AP Positioning for Estimating People Flow as Origin Destination Matrix for Smart Cities

Pierfrancesco Bellini, Daniele Cenni, Paolo Nesi
Distributed Systems and Internet Technology Lab, http://www.disit.dinfo.unifi.it ,
Department of Information Engineering, http://www.dinfo.unifi.it
University of Florence, http://www.unifi.it, Florence,Italy phone +39-3355668674
Email: {pierfrancesco.bellini, daniele.cenni, paolo.nesi}@unifi.it

Abstract— Controlling and regulating people flows and the access to the city services are major topics in the context of Smart City management. Flow surveillance provides valuable information about city conditions, useful for not only monitoring and controlling the environmental conditions, but to optimize the exploitation of various city services. In this context, it is mandatory to develop tools for assessing people flow. This paper presents a methodology for an effective placement of counter sensors, to model flows with a statistically significant precision rate. Comparative analyses are conducted with respect to real data (i.e., cab traces) of the city of San Francisco. Several different placing methodologies of Wi-Fi access points have been tested and compared, to minimize the cost of AP installation. The research work described in this paper has been conducted in the scope of the EC Horizon 2020 funded project Resolute (http://www.resolute-eu.org) and for Sii-Mobility.
Keywords: people flows; smart city; WiFi Access Point location

I. Introduction

The optimization of services for the citizens is one of the most challenging activities of the Smart Cities. City services can be related to mobility, government, energy, cultural events, commercial, environment, etc. Among the services, mobility is a commodity; thus, transportation and mobility analyses are valuable aspects always considered for an effective definition of Smart City. According to [Giffinger et al. , 2007], Smart Mobility is among the key factors of a modern Smart City, including local and international accessibility, availability of ICT infrastructures, sustainable, innovative and safe transport systems.

[Caragliu et al., 2009] include traditional transport communication infrastructures among the essential requirements for Smart Cities.

In the context of mobility, traffic/flow analysis is a major prerequisite for planning traffic routing. Thus, it is a central part of the so called Intelligent Transportation Systems (ITS) for managing public transportation. Traffic flow analysis is commonly used to ease the transportation management, for regulating the access control to the cities, for Smart Parking, for traffic surveillance providing information about road conditions and travel, or for monitoring and controlling the environmental conditions, such as harmful emissions (e.g., CO_2, PM10, ozone). The European Commission indicates, among the main topics that should be considered with special attention in the framework of the CARS 2020 process, the implementation and promotions of ITS, including Smart Mobility [CARS 2020]. Some of the techniques adopted for traffic monitoring and management can be declined for people flow analysis and support in the city. It is very important to know the movement of people within a certain precision, and detecting where and how people are crossing the city and using its services by using different kind of moving solutions: car, bike, walking, taxi, car sharing, buses, tram, etc., targeting services into the city [Bellini at al., 2014]. Typically, the telecom operators are not capable to provide this kind of information. They know the number of people connected to each cellular cell at a given time slot during the day, and not how people move in city.

At this regard, specific tracking services for mobile phone IDs are needed and, when applied, the citizens have to be informed via an informed consent (e.g.,

1

terms of use, policy privacy). In order to measure/derive the typical people flow, the estimation of the so called OD matrix (Origin Destination Matrix) could be needed. The OD matrix presents on both axes the city zones, while the single element may contain the number of people (or the probability) of passing from the zone of origin and reaching the zone of destination, in the day or in a given time window. The OD matrix estimation via observations is very relevant for traffic flow prediction and management, in particular for (i) planning optimized routes predicting shortest and viable paths, (ii) providing info-traffic services on desktop or mobile devices, via the so called Advanced Traffic Management Systems (ATMS).

Most of the OD matrixes can be time dependent, and thus their dynamic real-time estimation may be needed, or at least the estimation of their values along the day and week, (e.g., typical sampling period is every 15 minutes). Moreover, real-time OD matrixes can be continuously estimated during the day or in a temporal window. On the other hand, OD matrixes can be sensitive to traffic conditions. Their values are of primary interest if they represent the maximum or at least sustainable values, disregarding values when the traffic system cannot sustain the traffic. Thus, pre-calculated OD matrixes can be used as default descriptors of the traffic conditions for plan estimation.

