IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 29, 2018, accepted May 7, 2018, date of publication June 7, 2018, date of current version July 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2845339

Using Social Media for Attendees Density Estimation in City-Scale Events

V. X. GONG[1], J. YANG[2], W. DAAMEN[1], A. BOZZON[1], S. HOOGENDOORN[1], AND G. J. HOUBEN[1]

[1]Delft University of Technology, 2628 CD Delft, The Netherlands
[2]eXascale Infolab, University of Fribourg, 1700 Fribourg, Switzerland

Corresponding author: V. X. Gong (x.gong-1@tudelft.nl)

This work was supported by the European Research Council and Amsterdam Institute for Advanced Metropolitan Solutions through the ALLEGRO project under Grant 669792.

ABSTRACT City-scale events attract large amounts of attendees in temporarily re-purposed urban environments. In this setting, the real-time measurement of the *density* of attendees stationing in—or moving through—the event terrain is central to applications, such as crowd management, emergency support, and quality of service evaluation. Sensing or communication infrastructures (e.g., sensor networks and mobile phones) can be deployed to estimate the number of attendees currently occupying an area. However, the adoption of these technologies is hindered by their cost or sensing resolution. There is evidence that social media data can provide a real-time and semantically rich insight into attendees' behavior during city-scale events. Their suitability as a data source for attendees density estimation is yet to be investigated. With this paper, we aim at filling this knowledge gap by studying how micro-posts harvested from social media can be used during city-scale events to estimate the density of attendees in a given terrain. To cope with issues of temporal and spatial resolution, we propose three classes of density estimation strategies (i.e. geo-based, speed-based, and flow-based) inspired by elements of pedestrian traffic flow theory that were successfully assessed during city-scale events. We study the performance of these strategies in the context of SAIL Amsterdam 2015 (Sail) and Kingsday Amsterdam 2016 (Kingsday), two city-scale events that attracted 2 and 1.5 million of attendees in the span of five days and one day, respectively. We defined four experimental terrains for the Sail event and one for the Kingsday event, and compare density estimates from social media data with measures obtained from counting systems and Wi-Fi sensors. Results show the potential of solutions embedding elements from pedestrian traffic flow theory, which yielded estimates with strong temporal correlations with the sensor observation, and limited mean errors.

INDEX TERMS Data science, social sensing, urban analytics, computational social science, traffic flow state, density estimation.

I. INTRODUCTION

As cities battle for global importance and influence, city-scale public events are becoming an important weapon of choice to foster tourism and economic growth. Olympic games, thematic exhibitions, and national celebrations are examples of city-scale events that take place in vast urban areas, and attract large amounts of attendees within short time spans. The scale and intensity of these happenings demand for technological solutions able to support relevant stakeholders (e.g. event organizers, public and safety authorities, attendees) with the monitoring of an event's state with respect to the crowd.

For instance, it is common for public authorities to monitor the amount of attendees present in a given event terrain, to promptly identify capacity issues and minimize the risk of incidents due to overcrowding – stampedes are more likely to occur in high-density crowds [1]. The estimation of attendees *density* requires a measurement infrastructure that is characterized by stringent requirements in terms of spatial resolution, temporal resolution, and accuracy. These measurement activities are typically performed by personnel operating on the event terrain [2]; the data they provide is however temporally scarce, spatially non-uniform, and often subjective.

Ad-hoc sensing infrastructures – such as counting system and Wi-Fi sensors – or pre-existing communication infrastructures – such as mobile phone networks – are an automatic solution for the real-time measurement of the amount of individuals and/or connected devices present in a given area [3]. Their widespread adoption is however constrained by economical and operational limitations. Counting system infrastructures are expensive to set-up and operate; their monitoring capability is limited to a fixed and relatively small area; as counting is performed by means of computer vision algorithms trained to recognize human faces, heads or shoulders, their accuracy decreases in non-standard operational conditions – for instance, when it becomes too crowded, or when adverse meteorological conditions force people to use umbrellas. The accuracy of Wi-Fi sensors is clearly dependent on issues such as technological penetration, technology of devices; and data from mobile communication may be only available at coarse-grained resolution due to privacy or technological limitations.

Social media data produced by platforms like Twitter or Instagram are increasingly used to study urban-related problems [4]–[6], and to monitor the on-line liveness of city-scale events [7], [8]. Their popularity is certainly due to their availability, ease of access, real-timeliness, and geographical annotation. On the other hand, social media data suffer from known limitations in terms of representativeness of the targeted population, and (spatial and temporal) sparsity. Intuitively, not all attendees feel compelled to share their experience on social media, or are active on such platforms; also, event areas are differently attractive; and the event is not equally engaging over time.

As a result, there is a lack of scientific knowledge about the suitability of social media as a data source for density estimation. In this paper, we aim at filling this knowledge gap by studying how micro-posts harvested from social media can be used during city-scale events to estimate the density of attendees stationing in – or moving through – a given terrain. We formalize the problem in a probabilistic framework, and calculate the likelihood of event attendees to be present in the targeted event terrain within a given time span. Inspired by methods of pedestrian traffic flow theory successfully tested in crowd monitoring applications [9], we propose 3 density estimation strategies: *geo-based*, *speed-based*, and *flow-based* strategy.

The assessment of the performance of these strategies in real-world settings is a challenge per-se, and it is often neglected in existing studies. This work contributes the results of an analysis performed on two large-scale sensing infrastructures, that we set-up in the city of Amsterdam during SAIL 2015 (Sail) - the largest free nautical event in the world, and King's Day 2016 (Kingsday) - the national King's birthday event, held once a year, and attracting millions of people. During the Sail event, we focused on 4 terrains located along a walking route close to where most tall ships were moored; during King's day, we focused on 1 terrain in the south of Amsterdam, covering a busy square between Station Amsterdam Zuid and World Trade Center (WTC) with various shops and restaurants around. These 5 event terrains are characterized by different morphology and relevance to the activities of both events.

We then compared the density values estimated from social media data with the measures obtained from the sensing infrastructure. Results show that the proposed density estimation strategies are able to cope with data sparsity issues typical of geo-referenced social media. Errors in density estimation are in the range of 1-2 order of magnitudes, but with strong temporal correlations with measures obtained from the sensing infrastructure. Finally, we show that density estimation is influenced by the characteristics (e.g. morphological and functional) and the traffic status of the monitored terrain. We stress the importance of a systematic comparison with real-world data, and the challenging nature of our experimental setting: in our work we are able to provide novel insights into the suitability of social media as a data source for density estimation, and to ground them against measurements from state-of-the art pedestrian traffic flow measurement infrastructures.

The remainder of this work is organized as follows: in Section II, related works are discussed. In Section III, we propose our method to tackle this problem, followed by experimental setup for two cases in Section IV. The results of experiments are presents in Section V and discussed in Section VI. The conclusions including future research of this article is in Section VII.

II. RELATED WORK

A growing number of studies investigates pedestrian behavior models aiming at developing systems to automatically identify overcrowding during city-scale events. Wirz *et al.* [2] propose a pedestrian-behavior model to infer crowd conditions in city-scale events based on GPS location traces. Blanke *et al.* [10] study crowd mobility dynamics in city-scale events using GPS data. Weppner and Lukowicz [11] study the problem of density estimation by Bluetooth scans with mobile phones. However, fewer works attempt to make use of social media data to provide insights into attendees' behavior during city-scale events, while numerous recent works [4], [6], [12], [13] provide evidence that social media data can give semantically rich insights into the spatio-temporal dynamics of urban areas.

Botta *et al.* [14] show evidence of a relationship between the number of attendees at a given location at a given time with their social activities. They performed a correlation analysis of the number of attendees in two cases, a football stadium and an airport, with regard to their social media usage on Twitter, mobile calls and SMS activities on 11 event days in a city. It showed that data generated through interaction between people can be used to extrapolate the number of people in a given location at a given time, which may be valuable for business and policy makers. However, the purpose of their work is slightly different from ours. In our work we also use social media as data source to estimate the number of attendees at a location during a given time period. In order to provide valuable information for crowd management, we target on a more fine-grained analysis, i.e. in an hourly basis and within several specific terrains. This also leads us to deal with social media sparsity during a short time and within a small space. Besides, we also looking into

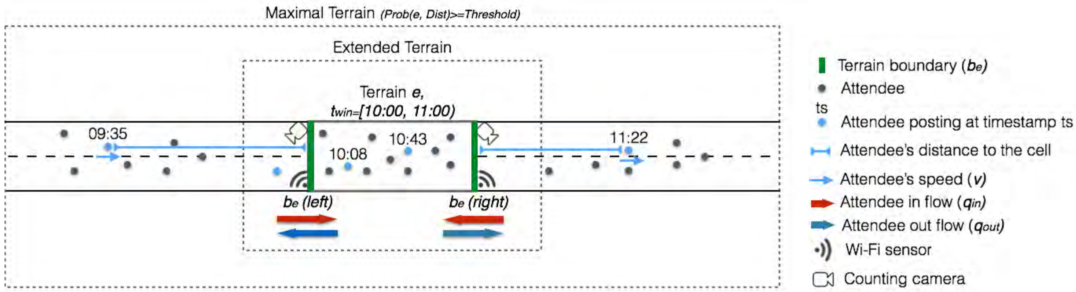


FIGURE 1. Illustration of geo-, speed- and flow-based density estimation methods. To estimate attendees density in the terrain during the time window t_{win} = [10:00, 11:00), geo-based density estimation method considers the number of users posting at least once within the terrain ($\widehat{k_1}$) or within the extended terrain ($\widehat{k_2}$) during t_{win}; the speed-based density estimation method ($\widehat{k_3}$) considers attendees travel speed to account for attendees that could potentially be present in the terrain during t_{win}, but that post on social media in a location within walking distance to the terrain; the flow-based density estimation method ($\widehat{k_4}$) further considers attendees flow information produced by the sensing infrastructure.

insights from social media data to interpret the estimation result.

