BACHELOR THESIS IN
ARTIFICIAL INTELLIGENCE

Radboud University

Using Traffic Sensors for Mapping Event Effects on Pedestrian Flow

Author:

Jaap van Bussel
(S4805879)

First supervisor:

Dr. W.F.G. HASELAGER
(Radboud University, Donders Institute)
W.Haselager@Donders.ru.nl

Second reader:

Dr. F.A. GROOTJEN
(Radboud University, Artificial Intelligence)
F.Grootjen@ai.ru.nl

January 29th 2021

Abstract

Today, the amount of traffic on the streets is more important than ever before. To combat the number of new infections of COVID-19, data must be gathered to make informed decisions. A good place to start is the city centre of our own city, Nijmegen. Because of the narrow streets and the high number of people in those streets, it is sometimes impossible to keep the desired distance from others. To help cities manage crowd control on a well-informed basis, it is useful to use sensor readings about the number of traffic participants. This can be used further to map how events affect the amount of people that are in the city centre. This is possible by comparing the number of people in an area in two periods are compared, that differ in respect to a selected criterion, such as an event (e.g. shopping Sundays, or the four days marches event). Comparing such periods gives insight into the effect of the event on the amount of people in the city. The goal of this thesis is to develop and evaluate a visualization tool that allows for such comparisons and to indicate ways to improve this tool in future research. The tool uses data which is generated by smart city sensors that have been placed in and around the city centre of Nijmegen. The visualisation is created by interpolating the data with the use of radial basis function interpolation. The tool enables the user to visualize the effect of a big variety of events on the amount of people in the city.

Contents

Introduction

During the past months, the government has introduced multiple measures because of COVID-19, with the goal of reducing the number of new infections. However, it is hard to see where and how much these rules have affected the behaviour of citizens. To allow the government to gain insight in the behaviour of pedestrians in the city centre of Nijmegen, the following research question was developed:
Is it possible to develop a digital visualization tool that allows users to observe the level of impact events have had where and when on the number of pedestrians in the city centre of Nijmegen?

A digital visualization tool, as intended in this thesis, is a program, which can be used by a user to clearly visualize data, which would otherwise be too abstract or would take a long time to comprehend. It helps users comprehend complicated data by visualizing it. Such tool could allow the city council to survey the streets and it could inform the council on whether the rules or the enforcement of the rules, that are devised to combat COVID-19, could be improved. By doing this, the fight against COVID-19 could become easier and the number of new infections per day could be decreased.
The tool can also be used to inform the council, so they can increase security during public events in the city centre. By making a comparison between a public event and the same event one year prior, the increase or decrease of the number of pedestrians becomes visible. This can give an indication of the number of people that can be expected to visit the next edition of the event at certain locations in the city centre. The city council could use this information to choose security guard locations.
The tool can also be used by shop owners to decide whether they should open their shop on a certain day that an event is planned. Using the tool, they can inspect the same event one year prior to see how much effect that event has had on the number of pedestrians. If the event has had a big negative effect on the number of pedestrians in the city, they could decide to close their shop for that day to prevent losses.

A lot of cities around the world have adopted a smart approach to city planning and of course, there are many options for observing traffic. City councils must decide which of these observation tools they implement, and they have to keep the privacy of their citizens in mind. The city council of Nijmegen has adopted 21 sensors made by the company Numina, which observe a total of 42 locations. These sensors are spread throughout the city centre of Nijmegen. They capture video footage of streets and using AI they count and save the paths of cars, busses, trucks, bikers and pedestrians. To ensure privacy of the traffic participants, the video footage is removed, but the counts and paths are stored. The digital tool uses the data gathered by these sensors and only the counts of pedestrians are used. The locations of the sensors can be seen in figure 8. The sensors that are available now, have been installed starting from January 2019, so the data used in this thesis is collected over two years. The result of the observations of the sensors is a dataset with a pedestrian count per sensor per day.

The tool that is created allows the user to select two time periods which consist of one or more days. The average number of pedestrians per day are computed of each period. These averages are then subtracted and interpolated on the map of Nijmegen, so the user can clearly see how much the second period differed from the first period. Because the average per day is used, longer periods of time can be compared with shorter periods of time. This allows the user to compare events that differ in length, like an eight-day festival compared to a ten-day festival, or a month with 30 days compared to a month with 31 days. If the absolute number of pedestrians would be taken with such comparison, the pedestrian count is biased towards the longer period. To prevent the bias, the absolute pedestrian counts are divided by the number of days.

In the context of COVID-19, events can be periods that measures were enforced, like the period the schools were closed or the period that face mask use was obligatory in public buildings. The tool created shows the influence of these events on the number of pedestrians in the city. However, the tool can visualize the effect of non-COVID-19 related events too, like festivals and market days. The tool can also visualize effects of longer events, like the effect of a certain season on pedestrian counts.

Related work

Traffic modelling, prediction & visualization

A lot of research has been done on traffic modelling and prediction. This thesis is mostly built on the foundation laid by Giržadas in his bachelor thesis on pedestrian traffic prediction and interpolation [1]. Giržadas implemented multiple machine learning models to predict pedestrian activity in the city centre of Nijmegen per hour, per sensor. The prediction is a number for every of the 42 sensor locations, which represents the number of pedestrians the model expected to be there at the selected timepoint. Unfortunately, none of the machine learning algorithms were better than the baseline prediction, which was the average per day of the week per hour. The best performing algorithm is the Support Vector Regression, with an R^2 of 0,669, which scored just below the baseline which has an R^2 of 0,722.

After the data were gathered and predicted, Giržadas interpolated the data on the map of Nijmegen using Radial Basis Function (RBF) interpolation [2], which will also be used in this thesis.

Interpolation is a type of estimation where new data points are constructed on the basis of known datapoints. This method is common in the field of Geographic Information Systems (GIS) and is among other things used for physical, demographic and socioeconomic spatial data mapping, but can also be used for traffic mapping. Because a map is two-dimensional, a two-dimensional RBF is used, but interpolation can also be done for one-dimensional or higher dimensional datasets. RBF interpolation was found to perform well by Zandi et al. [3]. They took soil samples and tested RBF, Ordinary Kriging and Inverse Distance Weighting. RBF had a relative Mean Absolute Error of 0,1645%, which was better than the other interpolation methods they applied.

Another approach to traffic mapping uses a graphical representation, in which edges represent streets and nodes represent crossings [4] [5]. An interactive example of this approach has been made of Enschede [6], which is shown in figure 1. In this map, streets are represented as coloured lines, where the colour of each line represents the number of pedestrians counted in the selected hour. Lines can be selected to see the exact count, as shown in figure 1 at ‘De Heurne’.

Figure 1, Interactive graphical map of Enschede.