These solutions are called Advanced Traveller Information Systems (ATIS). In practice, we would like to measure the typical people flows from the several zones of the city. In the context of traffic flow, some methods make use of parametric estimation techniques (e.g., Maximum Likelihood, Generalized Least Squares, Bayesian inference). ML methods minimize the likelihood of observing the OD matrix and the traffic counts. Other methods based on traffic count include Combined Distribution and Assignment (CDA) [Cascetta et al., 2001], Bi-level Programming [Doblas et al., 2005], [Kim et al., 2001], Heuristic Bi-level Programming [Lundgren et al., 2008], Path Flow Estimation (PFE) [Nie et al., 2005], or Neural Networks [Gong, 1998]. For example, [Ashok and Ben-Akiva, 2000] used a Kalman filtering technique to update the OD matrix. Time dependent offline estimation deals with time-series of traffic counts. Typically, building an OD matrix for mobility requires installing devices to count the single vehicle in the traffic (and

eventually record the speed of each vehicle) on the road. A traffic counter is a device that records vehicular data (i.e., speed, type or weight). At this regard, the US Federal Highway Administration defines three main traffic counting methods: human observation (manual), portable traffic recording devices and permanent automatic traffic recorders (ATR). Thus, at level of traffic flow observation several different techniques are used: video cameras, pneumatic road tubes, piezo-electric sensors embedded in the roadway as inductive loop detectors, magnetic sensors and detectors, microwave radar sensors, Doppler, passive infrared sensors, passive acoustic array sensors, ultrasonic sensors, laser radar sensors. Most of these sensors use intrusive technologies and require pavement cut; in some cases lane closure is required, the devices are sensitive to environmental conditions and require an expensive periodic maintenance.

Some of the above mentioned techniques can be used to produce vehicle classification (e.g., rural cars, business day trucks, through trucks, urban cars). Recently, other techniques have been adopted as RFID, Bluetooth, Real Time Location System (RTLS) and Wi-Fi access points [Danalet et al., 2012], [Patil et al., 2015]. In some cases, the position of vehicle can be monitored from the GPS position of mobile devices installed on the vehicle, or simply by using smartphone navigators (e.g., Google Maps, TomTom, Waze), thus providing crowd sources positions and velocity of the vehicles. In these two cases, the track of position is agreed with the users that install the device or run the mobile application on the smartphone or navigator. RFID is quite unsuitable to detect devices because of the small range of action. Bluetooth can be more suitable but expensive, since specific station to collect the passage is needed. Wi-Fi access points are less reliable in detecting the presence of motorized sources with respect to physical devices, and GPS methods.

We decided to use Wi-Fi Access Points as counter devices: given the high distribution of mobile devices and the low cost of a Wi-Fi AP, and the fact that a large number of APs is already installed in the city. This solution is quite cheap and easy to implement, also considering that many municipalities offer free Wi-Fi connectivity, and the needed coverage can be easily obtained. Therefore, the identification of the best placement of Wi-Fi Access Points, as detectors

2

for measuring the OD matrix is very important. The positioning of APs aims at limiting the costs by obtaining acceptable measures for the OD matrix. That means to obtain an OD matrix with good approximations with respect to a massive and systematic measuring of the whole city flow by humans and/or specific sensors. This paper presents a study and methodology for the positioning of Wi-Fi APs according to the zone adopted for the OD matrix needed estimation. The proposed study and solution has been validated by exploiting the data set introduced in [Piorkowski et al., 2009] which covers cab mobility traces, collected in May 2008 in San Francisco.

The proposed approach assesses the present distribution of Wi-Fi Aps on San Francisco to collect flows data with mobile device discovering. On the other hand, the proposed method could be used to suggest better positioning and/or completing the positioning of the city Wi-Fi APs distribution to be used also as sensors, for the OD matrix estimation. The research work described in this paper has been conducted for RESOLUTE EC Horizon 2020 project (http://www.resolute-eu.org).

This paper is structured as follows. In Section II, the definition of the OD matrix is reported together with the analysis of data collected in San Francisco. Section III presents the proposed approaches for positioning the APs and their comparison. In Section IV, results are analysed. Conclusions are drawn in Section V.