Liang *et al.* [15] establish a model to calculate the volume of event attendees through social media, considering the number of check-in users and the duration of their stay in an event. Their model uses check-in and check-out number of social media users to estimate population. The check-in number of people is calculated by the number of posts sent from a location. While the check-out number of people is calculated through the amount of check-in people with the length of duration each people stay in the event. The duration time is estimated using timestamps between multiple posts sent by one user. The advantage of this model is that it transfers a population modeling problem into a temporal duration estimation problem making use of timestamp information of multiple posts sent by one user. Similar to our method, to tackle the social media sparsity authors make use of the duration information to estimate an emission rate, i.e. a probability of a person sending a post during an event in a crowd. However, using the duration information as signal for estimation population of a crowd will introduce bias as fewer people sent multiple posts in one day, which reduces the precision of the estimation. To avoid this risk, in our method, instead of using the duration information, we construct the probability by loosing the temporal and spatial limitation to count people nearby.

Georgiev *et al.* [16] further investigate factors which influence people participating in an event using social media data. It shows evidence that friends' co-attendance and the popularity of the event are dominating factors. In our work, we further interpret results using profile information derived from social media data, such as age, gender, city-role, and PoI preference of users.

III. ESTIMATING ATTENDEES DENSITY FROM SOCIAL MEDIA DATA

This section introduces the problem of attendees density estimation, and presents our proposed solutions. First, we introduce concepts from pedestrian traffic flow theory useful in the context of density estimation. Then, we describe three classes of density estimation strategies, namely: 1) *geo-based* strategies, operating only on social media data; 2) *speed-based* strategies, which estimate density by considering the travel speed (i.e. distance covered per unit of time) of attendees on the event terrain; and 3) *flow-based* strategies, that consider travel flow information (i.e. number of attendees passing a reference point per unit of time).

A. PEDESTRIAN TRAFFIC STATE VARIABLES

In pedestrian traffic flow theory [17]–[19], one of the fundamental characteristics of a moving population, from a macroscopic point of view, is the average flow $q = vk$. Given the average walking speed v (m/s) and the average density k (P/m^2), the flow q (P/ms) is defined as their product.

Density is a property related to a terrain where the event takes place, i.e. a shaped space formed with boundaries defined by a set of coordinates. To simplify the discussion, we assume event terrains to have rectangular shapes as in Fig. 1. Consider an event terrain e having area A_e. The density is defined as the number of attendees P per unit area of the event terrain at a certain moment in time ts, and is formalized as follows [20]:

$$k(e, t_s) = \frac{\mathcal{P}(t_s)}{A_e} \tag{1}$$

$\mathcal{P}(t_s)$ denotes the number of attendees at the terrain e at t_s.

Speed is the distance of attendees' movement per unit time. Consider an attendee crossing a whole terrain e during the time window $[t_1, t_2)$, the speed is formally defined as:

$$v(e, t_1, t_2) = \frac{L_e}{|[t_1, t_2)|} \tag{2}$$

where L_e is the distance covered by the attendee when moving through the terrain e. When considering multiple attendees moving through a terrain in different time windows, we could obtain a distribution of speed as a property associated to the terrain, denoted as $\mathbb{V}(e)$.

For an event terrain e, the net flow of attendees traversing a terrain boundary b_e during the time window $[t_1, t_2)$ is defined as:

$$q(t_1, t_2) = \frac{\mathcal{P}_{in}(t_1, t_2) - \mathcal{P}_{out}(t_1, t_2)}{|[t_1, t_2)|} \tag{3}$$

$\mathcal{P}_{in}(t_1, t_2)$ and $\mathcal{P}_{out}(t_1, t_2)$ denote, from t_1 to t_2, the number of attendees moving into the terrain through this boundary, and the number of attendees moving out the terrain, respectively. A flow value $q(t_1, t_2) > 0$ indicates that through b_e the number of attendees entering the terrain exceed the attendees that exit it from t_1 to t_2; otherwise, $q(t_1, t_2) < 0$.

B. GEO-BASED DENSITY ESTIMATION

Density, as defined in Eq. 1, can be measured using traditional sensing infrastructures (e.g. counting systems and Wi-Fi sensors) by means of state-of-the-art methods [9].

The sparse nature of social media data, however, calls for different ways to measure density. Intuitively, given an arbitrary event terrain e (e.g. a square, a venue), the amount of people performing social media activity at a given time instant t_s is normally rather small. To account for such sparsity, we modify the definition of density by considering it a property associated to a *time span* $t_{win} = [t_{start}, t_{end})$. We therefore formalize density measured through social media data as follows:

$$\widehat{k_1}(e, t_{win}) = \frac{|\{u|\forall u \in U, p_u(t_{win}) \geq 1\}|}{A_e} \tag{4}$$

where U is the set of event attendees generating social media activities at the location of the event terrain and $p_u(t_{win})$ denotes the number of posts the social media user u post in t_{win}. The density $\widehat{k_1}$ of a terrain e in the time window t_{win} is therefore calculated as the number of users posting *at least one* micro-post in the targeted area during the considered time window. Considering sparsity of geo-referenced social media data, we choose a time window of one hour. Fig. 1 shows an example estimating the density of the terrain for time window $t_{win} = [10{:}00, 11{:}00)$ considering social media sparse. We leave the investigation of density estimation in shorter time windows to future work.

While increasing temporal boundaries for density calculation, the previous definition puts a very strict constraint on the geographical boundary of the terrain of interest. Attendees could perform social media activity in close proximity to the terrain area. Their communication device could also introduce localization errors due to technical[1] or environmental (e.g. signal blockage, proximity to tall buildings) issues. These errors can range from dozens of meters[2] to even more than 100 meters.[3,4]

To account for such uncertainty, we consider a second definition of density where the boundaries of the considered terrain area are extended by 111.32 meters[5] in each direction. The resulting density measurement is expressed as:

$$\widehat{k_2}(e, t_{win}) = \frac{\{u|\forall u \in U, p_u(t_{win}) \geq 1\}}{A_e^{extend}} \tag{5}$$

C. SPEED-BASED DENSITY ESTIMATION

Though the second definition in the previous section accounts for attendees sent posts in the terrain e or in the extended terrain e during the time span of interest, it does not account attendees who could have been active before entering e, or after leaving it. By considering attendees travel speed, it is possible to account for people that could potentially be present in e in the time span of interest, but posted on social media in a location within walking distance.

Pedestrian speed is known to approximately follow a normal Gaussian distribution [21]. City-scale events can be very crowded: with lots of activities taking place on the event terrains, the motion of pedestrian can be relatively slow. This is the experimental conditions in the "Precinct" scenario of where $\mathbb{V}(e) \sim N(0.97, 0.21^2)$ [21]. We therefore use this result as the assumed pedestrian speed distribution in our study. We leave the robust analysis with respect to the assumption of parameters as well as the assumption in different terrains as future work. We include a parameter Δt that constrains the temporal scope of our model: only users whose posts are detected in the time span $[t_{start} - \Delta t, t_{end} + \Delta t)$ (where $t_{win} = [t_{start}, t_{end})$) are to be considered. As an example, for the terrain in Fig. 1 and the time window $t_{win} = [10{:}00, 11{:}00)$, we consider an extended time span $[09{:}30, 11{:}30)$ (i.e. $\Delta t = 30$ minutes) to account for attendees' travel speed. Attendees posting during this time span, e.g. posting at 09:35, could be present in the terrain during $[10{:}00, 11{:}00)$, are therefore included in the density estimation.

Given the speed distribution and the scoped amount of time, attendees active on social media outside the terrain e before t_{start} (respectively, after t_{end}) will have a probability of being in e within t_{win} that is related to their distance.