Another option for a representation of hourly pedestrian counts is shown in figure 2, where it has been used in the city of Melbourne [7]. Here the data are not interpolated, but the pedestrian count is only shown where these data are collected. The vertical lines above the dots give an indication of the absolute number of pedestrians during the selected hour. The line above Flinders Ln-Swanston St (W) shows a pedestrian count of 3000. Furthermore, the dots are colour-coded to represent their relation to the average of the 4 weeks previous to the selected moment. A yellow, white and blue dot represent more than average, around average and less than average respectively. When a dot is selected, a graph is shown with three lines that show the day's, four weeks' and year's counts per hour.

Figure 2, Dot map of Melbourne.

Events & COVID-19 pedestrian analysis

Unfortunately, only little research has gone into the effect of events on pedestrian count and flow. Researchers working for the Nederlands Studiecentrum Criminaliteit en Rechtshandhaving (NSCR) have researched the effects of obligatory face mask use in Rotterdam and Amsterdam [8]. The NSCR visually inspected camera footage of the zones where face mask use was obligatory and they found that the obligation had no effect on the number of pedestrians.

It is evident that a lot of research has gone into traffic modelling, prediction and visualization. Many different approaches have been conceived to visualize the data to make it clear at just a glance. The opposite is true for the research that has gone into the effects of events on traffic. There is still much to learn in this field, which is why this thesis will attempt to answer the research question: Is it possible to develop a digital visualization tool that allows users to observe the level of impact events have had where and when on the number of pedestrians in the city centre of Nijmegen?

This will be attempted by making a tool that compares two periods that a user can select. One period has a criterion (i.e. event) and the other period does not. The difference between the pedestrian counts of both periods is visualized, so the effect of the criterion on the number of pedestrian traffic is visible. This could, for example, give the city council information about how much the COVID-19 measures have helped to make people stay inside their homes. The tool could also be used by shop owners to make a more informed decision on whether they should open or close their shops on certain days on which an event is planned. In the next sections the data and development of the tool will be discussed, after which three results that are generated by the tool are presented to show the uses of the tool. Lastly the tool is reviewed and its weaknesses, possible improvements and future research in the field are discussed.

Research

In this section the processes of the tool will be discussed. It was developed with Python.

Data

The data which are used for the visualization are in a two-dimensional dataset, retrieved from the Numina server with the use of Giržadas' data retriever class [1]. The dataset has the number of sensors on one axis (42) and the number of hours on the other. The data used in the examples in the results section cover January 11th 2019 – November 30th 2020 which is 16560 hours, but every 24 hours are summed, so the dataset has 16560/24 = 690 days on the time axis.
In theory, shorter periods (hours or minutes) could also be compared, but because of the high variability of the data in short periods, the findings of the tool will be less meaningful than when longer periods (days) are compared. For example, if there is a quiet day with only few pedestrians per hour (around 100), but with one peak (500) and another busy day with many pedestrians per hour (around 500), but with one low point (100) and a comparison is made with hours and the user accidentally selects the peak and the low point, the visualization will show that the hour on the busy day has 400 pedestrians less than the hour on the quiet day. This will give an incomplete picture of the real situation, because on the busy day there are about 400 pedestrians more at all other hours than on the quiet day. To prevent aforementioned situations, it was decided to sum the hours and only allow for comparisons of at least one day.
To see how an event has had impact on the pedestrian counts, two periods are selected. In the following example the periods are the two first days of November of 2019 and 2020 (1-11-2019 to 2-11-2019 and 1-11-2020 to 2-11-2020). The data of these periods are isolated from the dataset. The pedestrian counts of each period are summed for every sensor and then divided by the number of days that the periods are. This results in two arrays with the average number of pedestrians per sensor. In figures 5 and 6 the pedestrian counts of each period are interpolated on the map of Nijmegen. At locations with a grey colour there were almost no pedestrians counted on average during that period and at purple coloured locations that count was 5000 or more. It becomes clear from these pictures that there were more pedestrians counted in 2019 than in 2020, because the colours are higher in figure 3 than in figure 4.

Interpolation

To create these visualizations of pedestrian counts, first a (pixel) grid is created with a size of 1600*950 pixels, which can be scaled by a scaling factor (SF), which is set to 1.0 by default. These dimensions are used, because the overlay image has that resolution. Setting the SF to 1.0 will result in an image with a high definition. When the scaling factor is 1.0 (1600 * 950 * SF^2 – 42 =) 1.519.958 data points have to be estimated by the interpolation method, which can take a long time (up to a few minutes, depending on the device). Setting the SF to a lower value can significantly decrease the time the interpolation process takes.
All pixels in the grid are set with a count, which later gets represented by the colours of the legend in figures 5 and 6. The pixels on the 42 locations of the sensors are set first. They are set with the data of the number of pedestrians of that sensor. Once the known pixels are set to the correct values, the remaining unobserved grid cells (1600 * 950 * SF^2 – 42) are estimated by RBF interpolation [2]. This method takes the distance (r) from other known pixel locations and uses a multiquadric formula to compute their values:

$$\sqrt{\frac{r^2}{\varepsilon} + 1} \tag{1}$$

where ε is the shape parameter, which determines how rapidly the values change between locations. With a high ε the values will change rapidly and with a low ε they will change gradually. Giržadas used a comparable method and used an ε value of 3, based on visual analysis [1]. Applying this method results in an image where the number of pedestrians is represented by colour. These images are shown in figure 3 and 4. To complete the visualization, an overlay image is placed over this image where only the streets are visible because they are transparent and the city blocks are non-transparent. The placement of the overlay image changes the image from figures 3 and 4 to the images from figures 5 and 6.
A comparable method was used in Giržadas' thesis. The approaches differ in legend, because Giržadas used hourly counts and in this thesis daily counts are represented. Giržadas interpolated the information of one hour, but here two periods are compared, and that comparison is interpolated, as will become clear in the next section.

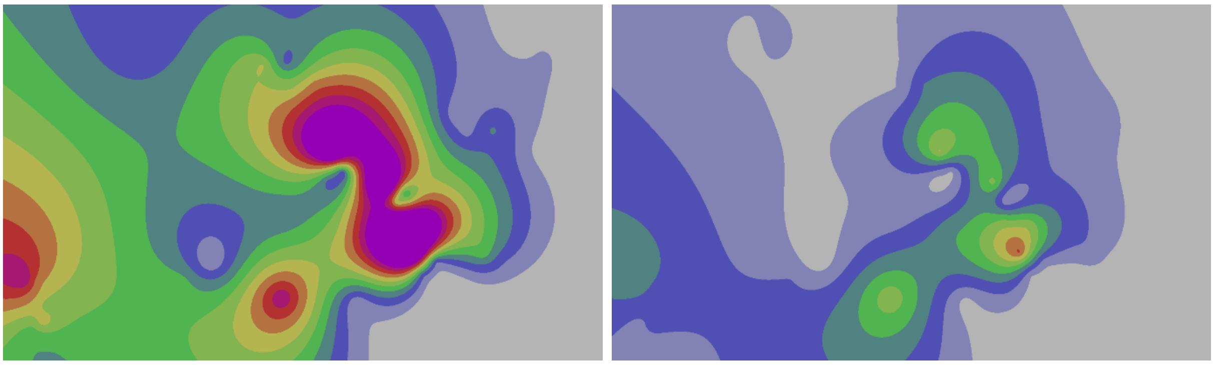

Figure 3 and 4, Pedestrian counts on Nov 1st and 2nd, 2019 and Nov 1st and 2nd, 2020 without overlay image.

