Abstract

City governments all over the world face challenges understanding mobility patterns within dense urban environments at high spatial and temporal resolution. While such measures are important to provide insights into the functional patterns of a city, novel quantitative methods, derived from ubiquitous mobile connectivity, are needed, offering decision makers better insights to improve urban management and planning decisions.

In this paper, we propose a model that uses WiFi probe request data to model urban mobility in a dense, mixed-use district in New York City. We collect probe request data over 29 access points of a public WiFi network in Lower Manhattan for one day, accounting for more than 1 million observations and over 60,000 unique devices. First, we aggregate unique entries per access point and per hour, showing that our method has the potential to collect the same type of data as other common methods, by detecting differences in usage activity patterns by time of day. We then use a spatial network analysis to identify edge frequencies of journeys between the network nodes, and apply the results to the road and pedestrian sidewalk network to identify usage and trajectories at the street segment level.

CCS Concepts

• Applied computing → Transportation;

KEYWORDS

Modeling Urban Mobility, WiFi Probe Data, Spatial Network Analysis, Big Data

1 INTRODUCTION

With an annual growth of 60 million new city dwellers every year [19], the world is experiencing a rapid population shift of people moving from rural areas into urban environments over the last several decades. Driven by technological innovations and increasing economic opportunities [5], this situation has led to a steady increase in motorized and pedestrian mobility in cities all over the world [12]. For city governments, this increased demand has lead to challenges in managing city services and infrastructure, and in maintaining quality-of-life standards for its population, as congestion and overcrowding of areas can negatively affect the city’s economy [16], sustainability [22] and its population’s health [8].

To address these challenges, city managers need to understand patterns of urban mobility to enable targeted and “smart” interventions to limit overcrowding, improve service delivery, and ensure effective emergency response. In many cases, methods to measure mobility dynamics focus on reporting traffic counts at specific points in the city at discrete times, typically using rather simple technology, such as pressure hoses or Piezo-electric sensors [15]. Installed at strategic locations of the city, such as at intersections or heavily-used roads, these “counting-gates” are limited in terms of scalability and real-time feedback, and can be cost-intensive when applied to large areas.

With the rise of remote and in-situ sensing technologies, the analysis of closed-circuit-television (CCTV) footage offers a more quantitative approach for computer scientists and urbanists to count not only motor, but also pedestrian traffic on a large scale by using computer vision algorithms [15]. As CCTV cameras have become a common way to observe people and traffic in cities, these methods offer a more scalable and less labor-intensive method to count traffic, being limited only by the number of camera locations.

However, these methods are limited to providing counts at specific, fixed locations for a specific time period. In doing so, they do not provide data about the routes and trajectories of pedestrians, constraining the potential to understand actual street and sidewalk usage over space and time. Current work in data mining aims to fill this gap by using mobile phone data to model urban mobility [2, 9]. As promising results are in terms of accuracy, they show limitations...
In terms of population representation by capturing only mobile users of a specific network provider and typically with low spatial granularity.

Data that are independent of specific network providers are able to capture a larger sample of the population at any given place and time. One example of this is smart device probe requests to WiFi access points (APs) in public urban space. With an increasing number of public WiFi APs and networks in cities, these networks can provide dense coverage across the citiescape, particularly at the neighborhood or district scale. Each AP continuously "senses" its surroundings in terms of potential users equipped with a WiFi-enabled mobile device, which are sending probe requests to available networks and proximate APs at a regular frequency.

As mobile devices are configured by default to scan the environment for WiFi APs automatically to minimize network charges, they regularly transmit packets of data to nearby APs. With an increasing penetration of WiFi connectible mobile devices, such as tablets or smartphones (64% of all U.S. citizens owned a smartphone in 2016 and approximately 80% in New York City)\(^1\), we hypothesize that WiFi probe data can be used to model urban mobility and its traces on a large scale.