Assume a user u to be active at a distance d from the event terrain of interest. We use *pdf* to denote the probability density function of traveling speed. Intuitively speaking, the user should have a speed of at least $\frac{d}{\Delta t}$ in order to reach the terrain e within Δt. Therefore the probability equals to the probability of $v = \frac{d}{\Delta t}$ in the inverse cumulative distribution function of speed distribution. This means that a social media user is more likely to reach the terrain within a certain time window when performing an activity with small distance from considered terrain. The probability of being in the terrain within Δt can be calculated as:

$$\begin{aligned} P_{\Delta t}(e, d) &= P(v(e) \geq \frac{d}{\Delta t}) \\ &= \int_v pdf(v(e) \geq \frac{d}{\Delta t}) \end{aligned} \tag{6}$$

[1] https://tnp.uservoice.com/knowledgebase/articles/1117027-gps-location-errors

[2] https://www.gps.gov/systems/gps/performance/accuracy/

[3] http://www.radio-electronics.com/info/satellite/gps/accuracy-errors-precision.php

[4] https://msu.edu/~brook/publications/prec_ag/oct1998.htm

[5] 111.32 meters are equivalent to a decimal degree precision of 3 decimal places: https://en.wikipedia.org/wiki/Decimal_degrees

Assuming that at the same location with distance d to the terrain there are $N(d)$ attendees active on social media, then $N(d) \times P_{\Delta t}(e, d)$ of them will possibly be in the terrain during t_{win}. When considering users at locations with different distances from the terrain, the number of users that could contribute to the density of the terrain in the considered time span can be calculated as:

$$\widehat{k_3}(e, t_{win}) = \frac{1}{A_e}\left\{\{u|\forall u \in U, p_u(t_{win}) \geq 1\}\right. \\ \left. + \int_d N(d) \times P_{\Delta t}(e, d)\right\} \quad (7)$$

D. FLOW-BASED DENSITY ESTIMATION

Data about attendee flows (i.e. number of attendees traversing the boundaries of a terrain per unit of time) could also be used to support attendees' density estimation. Such flow information can be obtained by counting systems and/or Wi-Fi sensors, as illustrated in Fig. 1. Values of $q(b_e, t_1, t_2)$ for other moments of time, such as the previous day, previous week, or during the event on the same day last edition, could be used to scale up attendees' density in the terrain by scaling the probability $P(e, d)$ in Eq. 6 before t_{start} (or after t_{end}) according to previous traffic conditions. To model this, we consider for each terrain boundary b_e the number of attendees 1) active before t_{start} ($[t_{start} - \Delta t, t_{start})$) and 2) after t_{end} ($[t_{end}, t_{end} + \Delta t)$). We use $c_{bf}(b_e)$ and $c_{af}(b_e)$ to denote the scaling factors for boundary b_e considering attendees active before t_{start} and after t_{end}, respectively. In addition, $N_{bf}(d)$ and $N_{af}(d)$ denote the number of social media users with distance d to the terrain before t_{start} and after t_{end}. The estimated density is calculated as follows:

$$\widehat{k_4}(e, t_{win}) = \frac{1}{A_e}\left\{\{u|\forall u \in U, p_u(t_{win}) \geq 1\}\right. \\ + \sum_e \left(c_{bf}(b_e)\int_d N_{bf}(d) \times P_{\Delta t}(e, d)\right. \\ \left.\left. + c_{af}(b_e)\int_d N_{af}(d) \times P_{\Delta t}(e, d)\right)\right\} \quad (8)$$

The scaling factor $c_{bf}(b_e)$ and $c_{af}(b_e)$ for each boundary b_e are calculated as in Eq. 9, to respectively account for activities performed before or after the considered time span. In the equation, $t_s = t_{start}$ and $t_e = t_{end}$.

$$c_{bf}(b_e) = \begin{cases} \frac{\mathcal{F}(b_e, t_s - \Delta, t_s)}{\int_d N_{bf}(d) \times P_{\Delta t}(e, d)}, & \text{if } \mathcal{F}(b_e, t_s - \Delta t, t_s) > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$c_{af}(b_e) = \begin{cases} \frac{|\mathcal{F}(b_e, t_e, t_e + \Delta t)|}{\int_d N_{af}(d) \times P_{\Delta t}(e, d)}, & \text{if } \mathcal{F}(b_e, t_e, t_e + \Delta t) < 0 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Let us first consider the case of attendees active outside the terrain during $[t_{start} - \Delta t, t_{start})$. The scaling factor $c_{bf}(b_e)$ assumes a positive value when $\mathcal{F}(b_e, t_{start} - \Delta t, t_{start}) > 0$, i.e. when, in the considered time period there are more attendees entering the terrain than leaving it. When, on the other hand, $\mathcal{F}(b_e, t_{start} - \Delta t, t_{start}) < 0$, i.e. there are more attendees leaving the terrain than entering it, their impact can be modeled as $c_{bf}(b_e) = 0$, that is, no additional attendees active on social media should be counted in estimating the density of the terrain during $[t_{start}, t_{end})$.

When attendees are active outside the terrain during $[t_{end}, t_{end} + \Delta t)$ (i.e. after the considered time span), the positive and negative of scaling factor $c_{af}(b_e)$ are the other way around.

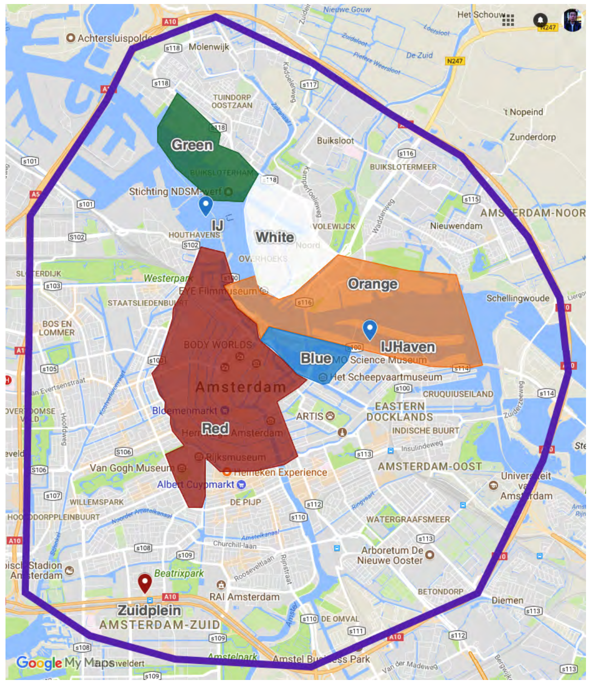


FIGURE 2. Location of targeted terrains in Sail 2015 and Kingsday 2016 in Amsterdam. Most of activities during the Sail event took place in 5 colored oceans (areas), i.e. Orange, White, Blue, Green and Red Oceans. Activities during Kingsday took place in the whole city of Amsterdam (area bounded by dark blue line). Marked locations indicate where the terrains considered in the research are located. Terrains of the Sail event are located around the IJhaven (Blue marker), while the terrain on Kingsday is located at Zuidplein (Red marker).

IV. EXPERIMENTAL SETUP

This section describes the experimental infrastructure designed and implemented in our work.

We performed our studies in the context of two events city-scale events, the SAIL Amsterdam 2015 nautical event (Sail) and Kingsday Amsterdam 2016 national holiday (Kingsday). First, we elaborate reasons for selecting these two events. Then, we provide a brief introduction of each event, and introduce their terrains focused upon in the experiment. Further, we detail the 4 experimental testing definitions. Finally, we introduce the sensor and social media data collection infrastructure, and the metrics used to compare the performance of our density estimation methods (working on social media data) against the density measurement performed through the sensing infrastructure, here interpreted as ground truth.

A. EVENT SELECTION

The areas affected by Sail and Kingsday are shown in Fig. 2. In the attempt of broadening the scope and validity of our work, we selected events sharing similar properties. Both Sail and Kingsday are 1) *city-scale* events taking place in the same urban environment; 2) *planned*, *temporally constrained*, and thoroughly organized (in contrast to seasonal events, such as Christmas shopping, or serendipitous events, like protests); 3) *popular*, as they are known to attract large crowds, regardless of weather conditions; and 4) *generalist*, and they attract diverse demographics. At the same time, the two events also have important differences, such as 1) *duration*, as Sail lasts for 5 days, ending in a weekend. While, Kingsday is a single-day event, and a public holiday, with celebrations starting from one day before the event day and last for day after it; 2) *topic*, being Sail a naval event (offering, for instance tall-ship exhibition, nautical history experience, fireworks show), while Kingsday is a recurrent national celebration, which offers a boat parade, free market and parties; 3) *event terrain*, with Sail activities centered around the IJhaven area (where ships docked), while Kingsday activities are scattered throughout the city.

More details about events and their terrains for this experiment are introduced in the following sub-sections.

1) THE SAIL AMSTERDAM 2015 NAUTICAL EVENT

SAIL [6] is the largest free nautical event in the world. It takes place every five years in the city of Amsterdam, being the largest public event in the Netherlands. It hosts tall ships from all over the world, moored in the eastern harbor of the city *IJHaven* (IJ harbour) and across the IJ river for attendees coming from all over the world to see and visit. The 2015 edition of SAIL took place from August 19 until August 23, and attracted in total more than 2 million attendees. A high-level view of the area of Amsterdam where the event took place is depicted in Fig. 3c.