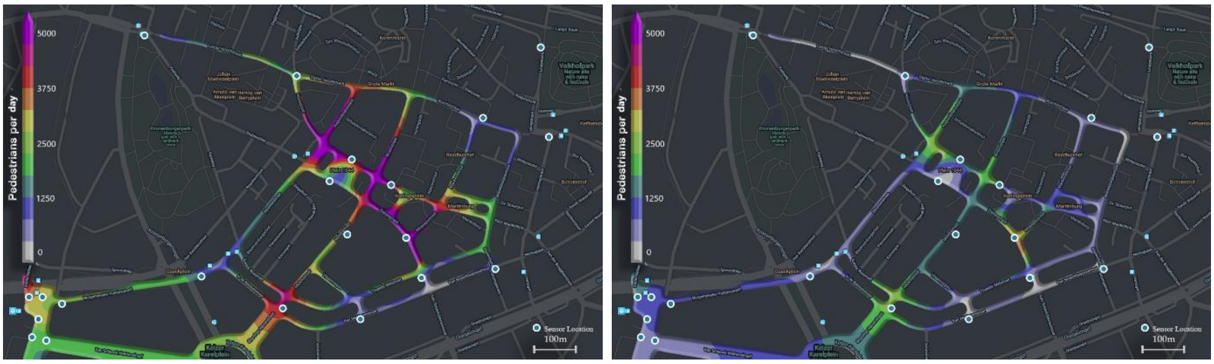

Figure 5 and 6, Pedestrian counts on Nov 1st and 2nd, 2019 and Nov 1st and 2nd, 2020 with overlay image.

These figures show the average number of pedestrians in both periods. In a blue zone for example, there are on average about 1000 pedestrians per day, at a yellow zone there are 3000 pedestrians and at a purple zone there are 5000 or more pedestrians on average. The sensor locations are represented by blue dots with white circles. There are 21 dots and sensors, but there are 42 sensor columns in the data. That is because each sensor can observe multiple areas in its view.

Event effect

Now that the average pedestrian counts are stored in an array for each period, the difference of these pedestrian counts can be computed. Subtracting one array with the other leaves one array with the differences in pedestrian count per sensor location. The data in this array are interpolated and combined with the overlay image, which gives figure 7 as the result. Figure 7 has to be interpreted differently than figure 5 and 6, because this figure shows the difference between figure 5 and 6. Figure 5 and 6 show the absolute average number of pedestrians ranging from 0 to 5000, whereas figure 7 shows the difference, so that range goes from -5000 to +5000. Most days do not exceed a pedestrian count of 5000, so this is chosen as the capacity.
Figure 7 is produced by interpolating the difference of the two arrays of the periods and not by subtracting the interpolation images from figure 3 and 4, because that would take double the amount of time to produce figure 7.

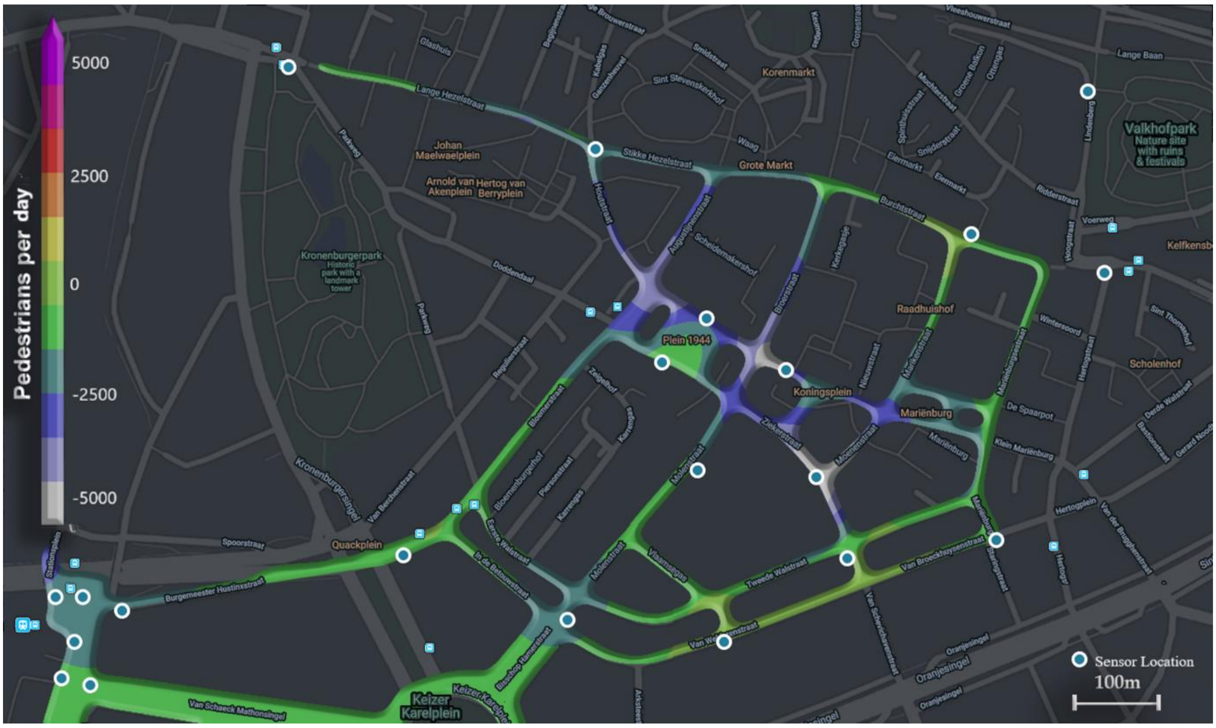

Figure 7, Difference pedestrian counts on Nov 1st and 2nd, 2019 and Nov 1st and 2nd, 2020.

This figure shows the comparison of November 1st and 2nd, 2019 and November 1st and 2nd, 2020. For example, in a light green area, there is no difference in pedestrian counts between both periods. An area that is blue shows there were 3000 less pedestrians in 2020 than 2019. Grey areas show that the count is more than 5000 lower at that location in 2020 than in 2019. If an area were purple, that would indicate that there were more than 5000 more pedestrians in 2020 than in 2019.

There are a few interesting areas that can be focussed on when observing the map of Nijmegen, because their locations could explain some of the pedestrian counts that are visible on the interpolation images. There are some shopping areas, some entertainment areas, supermarkets and a station which are indicated in figure 8.

The differences at most locations in figure 7 are not very big, because most areas are green, which indicates a difference of 1000 between 2019 and 2020. However at Plein 1944, Ziekerstraat and Broerstraat big differences are visible. These areas are grey and light blue.

