi tracking in and around shops". Downloaded from (https://autoriteitpersoonsgegevens.nl/en/news/dutch-dpa-investigates-wifi-tracking-and-around-shops)

Bakker W. (2016) "The analysis of passive Wi-Fi tracking" University of Twente, Faculty EEMCS, the Netherlands. Downloaded from (http://essay.utwente.nl/71218/)

Ballon P., Glidden J., Kranas P., Menychtas A., Ruston S., and Van Der Graaf S. (2011) "Is there a Need for a Cloud Platform for European Smart Cities?" in "*eChallenges e-2011 Conference Proceedings*" Florence, Italy

Borcher J.G., Van Amersfoort J.M.M., Berendsen H.J.A., Druijven P.C.J., Kouwenhoven A.O., and Scholten H. (2004) "The use of maps in the exploration of geographic data" *Geographical Studies* 326, Netherlands

Bussche D. and Van de Coevering P. (2015) "Bikeprint – in depth analysis of cyclist behaviour and cycle network performance using extensive GPS-track data" *European transport conference*, Frankfurt, Germany

Datatilsynet (2016) "Tracking in Public Spaces. The use of WiFi, Bluetooth, beacons and intelligent video analytics" *The Norwegian Data Protection Authority*. Downloaded from (https://www.datatilsynet.no/English/Publications/tracking-in-public-spaces/)

Deak G., Curran K., and Condell J. (2011) "Filters for RSSI-based measurements in a Device free Passive Localization Scenario" *International Journal on Image Processing & Communications* Vol. 15-1 (pp.23-34)

Deakin M. (2014) "Smart Cities. Governing, modelling and analyzing the transition" Routledge, New York.

Department of Business for Innovation & Skills (2013). "Smart Cities".UK.

Department of Economic and Social Affairs (2015) "World Urbanization Prospects" United Nations. New York.

Dodge S., Weibel R., and Lautenschütz A. (2008) "Towards a taxonomy of movement patterns" *Information Visualization* vol.7-3,4 (pp. 240-252)

Duynstee C., Haayen M., Kyritsis D., Ortega-Cordova L., and Samat S. (2016) "Smart City Dordrecht: Identification of Pedestrian Movement Patterns with Wi-Fi Tracking Sensors" TU Delft. Faculty of Faculty Architecture and The Built Environment. Downloaded from (uuid:1dcf6c6b-67c4-4f80-bec9-eeba9c26b939)

European Digital Rights (2015) "WiFi tracking and the ePrivacy Directive in Denmark" Downloaded from from (https://edri.org/wifi-tracking-eprivacy-directive-denmark/)

European Commission. (2013) “European Innovation Partnership on Smart Cities and Communities” in *Strategic Implementation Plan*

European Commission (2017) “Mobility and Transport. Road Safety” Downloaded by (https://ec.europa.eu/transport/road_safety/topics/vehicles/vehicle_categories_en)

Goldreich Od. (2010) “P, NP, and NP-Completeness: The Basics of Computational Complexity” Cambridge University Press

Grimaldi R. (2003) “Discrete and Combinatorial Mathematics: An Applied Introduction” Pearson Editions

Gudmundsson, J., Laube, P., and Wolle, T. (2008) “Movement Patterns in Spatio-Temporal Data” in Shekhar, S., Xiong, H., (eds) *Encyclopedia of GIS*, Springer, Heidelberg.

Hallenbeck M., Rice M., Smith B., Cornell-Martinez C., and Wilkinson J. (1997)“Vehicle volume distributions by classification” Washington State Transportation Center, Downloaded by (http://depts.washington.edu/trac/bulkdisk/pdf/VVD_CLASS.pdf)

Henniges, R. (2012). “Current approaches of Wifi Positioning” TU-Berlin, Germany

Hofmann B., Lichtenegger H., and Wasle E. (2008) “GNSS - Global Navigation Satellite Systems” Springer, Wien New York

Iliadi A. (2016) “A study into the joint activity-route choice behaviour of pedestrians during large scale events, based on revealed preference data SAIL event, case study”. Master thesis. TU Delft, the Netherlands

Jung W. R., Bell S., Petrenko A., and Sizo A. (2012) “Potential risks of WiFi-based indoor positioning and progress on improving localization functionality” in *Proceedings of the Fourth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness* (pp.13–20), Redondo Beach, California

Komninos N., Pallot M., and Schaffers H. (2013) “Special Issue on Smart Cities and the Future Internet in Europe” in *Journal of the Knowledge Economy* Vol.4-2, Springer

Kwan M.P. (2004) “GIS methods in time geographic research: Geocomputation and geovisualization of human activity patterns” *Geografiska Annaler*, Series B, Human Geography 86, (pp. 267–80)

Lui, G., Gallagher T, and Binghao L.(2011) “Differences in RSSI readings made by different Wi-Fi chipsets: A limitation of WLAN localization” International Conference on Localization and GNSS (ICL-GNSS), Downloaded from (http://ieeexplore.ieee.org/document/5955283/?tp=&arnumber=5955283)

Meneses F. and Moreira A. (2012) “Large scale movement analysis from WiFi based location data” in *Proceedings of the 3rd International Conference on Indoor Positioning and Indoor Navigation*, Sydney, Australia

Miller M. (2015) “The Internet of things: how smart TVs, smart cars, smart homes, and smart cities are changing the world” *Que*. Indianapolis, Indiana, USA

Musa A. and Eriksson J. (2012) “Tracking Unmodified Smartphones Using Wi-Fi Monitors” SenSys Vol.12, Toronto, Canada

Oldreich Od. (2010). “P, NP, and NP-Completeness: The Basics of Computational Complexity” Cambridge University Press

Orellana D., Wachowicz M., De Knegt H. Ligtenberg A., and Bregt A. (2010) “Uncovering patterns of suspension of movement” in *Proceedings of the Sixth International Conference on Geographic Information Science,* Zurich, Switzerland

Orellana D., Ligtenberg A., Wachowicz M. and Bregt A. (2012) “Exploring visitor movement patterns in natural recreational areas” *Tourism Management* 33: (pp.672-682) Elsevier

Radu V. and Marina M. K. (2013) “HiMLoc: Indoor smartphone localization via dead reckoning with selective crowdsourced WiFi fingerprinting” in *Proceedings of the Fourth International Conference on Indoor Positioning and Indoor Navigation*, Montbeliard-Belfort, France

Rastogi R., Ilango T., and Chandra S. (2013) “Pedestrian Flow Characteristics for Different Pedestrian Facilities and Situations” *European Transport*, Issue 53, Trieste, Italy

Retscher, G., and Fu, Q. (2008) “Active RFID trilateration for indoor positioning” *Coordinates*, Downloaded from (http://mycoordinates.org/active-rfid-trilateration-for-indoor-positioning/all/1/)

Robinson R. (2016) “Why Smart Cities still aren’t working for us after 20 years. And how we can fix them.” Downloaded from (https://theurbantechnologist.com/2016/02/01/why-smart-cities-still-arent-working-for-us-after-20-years-and-how-we-can-fix-them/)

Rose I., and Welsh. M. (2010) “Mapping the urban wireless landscape with Argos” SenSys '10, in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, Zürich, Switzerland

Stevens M. (2007) “On Collisions for MD5”. Master thesis. Eindhoven University of Technology, the Netherlands

Thiagarajan A., Ravindranath L., Balakrishnan H., Madden S., and Girod L. (2011) “Accurate, low-energy trajectory mapping for mobile devices” in *Proceedings of the Eighth USENIX Conference on Networked Systems Design and Implementation*, (pp. 267-280) Boston, Massachusetts, USA

U.S. Department of Transportation (2016) “Traffic Monitoring Guide”, Washington, USA, Downloaded from (https://www.fhwa.dot.gov/policyinformation/tmguide/)

Van der Spek. S., Van Langelaar M., and Kickert C. (2013) “Evidence-based design: satellite positioning studies of city centre user groups” in *Proceedings of the Institution of Civil Engineers,* Vol. 166, Issue DP4

Van der Spek S., Verbree E., Quak C.W., Groeneveld IJ.D.G., Sulzer R., Theocharous E., Tryfona M.S., Willems O.T., and Xu Y. (2016) “Track-id: Activity Determination based on Wi-Fi Monitoring” in *Proceedings of the 13th International Conference on Location Based Services*. University of Technology, Vienna, Austria

Van Loenen B., Kulk S., and Ploeger H. (2016) “Data protection legislation: A very hungry caterpillar. The case of mapping data in the European Union” *Government Information Quarterly,* Vol.33: (pp. 338–345), Elsevier

Vanhoef M., Matte C., Cunche M., Cardoso L., and Piessens F. (2016) “Why MAC Address Randomization is not enough: An Analysis of Wi-Fi Network Discovery Mechanisms” in *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*. ACM, New York, USA

VCA-Vehicle Type Approval UK (2017) “Definition of vehicle categories” Downloaded from (http://www.dft.gov.uk/vca/vehicletype/definition-of-vehicle-categories.asp)

Verbree E., Van der Spek S.C, Kalogianni E., Sileryte R., Lam M., Zhou K., and Van der Ham M. (2015) “Passive WiFi Monitoring of the Rhythm of the Campus” in *Proceedings of the 18th AGILE International Conference on Geographic Information Science*, Geographics Information Science as an Enabler of Smarter Cities and Communities, Lisboa, Portugal

Zervos M. (2013) “Multi-camera face detection and recognition applied to people tracking” Ecole Polytechnique Federal, Lausanne, Switzerland

Zhou M., Tian Z., Xu K., Yu X., Hong X., and Wu H. (2014) “SCaNME: Location tracking system in large-scale campus Wi-Fi environment using unlabeled mobility map” *Expert Systems with Applications,* Vol.41: (pp. 3429–3443), Elsevier

Appendices

Appendix A. Limitations, suggestions for Meshlium sensors

At the end of this study, based on the literature review and the experience from the use of the Libelium Meshlium sensors for the implementation of the Wi-Fi monitoring system, it is possible to draw some conclusions about these devices and their attributes. In general, it can be noted that these devices are very well constructed and their technical characteristics allow for outdoor use regardless of weather conditions. However, there are some parameters which need to be investigated more extensively and improved, especially in cases with a similar purpose.

First of all, it is crucial to eliminate the problems in the timestamp of sensors. During the placement of sensors in the research area as well as after downloading the data, the time zone was checked and synchronized using the same external clock. However, in many cases the time of the devices was wrong, as it was significantly earlier or later than the actual time. After downloading the data, the time was corrected, but the same problem appeared during the next efforts with various deviations from the actual time. Based on this problem and in combination with the need for full time synchronization of the sensors for this kind of application, it is clear that this time error significantly affects the analysis procedure and the final outcome. Despite the fact that the movement of the devices is recorded throughout the day, the relevant orders in the area can be wrong due to these time errors. In order to solve this problem, a different kind of time can be used as the timestamp of the records. As research has shown, there are two kinds of time in each device. The first is the system time which is the problematic one and is used as the record time, while the second is the hardware clock that remains always synchronized. Thus, through the direct use of the latter time or the computation of the difference between the wrong and correct time in each sensor, this problem can be overcome.

Furthermore, some more suggestions can be made which can lead to notable improvements with regard to the applicability of sensors. Despite the fact that it is possible to automatically hash the recorded MAC addresses during the scanning of the research area, it is necessary to ensure that all devices hash the same device in the same way and use the same hashed MAC address. Apart from the different sensors, the identical hashing of each device should also be applied for different days. Otherwise, it would be impossible to investigate users' movement behavior throughout the day or its repeatability. Moreover, the replacement of MD5 with another and newer hashing algorithm is suggested as based on the literature review collisions is possible to be found in MD5 and thus is severely compromised (*Stevens, 2007*).

Despite the ability of Meshlium sensors to record the RSSI of the transmitted signals, another significant parameter, which could be very useful, is the simultaneous storing of the signal direction. In this way, in combination with the estimated distance from the sensor, it would be feasible to compute the relevant device location at the time of the record as well as to use angle or direction threshold during the data collection procedure. Last but not least, the use of an internal battery as well as a base which rotates to various directions can offer a higher degree of independency regarding the choice of the sensors' location and a more precise selection of the scanned area too.

Appendix B. Zero-level test

Average RSSI values at different locations around each sensor. Google earth figures are used as background layer.