II. Origin Destination Matrix and Reference Data

According to the introduction, the main goal of the presented research has been to identify and validate a method and tool for the positioning of Wi-Fi APs, with the aim of using them as sensors/observers for estimating the OD matrix. In this context, the OD matrix representing flows among the zones of the city, considered as zip codes z is defined as

$$OD_{n,n}= \begin{pmatrix} z_{1,1} & \cdots & z_{1,n} \\ \vdots & \ddots & \vdots \\ z_{n,1} & \cdots & z_{n,n} \end{pmatrix} \qquad (1)$$

where $z_{i,j}$ represents the total number of traffic counts from z_i to z_j (i.e., how many cabs moved from z_i to z_j) defined as

$$z_{i,j} = \sum_{t \in T} n_t(i,j) \qquad (2)$$

and T is the set of unique cab traces, $n_t(i, j)$ is the number of traffic counts from z_i to z_j for trace t.

The validation of any AP positioning in the city for the people count (or traffic flow) is not a trivial task. In principle, one should be capable to install the APs in the city in certain positions and demonstrate, by measuring on the real context, that their position and data collected is strongly correlated with the effective number of people and thus of flows among the different areas of the city. This approach is very expensive and unfeasible for a number of configurations. Therefore, we adopted an indirect method described as follows.

The data set introduced in [Piorkowski et al., 2009] includes the trace flows of cabs in the area of San Francisco, collected in May 2008. The dataset reports all the cab traces in that area, providing fine GPS positioning of them. In **Figure 1**, the trace flows are reported on the city map (for a particular day in the time range 8:00 am - 9:00 am). The data set consists of 446,079 traces based on about 11.2 millions of single GPS points collected by cab movement, not only in the downtown of San Francisco, but spanning on the whole city neighbourhoods. On the other hand, the areas at higher density are those in the downtown, coherently with what one could have from the movements of pedestrians in the city. This area can be identified with about 13 central zip code areas.
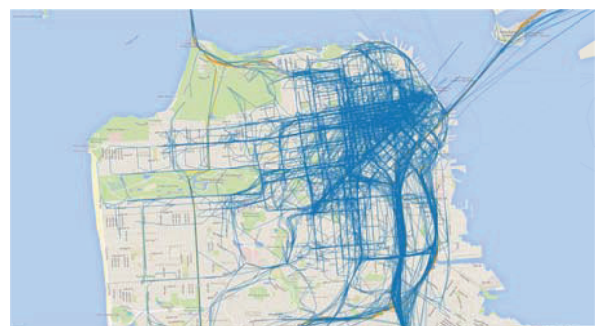

Figure 1. Trace flows in San Francisco on a working day of May, 8:00 AM - 9:00 AM

In order to perform an effective data analysis and visualization, some web tools for viewing and comparing flows in different scenarios were developed. At this regard, OD matrix and thus flows

3

among zip areas are represented with a chord diagram to put in evidence single and aggregate contributions to the total flow count among the various city zones (in **Figure 2,** the chord diagram is reported for the central part of the city with 13 zip code areas). An interactive version of this tool is accessible at http://www.disit.org/6694.

On the user interface, the user can select a time interval on the day to visualize its related chord diagram, which is constituted by circular sectors, each of them representing a city area; passing the mouse over a sector provides additional information about the traffic counts originated from it towards other zip areas. In this manner, it is possible to depict in a compact and intuitive way the traffic flows among the various zones. Moreover, for San Francisco we know the structure of the city and the position of the APs in the downtown (see **Figure 3).** The positions have been taken from OpenWiSpots, http://www.openwispots.com , with GPS positioning. They consist of 494 Wi-Fi Aps offered by city services, from a total of 983 APs, (e.g., coffee shops, hotels, restaurants, libraries, bars, bookstores, grocery stores). Therefore, we supposed to use the Wi-Fi network of San Francisco to estimate the passage in the city by the mobile phones according to their MAC address.
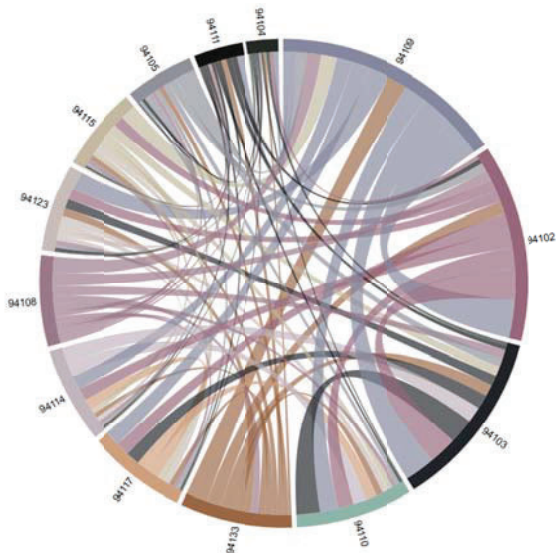