This shows that there were between 3000-5000+ less pedestrians in 2019 than in 2020 at those locations. In figure 8 it can be seen that these locations are near shopping areas (black with green circle). These areas are generally very busy as can be seen in figure 5 (more than 5000 pedestrians per day). And although these areas are also the peaks in figure 6 (more than 2500 pedestrians per hour), the difference between both periods was so big, that the difference map of figure 7 shows grey at those areas. In conclusion: figure 7 shows there were less pedestrians at the start of November 2020 than one year prior, especially in areas that have a lot of shops, which is most likely caused by COVID-19 and the measures the government has taken to combat the virus.

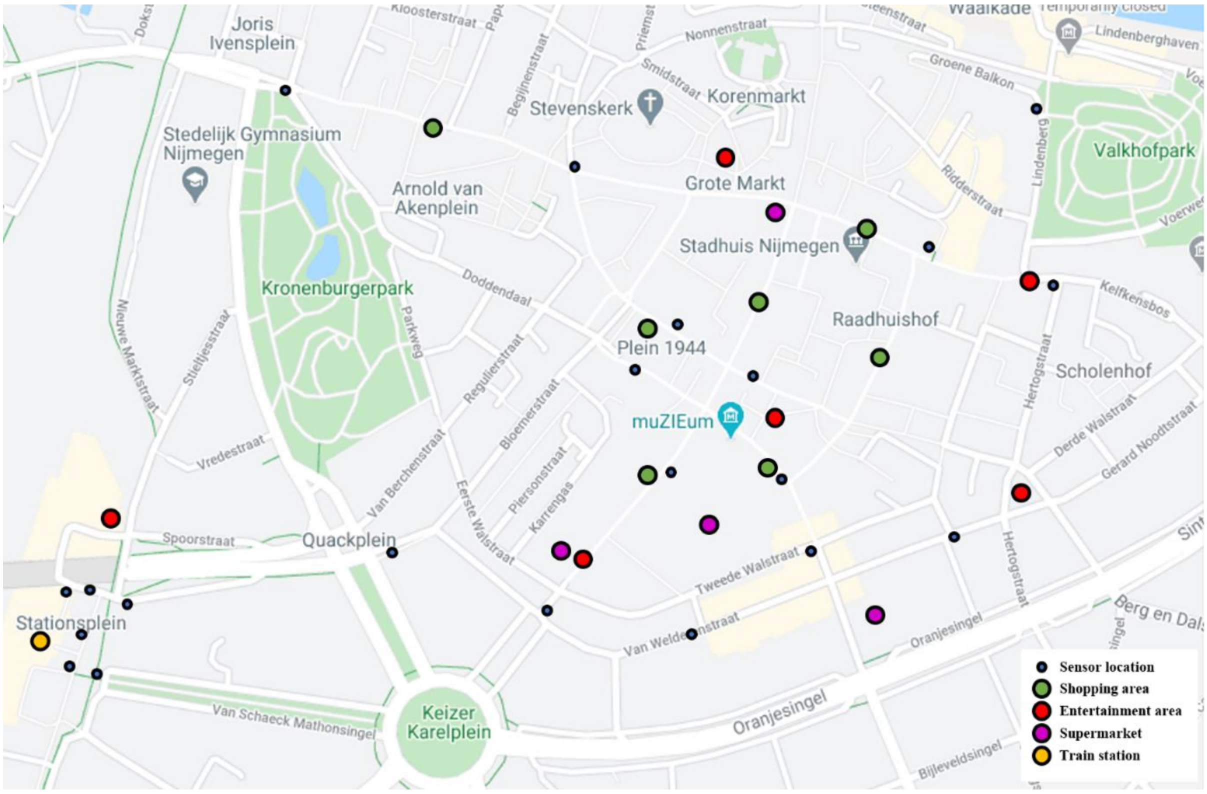

Figure 8, Important areas and locations in the city centre of Nijmegen.

Outlier estimation

The maps that are generated by the tool give a lot of insight in the activity of pedestrians. However, there are a few problems with the data. There are a lot of zeros in the data (8556 out of 29400 = 29,1%), because not all sensors were installed at the same time and some days sensors were turned off. There even were some days that all sensors were turned off and no pedestrians were observed. Next to the zeros in the data, there are also non-zero outliers in the data, which can be caused by sensors being turned off by half a day, so they do observe pedestrians, but their counts are only part of the real data.
Dropping the sensors with some zeros is not a good idea, because a lot of valuable non-zero data will also be lost. Dropping days with zeros is also not a possibility, because it would leave gaps in the selectable periods of the tool, which could lead to missing information. If the dropped dates are included in a selected period, the counts of the dropped dates are not included in the average of the period, so the average will be wrong.
Keeping the zeros in the data is also a bad idea, because then the maps will contain zones that have zero pedestrians, even though there were pedestrians at those locations.
To prevent dropping valuable data and having zeros in the data, it was decided to predict the zeros in the data. This prediction is based on the daily average and the sensor average:

$$datapoint\ prediction = \frac{sensor\ average * daily\ average}{overall\ average} \tag{2}$$

Using this method, the day and sensor information are preserved. This computation ensures that, for example, a busy day in the city is considered when filling in the zeros. An example of this computation can be seen in formula 3, which takes the averages from table 1, which has three days and three sensors. The counts in this table are a lot lower than the real data. This is done to make the computations simple and clear. The real data are stored in a similar fashion. As can be seen in the table sensor 2 was turned off on day 2. At datapoint X 0 pedestrians were observed, so this datapoint has to be predicted. The formula predicts a count of 2,21 pedestrians.

Sensor / Day	1	2	3	Average
1	4	2	3	9/3 = 3
2	7	X (0)	1	8/2 = 4
3	4	2	6	12/3 = 4
Average	15/3 = 5	4/2 = 2	10/3 = 3,33	29/8 = 3,625

Table 1, Example of datapoint prediction.

$$datapoint\ prediction\ (X) = \frac{2 * 4}{3{,}625} = 2{,}21 \tag{3}$$

Unfortunately, there were days during which all sensors were turned off, so computing a daily average is not possible for that day. In these situations, the global daily average is used instead of the formula. Even though this is a good estimation in theory, in practice this estimation does not perform that well. The relative mean absolute error (MAE) is 120,5%. For example, a day with 500 pedestrians at a certain location actually gets estimated at 500 * 220,5 / 100 = 1102,5. The reason for this poor performance is the fact that there are a lot of outliers in the data, which are used to compute the prediction. Days without pedestrians are not included for the estimation, but days with more than zero pedestrians are. That means that when a sensor has been off for half a day that data are still included in the estimation, so the estimations are much lower than the correct data. When the outliers are excluded from the estimator evaluation by including a cut-off at (at least) 300 pedestrians per day for all sensors, the relative MAE is reduced to 28,8%, which is not that bad, but still leaves a lot of room for improvement. The problem with this cut-off, is that it is based on visual inspection of the data and not on computations. A good estimation does not necessarily mean a good cut-off, because they are independent of each other. A good cut-off would find outliers per sensor instead of applying a uniform cut-off for all sensors. In future research a solid outlier-distinguishing-step could be created, so the prediction also works better and the relative MAE can be reduced even further. The prediction which is currently done by formula 2 could also be done by a machine learning algorithm which has been done by others, like Siddiquee and Hoque [10], who have predicted hourly traffic trends with an MAE of 12.67% using an artificial neural network.