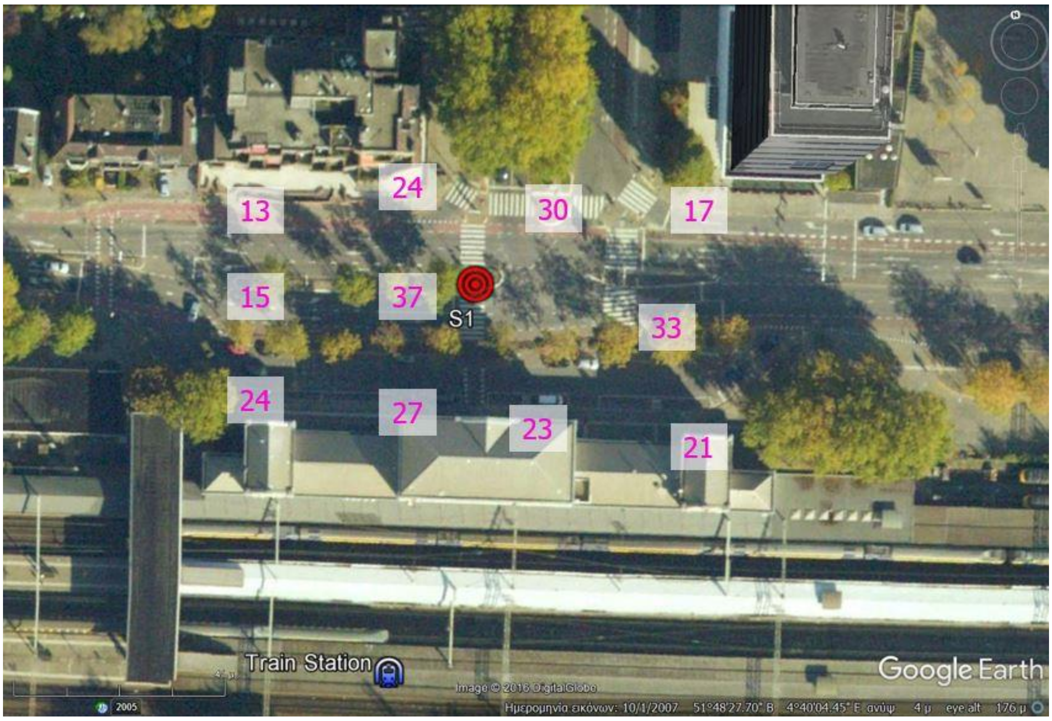


Figure 84: Average RSSI values around sensor 1

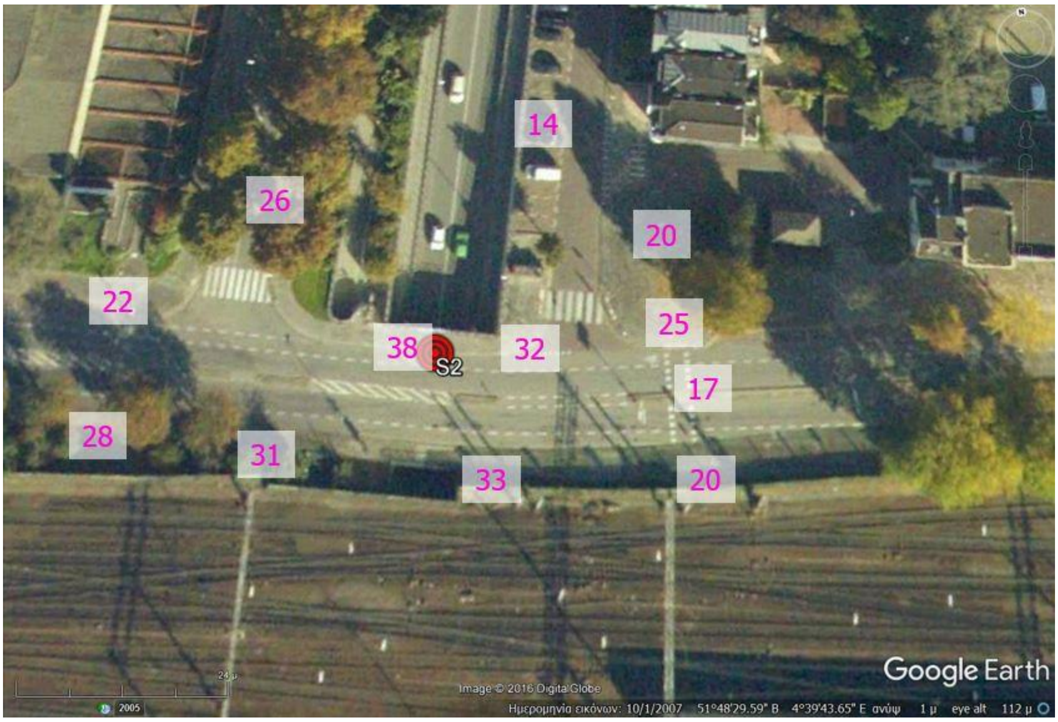


Figure 85: Average RSSI values around sensor 2

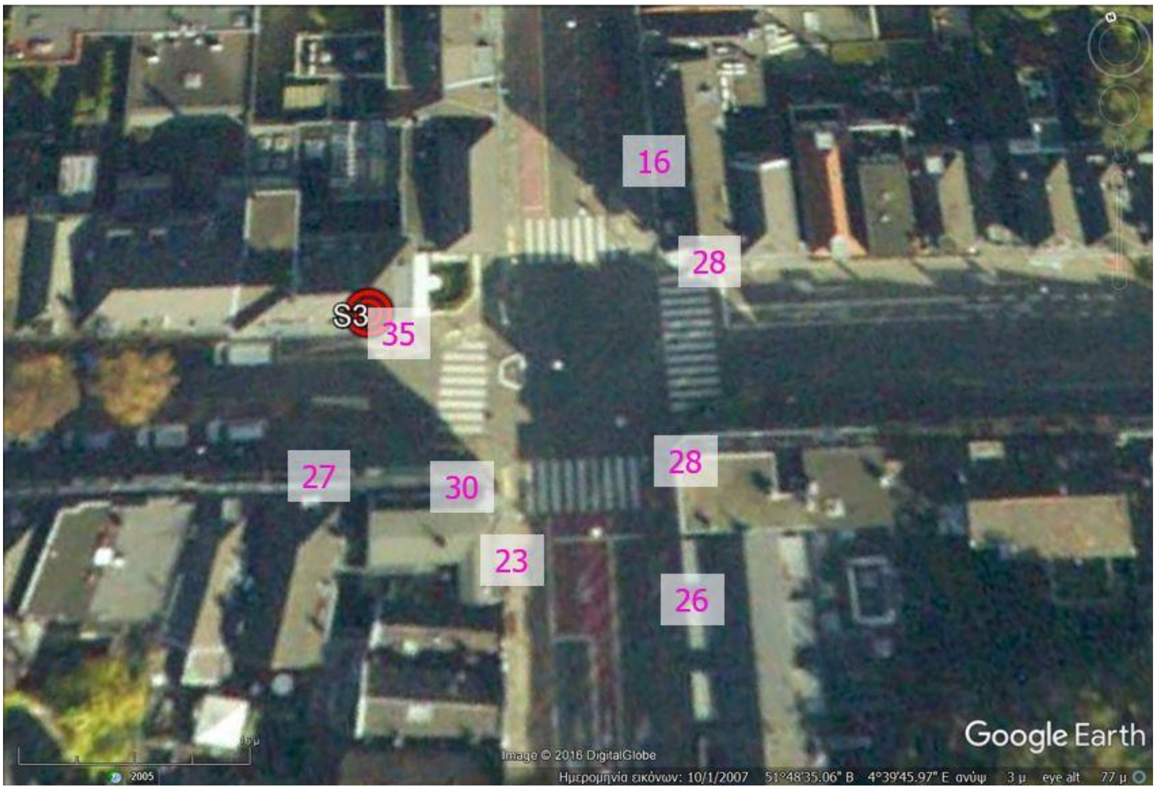


Figure 86: Average RSSI values around sensor 3

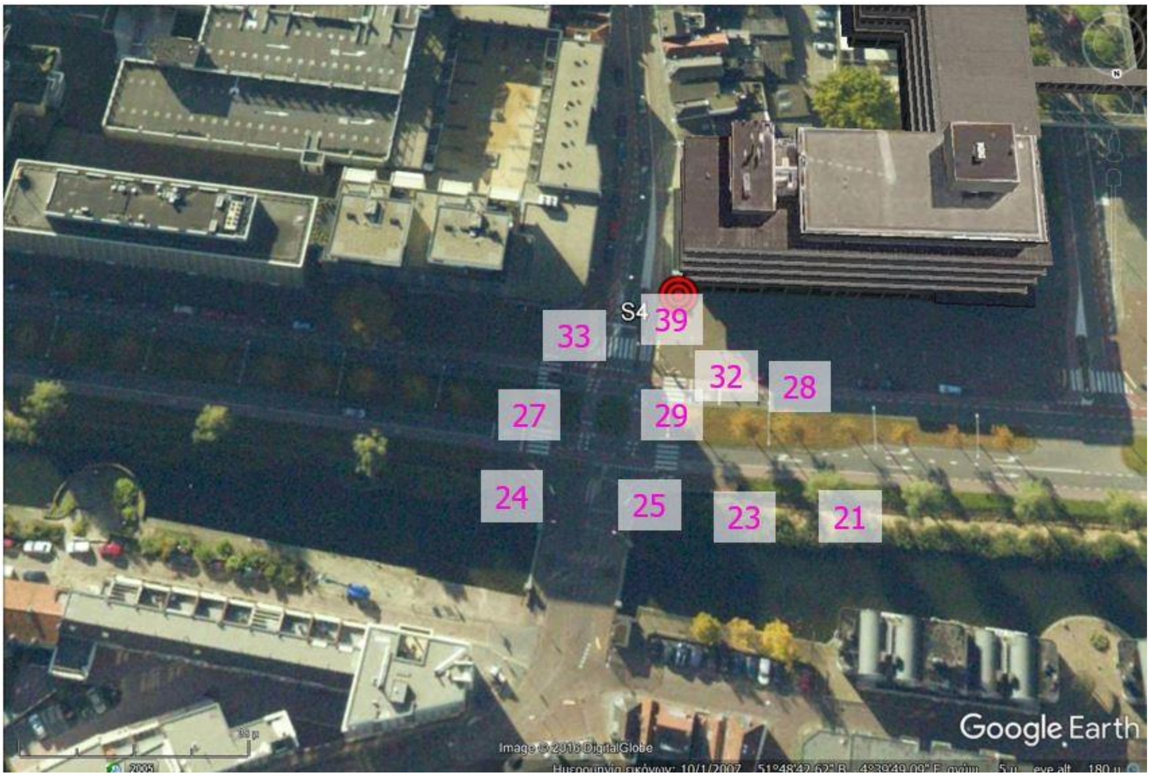


Figure 87: Average RSSI values around sensor 4

Figure 88: Average RSSI values around sensor 5

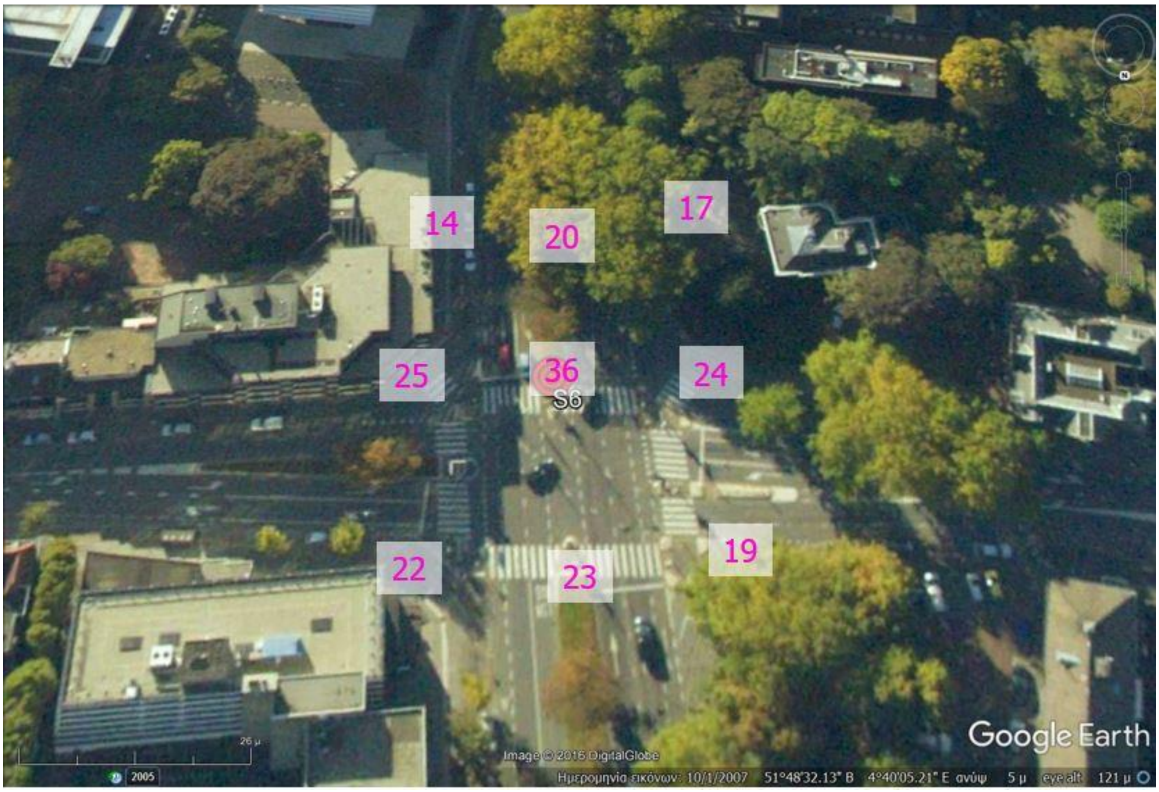


Figure 89: Average RSSI values around sensor 6

Figure 90: Average RSSI values around sensor 7

Figure 91: Average RSSI values around sensor 8

Appendix C. Diagrams - Tables

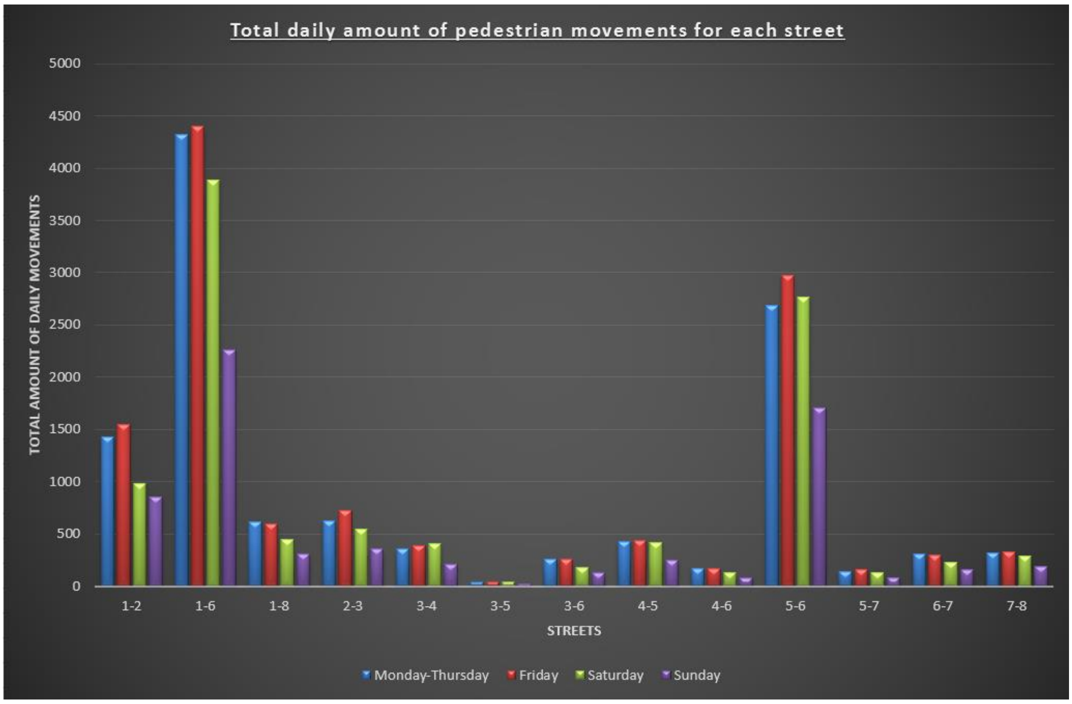


Figure 92: Daily amount of pedestrian movements for each street

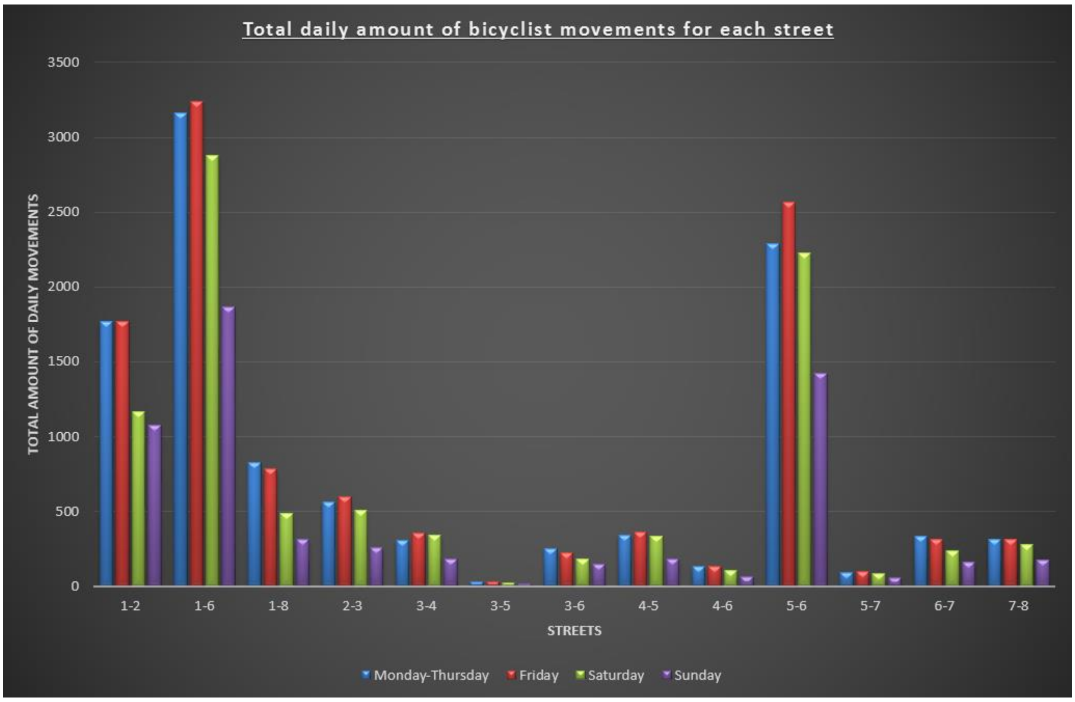


Figure 93: Daily amount of bicyclist movements for each street

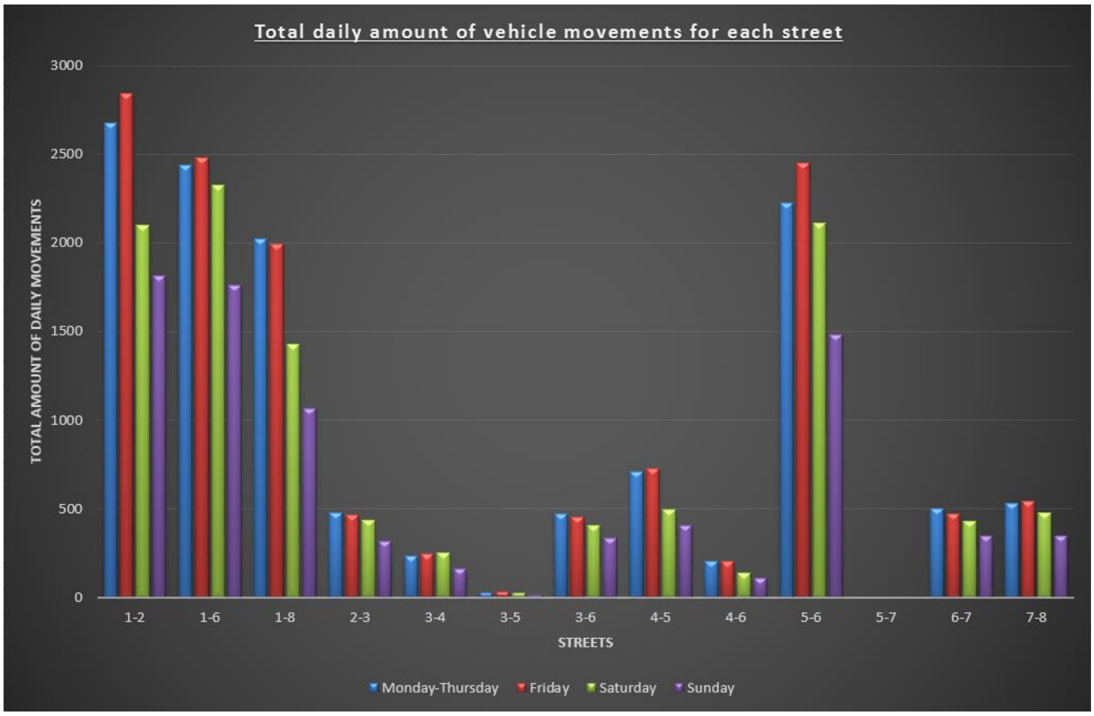


Figure 94: Daily amount of vehicle movements for each street

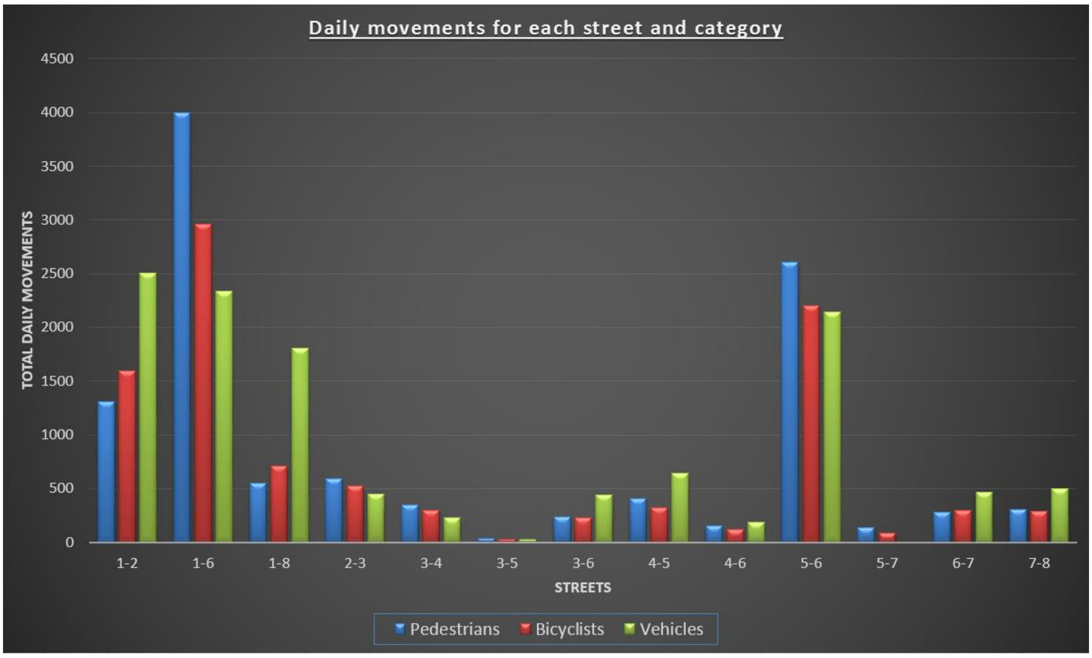


Figure 95: Daily amount of movements for each users' category

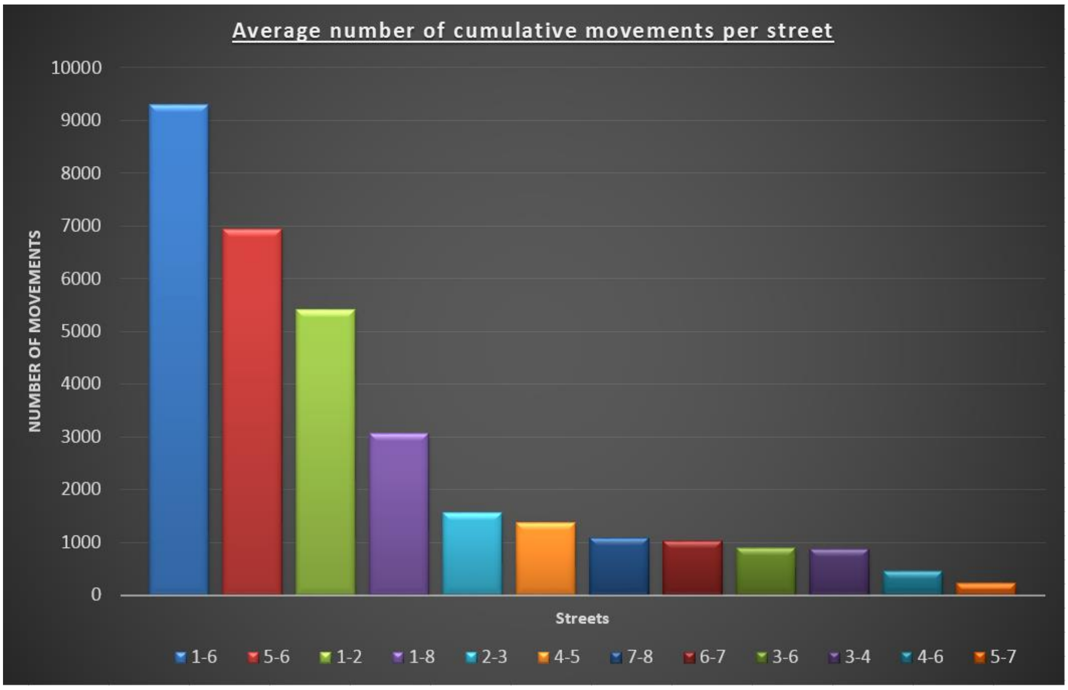


Figure 96: Average daily number of cumulative movements per street

Patterns of 3 sensors

Pedestrians Patterns	Daily Amount	Relative% (absolute%)	Bicyclists Patterns	Daily Amount	Relative% (absolute%)	Vehicles Patterns	Daily Amount	Relative% (absolute%)
165	1014	40.1 (8.2)	**165**	884	34.4 (8.2)	**561**	646	18.2 (4.7)
561	796	31.4 (6.4)	**561**	723	28.2 (6.7)	**165**	576	16.3 (4.2)
234	96	3.8 (0.8)	**234**	141	5.5 (1.3)	**812**	478	13.5 (3.5)
612	68	2.7 (0.6)	**218**	102	4.0 (1.0)	**218**	251	7.1 (1.8)
167	55	2.2 (0.5)	**812**	85	3.3 (0.8)	**542**	177	5.0 (1.3)
216	55	2.2 (0.5)	**167**	62	2.4 (0.6)	**816**	144	4.1 (1.1)
432	44	1.8 (0.4)	**216**	56	2.2 (0.5)	**234**	144	4.1 (1.1)
164	41	1.6 (0.3)	**618**	49	1.9 (0.5)	**612**	125	3.5 (0.9)
461	38	1.5 (0.3)	**461**	46	1.8 (0.4)	**456**	121	3.4 (0.9)
456	38	1.5 (0.3)	**432**	46	1.8 (0.4)	**216**	116	3.3 (0.9)

Table 13: The most frequently used movement patterns of 3 sensors for each user category for Friday

Pedestrians Patterns	Daily Amount	Relative% (absolute%)	Bicyclists Patterns	Daily Amount	Relative% (absolute%)	Vehicles Patterns	Daily Amount	Relative% (absolute%)
561	871	38.9 (8.3)	**165**	826	38.2 (9.3)	**561**	615	20.7 (5.4)
165	767	34.2 (7.3)	**561**	653	30.3 (7.4)	**165**	536	18.0 (4.7)
234	146	6.5 (1.4)	**234**	133	6.2 (1.5)	**812**	328	11.0 (2.9)
432	69	3.1 (0.7)	**432**	90	4.2 (1.0)	**218**	276	9.3 (2.4)
612	46	2.1 (0.4)	**456**	55	2.5 (0.6)	**542**	203	6.8 (1.8)
461	46	2.1 (0.4)	**618**	55	2.5 (0.6)	**612**	141	4.7 (1.2)
456	46	2.1 (0.4)	**216**	47	2.2 (0.5)	**618**	125	4.2 (1.1)
654	35	1.5 (0.3)	**461**	47	2.2 (0.5)	**216**	115	3.9 (1.0)
216	35	1.5 (0.3)	**164**	39	1.8 (0.4)	**816**	104	3.5 (0.9)
563	35	1.5 (0.3)	**816**	39	1.8 (0.4)	**236**	99	3.3 (0.9)

Table 14: The most frequently used movement patterns of 3 sensors for each user category for Saturday

Pedestrians Patterns	Daily Amount	Relative% (absolute%)	Bicyclists Patterns	Daily Amount	Relative% (absolute%)	Vehicles Patterns	Daily Amount	Relative% (absolute%)
165	539	37.6 (8.1)	**561**	471	33.0 (8.0)	**561**	533	23.6 (6.2)