The event organizers predefined several walking routes for the attendees to follow. A detailed map of the SAIL event, its routes, and its point of interest is available on the event website.[7] The routes included streets facing the ships' docking areas. Each street is characterized by different morphology (length and width of attendees routes), facilities (e.g. toilets, information desks) and exposure to the main attractions. The main route, called *Orange* route, started from the Amsterdam Centraal station (*Ruijterkade*); it then proceeded east towards the end of the IJHaven passing by the *Veemkade*; to continue north around the *Java Eiland*, first traversing the *Javakade*, and then heading back through the *Sumatrakade*. The streets in proximity to the main attractions hosted stages (e.g. from sponsors) and markets. Buildings close to the event hosted concerts and other initiatives, and, in general, the part of the city nearby the IJHaven transformed to accommodate the event and its attendees.

The weather has been warm and dry for the whole duration of SAIL 2015. The programme included events spanning all five days. August 19 was mainly characterized by the *SAIL-in* parade: the first ships started at 10:00 in IJmuiden and arrived around 14:00 in Amsterdam, while the last ships entered Amsterdam around 17:00. All tall ships entered Amsterdam via the North See Canal, to then dock in the IJHaven. During the following three days, the tall ships were open for visits from 10:00 till 11:00. They then departed on August 23 during the closing *SAIL-out* events. Every day, a firework exhibition took place in the IJHaven around 11:00.

The authors were active in the crowd control room of SAIL 2015, and therefore could witness the evolution of the event. The fourth day (Saturday) was expected to be most crowded, mainly because of locals having their day off. Some crowd management measures have been applied, especially on Saturday afternoon. The *Veemkade*, where most of the tall ships were anchored, was very crowded, with queues forming to access the tall ships. Around stages and other points of interest, people stood still to enjoy music, to have social interactions with other attendees, or to consume food and drinks. Also, the *Javakade*, where people walk through narrow pedestrian bridge and watch tall ships docked in IJHaven, was very crowded.

We focused on four event terrains in Sail for this experiment, highlighted in Fig. 3b:

- **Terrain 1: Ruijterkade** (Blue. Length: 657m. Width: 109m. Area: 6.12ha): the terrain is located at the north of the Amsterdam Centraal station. It continuously serves people using public transport services (the train station, or ferries directed to the northern part of Amsterdam). During SAIL, it served as a main access point to the event. The terrain hosted no relevant points of interest.
- **Terrain 2: Veemkade** (Turquoise. Length: 485m. Width: 71m. Area: 3.41ha): main terrain of the event, where most of the ships were docked. The area hosts offices, bars and restaurants, and some private residence. The terrain gave access to the majority of docked boats.
- **Terrain 3: Javakade** (Red. Length: 617m. Width: 78m. Area: 4.80ha): located on the Java Island, the street directly faces the IJHaven. The terrain is residential, with no recreational businesses. Small pedestrian bridges connect areas separated by canals. The terrain gave access to several docked boats.
- **Terrain 4: Sumatrakade** (Green. Length: 253m. Width: 56m. Area: 1.38ha): located on the Java Island, facing the IJ. The terrain hosted less attractions, compared to the previous two terrains and gave access to only few boats.

[6] https://www.sail.nl/EN-2015
[7] https://www.sail.nl/media/644212/sail_perskaart_1400_990.pdf

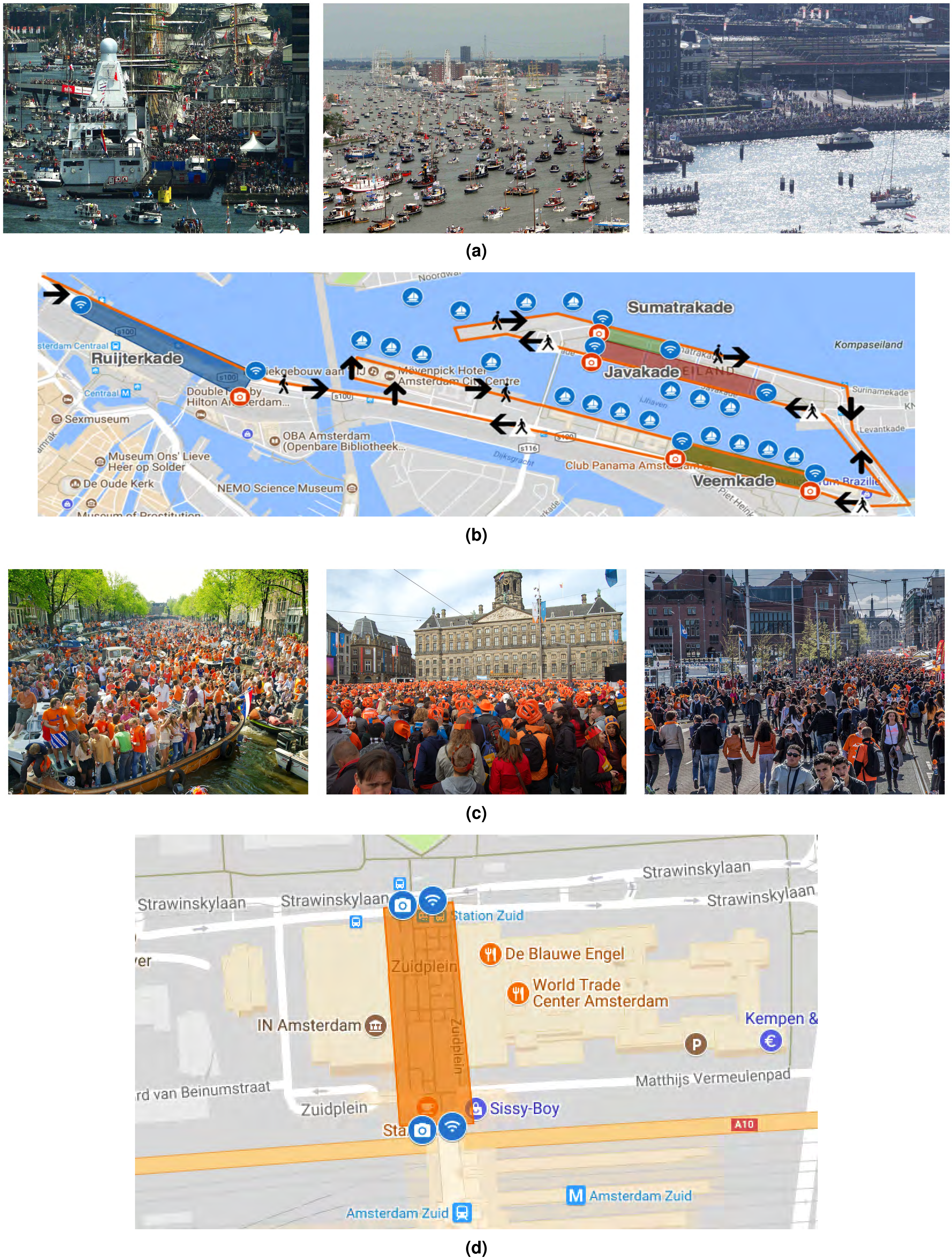


FIGURE 3. Sail Amsterdam 2015 and Kingsday Amsterdam 2016 selected for the experiment. (a) Pictures from Sail event. (b) Terrains of Sail event. (c) Pictures from Kingsday event. (d) Terrain of Kingsday event.

During the event, all locations were devoted to pedestrian and bicycles. Cyclist traffic was reduced during the more crowded hours.

2) THE KINGSDAY AMSTERDAM 2016 EVENT

Kingsday is a national holiday held each year in April 26th in major cities in the Netherlands. It is the birthday of King Willem-Alexander, celebrated with joyful open air festivities. People join this yearly event with their families and friends. In 2016, the King's day celebration attracted more than 1.5 million people in Amsterdam, including Dutch tourists and an organic amount of foreign tourists.

Though a one day public holiday, Kingsday is certainly not a day of rest. The celebrations start on the eve of King's day - named as King's night. Parties, music, and carnival atmosphere continuing throughout the city till the end of the big day. Following the King's Night, the major activities taking place on King's day are free market, boat parade, and gay parties. On King's day morning from 6:00 onwards, the citywide street market in Amsterdam facilitates attendees into trading of their secondhand wars on the streets and in the parks, creating one of the world's largest flea markets. South Amsterdam has the biggest market. In the Jordaan, a crowded market is carried out with folk singers music. Markets in the Vondelpark are dedicated for kids to trade their toys or clothes. From 13:00 onwards, canals are packed with boat parties, with boats sailing along the canals throughout the city with great party vibrations on it. Various street parties and sub-events are carried out in the city with everyone wearing orange. Gay parties are held around Westermarkt and Reguliersdwarsstraat. Besides parties, several big museums are open for people who would like to experience the culture and history.

Kingsday activities occur in the whole city. Pedestrian areas nearby transportation hubs are particularly crowded as people were gathering there and enjoying various activities. We focused on one terrain shown in Fig.3d.

