ARTICLE IN PRESS

Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Digital footprints: Using WiFi probe and locational data to analyze human mobility trajectories in cities

Martin W. Traunmueller[a,*], Nicholas Johnson[b,*], Awais Malik[c,*], Constantine E. Kontokosta[c,*]

[a] New York University Center for Urban Science & Progress, United States
[b] University of Warwick & New York University Center for Urban Science & Progress, United States
[c] Dept. of Civil & Urban Engineering, Center for Urban Science & Progress, United States

ARTICLE INFO

Keywords:
Modeling urban mobility
WiFi probe data
Spatial network analysis
Big data

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 to provide policy-makers with better insights to improve urban management and planning decisions. In this paper, we develop a model that uses large-scale WiFi probe request data to model urban mobility trajectories in dense urban environments. We collect probe request data from a public Wifi network with 54 access points in the Lower Manhattan section of New York City over one week, accounting for more than 30 million observations and over 800,000 unique devices. First, we aggregate unique entries per access point and per hour, demonstrating the potential to use WiFi data to approximate local population counts by type of user. We then use a spatial network analysis to identify edge frequencies and directions of journeys between the network nodes, and apply the results to the road and pedestrian sidewalk network to identify usage intensity levels and trajectories for individual street segments. We demonstrate the significant potential in the use of WiFi probe request data for understanding mobility patterns in cities, while highlighting non-trivial issues in data privacy raised by the growing availability of public WiFi networks.

1. Introduction

With an annual growth of 60 million new city dwellers every year (U. WHO, 2010), 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 (Dargay, Gately, & Sommer, 2007), this situation has led to a steady increase in motorized and pedestrian mobility activity in cities all over the world (Millard-Ball & Schipper, 2010). 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 (Sweet, 2014), sustainability (Zhao, 2014) and its population's health (Hansson, Mattisson, Bjoerk, Oestergren, & Jakobsson, 2011).

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 technologies (Slack, 2017) that are limited in terms of scalability and real–time feedback, and that 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 using computer vision machine learning techniques offers a new, and increasingly popular, approach for computer scientists and urbanists to count not only motor, but also pedestrian traffic on a large scale (Slack, 2017).

However, these “counting–gate” methods are limited to traffic counts at specific locations for a specific time period, and thus they do not offer data about trajectories of pedestrians between them. Current work in data mining aims to fill this gap by using mobile phone data to model urban mobility (Calabrese, Diao, Di Lorenzo, Ferreira Jr., & Ratti, 2013; Jiang et al., 2016), but shows limitations in terms of population representation by capturing only mobile users of a specific network provider, and typically with low spatial granularity. In

* Corresponding authors at: New York University, Center for Urban Science & Progress, 370 Jay St, 12th Floor, Brooklyn, NY 11201, United States.
E-mail addresses: martin.traunmueller@nyu.edu (M.W. Traunmueller), nicholas.johnson@nyu.edu (N. Johnson), awais.malik@nyu.edu (A. Malik), ckontokosta@nyu.edu (C.E. Kontokosta).

https://doi.org/10.1016/j.compenvurbsys.2018.07.006
Received 12 December 2017; Received in revised form 21 July 2018; Accepted 21 July 2018
0198-9715/

Please cite this article as: Traunmueller, M.W., Computers, Environment and Urban Systems (2018), https://doi.org/10.1016/j.compenvurbsys.2018.07.006

addition, computer vision techniques create significant concerns around confidentiality and privacy, as facial recognition methods become more widely applied.