561	448	31.2 (6.7)	**165**	424	29.7 (7.2)	**165**	345	15.3 (4.0)
234	91	6.4 (1.4)	**234**	148	10.4 (2.5)	**812**	208	9.2 (2.4)
812	67	4.7 (1.0)	**236**	43	3.0 (0.7)	**218**	168	7.4 (1.9)
432	51	3.6 (0.8)	**432**	39	2.7 (0.7)	**456**	96	4.3 (1.1)
216	36	2.5 (0.5)	**461**	31	2.2 (0.5)	**618**	91	4.0 (1.1)
612	32	2.2 (0.5)	**654**	27	1.9 (0.5)	**167**	86	3.8 (1.0)
167	32	2.2 (0.5)	**765**	27	1.9 (0.5)	**216**	81	3.6 (0.9)
578	24	1.7 (0.4)	**761**	27	1.9 (0.5)	**816**	81	3.6 (0.9)
164	24	1.7 (0.4)	**781**	27	1.9 (0.5)	**612**	76	3.4 (0.9)

Table 15: The most frequently used movement patterns of 3 sensors for each user category for Sunday

Patterns of 4 sensors

Pedestrians Patterns	Daily Amount	Relative% (absolute%)	Bicyclists Patterns	Daily Amount	Relative% (absolute%)	Vehicles Patterns	Daily Amount	Relative% (absolute%)
1654	115	23.9 (0.9)	**1654**	176	32.1 (1.6)	**4561**	166	20.6 (1.2)
4561	115	23.9 (0.9)	**5618**	113	20.5 (1.0)	**1654**	158	19.6 (1.2)
5618	50	10.4 (0.4)	**4561**	92	16.7 (0.9)	**8165**	117	14.5 (0.9)
2165	50	10.4 (0.4)	**8165**	63	11.5 (0.6)	**5618**	109	13.6 (0.8)
8165	43	9.0 (0.4)	**2165**	35	6.4 (0.3)	**2167**	64	7.9 (0.5)
7612	29	6.0 (0.2)	**5612**	28	5.1 (0.3)	**7812**	49	6.1 (0.4)
5612	22	4.5 (0.2)	**2365**	14	2.6 (0.1)	**7612**	45	5.6 (0.3)
2345	22	4.5 (0.2)	**5632**	14	2.6 (0.1)	**2345**	45	5.6 (0.3)
2367	22	4.5 (0.2)	**7632**	14	2.6 (0.1)	**2165**	23	2.8 (0.2)

Table 16: The most frequently used movement patterns of 4 sensors for each user category for Friday

Pedestrians Patterns	Daily Amount	Relative% (absolute%)	Bicyclists Patterns	Daily Amount	Relative% (absolute%)	Vehicles Patterns	Daily Amount	Relative% (absolute%)
5618	82	19.6 (0.8)	**1654**	99	21.8 (1.1)	**5618**	132	21.1 (1.2)
4561	82	19.6 (0.8)	**4561**	99	21.8 (1.1)	**8165**	124	19.7 (1.1)
2165	64	15.2 (0.6)	**5618**	81	17.9 (0.9)	**4561**	107	17.1 (0.9)
2345	55	13.0 (0.5)	**8165**	75	16.7 (0.9)	**1654**	70	11.2 (0.6)
5612	55	13.0 (0.5)	**2345**	70	15.4 (0.8)	**2345**	50	7.9 (0.4)
8165	46	10.9 (0.4)	**2165**	17	3.8 (0.2)	**5612**	33	5.3 (0.3)
1654	36	8.7 (0.3)	**2167**	12	2.6 (0.1)	**7612**	29	4.6 (0.3)
						2167	25	3.9 (0.2)
						2165	25	3.9 (0.2)
						2367	21	3.3 (0.2)

Table 17: The most frequently used movement patterns of 4 sensors for each user category for Saturday

<u>Pedestrians</u> Patterns	Daily Amount	Relative% (absolute%)	<u>Bicyclists</u> Patterns	Daily Amount	Relative% (absolute%)	<u>Vehicles</u> Patterns	Daily Amount	Relative% (absolute%)
5612	56	20.6 (0.8)	**1654**	68	20.9 (1.1)	**5618**	90	16.7 (1.1)
5432	40	14.7 (0.6)	**4561**	61	18.6 (1.0)	**4561**	76	14.0 (0.9)
4561	40	14.7 (0.6)	**5618**	53	16.3 (0.9)	**8165**	72	13.3 (0.8)
5618	32	11.8 (0.5)	**2165**	38	11.6 (0.6)	**1654**	65	12.0 (0.8)
8165	24	8.8 (0.4)	**8165**	38	11.6 (0.6)	**7612**	50	9.3 (0.6)
1654	24	8.8 (0.4)	**2365**	30	9.3 (0.5)	**2167**	43	8.0 (0.5)
2345	24	8.8 (0.4)	**7612**	23	7.0 (0.4)	**2345**	43	8.0 (0.5)
2367	16	5.9 (0.2)	**2167**	15	4.7 (0.3)	**2367**	22	4.0 (0.3)
4321	16	5.9 (0.2)				**7812**	18	3.3 (0.2)

Table 18: The most frequently used movement patterns of 4 sensors for each user category for Sunday