- **Terrain: Zuidplein**: the terrain is the forecourt of the station Amsterdam Zuid. It is a popular pedestrian square located between Station of Amsterdam Zuid and the Strawinsky Avenue surrounded by the World Trade Center (WTC) in the south of Amsterdam. Around the square, there are various shops, sandwiches and other amenities, attracting lots of people. It is a major pedestrian terrain connecting Amsterdam OUD-Zuid, with the CBD area, and Station of Amsterdam Zuid. Nearby, there are two large events in the RAI and the Olympic stadium, which generates large pedestrian flows through this station.

B. EXPERIMENTAL CONDITIONS

We investigate in this paper the properties and performance of the following density estimation methods:

- $\widehat{k_1}$: geo-based density estimation, considering the exact geographical boundaries of the targeted terrain;
- $\widehat{k_2}$: geo-based density estimation, considering the extended boundaries of the targeted terrain;
- $\widehat{k_3}$: speed-based density estimation, using the pedestrian speed distribution suggested by [21] to calculate the probability of social media activities to occur in the targeted terrain;
- $\widehat{k_4}$: flow-based density estimation, using flow estimated through the sensing infrastructure to scale the probability of social media activities.

All methods estimate density from social media data on an hourly basis.

C. DATA COLLECTION

Our experiment took place during the first four days of the SAIL event, and the whole day of the Kingsday event, focusing on the terrains introduced in the previous sections, i.e. the Ruijterkade, Veemkade, Javakade, Sumatrakade for the Sail event, and the Zuidplein for the Kingsday event.

We now describe the sensing infrastructure and social media data processing framework employed to collect experimental data.

1) SENSING INFRASTRUCTURE

Each targeted terrain has been equipped with counting systems and Wi-Fi sensors, as depicted in Fig. 3b and Fig. 3d. Counting systems ran computer vision algorithms on video feeds to count the amount of individual heads crossing a pre-defined cross-section in the street. The counting system provided every minute flow measurements in both directions (inflow and outflow), and had an accuracy of 92%-98%, depending on density conditions. Wi-Fi sensors detected the presence of mobile devices located in their proximity. For each device, the sensor hashed and stored its identifier, as well as its first and last detection time. We estimated that about one third of the counts from counting systems were identified by Wi-Fi sensors. The matching rate between two adjacent Wi-Fi sensors was 3% - 4% of the total flow at the cross-section [9].

TABLE 1. Sensing infrastructure and social media monitoring on targeted terrains.

	Terrain	Counting systems	Wi-Fi	Social media	
				Twit.	Inst.
Sail	Ruijterkade	Single	Both	Y	Y
	Veemkade	Both	Both	Y	Y
	Javakade	Single	Both	Y	Y
	Sumatrakade	Single	Both	Y	Y
Kingsday	Zuidplein	Both	Both	Y	Y

Twit. = Twitter, Inst. = Instagram
Single: the sensor is equipped on one boundary of this terrain.
Both: the sensor is equipped on both two boundaries of this terrain.
Y: data of this social media network is collected in this area.

The *Veemkade* terrain in the Sail and the *Zuidplein* terrain during Kingsday featured a counting system and a Wi-Fi sensor for both considered boundaries. Other terrains had only a single boundary equipped with both sensing devices. Table 1 lists counting systems and Wi-Fi sensors for each terrain.

For boundaries without counting systems, the amount of attendees traversing the cross-section (and the related flow information) has been estimated from Wi-Fi sensors, using the counting-to-Wi-Fi ratio calculated from the other boundary. This infrastructure has been tested and validated in previous studies on pedestrian traffic monitoring [9], and we consider it sufficiently reliable for the purposes of our study.

2) SOCIAL MEDIA DATA COLLECTING & PROCESSING FRAMEWORK

We employed SocialGlass [22], [23], an existing social media retrieval and enrichment framework, to listen from Twitter and Instagram streams for geo-located posts created within the city of Amsterdam during the first four days of SAIL 2015; for Kingsday 2016, we included the day of the event but also the previous and following days, for a total of 3 days of observation. We included in the analysis only geo-located posts, to maximize the spatial accuracy of the retrieved social media data. The inclusion of posts that are not geo-localized but related to the event (and, therefore, potentially localizable) is left to future work.

For each post, the latitude, longitude, timestamp, content, as well as the user id, are collected and stored in a database for further filtering and aggregation. Then, a *density estimation* module assigned each post to a targeted event terrain. Given as input a shape-file of the terrains, the module assesses the time and location of each post and user for each density estimation strategy. With $\widehat{k_1}$ and $\widehat{k_2}$, posts were assigned according to the geo-boundaries of the terrains. With $\widehat{k_3}$ and $\widehat{k_4}$, posts were assigned according to the geo-boundaries of possible routes that could lead to the terrains.

TABLE 2. Descriptive statistics of social media data captured by geo-, speed- and flow-based density estimation methods.

			Twitter		Instagram	
			#User	#Post	#User	#Post
$\widehat{k_1}$	Sail	Ruijterkade	16	24	25	28
		Sumatrakade	4	4	1	1
		Javakade	23	36	285	343
		Veemkade	6	22	61	86
	Kingsday	Zuidplein	2	2	4	4
$\widehat{k_2}$	Sail	Ruijterkade	19	31	38	45
		Sumatrakade	24	32	284	341
		Javakade	24	38	286	345
		Veemkade	10	28	76	104
	Kingsday	Zuidplein	4	5	20	21
$\widehat{k_3}/\widehat{k_4}$	Sail	Ruijterkade	349	717	2662	3577
		Sumatrakade	205	283	1113	1381
		Javakade	193	308	1026	1362
		Veemkade	466	877	3340	4554
	Kingsday	Zuidplein	191	355	3996	4925

Table 2 reports descriptive statistics about the number of geo-located posts and unique users identified for terrains during the two events. A manual inspection of all the posts from the event terrain showed that a high percentage of them referenced the event.

The basic density estimation strategy $\widehat{k_1}$ captured a limited amount of social media activities. This is to be expected, considering the generally low fraction of posts that are also geo-located – especially in Twitter, where geo-located posts are rare (around 1% frequency) [24]. *Sumatrakade*, the less attractive terrain, featured the least amount of posts. *Javakade* and *Veemkade* were the most popular, especially in terms of Instagram posts and users. This is also to be expected, given their proximity and access to tall ships and other points of interest. In Instagram, where geo-located posts are less sparse than in Twitter, *Ruijterkade* featured less posts than *Javakade* and *Veemkade*, indicating that attendees had less reasons to take pictures from that transit terrain. *Ruijterkade* has been comparably popular to *Javakade* and *Veemkade*; this is likely due to the proximity to the central station, a point of interest that attracts a lot of "check-in" posts from tourists and commuters. With other estimation strategies, the amount of captured social media activity and users increases up to one order of magnitude, from more than 300 users to around 4000 users. *Sumatrakade* featured the largest relative increase, due to its close proximity to *Javakade*.

3) COMPARISON METRICS

Density values are compared with three metrics commonly used in time series analysis: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) [25], and Pearson temporal correlation [26]. MAE measures the mean of absolute difference between two time series: a small distance would indicate similar time series in terms of magnitude. MAPE measures the mean of relative difference between two time series. The attendees density in a given event terrain greatly varies over time. Also, a city-scale event is not equally interesting through its whole duration. We therefore expect variations in the amount of attendees that feel compelled to share their experience on social media. Pearson temporal correlation computes the temporal correlation of two time series: a larger correlation would indicate the two time series have similar evolution patterns over time. The Pearson temporal correlation requires the time series data to follow a normal distribution [26]. We verified this condition for all density distributions using the Kolmogorov-Smirnov test [27].

V. RESULTS

This section presents and compares the density ($Persons/M^2$) estimation performance of the four considered methods. We first present the estimated densities; then, we assess their accuracy by comparing the calculated figures against density measured by the sensing infrastructure. Finally, we perform a sensitivity analysis on the Δt parameter of the speed- and flow- based models.

A. RESULTS OF DENSITY ESTIMATION

Density estimates and sensor measurement are calculated on an hourly basis. The technique used to process sensor

TABLE 3. Density of people (#Persons/M²) estimated by geo-, speed-, and flow-based estimation methods based on social media data, compared with sensor data.