Weather

Weather conditions have a big impact on the number of people that go outside [9]. So, when making a comparison between two days or short periods, the weather conditions must be considered. If weather effects are not considered and a criterion’s influence on pedestrian counts is investigated, the outcomes of that investigation could be false.

The weather of both periods is not gathered or integrated in the tool, so the users must consider and incorporate the weather data themselves. This is not always necessary when comparing longer periods, because of the irregular nature of weather conditions, which cause rainy days to get rectified by sunny days for example. However, when comparing a month in the summer with a month in the winter, weather conditions do have to be considered, because on average, slightly more people go into the city centre in the summer than in the winter (difference of 0-1000 pedestrians per day), as shown in figure 9, which has been created by the tool. It shows how many pedestrians were counted in the winter of 2019 (11-01-2019 to 21-03-2019) subtracted by the summer of 2019 (21-06-2019 to 21-09-2019). At most locations there is either no difference or it is very small. Light green areas, at Plein 1944 for example, show no difference and green areas, at Koningsplein for example, show that the pedestrian count during the winter is 1000 less than in the summer. There is one area, between Keizer Karelplein and the Molenstraat, where the difference is very big. This area is coloured blue, so there were 3000 pedestrians less in the winter than in the summer at this location. This can partially be explained by the fact that this is in the centre of an entertainment area, where during summer months people go outside during the day as well as the night, so the pedestrian counts can be expected to be higher during the summer than the winter in this area. However, this is not the case for other entertainment areas, which is unexpected.

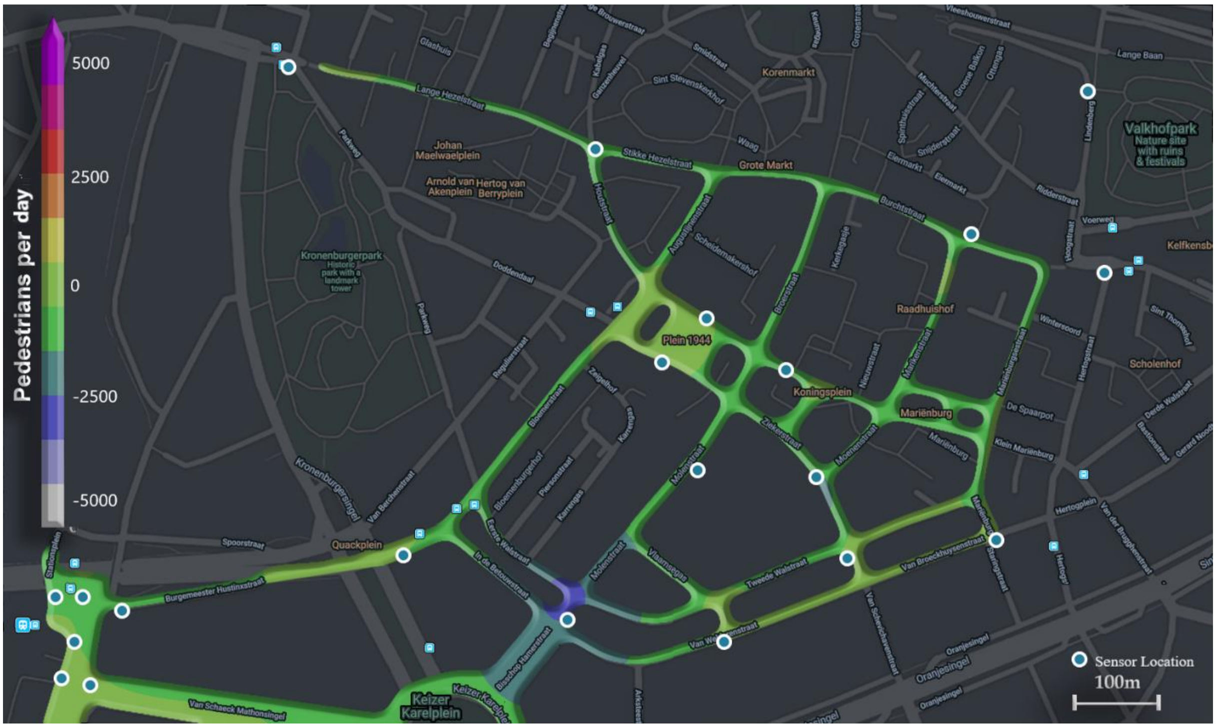

Figure 9, 11-01-2019 to 21-03-2019 minus 21-06-2019 to 21-09-2019.

Results

In the sections below, three examples of the use of the tool are shown, to show it can be used to see the effect of events on the number of pedestrians in Nijmegen.

Vierdaagsefeesten

The first example compares pedestrian activity during the ‘Vierdaagsefeesten’ of 2020 and 2019, which is a big event in Nijmegen that is held in the week of the International Four Days Marches Nijmegen. The duration of both events is eight days.
Because of COVID-19 it can be expected that there are a lot more pedestrians during the time of the ‘Vierdaagsefeesten’ of 2019 than of 2020. In 2020 the festivals were cancelled and only a few bars organised activities, so the figure should show a lot of blue or light blue areas, especially at entertainment areas like the Grote Markt, Molenstraat and Faberplein (see figure 8), because these colours show that there are between 5000+ and 3000 pedestrians more in 2019 than in 2020.
If a street with a sensor was part of the walking route in 2019, it can be expected that the colour of that street is white, because thousands of people take part in the Four Days Marches and they all pass the sensor. The only street in the map that was part of the route was the Van Broeckhuysenstraat in the southeast corner of the map, which was used in the walking route on one of the four days.