<u>Occupancy patterns</u>

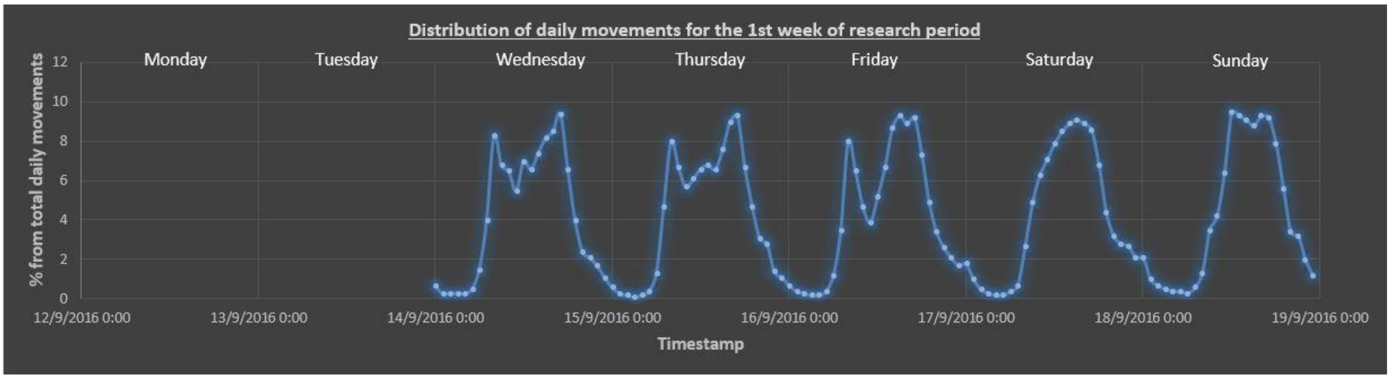


Figure 97: Distribution of daily movements for the first week of the research period

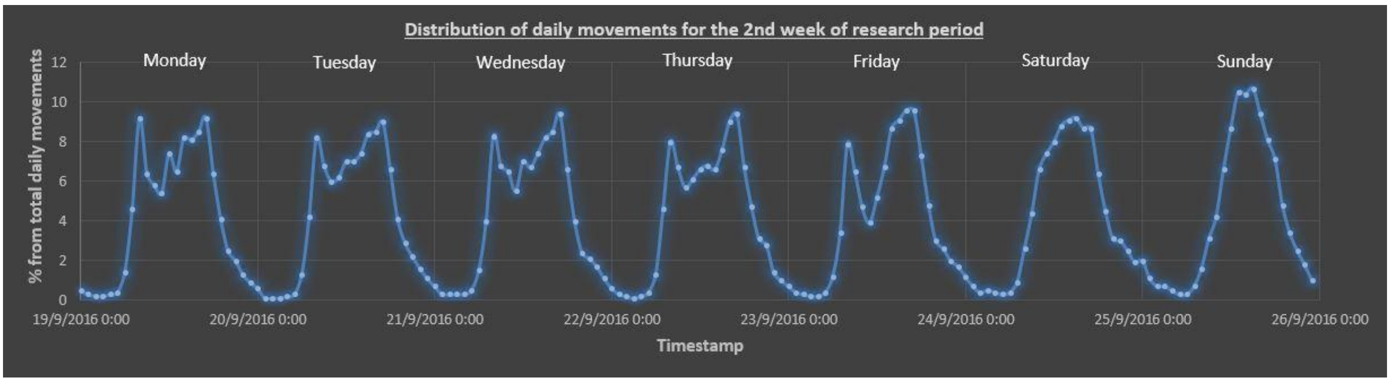


Figure 98: Distribution of daily movements for the second week of the research period

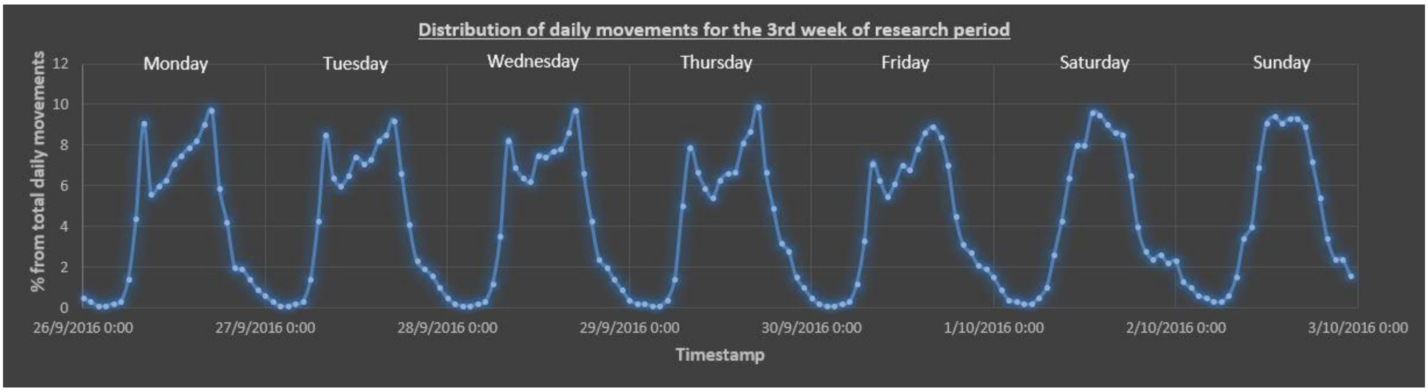


Figure 99: Distribution of daily movements for the third week of the research period

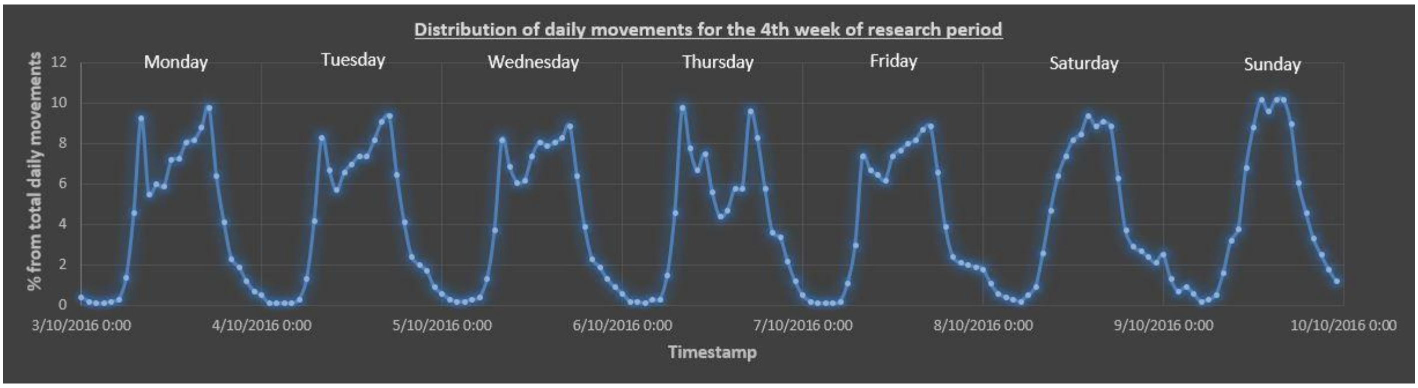


Figure 100: Distribution of daily movements for the fourth week of the research period

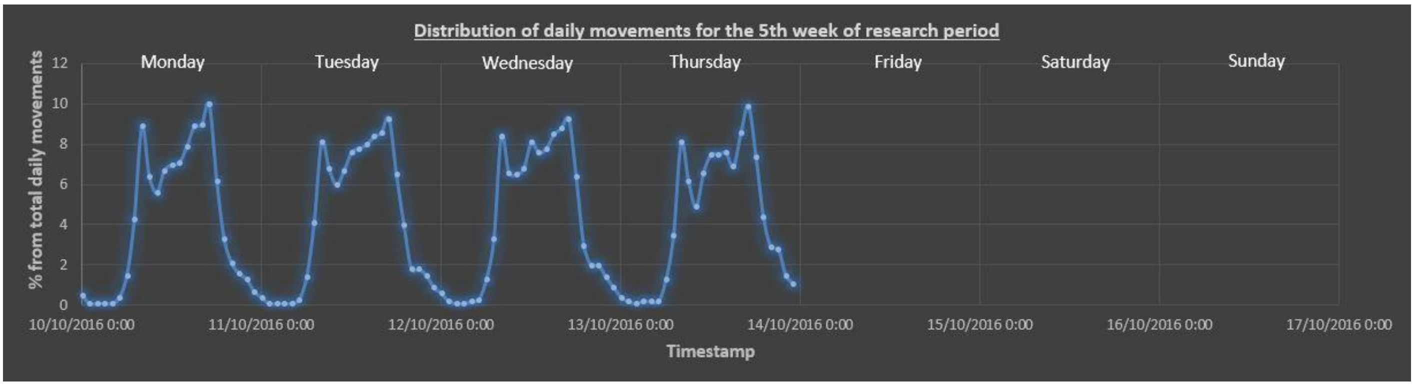


Figure 101: Distribution of daily movements for the fifth week of the research period

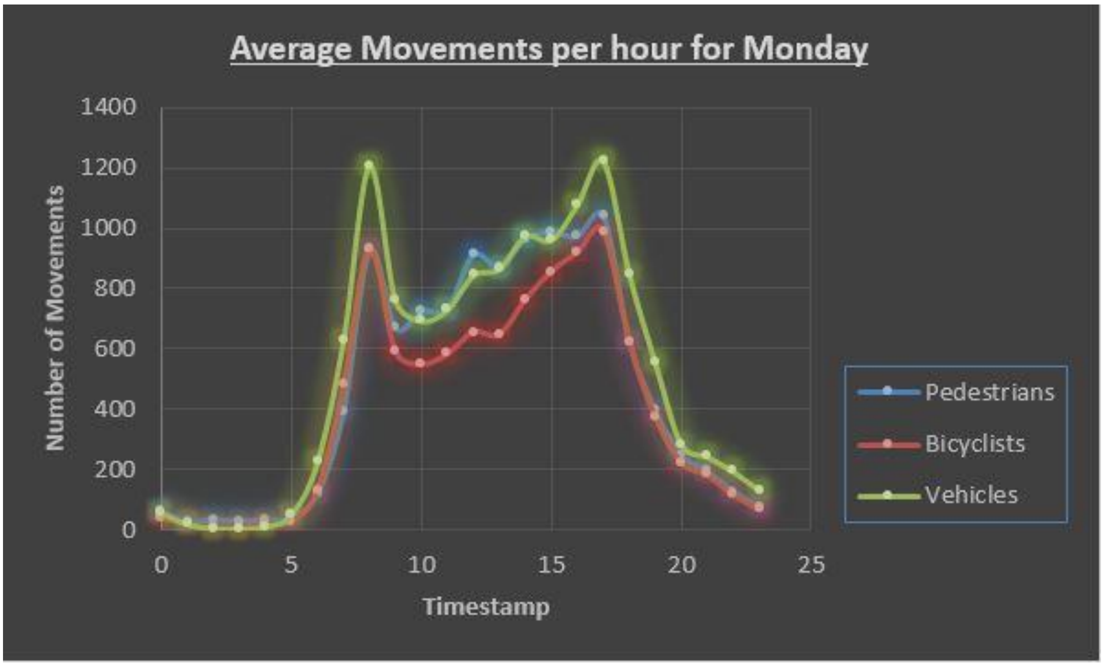


Figure 102: Average number of movements for each user category on Monday

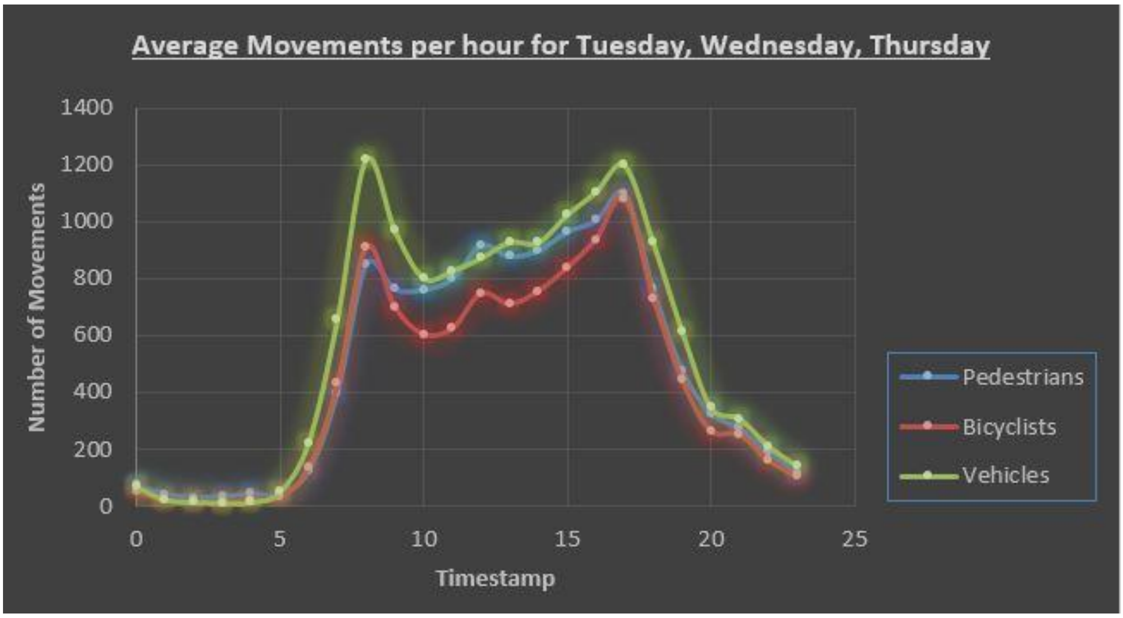


Figure 103: Average number of movements for each user category from Tuesday till Thursday

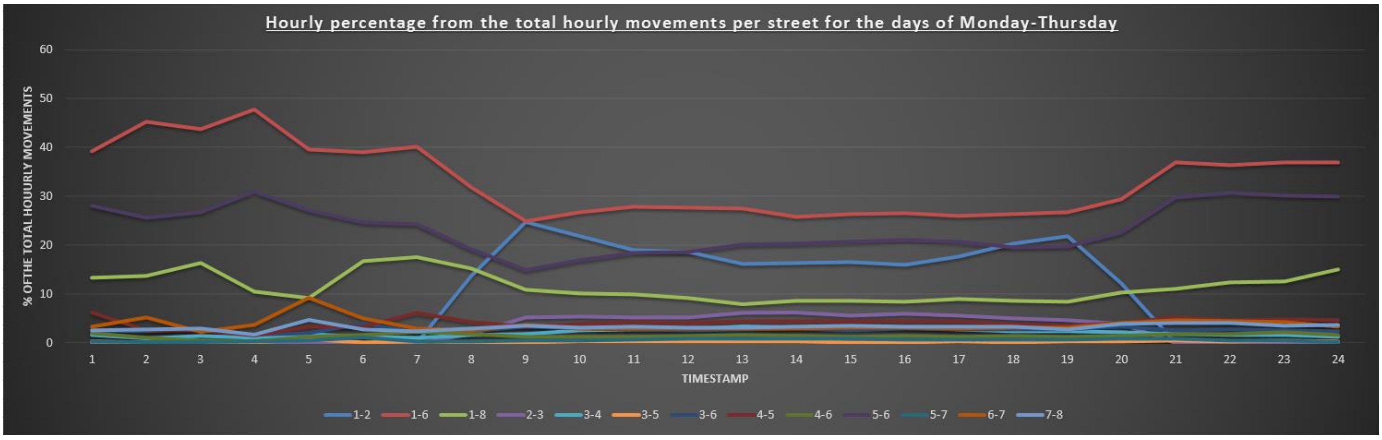


Figure 104: Hourly percentage from the total hourly movements per street for the days of Monday-Thursday