In this paper, we use a dataset of WiFi probe requests collected by 29 WiFi APs over the duration of one day in an area of Lower Manhattan in New York City, NY. First, we show how WiFi probe data can be used to report counts at each AP, similar to "counting gates" methods as described above, and used to understand localized activities. Second, we conduct network analysis to describe a spatial network that can be applied to street and sidewalk segments. We demonstrate how these data can be used to analyze common paths of travel and trajectories, indicating the intensity street activity over time.

We begin by presenting current literature on measuring urban mobility, and then present our data and data processing steps. We then introduce our methodology and present results for urban mobility counts and traces. We conclude with an in-depth discussion of the findings, including limitations and applications to city management and planning.

2 LITERATURE

The most commonly used method to capture urban mobility by city agencies is the installment of "counting–gates" at pre-defined places, such as intersections or heavily–used main roads. While technology has improved over the years, the method has remained relatively the same by using, for instance, pneumatic road tubes, Piezo–electric sensors or infrared sensors \(^{15}\) to count primarily motor traffic.

While these methods offer an easy way to quantify the number of cars on a street, they are rather expensive to run due to installation and service charges, compared to the output they provide. Offering traffic aggregations for specific geographic locations within a specific time frame, such as per hour or per day, also create limitations in terms of temporal and geographical scalability.

More current work uses advances in computer vision to analyze closed–circuit–television (CCTV) feeds to count motorized and pedestrian traffic at lower costs. In doing so, researchers and city governments are now able to count traffic at places with CCTV–coverage, such as on intersections for instance, by applying computer vision algorithms, such as blob–detection\(^2\). Focusing primarily on motor traffic, this approach has been extended over the years to also count pedestrians\(^3\).

The analysis of CCTV footage offers effective ways to aggregate traffic quantitatively and is only limited by the number of CCTV–camera locations in a city. As the usage of CCTV cameras in the urban environment is growing due to congestion and security concerns, the method becomes increasingly applicable to count traffic on a large scale. However, in focusing on traffic counts, it does not offer any insight into the routes people take between their locations, and hence do not provide information about usage frequency on road level.

The increasing availability of open data has offered researchers novel opportunities to study traffic routes, in particular for public transport, on a large scale following a data mining approach. In doing so, metro journeys\(^4\), the use of public bike sharing schemes \(^{20}\) or GPS traces of taxis \(^{6}\) for instance have been visualized and the time-dependent frequencies of routes through cities detected.

While the results of such studies can contribute to the efficiency of public transport systems, these open data sources do not include information about the population who do not use public transport. As many people in U.S. cities travel by car or increasingly walk \(^{13}\), using these data sources excludes a large portion of the urban population and is therefore not representative of the entire urban population.

A data source that includes these populations are call detail records (CDR). With the increased use of mobile phones over the last decade, CDR data from mobile phone providers have become a popular source for urban mobility research. For instance, Yuan and Raubal \(^{21}\) extracted dynamic mobility patterns in urban areas using a ‘Dynamic Time Wrapping’ algorithm, and were able to classify areas according to the patterns.

Calabrese et al. \(^2\) combined mobile phone traces and odometer readings from annual vehicle safety inspections to map mobility as averaged individual total trip lengths for the case of Boston. In doing so, researchers found for instance, that the 2 most important factors for regional variations in mobility are accessibility to work and non–work destinations, while population density and mix of land–use showed less significance.

Other work uses CDR data to model urban flow instead. Gonzalez et al. \(^7\), for instance, studied 100.000 mobile phone user trajectories over 6 months and found, that human trajectories show a high degree of temporal and spatial regularity. Furthermore, findings suggest that humans follow rather simple reproducible mobility patterns.

These works show the opportunities of using CDR data to research urban mobility on large scale. However, at the same time telecommunication data is highly confidential and often difficult to access for researchers. One possible way to gain access to such data is to take part in a data mining challenges, such as the Data

---

\(^2\)http://www.trafficvision.com
\(^3\)http://www.placemeter.com
\(^5\)http://www.trafficvision.com
\(^6\)http://www.placemeter.com
\(^7\)http://wgallia.com/#!underground
for Development challenge\textsuperscript{5} and the Big Data Challenge\textsuperscript{6}, where providers make parts of their data publicly available. However, as available data is being pre-processed, its accuracy often suffers due to unknown data processing steps \textsuperscript{17}. Furthermore, as far as data is provided by various mobile phone providers, it offers a low representation of urban population by keeping out various groups of people that use other providers, or people that do not have a cell phone contract, such as the elder and poor, or use Pay–As–You–Go options.