Figure 2. San Francisco OD matrix as chord among the 13 central ZIP areas of the city (real cab flows)
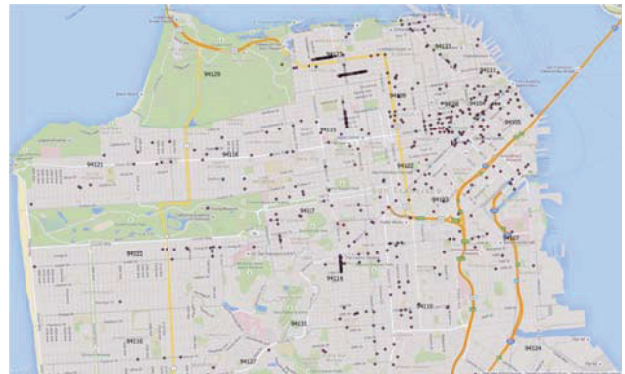

Figure 3. Distribution of real Wi-Fi APs in San Francisco

This solution can be implemented by collecting the first discovery by mobile of the Wi-Fi AP that can stream them as log data to a central server that also anonymizes the MAC. In alternative, some of the APs or aggregators of APs may compute the anonymization algorithm, based on a hash code of the identifiers. Once detected the passages of flow on the APs, the OD matrix can be computed. As a first approximation, we assumed to have the possibility of detecting the flows by using the present APs, by capturing the real traces passing within a distance of 25 meters from the AP position. The proposed approach is a sort of partial simulation based on real data about traffic flow, that is more realistic than producing fully simulated data. It is obvious that the real data captured by the APs would be probably only a part of the real traffic of people passing close to them. On the other hand, it is reasonable to have the simulated measures performed as strongly correlated to the real effective numbers.

As a general consideration, only 1,470,091 points were found to intersect with the real APs positions, which downtown are 1,418,207 with respect to 494 APs. Therefore, in this manner, we assessed the available distribution of Wi-Fi APs in San Francisco, in order to collect people flows data with their mobile device. Once obtained the observations by intersection of traces with respect to the APs areas, an estimated OD matrix on the basis of the APs has been obtained as reported by the chord diagram in Figure 4a. In Figure 4b, the matrix of difference from the OD matrix of Figure 2 and that of Figure 4a is reported, the differences from back abnd forward are not perceivable.
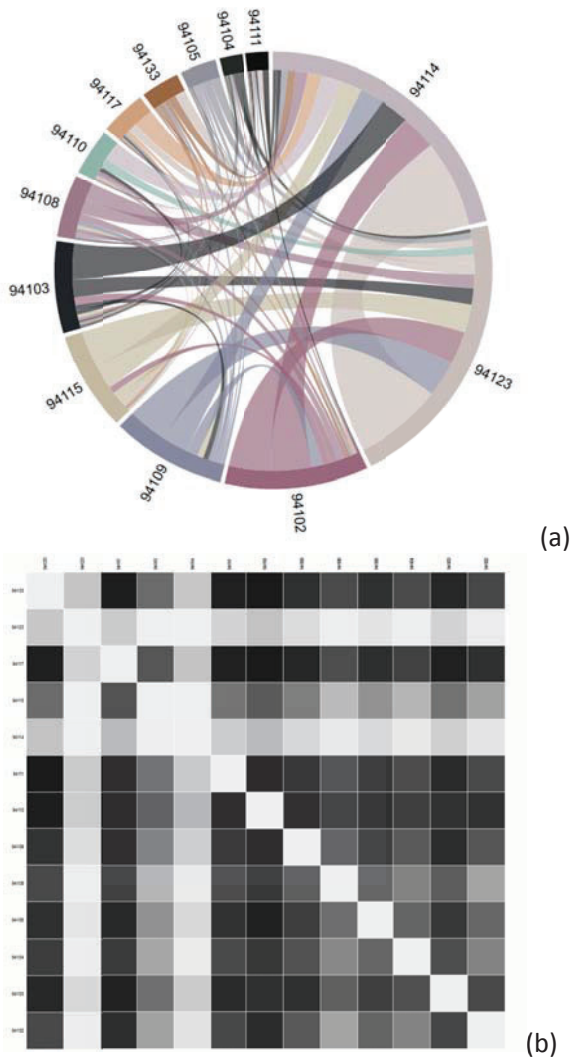
4