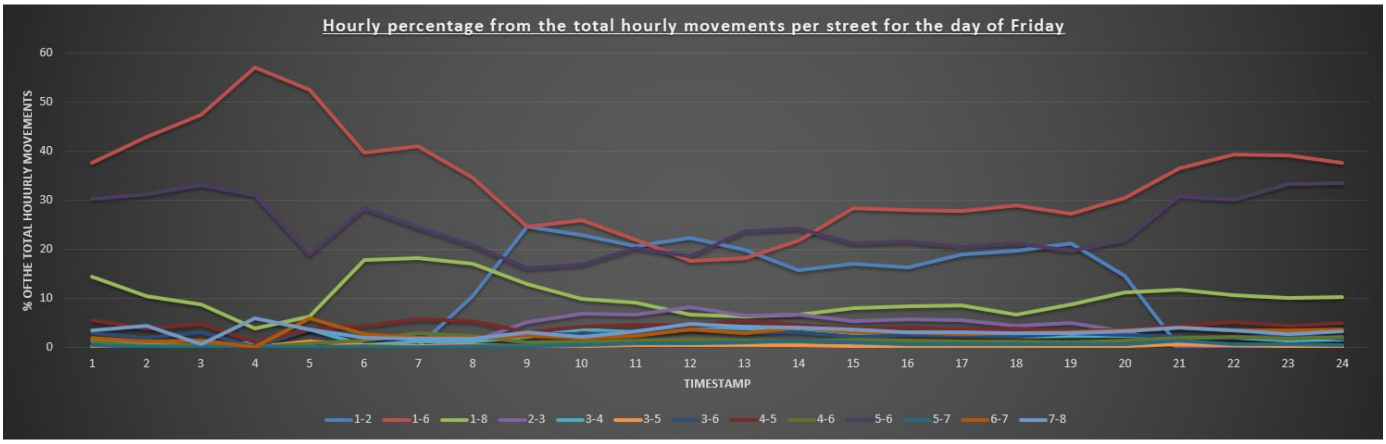


Figure 105: Hourly percentage from the total hourly movements per street for the day of Friday

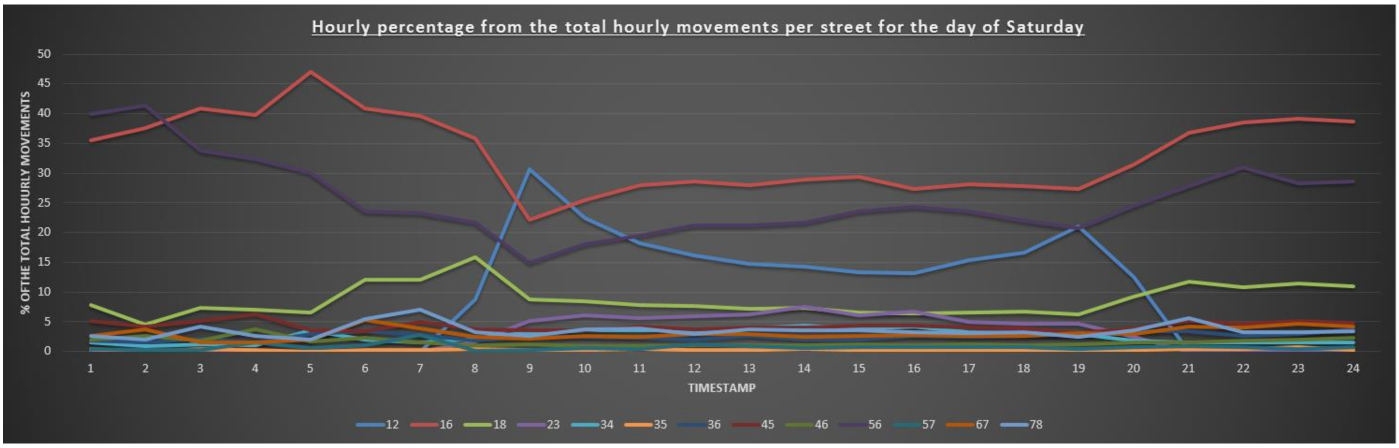


Figure 106: Hourly percentage from the total hourly movements per street for the day of Saturday

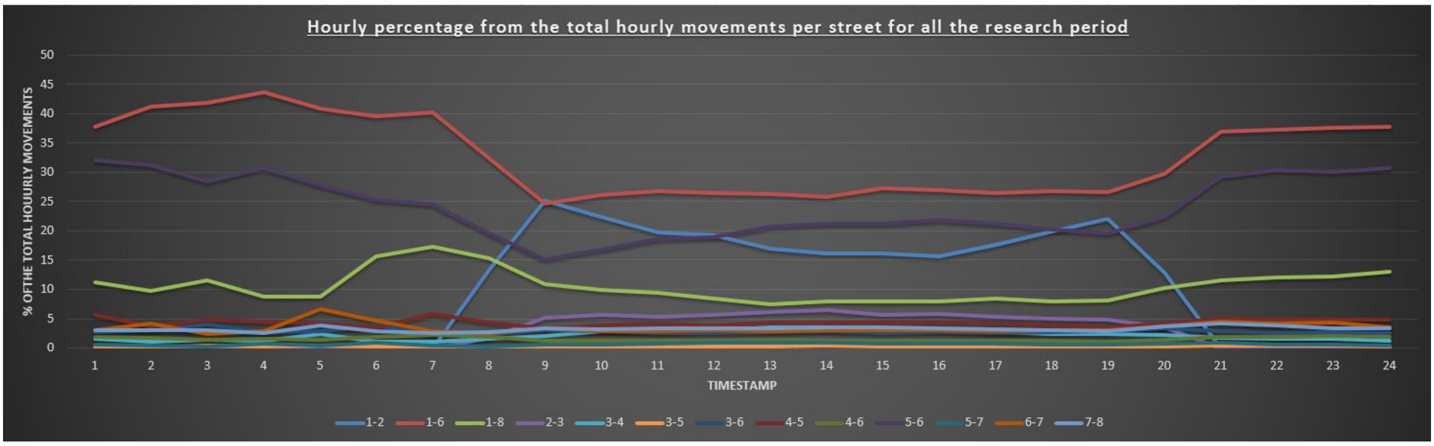


Figure 107: Hourly percentage from the total hourly movements per street for the average day

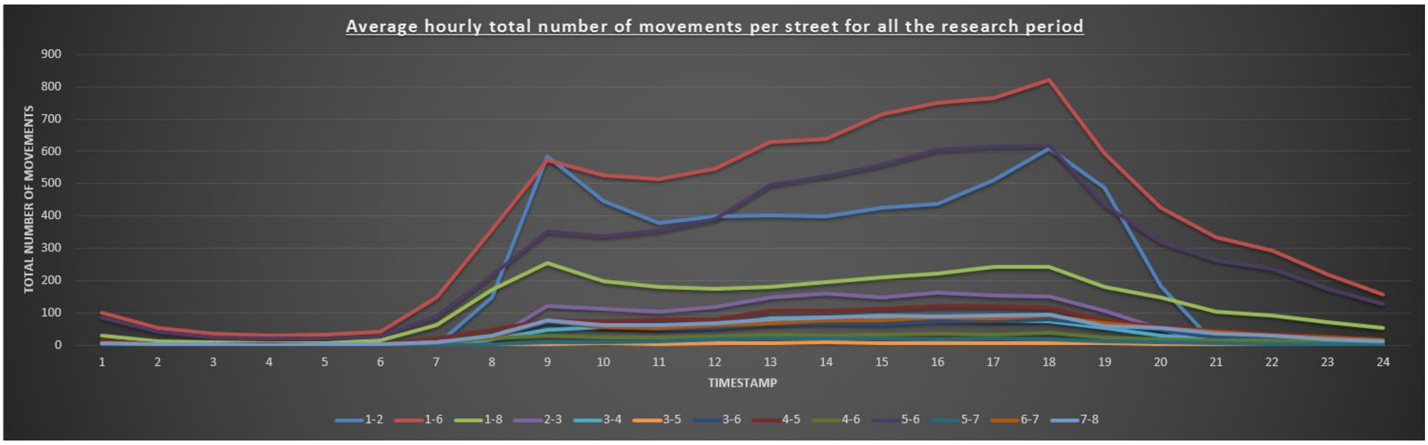


Figure 108: Distribution of the total number of movements of each street over the average day

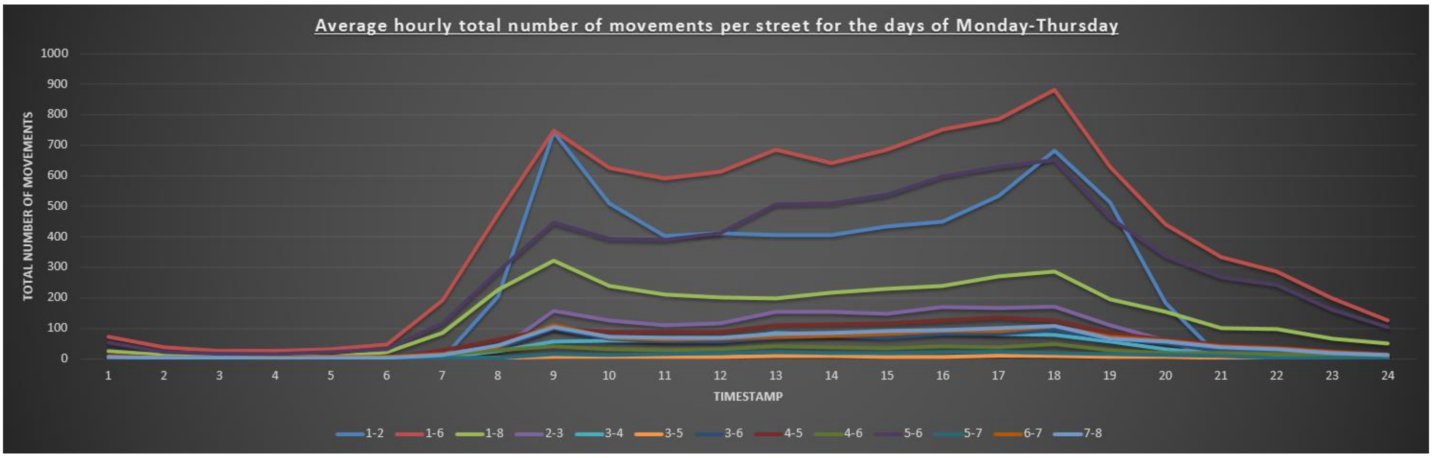


Figure 109: Distribution of the total number of movements of each street over the days Monday-Thursday

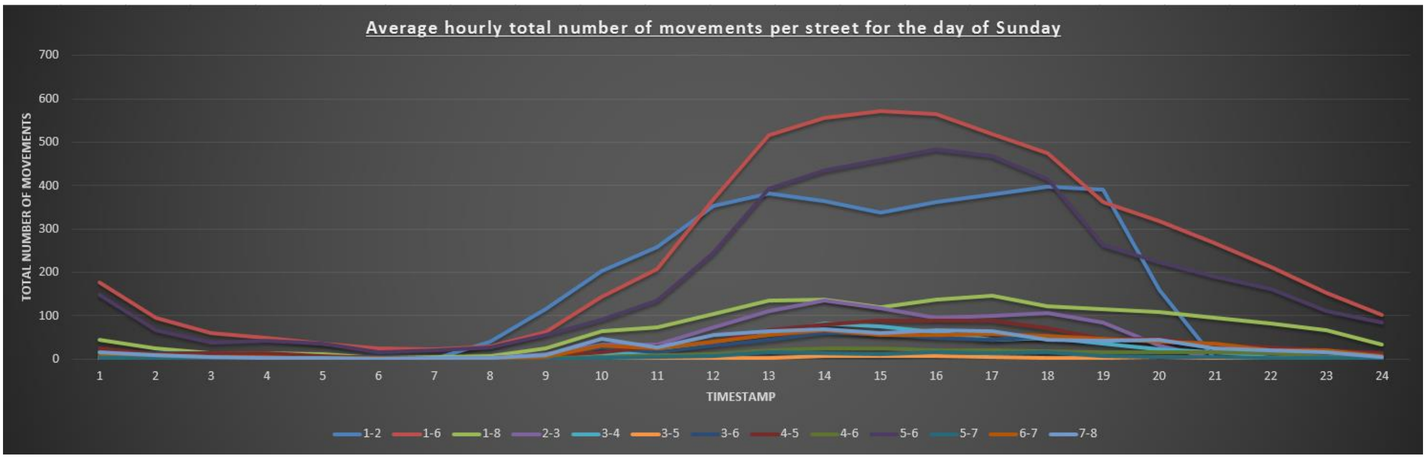


Figure 110: Distribution of the total number of movements of each street over the day of Sunday

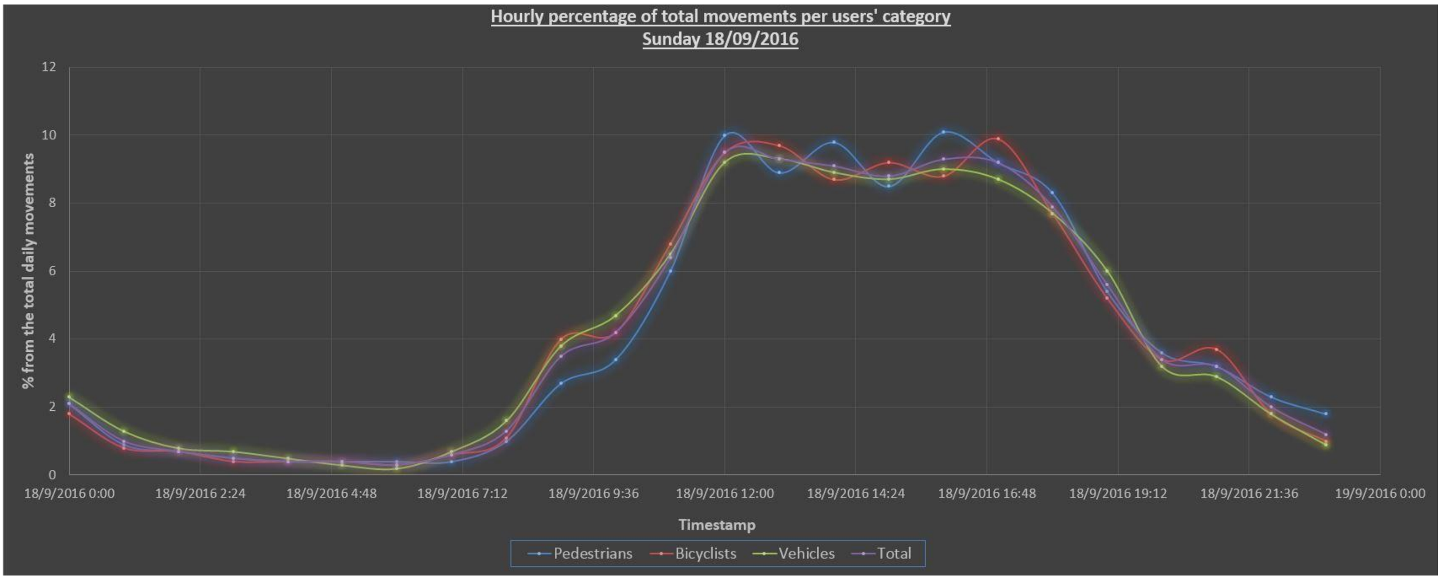


Figure 111: Distribution of the daily number of movements of each user category as well as of their sum on Sunday 18/09/2016