A data set that includes those groups is Wifi probe data by being provider independent. By default, a mobile device is configured to steadily scan its environment for available Wifi APs and hence, continues to transmit data, including time, geolocation and the device MAC address. Such data has been used by researchers to predict user destinations and routes through a probabilistic approach that uses visited destinations according to Wifi traces \textsuperscript{4}. More recent work \textsuperscript{14} uses Wifi data that collected through a large scale study, to capture human mobility for the case of Copenhagen. Based on a previous scientific study, the data contains a high level of information about the users and uses data of logged–in users only. In reality, Wifi probe data that has been collected ‘in the wild’ is less detailed and the number of people logging–in to Wifi networks is limited. These circumstances lead to questions about the suitability of this approach for quantitative mobility modeling.

In a recent study\textsuperscript{7}, London’s public transport provider ‘Transport for London’ (TfL) uses Wifi probe data collected from APs of subway station within its network to model passenger flow of its customers. While TfL’s ‘Oyster’ travel card data allowed the detection of traveler journey’s origin and destination, the data source did not provide information about the route the passenger took in between them, as for instance, where passengers change trains. Adding Wifi probe data to the analysis shed light into this limitation, as researchers were now able to detect the different routes and train changes people took to reach their destination. Furthermore, by using Wifi probe data from its network, TfL was also able to track people’s movement in stations, so to detect busy platforms and dispatch staff accordingly.

This study shows the potential of Wifi probe data, collected by public transport’s network provider, to model mobility within a subway network and within its stations. In this paper we show how similar data, collected by public Wifi APs, can be used to model urban mobility on large scale in a densely populated area of New York City instead.

3 METHOD AND RESULTS

3.1 Dataset description

The method we are proposing requires access to a Wifi probe data set, that contains information about detected client devices and their proximity to surrounding Wifi APs at a specific time. For this study we use a dataset provided by the ‘Alliance for Downtown New York\textsuperscript{8}, the management for Downtown–Lower Manhattan’s Business Improvement District, covering an area within Manhattan’s Lower East Side in New York City (Figure 1).

With its mix of high density commercial (colored red) and residential (colored yellow) buildings,\textsuperscript{9} Lower Manhattan represents a very dynamic and diverse environment, which makes it suitable for our urban mobility testing, as different groups of people are expected to activate the area at different times throughout the day.

Our dataset covered observations from 29 Wifi AP locations throughout the study area along Water and South Street, from Broad Street in the South–West to Fulton Street in the North–East, and included APs on two of the East River piers, as shown in Figure 1. Data was collected on April, 18th 2016 between 4.00 AM and 6.00 PM, and includes a total of 1,012,519 data points. Every data point includes the MAC address of the observing AP (apMac), the client device’s MAC address (clientMac), the device’s received signal strength indicator as seen by AP (rssi) and the observation time (seenTime). Locations for each AP were stored in a separate shapefile by the provider, containing street name and nearest building number to the Wifi AP, which we visualized using mapPLUTO for NYC.\textsuperscript{10}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Location and land–use map of Wifi access points (APs) on Manhattan’s Lower East Side / New York City, as used in this study. Wifi APs are marked as black dots; red areas are commercial use, yellow are residence areas and purple are industrial areas.}
\end{figure}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
Column 1 & Column 2 \\
\hline
A & B \\
C & D \\
\hline
\end{tabular}
\caption{Table of values}
\end{table}

\footnotesize{\textsuperscript{5}http://www.d4d.orange.com/en/Accueil
\textsuperscript{6}http://www.telecomitalia.com/tit/en/bigdatachallenge.html
\textsuperscript{7}https://tfl.gov.uk
\textsuperscript{8}http://www.downtownny.com
\textsuperscript{9}http://maps.nyc.gov/doit/nycitymap/
\textsuperscript{10}https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page}
3.2 Data pre-processing