(a)

(b)

Figure 4. (a) Chord diagram of flow counts with real Wi-Fi APs in the city centre; (b) Difference matrix among OD matrices of real flows and estimated with real Wi-Fi APs in the city centre

The difference matrix compares the real trafc flow data with respect to the flow that would be estimated by using the present APs distribution in the city. The difference values are reported with a grayscale (the higher the difference, the higher the darkness of the matrix element). The two distributions of Origin-Destination are uncorrelated (a correlation of 0.12 has been measured, see Table I). This result demonstrates the unsuitability of the present distribution of the APs for collecting flows. On the other hand, their positioning was not performed with the aim of measuring and observing the flows.

III. Adopting AP Positioning Models

A different positioning of the APs in the zip areas could create better correlation and thus better precision for the estimation of flows in the city. To this purpose, a number of different methods for AP positioning and thus for flow observations have been adopted and tested, taking them from the literature of the classical traffic flow observations strategies by humans. We started by creating a uniformly distribution grid of APs, placed at the middle of each street. In all cases, each AP was considered as a circular buffer with 50 m of diameter. The resulting AP set, consisting of 14,959 APs (a number of devices that is surely too high to be affordable), was further reduced using different strategies as reported in the following. Moreover, the reduction is also reasonable since a uniform distribution in all the zones of the city is not feasible. There are many zones in which the flows are very low, at least in the simulation data taken into account. On the other hand, the positioning of the APs in low flow areas is not efficient.

Also, a flow prediction strategy should be able to tell where to place traffic sensors, and how many sensors to use, providing a tuning strategy for selecting the required set of sensors, with the aim to minimize the number of traffic sensors and the costs of periodic maintenance of the monitoring infrastructure. In this section, we provide some alternative strategies of AP placement. The possible cases for distribution are the following.

a) *Random APs*: identification of the streets with the highest trace flow rate (those that have at least 30,000 traces) and then randomly selecting 400 APs from the AP grid described above (see Figure 6a for the OD matrix). This set of APs is a subset of the set selected in case (b).

b) *High Traffic APs*: identification of the streets with the highest trace flow rate (those that have at least 3000 traces) and then selecting all the APs intersecting those traces, thus resulting in 804 APs (see Figure 6b for the OD matrix).

c) *High Traffic APs (zip boundaries)*: identification of the streets with the highest trace flow rate (those that have at least 3000 traces) and then starting from the 804 APs of case (b) and considering only those within 300 m of the zip boundaries, thus resulting in 448 APs (see Figure 6c for the OD matrix). This set of APs is a subset of the set

5

Model		Coefficient	Std. Error	t-statistic	p-value	Correlation	# APs
Real APs	β	280393.858	19874.972	14.108	0.000	0.446	983
	α	9.448	0.543	17.400	0.000		
Real APs (cc)	β	1598664.580	116546.825	13.717	0.000	0.12	494
	α	1.714	1.141	1.502	0.135		
(a) Random APs (cc)	β	690144.338	75267.849	9.169	0.000	0.835	400
	α	52.921	2.813	18.816	0.000		
(b) High Traffic APs (cc)	β	684144.945	52950.289	12.921	0.000	0.915	804
	α	10.942	0.389	28.114	0.000		
(c) High Traffic APs (bn, cc)	β	1101641.803	86354.599	12.757	0.000	0.687	448
	α	13.586	1.159	11.727	0.000		
(d) High Traffic APs 400 (cc)	β	810743.094	70801.471	11.451	0.000	0.835	400
	α	24.429	1.297	18.829	0.000		
(e) High Traffic APs (bn, cc)	β	748987.390	58260.615	12.856	0.000	0.892	400
	α	39.960	1.634	24.453	0.000		