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 cityscape, particularly at the neighborhood or district scale. Each AP continuously "senses" its surroundings in terms of potential users equipped with WiFi–enabled mobile devices, which send probe requests to available networks and proximate APs at regular frequencies. With the increasing market penetration of WiFi connectible mobile devices, such as tablets or smartphones (64% of all U.S. citizens and approximately 80% in New York City owned a smartphone in 2016 (Smartphone Users, n.d.)), computer and urban scientists have begun to use WiFi probe data with the aim to understand human behavior and mobility (Kontokosta & Johnson, 2017). However, while many of the large–scale studies focus on indoor activities (Abedi, Bhaskar, & Chung, 2014; Meneses & Moreira, 2012), less has been done discussing every–day movement patterns in open public spaces at the neighborhood scale using Wifi probe data. This has largely been the result of the lack of data available at necessary spatial and temporal granularity, and the computational challenges in processing these data. Mobility data, however, must be handled with appropriate data management and access protocols, as concerns about data privacy become paramount. In addition, using such data can also raise equity issues, as sampling bias caused by differential access and technology adoption rates can exclude certain demographic groups (Kontokosta, Hong, & Korsberg, 2017).

In this work, we hypothesize that WiFi probe data can be used to analyze outdoor mobility and human trajectories in a large and densely populated urban area at high spatial resolution and temporal frequency. We use a dataset of WiFi probe requests collected by 54 public-access WiFi APs over the duration of one week in Lower Manhattan in New York City, NY, collected through the" Quantified Community" urban test–bed (Kontokosta, 2016). First, we show how WiFi probe data can be used to report hyperlocal, real-time counts at each AP, similar to "counting gates" methods described above, and be used to understand localized population segmentation. 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 of street activity over time. We begin by presenting recent literature on measuring urban mobility, and then present our data and data processing steps. We introduce our methodology and describe our results for pedestrian counts and trajectories. We conclude with an in–depth discussion of the findings, including limitations, privacy concerns, and applications to city management and planning.

2. Literature review

2.1. Capturing urban mobility

The most commonly used method to capture urban mobility by city agencies is the installment of "counting–gates" at pre–defined locations, 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 (Slack, 2017) to count primarily motor traffic. While these methods offer an easy way to quantify traffic aggregations on a street for a specified time period, such as per hour or day, they are limited in terms of temporal and geographical scalability and rather expensive to run due to installation and service charges, compared to the output they provide.

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, like high–volume intersections, by applying computer vision algorithms, such as blob–detection (Trafficvision, n.d.). Focusing primarily on motor traffic, this approach has been extended over the years to also count pedestrians (Placemeter, n.d.). The analysis of CCTV footage offers effective ways to aggregate traffic quantitatively and is only limited by the number of CCTV–camera locations (with appropriate resolutions and fields of view) 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 provides little ancillary information about activity patterns, and hence do not generate critical information for city managers.

The increasing availability of open data has offered researchers novel opportunities to study traffic routes, in particular for public transport, on a large scale using a data mining approach. In doing so, metro journeys (Tfl Study, n.d.), the use of public bike sharing schemes (Woodcock, Tainio, Cheshire, & Goodman, 2014), or GPS traces of taxis (Ferreira, Poco, Vo, Freire, & Silva, 2013), 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 (Milne, 2014), using these data sources excludes a large portion of the urban population and are therefore not fully representative. The focus on individual transport modalities also limits valuable information about human behavior and activity at the micro- and meso-scales in various urban environments.

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 & Raubal, 2012) extracted dynamic mobility patterns in urban areas using a 'Dynamic Time Wrapping' algorithm, and were able to classify areas according to the observed patterns. (Calabrese et al., 2013) 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 two 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 flows. (Gonzalez, Hidalgo, & Albert-Laszlo, 2008), for example, studied 100.000 mobile phone user trajectories over six 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 studies demonstrate the opportunities for using CDR data to study human mobility at the urban scale. However, at the same time, telecommunication data can be sensitive 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 (Competition Example, n.d.) where providers make parts of their data publicly available. However, as available data are pre–processed, their accuracy often suffer due to unknown data processing steps (Traunmueller, Quattrone, & Capra, 2014). Furthermore, when data are provided by individual mobile phone providers, they offer limited representativeness of the urban population by excluding various groups of people that use other providers, or people that do not have a cell phone contract, such as the elderly or lower–income populations, or those that use Pay–As–You–Go options.