Data, such as our WiFi probe data set, is very sensitive through the provision of MAC addresses of clients. Combined with logged WiFi AP locations, such data can lead to privacy issues as it opens opportunities for tracking individual people. To ensure anonymity, we first de-identified our dataset by replacing MAC addresses of clients with an anonymous identifier, consisting of a unique incremental integer starting from 1. In doing so, the data could not be traced back to the individual and hence, would eliminate many underlying privacy concerns. (The study received IRB approval with IRB No.: #IRB-FY2017-526.)

With the de-identified dataset, we next checked for uniqueness of data points and possible missing input to ensure its completeness for our analysis. In doing so, we noticed that 74 data points were counted multiple times, and 428 data points were missing geographical or temporal information. To provide completeness, we removed these data points from the dataset.

We also observed that some of the client devices were captured by multiple WiFi APs at the same time. As this would lead to spatial inaccuracy for our results, we removed multiple entries from our dataset, ensuring to keep the entry for the client’s closest AP only. The rssi information provided in the dataset provides a useful indicator for this purpose: By keeping the entry with the highest rssi-value, we keep the closest WiFi AP location and excluded the others from the dataset.

Furthermore, we examined the WiFi APs each client was connected to and observed that while most clients showed connections to various APs, indicating movement between the locations, some clients were only captured by a single AP during the entire study period. As people might pause for a while on the go, we removed data points that showed no movement for longer than 20 minutes, as suggested by [1]. These device connection patterns are indicative of being stationary and hence would skew the mobility model by increasing an AP’s use frequency. As a matter of fact, most of the deleted clients were detected by the same AP throughout the whole day. The removal of detected faulty data – especially through the stationary devices – resulted in a total of 259,444 entries, collected from 60,970 unique clients. Overall, we removed 74.37% of the observations from our raw data.

After cleaning the data, we joined location and observation data sets by assigning AP addresses to each observation, according to AP ID number, and mapped their locations.

3.3 Aggregation

Having cleaned the dataset from multiple entries and assigned AP locations, we were now able to build a model of urban mobility. As described in the literature section, the most common way for city governments to model urban mobility is the installation of "counting-gates", providing aggregated traffic counts at specific locations for a defined temporal unit, such per hour or weekday. By scanning and detecting WiFi-enabled devices in hyper-local environments, we show now how WiFi APs can be used in a similar way to count the number of pedestrians at a given time. Therefore, as a first step, we aggregate the number of entries per AP and per hour to detect differences, such as low and peak hours, in frequencies of counts over the day.

Looking at the frequencies of entries for each AP we observe a long-tail distribution: while the most-frequented AP shows 137,087 entries, the least-frequented AP showed 1601 entries; the overall mean frequency among observed APs was 31,320 (1 Quartile = 13,520, 3 Quartile = 42,810).

In Figure 2 we visualize the entry distribution per hour for observed time frame, with the most-frequented AP reaching up to over 17,000 entries in the hour between 3PM-4PM, reflecting the above outcome.

While we observe differences in frequencies between WiFi APs, as, for instance, some APs are used more frequently than others, we can see a similar temporal pattern of frequency rise towards the afternoon across all APs in the network. Usage is generally low before 6AM, as expected, and then increases after 9AM, reaching a first peak at noon (12PM-1PM) and the main peak between 3PM and 4PM.

These results are rather surprising as we would have expected a first peak in WiFi usage in the morning rush-hours (around 8AM), in addition to peaks at noon and in the late afternoon. However, we interpret this outcome as reflection of the social dynamics of this area, resulting from the mix of land-use, which is mostly commercial and residential as shown in Figure 1, and the fact that Lower Manhattan is very popular among visitors and tourists from all over the world. Activity in the area begins around 6AM, with commuters starting to go to work. Activity picks up at 9AM when, besides commuters, residents and visitors start their routine. Around noon, employees from local businesses have their lunch break, often in nearby restaurants. They go back to work until 3PM-4PM, when they commute back home.
To see how this pattern is reflected spatially, we next mapped these aggregations according to AP locations per hour. In Figure 3 we show aggregated frequencies for WiFi APs (indicated by colors) per hour, starting from 5AM (top left map) until 5PM (bottom right map). We observe geographic differences in usage frequencies by AP location depending on time of day. The general increase in AP usage in the afternoon, which we have observed in the temporal analysis, is also visible here and indicated by circle areas for each AP location – the larger the radius, the higher the number of aggregated clients.