Table I -- COEFFICIENTS — REAL APS, CC = CITY CENTER, BN = ZIP BOUNDARIES (WITHIN 300 METERS)

selected in case (b).

d) *High Traffic APs (top 400)*: identification of the streets with the highest trace flow rate (those that have at least 3000 traces) and then starting from the 804 APs of case (b) and considering only the top 400 APs (see Figure 6c for the OD matrix). This set of APs is a subset of the set selected in Case b.

e) *Real augmented with selected High traffic APs*: the real distribution of the AP in San Francisco's downtown was integrated with the top 300 AP from case (d) with the highest traffic rate. This set was then cleaned up by removing those APs that were found to be at a distance less or equal than 50 m from the real APs, and removing also intersecting APs, thus resulting in 400 APs (221 real APs, 179 selected high traffic APs).

The resulting OD matrix for these distributions of APs has been estimated by computing the intersections between the real cab measures with the placed APs, according to a capturing range of 25 m radius. The OD matrix for this configuration was generated by evaluating the traffic counts among the various APs, grouped by the zip code they belong to. The chord diagrams of these scenarios are reported in Figure 6.

The interactive versions of the chords diagrams in which it is possible for each couple of locations to see the effective flows (in a way and in the other, for a given time slot of the day) are accessible on http://www.disit.org/6694.

IV. Experimental Results Analysis

A comparative analysis of traffic flows was conducted, using the above cited set of cab traces, consisting of 11,219,955 unique detections from 536 cabs, with respect to the above described scenarios. With the above assumptions the real set of APs placed in the city centre was used to sample the original data set, by calculating the APs intersections with the cab traces. The OD matrix was calculated from the sampled data set (considering each city zip code as a separate area), reporting the traffic counts among every city area.

This procedure was repeated by choosing the APs with a pseudo random technique, and by placing the APs only in the roads with the biggest amount of traffic. After that, a comparative statistical analysis was conducted for each configuration (see Table I). The traffic flow outcome is predicted with a linear regression, finding the parameters that best fit the data in the linear model