		Sensor	$\widehat{k_1}$	$\widehat{k_2}$	$\widehat{k_3}$	$\widehat{k_4}$
Sail	Ruijterkade	2.269e-1±2.108e-1	2.266e-5±1.072e-5	2.639e-5±1.596e-5	1.467e-3±1.106e-3	2.935e-2±2.803e-2
	Veemkade	2.586e-1±2.589e-1	4.924e-5±3.002e-5	5.762e-5±3.683e-5	3.256e-3±2.643e-3	2.606e-2±1.902e-2
	Javakade	1.727e-1±1.486e-1	1.384e-4±1.174e-4	1.370e-4±1.178e-4	7.495e-4±6.165e-4	1.150e-1±1.319e-1
	Sumatrakade	1.018e-1±1.025e-1	7.246e-5±0.000e-5	5.611e-4±4.075e-4	2.803e-3±2.331e-3	3.619e-2±.3401e-2
Kingsday	Zuidplein	3.036e-1±3.583e-1	1.302e-5±5.043e-5	4.557e-5±8.413e-5	8.998e-3±6.163e-3	1.352e-1±1.875e-1

Density of people estimated using social media is on hourly basis according to definition. 4
Density of people measured using sensor data is the mean of density in the same time window as the social media based methods calculated.

measurements is described in previous work [9]. In $\widehat{k_3}$ and $\widehat{k_4}$ Δt is set to 30 minutes. Flow values in $\widehat{k_4}$ are obtained averaging, for each boundary, flow data produced during the Δt preceding the considered time window. Table 3 reports the density ($\mu \pm \sigma$) estimated by the four methods, and measured with sensors for the four SAIL terrains and the Kingsday terrain. Fig. 4 shows for each of the considered terrains the temporal evolution of the estimated densities, to compare them with the density measured with sensors. Estimations from geo-based methods $\widehat{k_1}$ and $\widehat{k_2}$ are 3-4 orders of magnitude lower than density measured by sensors. This is due to the sparsity of social media data within the terrain areas, and in the considered time frame.

Loosening the temporal and spatial constraints, $\widehat{k_3}$ estimates densities 2-3 orders of magnitude lower than the densities measured with the sensing infrastructure. Finally, $\widehat{k_4}$, which uses flow information to scale the density estimated by $\widehat{k_3}$, reaches 1-2 magnitude orders lower than density measured with sensors. In the following, we discuss the result using metrics in more detail.

1) MEAN ABSOLUTE (PERCENTAGE) ERROR

Table 4 (lines 2–11) reports the MAE and MAPE of each density estimation strategy, compared with measures based on sensor data. Geo-based density estimation methods $\widehat{k_1}$ and $\widehat{k_2}$ feature poor performance, with estimation errors up to 99%. The speed-based method $\widehat{k_3}$ provides slightly better performance, with an average 94% error. $\widehat{k_4}$ is the best in the pool, with an average error of 74%, decreasing to 56% in the *Javakade* terrain.

The results of $\widehat{k_3}$ and $\widehat{k_4}$ are very promising, despite the relatively large absolute difference w.r.t. sensor data. Geo-located posts represent only a fraction of all the posts, especially in the Twitter platform [24]. What is more, social media have a relatively small penetration rate in the overall population.[8] Despite this, the method is well capable to estimate densities.

2) SPEARMAN TEMPORAL CORRELATION

The density measured with sensor data in Fig. 4 shows daily patterns for all terrains in two cases, reaching a peak between 14:00 and 16:00, and minimum between midnight and 6:00.

Missing values are due to maintenance or disruptions with the sensing infrastructure. Density estimated with social media data, shows a distinct temporal pattern for each density estimation method.

In the following we analyses the performance of each method, by visually comparing the density curves in Fig. 4, and by commenting on the Pearson temporal correlations shown in Table 4 (line 12-16). Due to sparsity issues, $\widehat{k_1}$ and $\widehat{k_2}$ fail to provide usable density estimates for all terrains, and in almost all time windows. The only exception is *Javakade*, where on August 21 and August 22 an increasing amount of attendees active in social media allowed for a continuous density curve, but featuring a weak temporal correlation ($\widehat{k_1}$= .296, $\widehat{k_2}$= .308; p-value < .05) with sensor data.

The speed-based density estimation method ($\widehat{k_3}$) produces density estimates for most of the hourly time windows and for all terrains. $\widehat{k_3}$ features strong and significant temporal correlation with the sensor density time series. The result shows the benefits deriving from the consideration of attendees that could potentially be present in the terrains, but that post at locations within walking distance from the target event terrain.

The flow-based density estimation method $\widehat{k_4}$ achieves best results. Peak hours with $\widehat{k_4}$ fall into the same range of sensor measures. This could be explained by the scaling effect of flow data, an hypothesis supported by the relevant improvement in terms of temporal correlation (> 0.1) that can be observed from *Javakade* and *Sumatrakade*. However, there are also exceptions such as the correlation for the *Veemkade* terrain decreases (<0.1).

Daily patterns could be observed in Fig. 4 for each terrain, reaching the minimum between 14:00 and 18:00, and the maximum between 7:00 and 11:00. These peak hours differ from those of sensor data.

However, during the active hours (11:00-20:00) of event days the performance is varying in different cells and days. This is particularly obvious in *Veemkade*, where $\widehat{k_4}$ estimation shows plateau while sensor estimation reaches a peak in the afternoon.

Fig. 4e shows the density estimation for the second case, Kingsday 2016, at terrain Zuidplein based on social media and sensor data. Similar to the first case, $\widehat{k_1}$ and $\widehat{k_2}$ fail to provide usable density estimation in all time windows. The speed-based density estimation method $\widehat{k_3}$ and flow-based

[8]Twitter, for instance, has a 17% reach in the Netherland (source https://www.statista.com/statistics/279539/twitter-reach-in-selected-countries/).

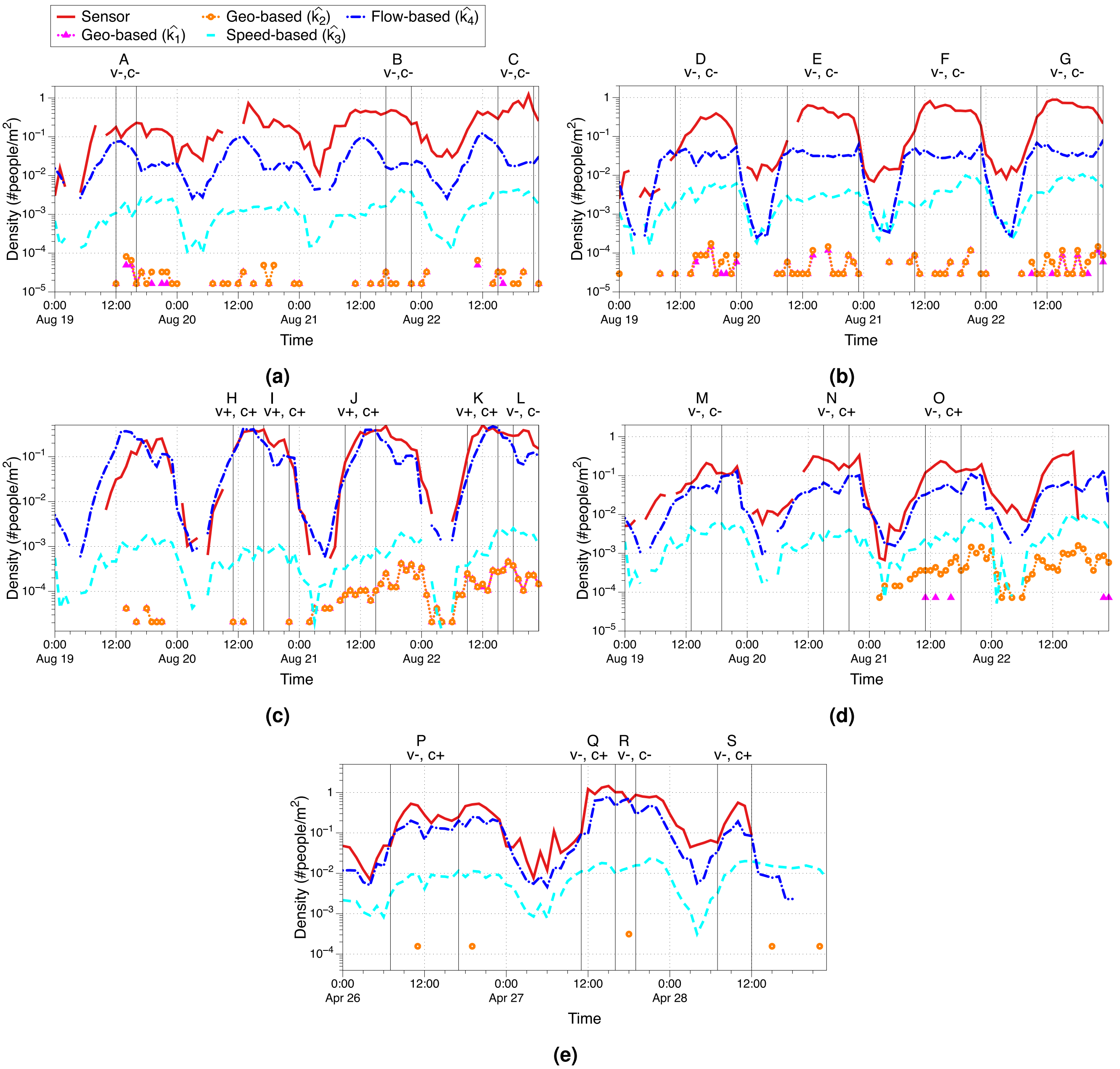


FIGURE 4. Evolution of density (P/m^2) estimates and sensor measurement during SAIL 2015 and Kingsday 2016. A to S denote the ID of selected periods which have similar or distinct value and temporal correlation listed in Table 5. "v+" denotes similar value. "v−" denotes distinct value. "c+" denotes similar temporal correlation. "c−" denotes distinct temporal correlation. (a) Terrain 1: Ruijterkade. (b) Terrain 2: Veemkade. (c) Terrain 3: Javakade. (d) Terrain 4: Sumatrakade. (e) Kingsday, Terrain 1: Zuidplein. Sensor data is only available till 12:00 April 28. Non-continuous lines of $\widehat{k_1}$ and $\widehat{k_2}$ are due to social media data sparsity.

density estimation method $\widehat{k_4}$ provide results for 3 days featuring strong and significant temporal correlation with the sensor density time series. They all clearly shows daily patterns during three days. $\widehat{k_4}$ featured better performance on both mean absolute percentage error and correlation compared with $\widehat{k_3}$ across all days. Density estimation by $\widehat{k_4}$ and sensor data on the second day (the day of the event) reaches the highest value among all three days, followed by the first day which is particularly active during the night. On the third day, $\widehat{k_3}$ features more stable estimation till the end of the day because the sensor data is only available till 12:00 on the third day, as such the $\widehat{k_4}$ is also affected by the lacking of flow information.