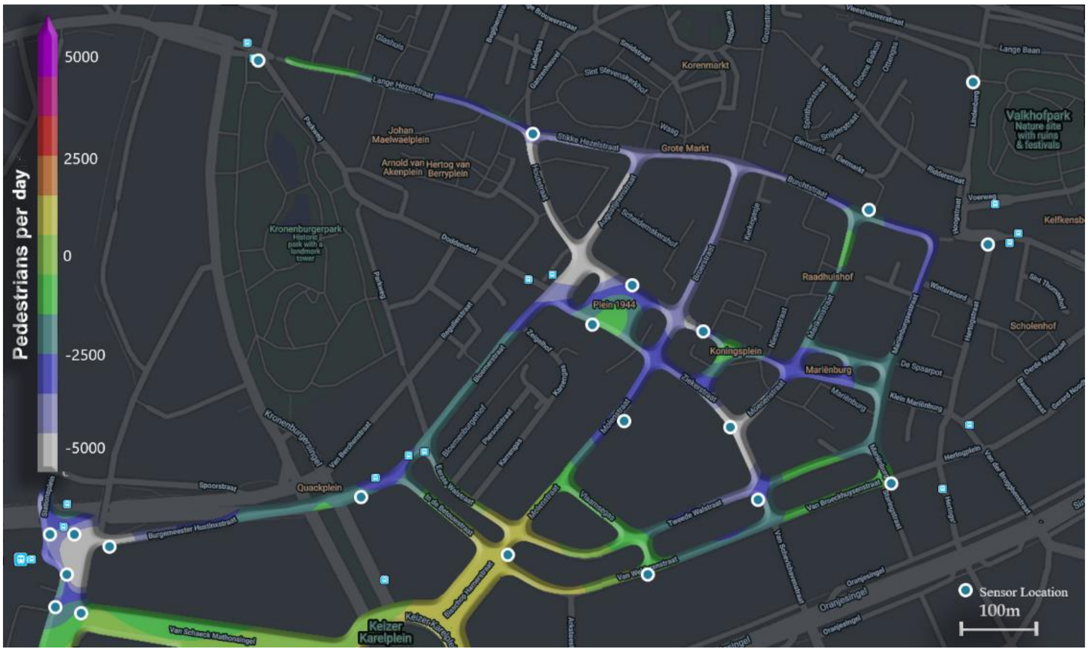

Figure 10, 18-07-2020 to 25-07-2020 minus 13-07-2019 to 20-07-2019.

There are a few places that are interesting in figure 10. At the station square, at the southwest corner of the map, there were more than 5000 pedestrians less than one year prior. This is an indication that citizens followed the rule correctly that forbids them to use public transport without necessity.
Something that is surprising, is the fact that there were about 1000 more people near Keizer Karelplein in 2020 than the year before. It is possible that there still were a lot of people in that area in 2020, since it is near an entertainment area and citizens of Nijmegen might still have wanted to go out during the Four Days Marches. It is also near two supermarkets which still have to be visited by citizens to get their groceries.

There is only a little difference in the number of pedestrians per day in the area of the Van Broeckhuysenstraat, which is also surprising, because it was part of the walking route in 2019, so the expectation was that there were far fewer pedestrians in 2020. The little difference might also be caused by the fact that the sensor in the Van Broeckhuysenstraat also is near an entertainment area or the sensor might have been observing another street than the street where the walking route lead the participants in the Four Days Marches.

The first lockdown

The second example shows that the tool can make a comparison between a short period and a long period of time. In this example the duration of the periods are one month and approximately two years, which is the period in which all available data have been collected. This comparison shows how much the specified month differs from the total average.
The first peak in new contaminations of the coronavirus lasted from approximately the entirety of April 2020. The comparison of this month with all available data, i.e. two years, will show the influence of the first peak in new contaminations on the number of people that go outside.
In figure 11 it is visible that there is a big decrease in the number of people in the city centre, especially in places that are normally crowded like Plein 1944 and Broerstraat there is a big decrease in the number of pedestrians, and there are no increased pedestrian counts at any of the sensor locations (no yellow or higher), so by observing this map it can be deduced that the request of the government to stay home has had a big impact on the number of people that go into the city centre of Nijmegen.
The specific periods in this example are from April 1st, 2020 to April 30th and the full time which the sensors have been available at the time of generating the example, which is January 1st, 2019 to November 30th, 2020.

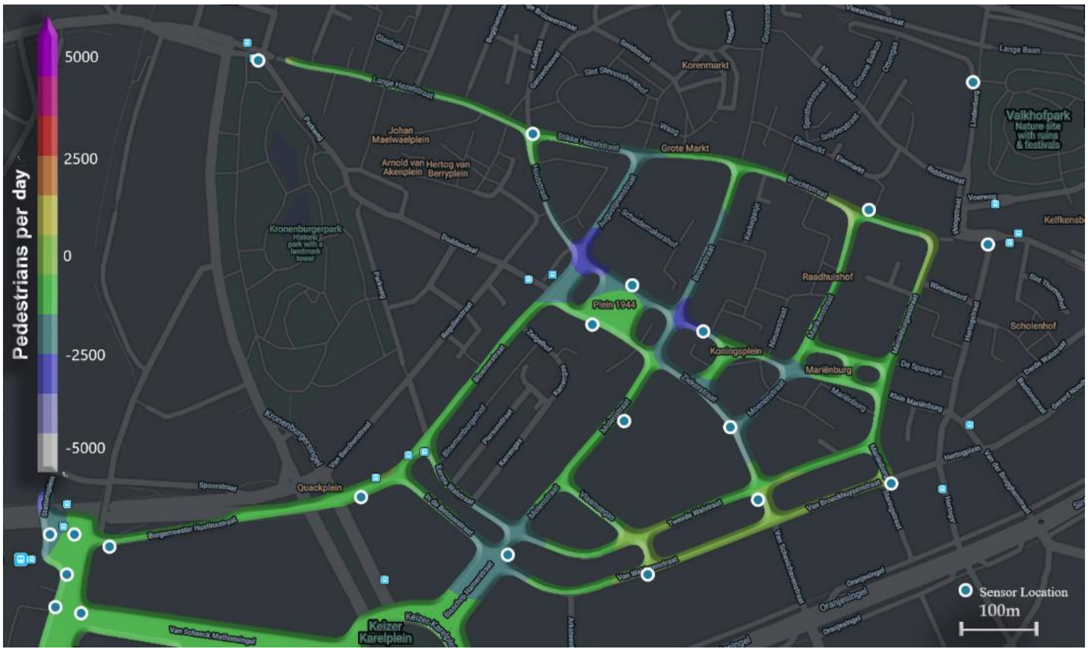

Figure 11, 01-04-2020 to 30-04-2020 minus 01-01-2019 to 30-11-2020.

The number of people counted by the sensors in the entirety of 2020 is a lot lower than the count in 2019, likely due to the COVID-19 measures enforced in 2020. Therefore, the current total average is biased towards lower pedestrian counts and does not represent the all-time average very well, which also represents data before the sensors were installed. The current total average, which is computed with the sensor data that has been available since 2019, will probably be a lot lower than the all-time average, which includes data from before 2019, but as more data are collected, this total average will become more representative of the all-time situation. When more data are collected in the following years, the influence of COVID-19 on the total average becomes smaller, whereas during the generation of these results, COVID-19 influences about half of the available data.

The figure above compares two long and different periods: one month and 23 months. If the user is interested in how a period compares to the all-time average, the example above shows that the tool allows them to visualize it. Comparing periods of different lengths can also be useful when comparing two events. For example an eight-day event like the Four Days Marches can be compared with a ten-day event like the introduction of the Radboud University, to see how the events influence the pedestrian counts. However, the tool can also be used to compare just two days, which is shown in the next example.

The first day of lockdown

In this example, the first day of the first lockdown is compared with approximately the same day one year prior. The first day of the lockdown is hard to pinpoint, since a lot of measures were introduced gradually. In this work March 16th is used as the first day of the lockdown, since the evening prior, the government announced that the catering industry and all schools would close, and everyone was meant to keep 1,5-meter distance from each other. March 16th, 2020 is a Monday, but March 16th, 2019 is a Saturday. It is unfair to compare those days, because there are a lot more people in the city on Saturdays, so the comparison is made with Monday, March 18th, 2019 instead. Both days have very similar weather conditions. They both had a maximum temperature of 12 °C and the minima were 6 °C and 7 °C for 2020 and 2019 respectively. In 2020 it was mostly dry and in 2019 it was mostly cloudy.