The temporal variance between morning (4AM-9AM) and afternoon hours (2PM-6PM) was found to be highest among APs that are located in the south–western part of the area, around Wall Street, compared to APs in the north–eastern corner, around Fulton Street. We furthermore observe that WiFi APs that are located on the East River piers slowly increase their usage throughout the morning and reach their peaks in the later afternoon, after 3PM.

Looking at the land–use of the area (Figure 1), this result is not surprising and reflects the area’s usage dynamics, as the south–western and north–eastern part of observed area attract different groups of people: while Wall Street and adjacent areas are known to be prime business locations in New York City, this area is most likely to attract commuters activating the area during office hours. The corner of Fulton / Water Street, on the other side, is mostly occupied by multi–family residences and hence, is most likely to attract residents. While commuters in commercial districts were detected by APs especially when coming or leaving the area, or during the lunch hour, the detection of local residents shows a more even temporal usage distribution throughout the day.

In summary these results show how WiFi APs can be used to measure aggregations representing population activity at given locations, offering a quantitative, less expensive alternative to common approaches of “counting–gates” as described in the literature. However, by measuring client aggregations at WiFi locations, the output offers only a limited picture as it does not include any information about where or how people move between these locations. Therefore this method is limited when it comes to evaluating actual street and sidewalk network usage.
Next we will show how the same Wi-Fi probe data can be used not only to generate aggregations at points in the city, but also to capture movements between those to model urban mobility at both the street segment level and on larger district scales.

### 3.4 User traces between Wi-Fi locations

To model urban mobility on road segment level, we first run a spatial network analysis to identify client traces between Wi-Fi APs. To do so, we filtered by client ID number and journey for our 60,912 users. A journey was defined as the series of visited Wi-Fi APs by each user throughout the day. In Figure 4 we show the network graph, as defined by all detected journeys, with edge frequencies between the Wi-Fi AP nodes. Frequencies showed a long–tail distribution: while the most–frequented journey show 2481 trips, the least–frequented journeys were single trips between two Wi-Fi APs; the overall mean frequency among observed journeys was 147.6 (1. Quartile = 11.0, 3. Quartile = 157.8). In the graph, we indicate the frequency of journeys between the nodes by the width of the edges – the wider an edge, the more journeys were detected.

We observe a high density of client journeys between the nodes located along Water Street, the main thoroughfare in Lower Manhattan and important traffic route for pedestrians, public transport and motor traffic, connecting the area from east to west.

We also observe a high density of client journeys between the nodes located along Water Street, the main thoroughfare in Lower Manhattan and important traffic route for pedestrians, public transport and motor traffic, connecting the area from east to west.

We observe a high density of client journeys between the nodes located along Water Street, the main thoroughfare in Lower Manhattan and important traffic route for pedestrians, public transport and motor traffic, connecting the area from east to west.

We also observe a high density of client journeys between the nodes located along Water Street, the main thoroughfare in Lower Manhattan and important traffic route for pedestrians, public transport and motor traffic, connecting the area from east to west.

We also observe a high density of client journeys between the nodes located along Water Street, the main thoroughfare in Lower Manhattan and important traffic route for pedestrians, public transport and motor traffic, connecting the area from east to west.

We also observe a high density of client journeys between the nodes located along Water Street, the main thoroughfare in Lower Manhattan and important traffic route for pedestrians, public transport and motor traffic, connecting the area from east to west.