B. Δt SENSITIVITY ANALYSIS

We now investigate how the performance of $\widehat{k_3}$ and $\widehat{k_4}$ density estimation methods changes with varying values of Δt, i.e. the model parameter controlling the temporal scope for

TABLE 4. Comparison between density measurement with sensor data and density estimates using geo- ($\widehat{k_1}$, $\widehat{k_2}$), speed- $\widehat{k_3}$, and flow-based $\widehat{k_4}$ methods.

			$\widehat{k_1}$	$\widehat{k_2}$	$\widehat{k_3}$	$\widehat{k_4}$
MAE	Sail	Ruijterkade	.260	.281	.228	.119
		Veemkade	.353	.354	.258	.233
		Javakade	.218	.216	.173	.081
		Sumatrakade	.156	.114	.103	.068
	Kingsday	Zuidplein	.304	.304	.297	.162
MAPE	Sail	Ruijterkade	.9999	.9998	.9879	.8474
		Veemkade	.9995	.9992	.9588	.8569
		Javakade	.9962	.9962	.9705	.5667
		Sumatrakade	.9995	.9764	.9344	.7198
	Kingsday	Zuidplein	.9999	.9997	.9529	.5235
Spearman correlation	Sail	Ruijterkade	.083	-.094	.655***	.698***
		Veemkade	.031	.123	.634***	.537***
		Javakade	.296*	.308*	.486***	.690***
		Sumatrakade	NA	.378*	.586***	.695***
	Kingsday	Zuidplein	.086	.183	.596***	.864***

MAE: Mean Absolute Error.
MAPE: Mean Absolute Percentage Error.
Spearman correlation marked with * and *** indicates p-value$<.05$ and p-value$<.001$, respectively.

micro-posts not created within a terrain of interest. We test values of Δt ranging from 5 minutes to 60 minutes, the length of the time window in this method. Results are shown in Fig. 5. The $\widehat{k_4}$ method is robust to variations of Δt, although optimal performance is achieved for $\Delta t > 20$ minutes. With $\widehat{k_3}$, the temporal correlation of the density estimated in all terrains increases with increasing values of Δt, to stabilize between 30 minutes and 40 minutes. Interestingly, variations are not consistent across terrains. *Veemkade*, for instance, is most affected by changes in the Δt parameter, especially in terms of temporal correlation. On the other hand, estimates in *Ruijterkade* are the most robust. We believe that such inconsistent behavior is due to differences in the properties of the terrains: *Ruijterkade* is a transit terrain, where attendees are less likely to stop during normal traffic conditions. Therefore, taking longer time frame into consideration does not significantly affect the amount of social media users accounted in the density calculation.

In the second case, the $\widehat{k_3}$ in Zuidplein is robust. However, the $\widehat{k_4}$ is not as robust as in terrains in the first case. It reaches the lowest mean absolute error when the value of Δt is around 30 minutes, then the mean absolute error is increased along with increasing of Δt, indicating that Zuidplein is more sensitive with regard to the variation of temporal scope. We account the result to the spatial characteristics of Zuidplein. As a pedestrian square, Zuidplein connects Amsterdam OUD-Zuid, CBD area and Station of Amsterdam Zuid, which is visited by a large number of people every day. However, there are several other streets and roads which also connect these places and are in parallel with the Zuidplein, such as Eduard van Beinumstraat, Beethovenstraat and Parnassusweg. Therefore, loosing temporal and spatial constraints will easily introduce errors in calculating number of people who passed Zuidplein instead of other ways, which consequently increases errors in the density estimation.

VI. DISCUSSION

This section discusses the result of density estimation of each terrain in two cases. In order to get more insights about similar or distinct density estimations, we also look into several factors (e.g. temporal, demographic factors) and discuss their influences.

The $\widehat{k_2}$ in *Javakade* and *Sumatrakade* provide similar density estimation on Aug 21 and Aug 22, the weekend days. The improved performance in *Sumatrakade* with $\widehat{k_2}$ may be explained by the contiguity of the terrain with *Javakade*. It indicates that on social media the density estimation is sensitive to surroundings.

The daily patterns observed using $\widehat{k_4}$ from social media data and sensor data are different, which could be explained by the different types of activities captured by the two infrastructures – respectively, pedestrian movement and social media communication. Intuitively, some time slots during the event are more worthy of communication than others (e.g. ships during good lighting conditions, fireworks); on the other hand, the amount of attendees visiting schedules are affected by other factors (e.g. time and day of the week). However, some of communication oriented activities, such as Fireworks (lasting a maximum of 30 minutes) at 11pm each day in Sail event, are not captured by $\widehat{k_4}$, i.e. no peaks around 11pm on $\widehat{k_4}$. This may account for the influence of the length of time window selected for this experiment, i.e. 1h duration of time window may neutralize the high crowds during the

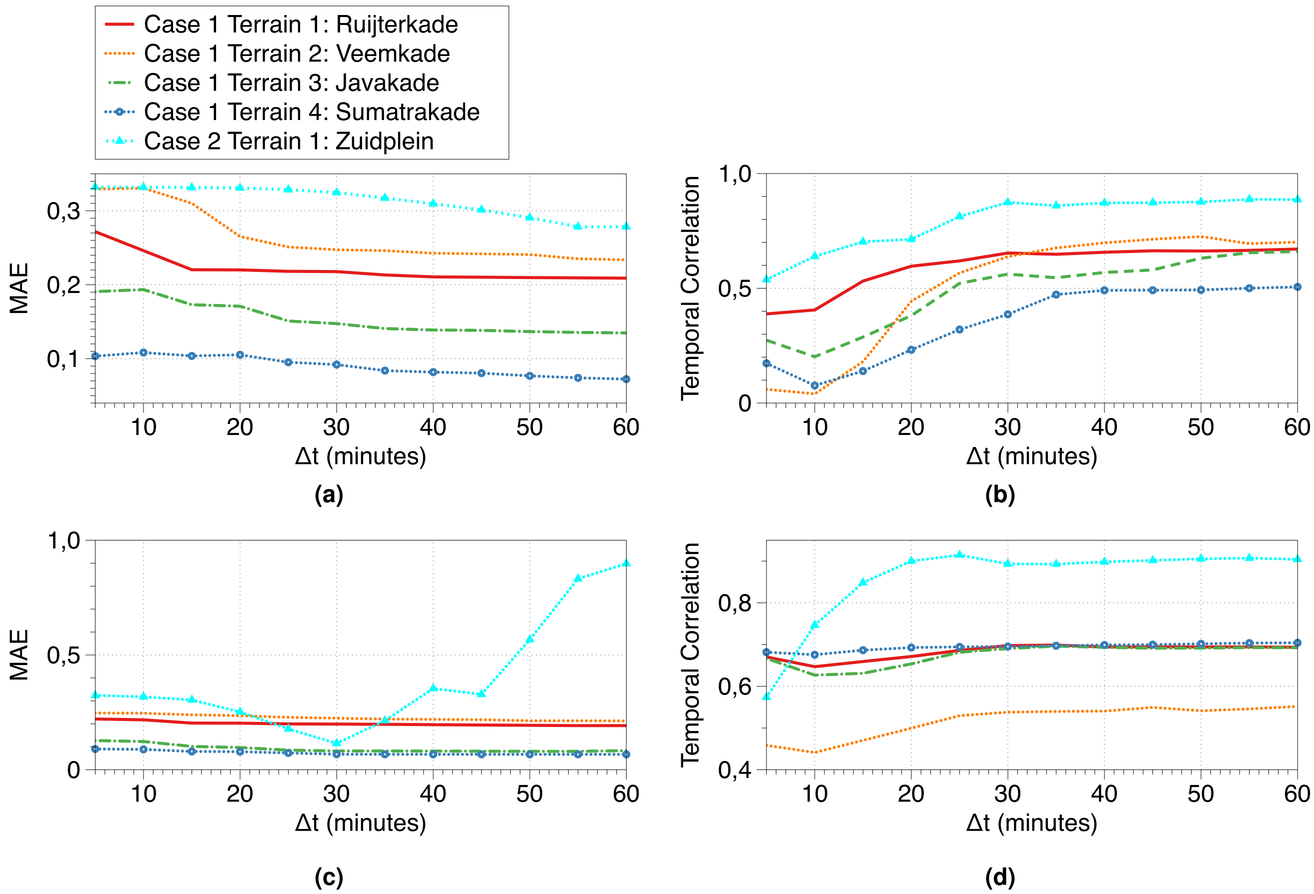


FIGURE 5. **The effects of Δt on the performance of speed-based density estimation method $\widehat{k_3}$ and $\widehat{k_4}$. (a) $\widehat{k_3}$: MAE. (b) $\widehat{k_3}$: Spearman correlation. (c) $\widehat{k_4}$: MAE. (d) $\widehat{k_4}$: Spearman correlation.**

fireworks and low flows towards the end of the hour. Thus, shorter time windows might capture these peaks.