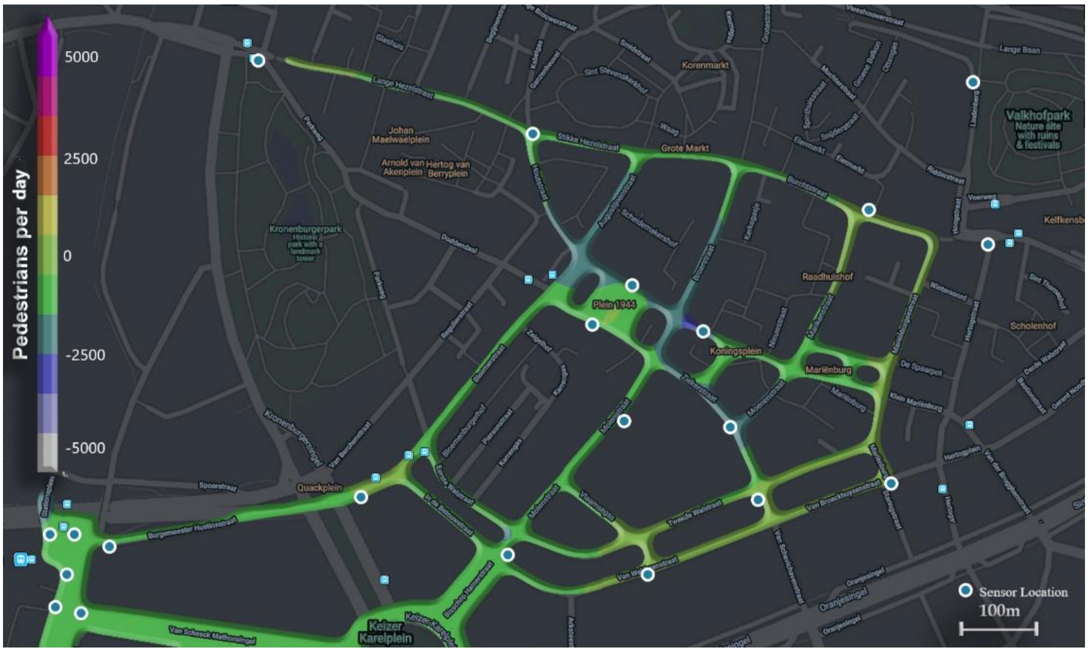

Figure 12, 16-03-2020 minus 18-03-2019.

It comes as no surprise that figure 11 and 12 are very similar, since both compare ‘normal days’ with ‘lockdown days’.

Conclusion

The goal of this thesis is to answer the following question:
Is it possible to develop a digital visualization tool that allows users to observe where and when events have had which level of impact on the number of pedestrians in the city centre of Nijmegen?

From the examples shown in the research section, it is evident that it is indeed possible to create a tool that allows a user to observe the effect of events on the number of pedestrians in the city centre of Nijmegen. The effect is visualized by creating a map by interpolating the difference in sensor data of the number of pedestrians of two periods, one containing the event, the other containing days comparable in various ways (e.g. weather), but without the event. This approach has here been used to generate maps that give insight into the efficiency of the measures enforced by the government regarding COVID-19 and the effect of other events on the number of pedestrians.
One important weakness of the developed tool is the big number of outliers and zeros, which are mostly caused by the fact that the sensors were installed at different points in time. This has caused the zero replacement estimation to be inaccurate (relative MAE = 120,5%). When an outlier threshold is placed the accuracy improves a lot (relative MAE = 28,8%), which shows that the estimation method does have potential. A graphical representation of the centre of Nijmegen instead of the interpolation method, could also improve the accuracy of the map, but more sensors would need to be placed to estimate pedestrian numbers with even higher precision.
Another opportunity that has become clear, is the addition of a system that incorporates weather conditions, when comparing two days or short periods. Combining these systems could create an even more powerful tool for visualizing the effect of events on the number of pedestrians in the city centre of Nijmegen.

Future research

Weather consideration

Luckily, the two days of the last example shown in the results section had very similar weather conditions, so a comparison is fair, since weather plays no role in the difference in pedestrian counts in this example. If the user does want to compare two days that differ a lot in weather conditions, there could be a solution which could be worked out further in future research.
This solution would involve collecting the average number of pedestrians per day for every combination of weather conditions. For example, the conditions could include temperature, ranging from -8 up to 40 °C with margins of 2°C, to reduce the number of conditions, and one Boolean for precipitation (none or rain, snow or hail). This results in

48 (°C difference total) / 2 (°C margins) * 2 (precipitation Boolean) = 48 conditions.

The averages of the conditions can be ranked. A higher number of pedestrians means a higher rank. The scale would go from 1 to 48. The result will be a list of averages per weather condition with a rank, which is based on the number of pedestrians. This is a data driven approach to weather condition ranking. The averages themselves determine their rank. A fictional possible example of the list is shown in table 1.

Average pedestrians per day per weather condition	Weather condition	Rank
5007,9	24 up to 26 °C, no precipitation	48
4812,2	22 up to 24 °C, no precipitation	47
…	…	…
2141,5	15 up to 17 °C, no precipitation	26
2119,8	19 up to 21 °C, precipitation	25
…	…	…
340,1	-6 up to -4 °C, precipitation	2
328,6	-8 up to -6 °C, precipitation	1

Table 2, fictional example of weather condition ranking system.

If the user wants to compare two days, for example on one day it is 20 °C and raining and on the other day it is 16 °C with no precipitation, the comparison would seem unfair at first glance due to the difference in weather conditions. However when looking up the weather conditions of both days in the table it is visible that the days are in rank 25 and 26 and the average pedestrians per day of their weather conditions are very similar, so a comparison would be fair. Weather can be excluded as a possible influence on the difference in pedestrian counts in this comparison. Therefore, the user can be more certain that the criterion (i.e. event) is the cause of the differences visible on the map that is created by the tool.

If two very different days are compared however, a comparison is more unfair and the effects of the criterion can be overshadowed by weather effects. For example, the city council wants to compare the effects of the first day in the Christmas holiday (day c) with the first day of the summer holiday (day s). On day c it is -3 °C and snowing, so it is in rank 2. On day s it is 20°C and raining, so it is in rank 25. While both days' weather conditions and ranks are very different, a good indication of their relative pedestrians counts can still be made, by looking at the number of pedestrians on day c and s relative to the average number of pedestrians of rank 2 and 25 respectively. If on day c 500 pedestrians are observed and on day s 2000 pedestrians are observed, there are more pedestrians on day c (500 / 340,1 = 1,47) than on day s (2000 / 2119,8 = 0,94), relative to their average number of pedestrians based on weather conditions. The quiet summer day with good weather gets penalized and the busy winter day with bad weather gets awarded using this system. These numbers could be considered when comparing days with different weather conditions to allow for a fair comparison, because the weather conditions' effect on the visualization are reduced.