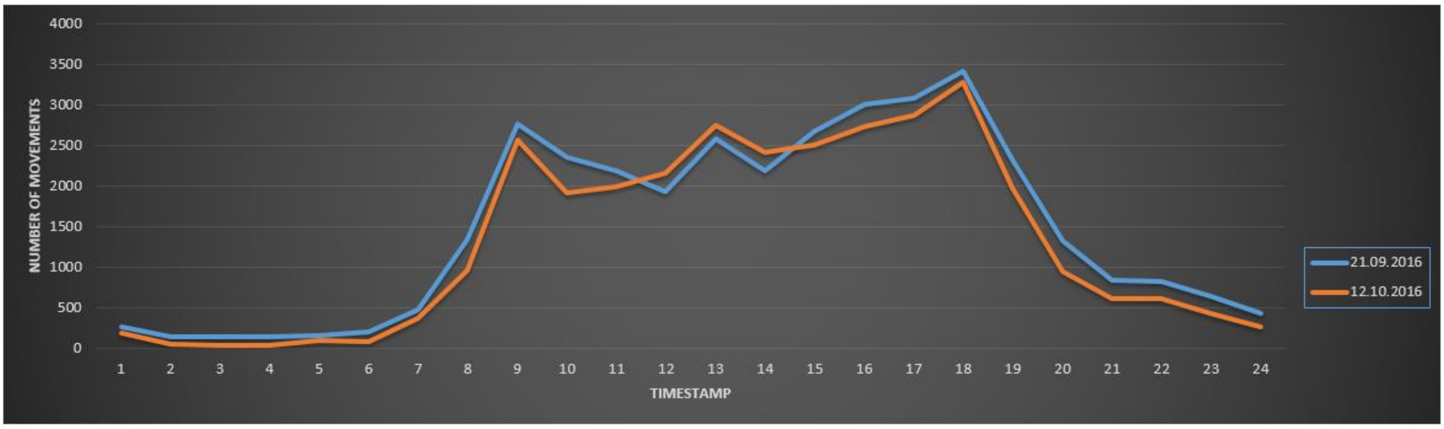


Figure 112: Cumulative number of movements in the research area for Wednesday 21/09/2016 (light blue) and Wednesday 12/10/2016 (yellow)

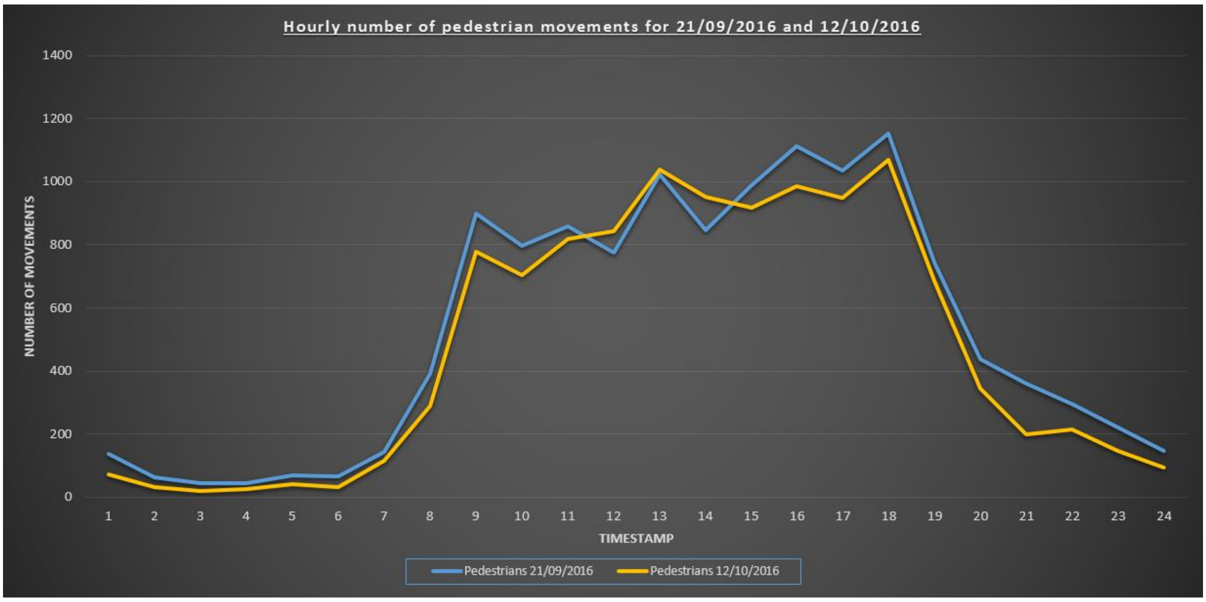
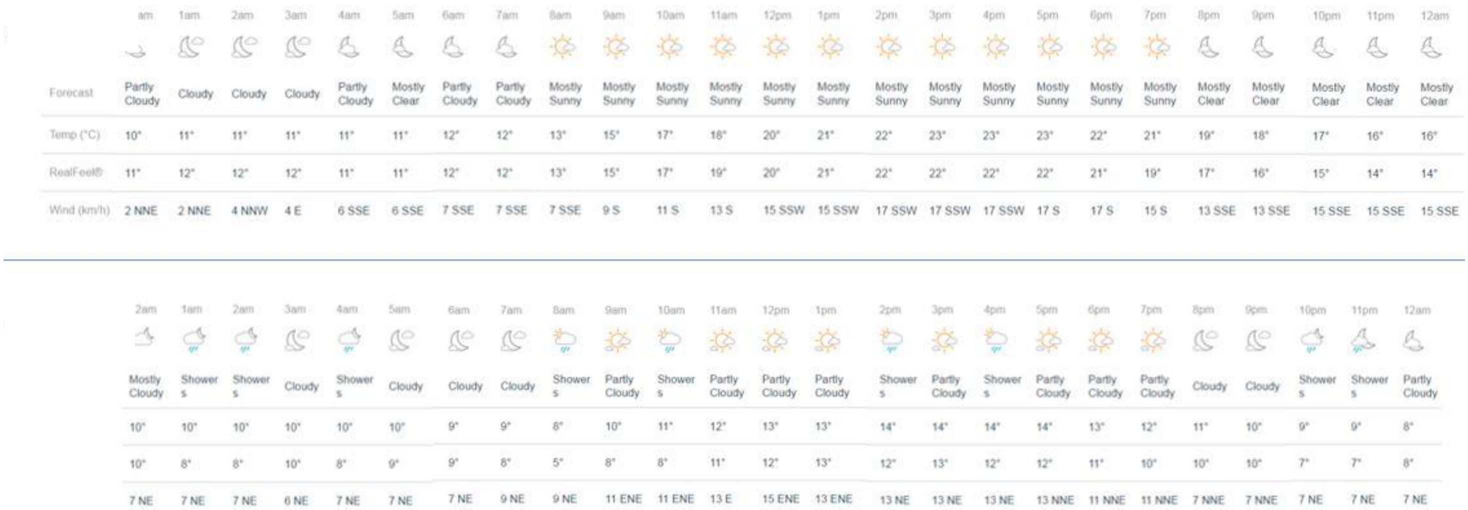


Figure 113: Hourly number of pedestrian movements in the research area on Wednesday 21/09/2016 (light blue) and Wednesday 12/10/2016 (yellow)



	am	1am	2am	3am	4am	5am	6am	7am	8am	9am	10am	11am	12pm
Forecast	Partly Cloudy	Cloudy	Cloudy	Cloudy	Partly Cloudy	Mostly Clear	Partly Cloudy	Partly Cloudy	Mostly Sunny	Mostly Sunny	Mostly Sunny	Mostly Sunny	Mostly Sunny
Temp (°C)	10°	11°	11°	11°	11°	11°	12°	12°	13°	15°	17°	18°	20°
RealFeel®	11°	12°	12°	12°	11°	11°	12°	12°	13°	15°	17°	19°	20°
Wind (km/h)	2 NNE	2 NNE	4 NNW	4 E	6 SSE	6 SSE	7 SSE	7 SSE	7 SSE	9 S	11 S	13 S	15 SSW

	1pm	2pm	3pm	4pm	5pm	6pm	7pm	8pm	9pm	10pm	11pm	12am
Forecast	Mostly Sunny	Mostly Sunny	Mostly Sunny	Mostly Sunny	Mostly Sunny	Mostly Sunny	Mostly Sunny	Mostly Clear	Mostly Clear	Mostly Clear	Mostly Clear	Mostly Clear
Temp (°C)	21°	22°	23°	23°	23°	22°	21°	19°	18°	17°	16°	16°
RealFeel®	21°	22°	22°	22°	22°	21°	19°	17°	16°	15°	14°	14°
Wind (km/h)	15 SSW	17 SSW	17 SSW	17 SSW	17 S	17 S	15 S	13 SSE	13 SSE	15 SSE	15 SSE	15 SSE

2am	1am	2am	3am	4am	5am	6am	7am	8am	9am	10am	11am	12pm
Mostly Cloudy	Shower s	Shower s	Cloudy	Shower s	Cloudy	Cloudy	Cloudy	Shower s	Partly Cloudy	Shower s	Partly Cloudy	Partly Cloudy
10°	10°	10°	10°	10°	10°	9°	9°	8°	10°	11°	12°	13°
10°	8°	8°	10°	8°	9°	9°	8°	5°	8°	8°	11°	12°
7 NE	7 NE	7 NE	6 NE	7 NE	7 NE	7 NE	9 NE	9 NE	11 ENE	11 ENE	13 E	15 ENE

1pm	2pm	3pm	4pm	5pm	6pm	7pm	8pm	9pm	10pm	11pm	12am
Partly Cloudy	Shower s	Partly Cloudy	Shower s	Partly Cloudy	Partly Cloudy	Partly Cloudy	Cloudy	Cloudy	Shower s	Shower s	Partly Cloudy
13°	14°	14°	14°	14°	13°	12°	11°	10°	9°	9°	8°
13°	12°	13°	12°	12°	11°	10°	10°	10°	7°	7°	8°
13 ENE	13 NE	13 NE	13 NE	13 NNE	11 NNE	11 NNE	7 NNE	7 NNE	7 NE	7 NE	7 NE

Figure 114: Weather conditions for Wednesday 21/09/2016 (above) and Wednesday 12/10/2016 (below)

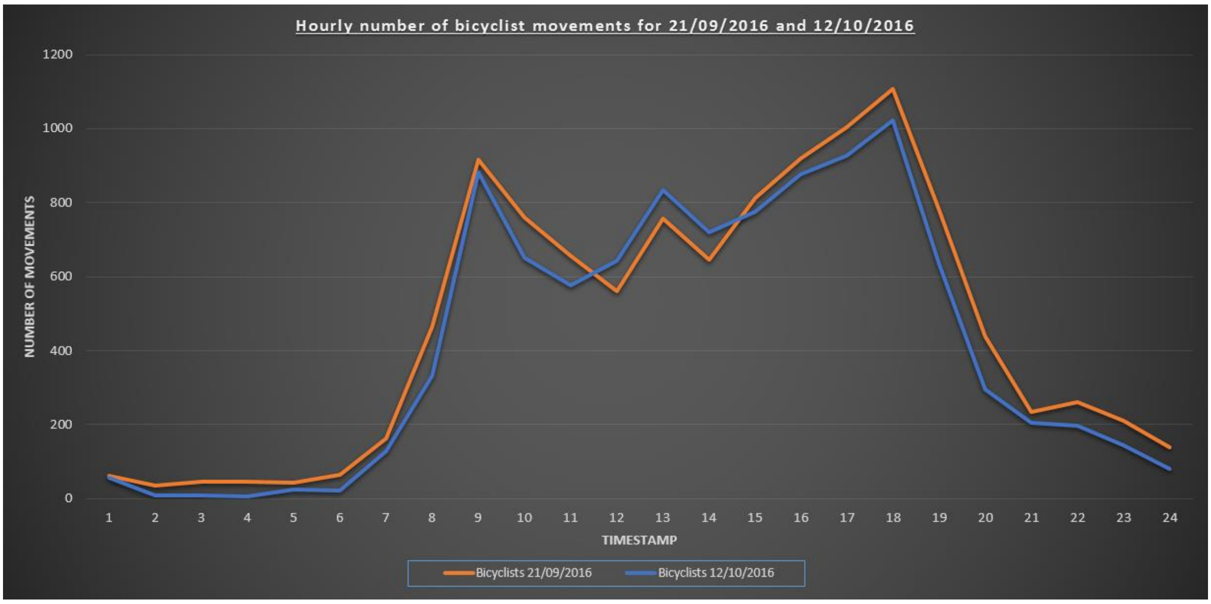


Figure 115: Hourly number of bicyclist movements in the research area on Wednesday 21/09/2016 (orange) and Wednesday 12/10/2016 (blue)

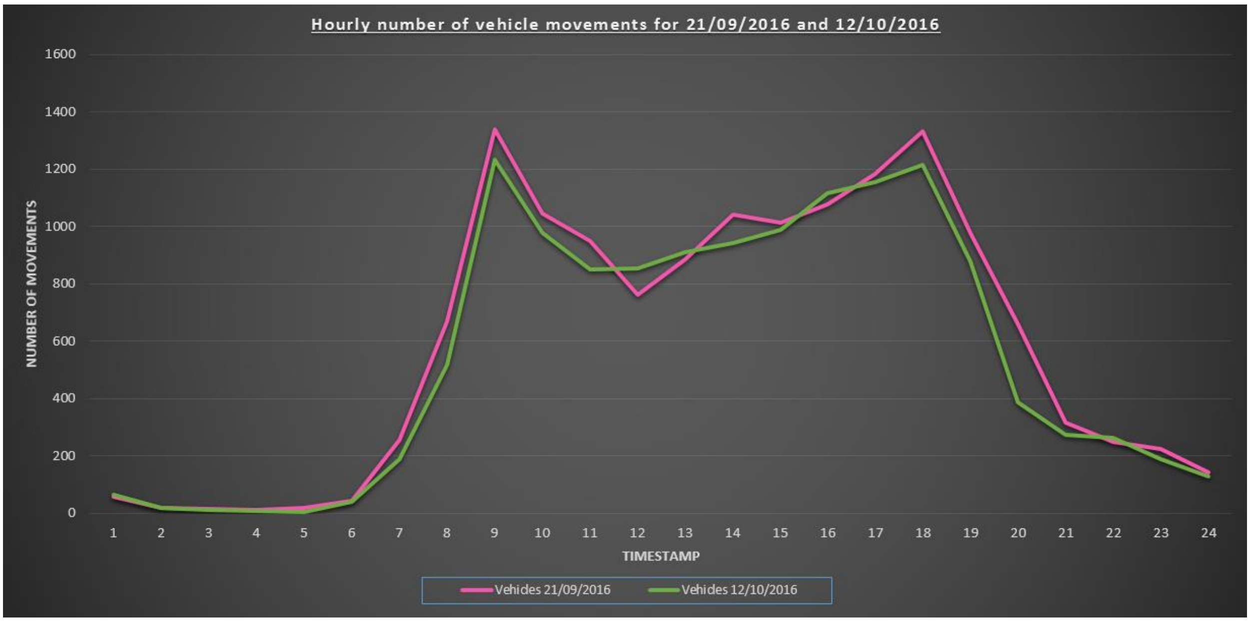


Figure 116: Hourly number of vehicle movements in the research area on Wednesday 21/09/2016 (pink) and Wednesday 12/10/2016 (light green)

Appendix D. Python Script

Input data

```
# data input
re=open("Initial records from sensors.txt", "r")
content=re.readlines()
#End of data input
```

Hashing of MAC addresses

```
#Get the list of records (lista) and create a list of the unique initial mac addresses.
def hasha(lista):
    import random
    mege=len(lista)
    listb=[9]
    for eacha in range(0,mege):
        helpa=lista[eacha]
        mac=helpa[2]
        listb.append(mac)
    listb=listb[1:] #till here: collect all the mac addresses
    listc=[8]
    megethos=len(listb)
    for each in range(0,megethos):
        a=listb[each]
        inside=0
        megeb=len(listc)
        for eachb in range(0,megeb):
            check=listc[eachb]
            if a==check:
                inside=1
        if inside==0:
            listc.append(a)
    listc=listc[1:] #till here: keep only the unique mac addresses
    megb=len(listc)
    listmacnew=newmac(megb) #request the production of megb new unique macs
    for eacha in range(0,mege):
        helpa=lista[eacha]
        macold=helpa[2]
        pos=-1
        for eachb in range(0,megb):
            check=listc[eachb]
            if check==macold:
                pos=eachb
        mactoget=listmacnew[pos]
        lista[eacha[2]]=mactoget #till here: replace the old mac addresses with the new ones
    return lista
#End of hasha.

#check the list of the new macs to be unique
def newmac(lista):
    listmac=[7]
    while (len(listmac))!=(lista+1):
        a=doamac
        same=0
        size=len(listmac)
        for i in range (0,size):
            check=listmac[j]
            if check==a:
                same=1
        if same==0:
            listmac.append(a)
    listmac=listmac[1:]
    return listmac
#End of newmac.

#Creation of new mac address.
def doamac():
    import random
    lista=[0,1,2,3,4,5,6,7,8,9,10,11]
    for i in range(0,12):
        a=random.randint(0,9)
        b=str(a)
        lista[i]=b
        d=':'

macnew=lista[0]+lista[1]+d+lista[2]+lista[3]+d+lista[4]+lista[5]+d+lista[6]+lista[7]+d+lista[8]+lista[9]+d+lista[10]+lista[11]
    return macnew
```

#End of doamac.