We also observe a high density of client journeys between the nodes located along Water Street, the main thoroughfare in Lower Manhattan and important traffic route for pedestrians, public transport and motor traffic, connecting the area from east to west.
ferry landing points at East River piers to offices in the commercial district.

Road usage in the residential dominated areas around Fulton Street of the north–west is found to be lower when compared to the areas of commercial uses, as expected. Looking at differences in scale of the urban environment, we can observe that in this area we find smaller block size, compared to the south–western part. As smaller blocks, crossed by more roads, offer more options in route–finding, the high frequencies as seen in Figure 4 might be explained by pedestrians spreading out over various routes.

In summary, our results show the potential of using Wifi probe data to model urban mobility quantitatively, even beyond aggregations of passers–bys at specific locations in the urban fabric. Next, we will discuss the results of our approach and its limitations and discuss applications and future steps of research.

4 DISCUSSION

4.1 Summary

In this paper, we have presented a novel method to model urban mobility on large scale, following a data analysis approach. The method requires access to Wifi probe data, including information about client activities and location information of its Wifi APs. From the dataset, we extracted aggregations about client frequencies per Wifi AP location, showing the method’s potential to capture similar data as commonly used “counting–gates” on a quantitative level and at low cost.

We then used the same Wifi probe data in a network analysis, combined it with open source road network data, and were able to evaluate journeys and their frequencies on road segment level between Wifi APs based on the shortest routes.

With respect to the area’s land–use pattern, both approaches allowed interpretations on how different groups of the urban population, such as workers or residents, occupy and use urban space at different times based on specific usage patterns. In doing so, we were able to detect temporal differences between routes within primarily commercial and residential areas and their connections to public transport, such as water taxi landings. This pattern recognition shows the method’s potential to identify not only how different population types move through localized areas in cities.

As the network of Wifi APs in cities is continuously becoming denser, the use of Wifi probe data to model urban mobility is becoming increasingly precise. Supported by the ongoing open data movement, an increasing amount of road network data for cities in different parts of the world is freely available and can be used for these purposes. Wifi probe data on the other hand is more difficult to access, but a variety of data mining challenges, such as Accenture’s Wifi Analytics Case Competition[12] or the MIT Big Data Living Labs[13] show a clear trend of Wifi providers and research facilities making their data available, at least at low spatial resolution, to the public. This development suggests that the proposed methodology will become increasingly applicable in the coming years.

4.2 Limitations

We acknowledge that this exploratory work is constrained by a number of limitations. First of all, using passively collected Wifi probe data, we have to be aware that our method is highly dependent on the overall number and network density of AP locations data has been collected from: The higher the density of APs, the more accurate the model’s performance becomes, we expect. In our case, we have shown the potential of the approach for cities with a relatively high penetration of Wifi APs, such as New York City. However, we do not know how the same approach would perform in other cities with a lower density of Wifi APs, as for instance, in less developed countries. As a matter of fact, as cities in these parts of the world show the highest rise in urban population[18] the method would contribute most to urban planning in such environments.

Besides AP network density, the method we propose is highly dependent on the accuracy of collected data. Accuracy could suffer, for instance, due to shielding problems (e.g. shielded by buildings) especially in densely built urban areas. Therefore researchers using this method need to be aware of what Wifi APs are able to capture, and what is meaningful to the purpose of modeling urban mobility. In our case, we also had to exclude a rather large part - three–quarters – of our data points that were identified as stationary devices (as for instance, a desktop computer in an office building) to ensure the model’s accuracy. By not changing their location, such devices are constantly detected by the same surrounding Wifi APs, and hence captured in the dataset throughout the day. Accurately detecting stationary devices is thus critical to the accuracy of the model.

By using Wifi probe data, we need to be aware of what it is able to represent, and what is unobservable. As described in the literature, we expect Wifi probe data to be more suitable to represent urban population compared to CDR data through being network–independent: As far as penetration of Wifi–capable devices in the U.S. is rather high (64% of all U.S. citizens owned a smartphone in 2016)[14], we can expect a high accuracy in outcome in the case of New York City. However, as promising these numbers are, they also indicate that our method still misses up to a third of the population, as we are not able to identify mobility of populations not carrying a smart device and are therefore not being captured by Wifi APs. This fact can lead to bias in the model outputs, particularly as it pertains to understanding the mobility of children, the elderly, and other population sub-groups less likely to possess a Wifi–enabled device.