Interpolation and sensor locations

The combination of the interpolation image with the overlay image makes it seem like only the streets' pedestrian activity is generated, however, as can be seen in figures 3 and 4, the interpolation method does not take the locations of streets and city blocks into account. RBF is used for estimating and visualizing geographic phenomena, but it is not perfect for pedestrian flow visualization, because it does not take into account the buildings that are in the way of pedestrians and it makes an estimation and visualization in which it seems like pedestrians can walk though buildings.

A more accurate visualization could be generated using the graphical representation of pedestrian flow. At this point however, this would be very hard to do, because the sensors are too clustered or too spread out throughout Nijmegen.

For an almost perfect graphical representation a sensor has to be placed on every street that has to be represented in the graph. The only error in this representation would be the error in the sensors' observations of pedestrians. The biggest problem with this method, is that it requires

a lot of sensors, as can be seen in figure 13. This is most likely the approach which was used for the map of Enschede in figure 1, as every street section in recorded streets has its own count. Streets in between the recorded streets do not show data, which indicates that no interpolation, graphical or otherwise, has been applied. Although it is likely they used this representation, this cannot be confirmed, because no additional information about the map has been released. Still, this graphical map of Enschede is an accurate example of what a graphical representation would look like when every street section is captured by a sensor.
Another option for a graphical representation is by placing the sensors at intersections. The values of roads connected to the intersections can then be estimated using a Gaussian process regression-based method [4]. Not every intersection has to have a sensor, because surrounding intersections can give information about the intersection in between. This method requires a lot less sensors to be placed, as every sensor gives information about the streets that are connected to its intersection. Figure 13 shows a possible configuration of the sensor locations in the Indische Buurt in Nijmegen, where no sensors have been placed yet. Spreading the sensors in ways similar to the spreading shown in both images, gives more information about the whole area when compared to a clustered sensor location configuration, as seen in the current setup at the station, which is shown in the bottom-left corner in figure 8. Sensors near the station could be spread out to, for example, the Grote Markt and the Waal quay, so more information is gathered instead of multiple sensors being pointed at one area. The sensor locations that give the maximal mutual information can be gathered by applying exploration strategy [4]. The further apart the sensors are, the less accurate the predictions will be of the streets in between sensors. It is also important to have readings of the peripheral areas of the city centre if that area is to be visualized, because if only busy places are observed in the city centre the prediction method will indicate high numbers of pedestrians in the peripheral areas as well, even though there are not that many pedestrians there.
Thus, using exploration strategy and a graphical representation will generate more accurate results than the current interpolation method, which does not take building locations into account. The sensor locations must be chosen so information is maximized.

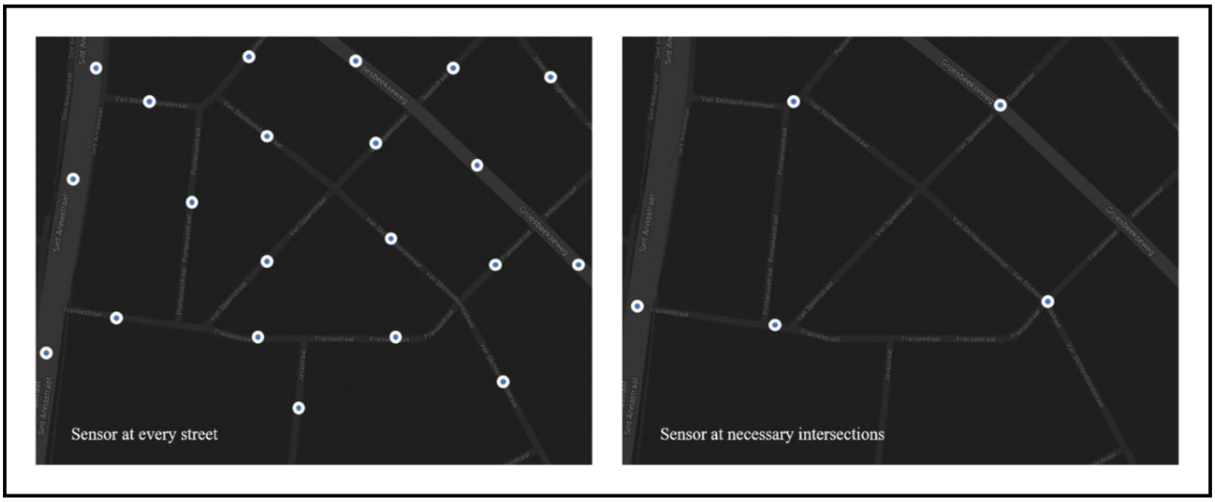

Figure 13, Indische Buurt Nijmegen. Left: every street: 21 sensors. Right: necessary intersections: 5 sensors.

Bibliography

[1] Giržadas, D. (2020). Adapting and Employing Smart City Sensor Data for Strategic Planning (bachelor's thesis). Radboud University, Nijmegen.

[2] Virtanen P et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3), 261-272.

[3] Zandi S., Ghobakhlou A., and Sallis P. (2011). Evaluation of spatial interpolation techniques for mapping soil pH. *19th International Congress on Modelling and Simulation*, 1153-1159.

[4] Liebig T., Xu Z., May M., & Wrobel S. (2012). Pedestrian quantity estimation with trajectory patterns. *Machine Learning and Knowledge Discovery in Databases*, 629-643.

[5] Bilotta, S., & Nesi, P. (2020). Traffic flow reconstruction by solving indeterminacy on traffic distribution at junctions. *Future Generation Computer Systems*, *114*, 649-660. https://doi.org/10.1016/j.future.2020.08.017.

[6] Kennispunt Twente. *Binnenstadsmonitor.* Retrieved from https://www.binnenstadsmonitorenschede.nl/druktebeeld-straten

[7] City of Melbourne. *Pedesetrian Counting System.* http://www.pedestrian.melbourne.vic.gov.au/#date=30-12-2020&time=1

[8] Lindegaard, M. R., Liebst, L. S., Thomas, J., Ejbye-Ernst, P., van Reemst, L. & van Doormaal, N. (2020). *Resultaten van de mondkapjesplicht: naleving, drukte en social distancing in Amsterdam & Rotterdam.* NSCR, Amsterdam. Retrieved from https://nscr.nl/app/uploads/2020/09/NSCRMondkapjesonderzoek.pdf

[9] Vanky, A. P., Verma, S. K., Courtney, T. K., Santi, P., & Ratti, C. (2017). Effect of weather on pedestrian trip count and duration: City-scale evaluations using mobile phone application data. *Preventive Medicine Reports*, *8*, 30-37. https://doi.org/10.1016/j.pmedr.2017.07.002.

[10] Siddiquee, M. S. A. & Hoque, S. (2017). Predicting the daily traffic volume from hourly traffic data using artificial neural network. *Neural Network World*, *27*, 283-294.