Preparation of data

```
#Preparation of data.
def preparation(contenta):
    mege=len(contenta)
    lista=[9]*(mege)
    listb=[9]*(mege)
    for eacha in range(0,mege): #convert the list of strings content to a list of lists of strings, called lista.
        helpa=contenta[eacha]
        helpb=helpa.split(';')
        lista[eacha]=helpb
    for i in range(0,mege): #convert the 1st, the 5th and the 7th element of each sublist from str to int.
        helpc=lista[i]
        for j in range(0,7):
            if j==0:
                helpc[j]=int(helpc[j])
            elif j==4:#apo edw...
                helpc[j]=int(helpc[j])
            elif j==6:
                helpc[j]=int(helpc[j][0])
        lista[i]=helpc
    for i in range(0,mege):#this loop convert the correct_timestamp from text to int and to an appropriate data format.
        helpc=lista[i]
        tr=helpc[5]
        a1=tr[0:4]
        a2=tr[5:7]
        a3=tr[8:10]
        a4=tr[11:13]
        a5=tr[14:16]
        a6=tr[17:19]
        a1=int(a1)
        a2=int(a2)
        a3=int(a3)
        a4=int(a4)
        a5=int(a5)
        a6=int(a6)
        import datetime
        dt=datetime.datetime(a1,a2,a3,a4,a5,a6)
        helpc[5]=dt
        lista[i]=helpc
    return lista
#End of preparation.
```

Data filtering & analysis

```
#Filter_1: Keep devices tracked from more than one sensor.
def morethanone(lmorathanone):
    listfiltered=[888]*1
    mege=len(lmorathanone)
    listhelp=[999]*1
    helpa=lmorathanone[0]
    listhelp[0]=helpa
    mac1=helpa[2]
    for i in range(1,mege):
        helpb=lmorathanone[i]
        mac2=helpb[2]
        if mac2==mac1:
            listhelp.append(helpb)#till here: we create step by step a list for each mac.
        else:
            different=0
            megethos=len(listhelp)
            helpc=listhelp[0]
            sensor1=helpc[6]
            for j in range(1,megethos):
                helpd=listhelp[j]
                sensor2=helpd[6]
                if sensor2!=sensor1:
                    different=1
            if different ==1:
                listfiltered.extend(listhelp)
            listhelp=[999]*1
            listhelp[0]=helpb
            mac1=helpb[2]#till here: check if each device is scanned by more than one sensor.
        if i==(mege-1):
            different=0
```

```
            megethos=len(listhelp)
            helpc=listhelp[0]
            sensor1=helpc[6]
            for j in range(1,megethos):
                helpd=listhelp[j]
                sensor2=helpd[6]
                if sensor2!=sensor1:
                    different=1
            if different ==1:
                listfiltered.extend(listhelp) #till here:
repeate the check for the last device.
    listfiltered=listfiltered[1:]
    return listfiltered
#End of Filter_1.

#Filter_2: Keep consecutive records in a period of 2
hours.
def everytwohours(leverytwo):
    listfiltered2=[888]*1
    megefiltered=len(leverytwo)
    listhelpa=[999]*1
    helpa=leverytwo[0]
    listhelpa[0]=helpa
    mac1=helpa[2]
    for i in range(1,megefiltered):
        helpb=leverytwo[i]
        mac2=helpb[2]
        if mac2==mac1:
            listhelpa.append(helpb)
        else:
            listhelpb=[777]*1
            megethos=len(listhelpa)
            helpc=listhelpa[0]
            starttime=helpc[5]
            sinolika=0
            ebala=0
            for j in range(1,megethos):
                helpd=listhelpa[j]
                nexttime=helpd[5]
                diarkeia=nexttime-starttime
                diarkeiadays=diarkeia.days
                diarkeiaseco=diarkeia.seconds

diarkeiatotalseseco=(diarkeiadays*24*3600)+diark
eiaseco
                if diarkeiatotalseseco<7200:
                    if ebala==0:
                        listhelpb.append(helpc)
                        listhelpb.append(helpd)
                        ebala=1
                    else:
                        listhelpb.append(helpd)
                else:
                    ebala=0
                starttime=nexttime
                helpc=helpd
            if len(listhelpb)>1:
                listhelpb=listhelpb[1:]
                listfiltered2.extend(listhelpb)
            listhelpa=[999]*1
            listhelpa[0]=helpb
            mac1=helpb[2]
        if i==(megefiltered-1):
            listhelpb=[777]*1
            megethos=len(listhelpa)
            helpc=listhelpa[0]
            starttime=helpc[5]
            sinolika=0
            ebala=0
            for j in range(1,megethos):
                helpd=listhelpa[j]
                nexttime=helpd[5]
                diarkeia=nexttime-starttime
                diarkeiadays=diarkeia.days
                diarkeiaseco=diarkeia.seconds

diarkeiatotalseseco=(diarkeiadays*24*3600)+diark
eiaseco
                if diarkeiatotalseseco<7200:
                    if ebala==0:
                        listhelpb.append(helpc)
                        listhelpb.append(helpd)
```

```
                ebala=1
            else:
                listhelpb.append(helpd)
        else:
            ebala=0
        starttime=nexttime
        helpc=helpd
    if len(listhelpb)>1:
        listhelpb=listhelpb[1:]
        listfiltered2.extend(listhelpb)
    listfiltered2=listfiltered2[1:]
    return listfiltered2
#End of Filter_2.

#Filter_3: Keep devices tracked continuously for a period smaller than 12 hours.
def longtotalperiod2(llongtotalperiod):
    listfiltered3=[888]*1
    listout=[777]*1
    megefiltered=len(llongtotalperiod)
    listhelpa=[999]*1
    helpa=llongtotalperiod[0]
    listhelpa[0]=helpa
    mac1=helpa[2]
    for i in range(1,megefiltered):
        helpb=llongtotalperiod[i]
        mac2=helpb[2]
        if mac2==mac1:
            listhelpa.append(helpb)
        else:
            megethos=len(listhelpa)
            helpc=listhelpa[0]
            starttime=helpc[5]
            sinolika=0
            outlier=0
            for j in range(1,megethos):
                helpd=listhelpa[j]
                nexttime=helpd[5]
                diarkeia=nexttime-starttime
                diarkeiadays=diarkeia.days
                diarkeiaseco=diarkeia.seconds
diarkeiatotalseseco=(diarkeiadays*24*3600)+diarkeiaseco
                if diarkeiatotalseseco<5400:
                    sinolika=sinolika+diarkeiatotalseseco
                else:
                    if sinolika>43200:
                        outlier=1
                    sinolika=0
                starttime=nexttime
                helpc=helpd
                if j==(megethos-1):
                    if sinolika>43200:
                        outlier=1
            if outlier==0:
                listfiltered3.extend(listhelpa)
            else:
                a1=listhelpa[0]
                a2=a1[2]
                a3=[a2]*1
                listout.extend(a3)
            listhelpa=[999]*1
            listhelpa[0]=helpb
            mac1=helpb[2]
        if i==(megefiltered-1):
            megethos=len(listhelpa)
            helpc=listhelpa[0]
            starttime=helpc[5]
            sinolika=0
            outlier=0
            for j in range(1,megethos):
                helpd=listhelpa[j]
                nexttime=helpd[5]
                diarkeia=nexttime-starttime
                diarkeiadays=diarkeia.days
                diarkeiaseco=diarkeia.seconds
diarkeiatotalseseco=(diarkeiadays*24*3600)+diarkeiaseco
                if diarkeiatotalseseco<5400:
                    sinolika=sinolika+diarkeiatotalseseco
```

```
            else:
                if sinolika>43200:
                    outlier=1
                sinolika=0
            starttime=nexttime
            helpc=helpd
            if j==(megethos-1):
                if sinolika>43200:
                    outlier=1
        if outlier==0:
            listfiltered3.extend(listhelpa)
        else:
            a1=listhelpa[0]
            a2=a1[2]
            a3=[a2]*1
            listout.extend(a3)
    listfiltered3=listfiltered3[1:]
    listout=listout[1:]
    return listfiltered3,listout
#End of Filter_3.
```

#Filter_4: Get a list of records and two times and keep the records that are within this timeslot.

```
def specificperiod(lspecificperiod,time1,time2):
    import datetime
    listfiltered=[888]*1
    mege=len(lspecificperiod)
    listhelp=[999]*1
    helpa=lspecificperiod[0]
    listhelp[0]=helpa
    mac1=helpa[2]
    for i in range(1,mege):
        helpb=lspecificperiod[i]
        mac2=helpb[2]
        if mac2==mac1:
            listhelp.append(helpb)
        else:
            listhelpb=[777]*1
            megethos=len(listhelp)
            for j in range(0,megethos):
                helpd=listhelp[j]
                recordtime=helpd[5]
                if recordtime<=time2:
                    if recordtime>=time1:
                        listhelpb.append(helpd)
            if len(listhelpb)>1:
                listhelpb=listhelpb[1:]
                listfiltered.extend(listhelpb)
            listhelp=[999]*1
            listhelp[0]=helpb
            mac1=helpb[2]
        if i==(mege-1):
            listhelpb=[777]*1
            megethos=len(listhelp)
            for j in range(0,megethos):
                helpd=listhelp[j]
                recordtime=helpd[5]
                if recordtime<=time2:
                    if recordtime>=time1:
                        listhelpb.append(helpd)
            if len(listhelpb)>1:
                listhelpb=listhelpb[1:]
                listfiltered.extend(listhelpb)
    listfiltered=listfiltered[1:]
    return listfiltered
#End of Filter_4.
```

#Filter_5: filter records with negative signal strength.

```
def nonegative(lmorathanone):
    listhelp=[888]*1
    megethos=len(lmorathanone)
    for i in range(0,megethos):
        a=lmorathanone[i]
        signal=a[4]
        if signal>0:
            listhelp.append(a)
    listhelp=listhelp[1:]
    return listhelp
#End of Filter_5.
```