In this work, we do not differentiate between pedestrian or motorist. As some roads are for motor traffic only, this might lead to inaccuracies in route generation. For instance, we observe in Figure 5 that roads along the East River (South Street and FDR Drive) are being heavily used by clients. As in reality parts of these routes are for motor traffic only, some of the generated routes might be misleading for individual cases. Furthermore, here we apply the shortest–route algorithm to generate our result. As people do not always follow the shortest route and may not always follow the street or road, we need to consider the application of other routing algorithms to be more suitable.

4.3 Implications

Recognizing these limitations, we can see the potential of this approach to benefit to a number of communities to model and research urban mobility. These communities include:

1. **Researchers** in urban and computational studies can use the method to model and analyze people flow and detect to what degree urban mobility patterns impact conditions in the urban environment, or the other way around. In doing so, the output can be used, for instance, to describe possible relationships to urban phenomena, such as crime, congestion or property values, by including other urban data sources, such as crime data, road construction data or house price data.

Other research can use the method to further discuss the relationship between urban mobility and local–use, and discuss opportunities to describe communities based on their movement patterns. By describing urban communities based on movement patterns, our method opens doors to investigate novel approaches to describe socio-economic patterns of a city. Results of such work would support recent work [10] suggesting WiFi data as dynamic alternative to census data.

2. **For city governments and decision makers** tools can be built on the top of our method, informing urban planning decisions. Such tools can, for instance, help to detect mobility patterns in a city over time, and how they would change in case of emergencies. The outputs of such tools can inform road improvement and construction work in order to avoid unnecessary traffic congestion. Other tools based on this method could inform planning decisions by modeling urban mobility on different road network variations to see how mobility flow would change and affect urban life. An optimized design solution can be defined and inform construction decisions.

3. **At the same time**, the method can also support local economy to increase its business activity. In doing so, the model could be used, for instance, to suggest locations that are considered to increase the revenue by providing a high rate of passing–by potential customers. Furthermore, as the method is able to describe different types of passers–by, such as residents or workers, it provides detailed insight for businesses into who the potential customers are, informing location decisions.

4. **For urban population**, mobile applications can be built based on our method, informing both motorized and pedestrian urban navigation in a dynamic way. Such application can, for instance, suggest routes with less car–congestion in real–time or suggest walking route preferences to avoid or seek crowds, depending on the user’s mood. In offering mood–defined routes, such applications have the potential to support urban walkability, impacting urban quality-of-life and the city’s sustainability.

5 CONCLUSION AND FUTURE WORK

In this paper we have presented a novel methodology in its early stages to model urban mobility at scale and at low cost by using WiFi probe request data, and were able to relate detected movement patterns to the local street network. As cities all over the world continue to grow dramatically, being able to measure urban mobility is key to providing improved quality–of-life for city dwellers. Our approach will contribute therefore to various communities in and outside of academia, to a city’s economy, and to the general public.

To do so, we will continue to improve the method by minimizing detected limitations and refining the methodology to increase its accuracy. Therefore, we will in a next step make an effort to differentiate between pedestrians and motor–traffic by using the the temporal information provided in the data as proxy for client’s movement speed. By combining output with additional information about the road types (if pedestrian–friendly or not), we will be able to inform routing procedures of the model accordingly and will provide a more accurate picture of urban mobility. To minimize flaws due to route generation, we will review and test routing algorithms other than the shortest–route.

Furthermore, we have shown in presented work how the method performs on a data set that captures 14 hours of one day only, for a test bed that is limited to 29 WiFi AP’s in New York City. To see how the method performs on larger scale, we will next include data capturing a longer time frame and a larger geographic area. These steps will include not only New York City, but also, to detect possible difference depending on built and social environment, in other cities, in other countries and of other cultures.

REFERENCES