$$y = \alpha x + \beta \qquad (3)$$

where x is the dependent variable or predictor (i.e., traffic counts as registered by the sensors), and y is the outcome (i.e., predicted traffic counts). Building
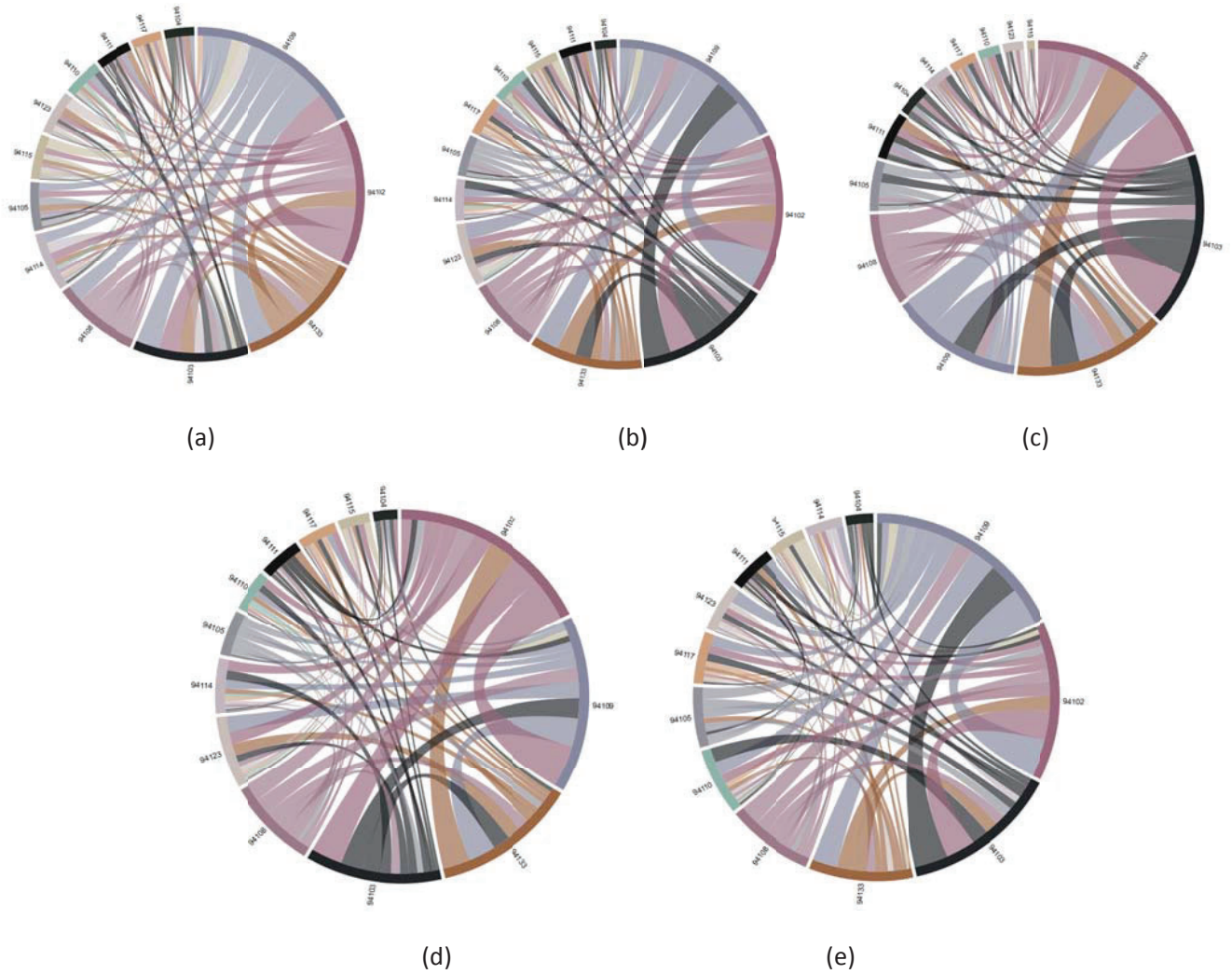
6

Figure 6. Chord diagram of flow counts. Cases as described in Table I: (a) Random APs; (b) High traffic APs; (c) High traffic APs (zip boundaries); (d) High traffic APs (top 400); (e) Real augmented APs

the model in (3) using the set of real APs gives a correlation of 0.446 (0.120 with real APs in the city downtown) with respect to the real traces. A number of cases have been assessed following the placement strategies described in Section III. In case (b), the APs have been placed on the roads with the highest traffic rate, producing a model with a correlation of 0.915, and of 0.835 using only the top 400 APs, as described in case (d); using random APs of case (a) gives a correlation of 0.835; using the APs only within 300 m from the areas' boundaries, described in case (c), gives a correlation of 0.687. It is clear from this data that using the real APs set produces noise and doesn't give a reliable model for the traffic counts. Randomly Distributing the APs gives a better correlation with the cab traces, while reducing the number of APs and considering only those in the proximity of each area, gives a good correlation while maintaining a limited number of APs. The set of real APs of case (e), integrated with some other APs and cleaned up from some unuseful or redundant elements (i.e., mutually intersecting APs), gives a correlation of 0.892). To visualize the results of the various OD models a web interface was developed, with the possibility to view the chord diagram for each computed configuration, for different time intervals of the day. This approach allowed to identify which are (i) the positions of the new APs to be added (i.e., 179) and (ii) the minimum set of APs already in place that have to be used for data acquisition (i.e., 229). The second point allows keeping limited the network bandwidth and the workload for the estimation of the OD matrix.