#Computation of average time at each sensor.

```
def readyforspeed(lspecificperiod):
```

```
listfiltered=[888]*1
import datetime
mege=len(lspecificperiod)
listhelp=[999]*1
helpa=lspecificperiod[0]
listhelp[0]=helpa
mac1=helpa[2]
for i in range(1,mege):
    helpb=lspecificperiod[i]
    mac2=helpb[2]
    if mac2==mac1:
        listhelp.append(helpb)
    else:
        listhelpb=[777]*1
        megethos=len(listhelp)
        helpd=listhelp[0]
        mac1=helpd[2]
        sensor1=helpd[6]
        recordtime1=helpd[5]
        recorda=helpd[5]
        dia=1
        sinolikosxoronos=0
        listhelpb.append(mac1)
        linkedwifi1=helpd[3]
        linkedwifia=helpd[3]
        linkedwifio=helpd[3]
        for j in range(1,megethos):
            helpd=listhelp[j]
            sensor2=helpd[6]
            recordtime2=helpd[5]
            linkedwifi2=helpd[3]
            if sensor2==sensor1:
                dia=dia+1
                diarkeia=recordtime2-recordtime1
                diarkeiadays=diarkeia.days
                diarkeiaseco=diarkeia.seconds
                diarkeiatotalseseco=(diarkeiadays*24*3600)+diarkeiaseco
                sinolikosxoronos=sinolikosxoronos+diarkeiatotalseseco
                recordtime1=recordtime2
                recordo=recordtime2
                linkedwifio=linkedwifi2
            else:
                listhelpe=[666]*1
                timeplus=sinolikosxoronos/dia
                time3=datetime.timedelta(seconds=timeplus)
                finaltime=recorda+time3
                listhelpe.append(sensor1)
                if sinolikosxoronos<=60:
                    listhelpe.append(finaltime)
                    listhelpe.append(linkedwifio)
                else:
                    r1=listhelp[(j-dia)]
                    r2=listhelp[(j-dia+1)]
                    r3=listhelp[(j-2)]
                    r4=listhelp[(j-1)]
                    t1=r1[5]
                    t2=r2[5]
                    t3=r3[5]
                    t4=r4[5]
                    d1=t2-t1
                    d12=d1.days
                    d13=d1.seconds
                    d14=(d12*24*3600)+d13
                    timeplus1=d14/2
                    time1=datetime.timedelta(seconds=timeplus1)
                    ave1=t1+time1
                    d2=t4-t3
                    d22=d2.days
                    d23=d2.seconds
                    d24=(d22*24*3600)+d23
                    timeplus2=d24/2
                    time2=datetime.timedelta(seconds=timeplus2)
                    ave2=t3+time2
                    linkedwifia=r1[3]
```

```
            linkedwifio=r4[3]
            listhelpe.append(ave1)
            listhelpe.append(linkedwifia)
            listhelpe=listhelpe[1:]
            listhelpb.append(listhelpe)
            listhelpe=[666]*1
            listhelpe.append(sensor1)
            listhelpe.append(ave2)
            listhelpe.append(linkedwifio)
          listhelpe=listhelpe[1:]
          listhelpb.append(listhelpe)
          sinolikosxoronos=0
          sensor1=sensor2
          recordtime1=recordtime2
          linkedwifi1=linkedwifi2
          dia=1
          recorda=recordtime2
          linkedwifia=linkedwifi2
          linkedwifio=linkedwifi2
        if j==(megethos-1):
          listhelpe=[666]*1
          timeplus=sinolikosxoronos/dia

time3=datetime.timedelta(seconds=timeplus)
          finaltime=recorda+time3
          listhelpe.append(sensor1)
          if sinolikosxoronos<=60:
            listhelpe.append(finaltime)
            listhelpe.append(linkedwifio)
          else:
            r1=listhelp[(j-dia)]
            r2=listhelp[(j-dia+1)]
            r3=listhelp[(j-2)]
            r4=listhelp[(j-1)]
            t1=r1[5]
            t2=r2[5]
            t3=r3[5]
            t4=r4[5]
            d1=t2-t1
            d12=d1.days
            d13=d1.seconds
            d14=(d12*24*3600)+d13
            timeplus1=d14/2

time1=datetime.timedelta(seconds=timeplus1)
            ave1=t1+time1
            d2=t4-t3
            d22=d2.days
            d23=d2.seconds
            d24=(d22*24*3600)+d23
            timeplus2=d24/2

time2=datetime.timedelta(seconds=timeplus2)
            ave2=t3+time2
            linkedwifia=r1[3]
            linkedwifio=r4[3]
            listhelpe.append(ave1)
            listhelpe.append(linkedwifia)
            listhelpe=listhelpe[1:]
            listhelpb.append(listhelpe)
            listhelpe=[666]*1
            listhelpe.append(sensor1)
            listhelpe.append(ave2)
            listhelpe.append(linkedwifio)
          listhelpe=listhelpe[1:]
          listhelpb.append(listhelpe)
          sinolikosxoronos=0
          sensor1=sensor2
          recordtime1=recordtime2
          linkedwifi1=linkedwifi2
          dia=1
          recorda=recordtime2
          linkedwifia=linkedwifi2
          linkedwifio=linkedwifi2
      if len(listhelpb)>1:
        listhelpb=listhelpb[1:]
        listfiltered.extend(listhelpb)
      listhelp=[999]*1
      listhelp[0]=helpb
      mac1=helpb[2]
    if i==(mege-1):
```

```
listhelpb=[777]*1
megethos=len(listhelp)
helpd=listhelp[0]
mac1=helpd[2]
sensor1=helpd[6]
recordtime1=helpd[5]
recorda=helpd[5]
dia=1
sinolikosxoronos=0
listhelpb.append(mac1)
linkedwifi1=helpd[3]
linkedwifia=helpd[3]
linkedwifio=helpd[3]
for j in range(1,megethos):
    helpd=listhelp[j]
    sensor2=helpd[6]
    recordtime2=helpd[5]
    linkedwifi2=helpd[3]
    if sensor2==sensor1:
        dia=dia+1
        diarkeia=recordtime2-recordtime1
        diarkeiadays=diarkeia.days
        diarkeiaseco=diarkeia.seconds

diarkeiatotalseseco=(diarkeiadays*24*3600)+diark
eiaseco
sinolikosxoronos=sinolikosxoronos+diarkeiatotalse
seco
        recordtime1=recordtime2
        recordo=recordtime2
        linkedwifio=linkedwifi2
    else:
        listhelpe=[666]*1
        timeplus=sinolikosxoronos/dia

time3=datetime.timedelta(seconds=timeplus)
        finaltime=recorda+time3
        listhelpe.append(sensor1)
        if sinolikosxoronos<=60:
            listhelpe.append(finaltime)
            listhelpe.append(linkedwifio)
        else:
            r1=listhelp[(j-dia)]
            r2=listhelp[(j-dia+1)]
            r3=listhelp[(j-2)]
            r4=listhelp[(j-1)]
            t1=r1[5]
            t2=r2[5]
            t3=r3[5]
            t4=r4[5]
            d1=t2-t1
            d12=d1.days
            d13=d1.seconds
            d14=(d12*24*3600)+d13
            timeplus1=d14/2

time1=datetime.timedelta(seconds=timeplus1)
            ave1=t1+time1
            d2=t4-t3
            d22=d2.days
            d23=d2.seconds
            d24=(d22*24*3600)+d23
            timeplus2=d24/2

time2=datetime.timedelta(seconds=timeplus2)
            ave2=t3+time2
            linkedwifia=r1[3]
            linkedwifio=r4[3]
            listhelpe.append(ave1)
            listhelpe.append(linkedwifia)
            listhelpe=listhelpe[1:]
            listhelpb.append(listhelpe)
            listhelpe=[666]*1
            listhelpe.append(sensor1)
            listhelpe.append(ave2)
            listhelpe.append(linkedwifio)
        listhelpe=listhelpe[1:]
        listhelpb.append(listhelpe)
        sinolikosxoronos=0
        sensor1=sensor2
        recordtime1=recordtime2
        linkedwifi1=linkedwifi2
```

```
            dia=1
            recorda=recordtime2
            linkedwifia=linkedwifi2
            linkedwifio=linkedwifi2
          if j==(megethos-1):
            listhelpe=[666]*1
            timeplus=sinolikosxoronos/dia

time3=datetime.timedelta(seconds=timeplus)
            finaltime=recorda+time3
            listhelpe.append(sensor1)
            if sinolikosxoronos<=60:
              listhelpe.append(finaltime)
              listhelpe.append(linkedwifio)
            else:
              r1=listhelp[(j-dia)]
              r2=listhelp[(j-dia+1)]
              r3=listhelp[(j-2)]
              r4=listhelp[(j-1)]
              t1=r1[5]
              t2=r2[5]
              t3=r3[5]
              t4=r4[5]
              d1=t2-t1
              d12=d1.days
              d13=d1.seconds
              d14=(d12*24*3600)+d13
              timeplus1=d14/2
time1=datetime.timedelta(seconds=timeplus1)
              ave1=t1+time1
              d2=t4-t3
              d22=d2.days
              d23=d2.seconds
              d24=(d22*24*3600)+d23
              timeplus2=d24/2

time2=datetime.timedelta(seconds=timeplus2)
              ave2=t3+time2
              linkedwifia=r1[3]
              linkedwifio=r4[3]
              listhelpe.append(ave1)
              listhelpe.append(linkedwifia)
              listhelpe=listhelpe[1:]
              listhelpb.append(listhelpe)
              listhelpe=[666]*1
              listhelpe.append(sensor1)
              listhelpe.append(ave2)
              listhelpe.append(linkedwifio)
            listhelpe=listhelpe[1:]
            listhelpb.append(listhelpe)
            sinolikosxoronos=0
            sensor1=sensor2
            recordtime1=recordtime2
            linkedwifi1=linkedwifi2
            dia=1
            recorda=recordtime2
            linkedwifia=linkedwifi2
            linkedwifio=linkedwifi2
        if len(listhelpb)>1:
          listhelpb=listhelpb[1:]
          listfiltered.extend(listhelpb)
        listhelp=[999]*1
        listhelp[0]=helpb
        mac1=helpb[2]
    listfiltered=listfiltered[1:]
    return listfiltered
#End of Computation of average time at each
sensor.

#Computation of movement speed of each device
and characterize it as pedestrian, bicyclist or vehicle
def movements(lmove):
    import datetime
    listhelpb=[888]*1
    megethos=len(lmove)
    start=0
    for k in range(0,megethos):
      a=lmove[k]
      if len(a)>=4:
        device=[a]
        listhelpb.append(device)
        start=0
```

```
else:
    if start==0:
        sensor1=a[0]
        time1=a[1]
        start=1
    else:
        sensor2=a[0]
        if (sensor1==1 and sensor2==2) or (sensor1==2 and sensor2==1):
            distance=410
        elif (sensor1==1 and sensor2==6) or (sensor1==6 and sensor2==1):
            distance=124
        elif (sensor1==1 and sensor2==8) or (sensor1==8 and sensor2==1):
            distance=372
        elif (sensor1==2 and sensor2==3) or (sensor1==3 and sensor2==2):
            distance=185
        elif (sensor1==3 and sensor2==4) or (sensor1==4 and sensor2==3):
            distance=218
        elif (sensor1==3 and sensor2==6) or (sensor1==6 and sensor2==3):
            distance=395
        elif (sensor1==4 and sensor2==5) or (sensor1==5 and sensor2==4):
            distance=398
        elif (sensor1==5 and sensor2==6) or (sensor1==6 and sensor2==5):
            distance=207
        elif (sensor1==5 and sensor2==7) or (sensor1==7 and sensor2==5):
            distance=414
        elif (sensor1==5 and sensor2==8) or (sensor1==8 and sensor2==5):
            distance=1220
        elif (sensor1==6 and sensor2==7) or (sensor1==7 and sensor2==6):
            distance=277
        elif (sensor1==7 and sensor2==8) or (sensor1==8 and sensor2==7):
            distance=204
        elif (sensor1==3 and sensor2==5) or (sensor1==5 and sensor2==3):
            distance=542
        elif (sensor1==4 and sensor2==6) or (sensor1==6 and sensor2==4):
            distance=600
        elif (sensor1==4 and sensor2==2):
            distance=630
        else:
            distance="problima"
            r1=str(sensor1)
            r2=str(sensor2)
            r3=r1+r2
            return r3
            break
        time2=a[1]
        start=1
        diarkeia=time2-time1
        diarkeiadays=diarkeia.days
        diarkeiaseco=diarkeia.seconds
diarkeiatotalseseco=(diarkeiadays*24*3600)+diarkeiaseco
        if distance==0:
            speed=0
        elif distance=="problima":
            speed="problima"
        else:
speed=(distance*36)/(diarkeiatotalseseco*10)
        if speed==0:
            device="stasimos"
        elif speed<=7:
            device="pedestrian"
        elif speed<20:
            device="bicyclist"
        elif speed>=20:
            device="vehicle"
        if (sensor1==3 and sensor2==2) and speed>7:
            device="bicyclist"
        if (sensor1==4 and sensor2==3) and speed>7:
```

```
                device="bicyclist"
            if (sensor1==5 and sensor2==7) and
speed>7:
                device="bicyclist"
            if (sensor1==7 and sensor2==5) and
speed>7:
                device="bicyclist"

b=[sensor1,sensor2,diarkeiatotalseseco,speed,devic
e]
            listhelpb.append(b)
            sensor1=sensor2
            time1=time2
    listhelpb=listhelpb[1:]
    return listhelpb
#End of Computation of movement speed.
#Compute & return the statistics from movements.
def statistic(lmorathanone):
    pedestrians=0
    bicyclist=0
    vehicle=0
    problima=0
    mege=len(lmorathanone)
    for i in range(0,mege):
        a=lmorathanone[i]
        if len(a)==5:
            b=a[4]
            if b=="pedestrian":
                pedestrians=pedestrians+1
            elif b=="bicyclist":
                bicyclist=bicyclist+1
            elif b=="vehicle":
                vehicle=vehicle+1
            elif b=="problima":
                problima=problima+1
    return pedestrians,bicyclist,vehicle,problima
#End of statistics.

#Counting of unique mac addresses.
def countuniquemacs(lin):
    listmacfilt=[888]*1
    helpline=lin[0]
    helpmacfilt=helpline[2]
    listmacfilt[0]=helpmacfilt
    megethos=len(lin)
    for i in range(1,megethos):
        helplineb=lin[i]
        helpmacfilt2=helplineb[2]
        if helpmacfilt2==helpmacfilt:
            continue
        else:
            listmacfilt.append(helpmacfilt2)
            helpmacfilt=helpmacfilt2
    length=len(listmacfilt)
    return length
#End of counting.
```

Appendix E. SQL Queries

Creation table with the data from sensor1 for the day 13102016

Create table sensor1d13102016

(

id_frame integer NOT NULL,

“timestamp” timestamp without time zone,

Mac character varying(17) NOT NULL,

Ssid character varying(32) NOT NULL,

Rssi character varying(3) NOT NULL,

vendor character varying(150) NOT NULL,

type character varying(10) NOT NULL,

ap character varying(17) NOT NULL,

sync integer NOT NULL DEFAULT 0,

sensor integer,

CONSTRAINT sensor1_pkey PRIMARY KEY (id_frame)

)

Add columns “correct_timestamp”

ALTER TABLE sensor1d13102016

ADD column “correct_timestamp” timestamp without time zone

Insert values

INSERT INTO sensor1d13102016 (correct_timestamp) VALUES

(′2016-10-13 10:32:35′),

(′2016-10-13 10:32:38′),

(′2016-10-13 10:32:40′),

Create table for the day of 13102016 containing the records from all sensors

Create table sensoralld13102016 as

(SELECT * from sensor1d13102016

UNION

SELECT * from sensor2d13102016

UNION

SELECT * from sensor3d13102016

UNION

SELECT * from sensor4d13102016

UNION

SELECT * from sensor5d13102016

UNION

SELECT * from sensor6d13102016

UNION

SELECT * from sensor7d13102016

UNION

SELECT * from sensor8d13102016);

Add primary key

ALTER TABLE sensoralld13102016

ADD CONSTRAINT sensorall_pkey PRIMARY KEY (id_frame,sensor, correct_timestamp)

Create table for sensor2 for the day of 13102016

Create table sensor2d13102016 as

(SELECT * from sensor2dall)

where EXTRACT (day from correct_timestamp)=13

AND EXTRACT (month from correct_timestamp)=10;

#Delete strange values

Delete from sensor2d13102016 where mac=′00:00:00:00:00:00′
or where correct_timestamp =′2000-01-01 00:00:00′

#Delete records before the research period

Delete from sensoralldall
Where correct_timestamp < ′2016-09-13 00:00:00′

Appendix F. Questionnaire

Questionnaire for the computation of the percentage of users who have enabled the Wi-Fi functionality at their devices. The same questionnaire was applied both on pedestrians and bicyclists.

Do you have any device with you with enabled the Wi-Fi or Bluetooth functionality?

- ○ **Yes**
- ○ **No**

How many devices do you have with enabled the Wi-Fi functionality?

1	2	3	4	5
☐	☐	☐	☐	☐

How many devices do you have with enabled the Bluetooth functionality?

1	2	3	4	5
☐	☐	☐	☐	☐

Appendix G. Photos from the installation of the Wi-Fi monitoring system


5 Sterrenburg


itelite


2
WIFI BT
TRACKER
Wi Fi


itelite


Dordts Cultureel Festi
Eddingtonweg 7 Dordrec


itelite
Hybrid


WIFI BT
TRACKER
Wi Fi


WIFI BT
TRACKER
Wi Fi


12 Zwijndrecht
ZUID HOLLAND
Hybrid
ARRIVA


PIKENEON
wij bezorgen u een goede naam
89-VS-VK
5 Sterrenburg

Reflection

This thesis addresses the identification of road modality and occupancy patterns with the use of a Wi-Fi monitoring system in a city environment. It investigates which user categories and occupancies can be identified, what the influence of the setup parameters on the final outcome is and proposes improvements that have to be further studied and tested in order to enhance the reliability of the system. The research was conducted from March 2016 to April 2017, along with the data collection procedure. The initial planning included timeslots for literature study, studying the passive Wi-Fi system and its applications, understanding the relevant benefits and limitations, designing the observation network, handling and processing data, and, finally, evaluating and analyzing the results and the setup parameters respectively.

The field of Geomatics Engineering includes practices concerned with the collection, manipulation and representation of information about the built and human environment. Streets contain geographical information which play an important role in numerous areas such as traffic monitoring, land management, public services, and urban development. This thesis focuses on the implementation of the Wi-Fi monitoring system in order to collect useful information about road modality and occupancy patterns, which is required for all the above areas. However, this information is limited as it is significantly difficult to collect.

The chosen method in this research noticeably aligns with the methodical line of approach in Geomatics involving data capture, storage, analysis, and visualization data from different sources, along with quality and representativeness control. The passive Wi-Fi monitoring system has been used in order to collect the appropriate data. As a second step, Postgres was used as the basic tool to store data in a database, process and analyze them with the use of sql queries and Python programming. Furthermore, QGIS and statistic tools were used for the visualization and validation of the outcomes. The conclusions drawn using all the above were examined with regard to the application of the Wi-Fi monitoring system in the research of the spatiotemporal behavior of the different user categories in the city environment as well as the relevant setup parameters which influence the final outcome.

In a wider context, the research and its results are directed towards a system which can provide almost real-time useful information about road modality in the streets of an urban area. Furthermore, formal outcomes about the overall behavior of the total set of users or of each category separately were reached, in this way enhancing the efforts for deeper investigation of people's spatiotemporal behavior. Thus, the method investigated in this research, in combination with the appropriate choice of setup parameters, can significantly support the urban planning and development procedure, leading to the improvement of the level of services.

The final step of this thesis is the investigation of the applicability of the Wi-Fi monitoring system to the computation of road modality and occupancy patterns as well as the influence of each setup parameter on the final outcome. There is definitely a prospect for continuing this research, as there are many technical and procedural problems which are interdependent and have to be studied simultaneously.

Dimitrios Kyritsis
MSc Geomatics

D.Kyritsis@student.tudelft.nl
Delft, 2017