7

V. Conclusions

In this paper, we presented a flow analysis using OD models of traffic counts. Considering public Wi-Fi access points is not reliable in determining an effective model for the city traffic flows. Instead, placement of sensors only in proximity to the boundaries of high traffic routes yields a statistically significant linear relationship, between the OD matrix trip of sampled and real data. Thus, it is possible to model the traffic behaviour with a limited amount of traffic counters (in this context Wi-Fi access points) with an acceptable precision. The proposed methodology is general and can be applied to different urban scenarios, in the context of Smart City traffic management. It makes use of Wi-Fi access points (AP) distributed across the city of San Francisco. Comparative analysis has shown an increased precision of the AP positioning validated by obtaining an OD matrix and assessing correlation with the actual OD taken from the original real data by considering cab flows from high traffic zones, positioned in proximity of the zones boundaries, with respect to random uniformly distributed or real APs (e.g., restaurants, hotels, schools).

The approach allows identifying which are the strictly needed AP to be added with respect to the AP that can be already in place in the city, to exploit the whole infrastructure of Wi-Fi also for people flow monitoring and assessment.

REFERENCES

[Danalet et al., 2012] Danalet, Antonin, Michel Bierlaire, and Bilal Farooq. "Estimating Pedestrian Destinations using Traces from WiFi Infrastructures." Pedestrian and Evacuation Dynamics 2012. Springer International Publishing, 2014. 1341-1352.

[Patil et al., 2015] Patil, P. H., and A. A. Kokil. "WiFiPi-Tracking at mass events." Pervasive Computing (ICPC), 2015 International Conference on. IEEE, 2015.

[Gigginger et al., 2007] Giffinger, Rudolf; Christian Fertner; Hans Kramar; Robert Kalasek; Nataa Pichler-Milanovic; Evert Meijers, Smart cities Ranking of European medium-sized cities, 2007 Available at: http://www.smartcities.eu/

[Caragliu et al., 2009] Caragliu A., Del Bo C., Nijkamp P., Smart cities in Europe, 3rd Central European Conference in Regional Science CERS, 2009

[CARS 2020] Report on the state of play of the outcome of the work of the High Level Group, European Commission, October 2014

[Cascetta et al., 2001] Cascetta, E. and Postorino, M. N. (2001) Fixed point approaches to the estimation of O/D matrices using traffic counts on congested networks, Transportation Science 35: 134-147

[Bellini at al., 2014] Pierfrancesco Bellini, Monica Benigni, Riccardo Billero, Paolo Nesi, Nadia Rauch, Km4City ontology building vs data harvesting and cleaning for smart-city services, Journal of Visual Languages & Computing, Volume 25, Issue 6, December 2014, Pages 827-839

[Doblas et al., 2005] Doblas, J. and Benitez, F. G. (2005) An approach to estimating and updating origin-destination matrices based upon traffic counts preserving the prior structure of a survey matrix, Transportation Research, Part B: Methodological 39: 565-591

[Kim et al., 2001] Kim, H., Beak, S. and Lim, Y. (2001) Origin-destination matrices estimated with a genetic algorithm from link traffic counts, Transportation Research Record 1771, Transportation Research Board, Washington, D.C.:156-163

[Lundgren et al., 2008] Lundgren, J. T. and Peterson, A. (2008) A heuristic for the bilevel origin-destination matrix estimation problem, Transportation Research, Part B: Methodological 42: 339-354

[Nie et al., 2005] Nie, Y., Zhang, H. M. and Recker, W. W. (2005) Inferring origin-destination trip matrices with a decoupled GLS path flow estimator, Transportation Research, Part B: Methodological 39: 497-518

[Gong, 1998] Gong, Z. (1998) Estimating the urban o-d matrix: a neural network approach, European Journal of Operational Research 106: 108-115

[Ashok and Ben-Akiva, 2000] Ashok, K. and Ben-Akiva, M. E. (2000) Alternative approaches for real-time estimation and prediction of timede-pendent origin-destination flows, Transportation Science 34: 21-36

[Piorkowski et al., 2009] Piorkowski, M.; Sarafijanovic-Djukic, N.; Grossglauser, Matthias, "A parsimonious model of mobile partitioned networks with clustering", Communication Systems and Networks and Workshops, 2009. COMSNETS 2009. First International, pp.1,10, 5-10 Jan. 2009

8