Density estimation using social media featured higher performance in the second case than in the first case. This could be attributed to the diverse fingerprints of events and terrains as activities during Sail enhanced distinction of pedestrian movement and social media communication more than activities during Kingsday in those terrains.

Results also show that during active hours (7:00-23:00), density estimation performance varies for different terrains and events. In order to get more insights into them, we selected a set of periods which have either very similar or very distinct density estimation through social media data compared to sensor data ($\widehat{k_4}$, flow-based strategy) according to Mean Absolute Error and Spearman Temporal Correlation shown in Table 5. For each period we derived information from the crowd for various aspects such as demographic (i.e. Age, Gender), role of people with regard to the city (i.e. resident, local tourist, foreign tourist) and PoI preference of people, extracted through the SocialGlass system.

During Sail event the density estimation during periods of H, I, J and K in *Javakade* reaches best performance, i.e. similar value and similar temporal correlation. We found that the gender distribution derived from social media is more equal in these periods compared with other period in the same terrain (i.e. L), or periods in other terrains (e.g. A, D, N). Results points toward a relationship between the gender distribution of social media users and the performance of density estimation. However, this does not hold in the second case, where periods of P, Q and S reach a similar correlation while having less distinct values but the gender distribution does not show obvious patterns. This result suggests that other factors, such as type of events and location of the terrain, also play a role in the performance of our methods.

With regard to periods D, E, F and G in *Veemkade* which show huge distinctions in density estimation with regard to the sensor based method, we found that there are more male residents. Recent research [28] found that male and resident social media users are less active during city-scale events. Thus the reverse observation may indicate that the representativeness of social media data w.r.t. the reality is decreased. Consequently, the performance of density estimation based on social media data is affected. *Veemkade* is the narrowest terrain on the route of Orange Route connecting Amsterdam Central Station with *Javakade* and *Sumatrakade*, and it hosted restoration services and other Point of Interest,

TABLE 5. Selected periods with similar or distinct MAE. and Temproal Correlation in density estimation based on sensor and social media data.

	Terrain	ID	Value	Temporal Correlation	Day	Period (hh-hh)
Sail (Aug. 2015)	Ruijterkade	A	-	-	19th	12-16
		B	-	-	21st	17-22
		C	-	-	22nd	15-23
	Veemkade	D	-	-	19th	11-23
		E	-	-	20th	09-23
		F	-	-	21st	10-23
		G	-	-	22nd	10-23
	Javakade	H	+	+	20th	11-15
		I	+	+	20th	17-22
		J	+	+	21st	09-15
		K	+	+	22nd	09-15
		L	-	-	22nd	18-23
	Sumatrakade	M	-	-	19th	13-19
		N	-	+	20th	15-20
		O	-	+	21st	11-18
Kingsday (Apr. 2016)	Zuidplein	P	-	+	19th	07-17
		Q	-	+	20th	11-16
		R	-	-	20th	16-19
		S	-	+	21st	07-12

ID: refers to the ID of periods shown in Figure 4.
Value: the value of estimated density.
Temporal Correlation: the temporal correlation of estimated density.
"+": denotes the similar value or temporal correlation.
"-": denotes the distinct value or temporal correlation.

where people would stop, stand still, and block or hamper the flow of attendees. These may lead to the result that more people are detected by sensors rather than from social media. Consequently, the density of people detected from sensors and social media is in different value and correlation during these periods.

The selected periods A, B and C which show both distinct value and temporal correlation are from *Ruijterkade*. We found that during these periods there are more female foreigners active in social media, visiting PoIs such as Art & Entertainment, Food and Shop & Services in this terrain. However, the pattern of their influences is not clear.

Density estimations during periods N and O in *Sumatrakade* show similar temporal correlation but distinct value. We found that proportion of gender and role of people derived from social media in these periods show diverse values, but their patterns are not obvious, which is similar to the periods in *Zuidplein* in the second case.

In *Zuidplein*, density estimations in periods of P, Q and S show similar temporal correlation but distinct values, while period R shows both distinct temporal correlation and value. We found that the proportion of gender, role and the PoI preference of people are diverse during these periods. However, the pattern of their impacts is not obvious.

Above insights of the selected periods indicate that demographics, role, PoI preference of crowd, type of events, location of terrains as well as other factors may affect density estimation performance using social media. To fully understand their impacts, it calls for future work on factor analysis on density estimation performance based on social media data.

VII. CONCLUSIONS

The density of attendees in an event terrain is an important measure of success and safety for city-scale events. In this paper we investigated the suitability of geo-referenced social media data produced during a city-scale event as a source for attendee density estimation. Social media have been used in a variety of contexts to analyses the amount of attendees at high temporal granularity, but low spatial granularity (e.g. city scale). However, due to the inherent geographical sparsity of geo-located social media data, the analysis of attendance at higher spatial granularity (e.g. street-scale) received less attention.

This paper proposes three density estimation strategies based on pedestrian traffic flow theory – respectively geo-, speed- and flow-based density estimation – that were successfully validated during city-scale events. When applied to geo-located social media sources for all strategies and additional flow data source for flow-based strategy, these strategies mitigate the spatial sparsity problem by considering traffic conditions (speed distribution and flow) to account for attendees that perform event-related social media activity outside an event terrain of interest. Thanks to a sophisticated sensing infrastructure deployed during SAIL 2015 and Kingsday 2016 in Amsterdam in the Netherlands, we assessed the performance of our methods on 5 event terrains characterized by different morphology and relevance to activities in both events. The flow-based method achieves promising performance in all terrains, both in terms of relative mean difference (from 20% to 250% improvement with regard to other methods) and temporal correlation (between .54 and .87). The speed-based method also features strong temporal correlation (between .49 and .65), but with higher estimation errors. Geo-based methods can yield useful results only when the amount of social media activity in the targeted terrain is sufficiently high.

We show that several factors play a significant role in terms of estimation accuracy and temporal correlation, such as the properties of a terrain, demographics, role and PoI preferences of the crowd. In Sail 2015, an attractive and trafficked terrain like *Veemkade* featured lower estimation accuracy and lower correlation than other terrains; a trafficked but less interesting terrain like *Ruijterkade* featured maximal temporal correlation but low estimation accuracy; a less trafficked terrain like *Javakade* featured higher estimation accuracy, but lower temporal correlation. Across all terrains, it is observed that maximal performance (i.e. higher temporal correlation and estimation precision than other terrains) is achieved with equal proportion of male and female in the crowd.

In the second case, Kingsday 2016, the trafficked terrain *Zuidplein* featured high correlation. The sensitivity analysis showed by loosing temporal and spatial constrains that the speed-based and flow-based methods achieve optimal performance when including users active at walking distance, and within 30-40 minutes from the temporal windows of observation. The characteristics of people counted for density estimation also affect the result. *Javakade* in the first case featured best performance with equally distributed gender of social media users than any other cells. Other factors, such as role and PoI preference of people, different types of events, also introduce influences on the result, but the patterns of their impacts are not clear.

The experimental result and the identification of influencing factors on the one hand help to avoid bias in applying this method for density estimation using social media, while on the other hand they call for future research in order to improve the estimation performance. In the next step we plan to take into consideration activity times of attendees, and investigate if the actual attendee speed distribution on the event terrain can be used for optimizing the density estimation. Further, we are going to zoom-in on the relation existing between traffic conditions and social media activity, to seek for stronger evidences of laws that relate attendees density with mobile online activity. We are also going to improve our estimation methods by using counting systems to provide speeds, using non-geo posts or posts with PoI information (e.g. from Facebook pages) in order to overcome data sparsity, using auto-filtering techniques to enhance posts filtering performance, and so on. We plan to compare the performance of our methods in various contexts of city-scale events, having different nature, size, and position in the city. Finally, we will explore the impact of factors, such as demographics, role, PoI preference of crowds, on the density estimation performance.

REFERENCES

[1] D. Helbing, L. Buzna, A. Johansson, and T. Werner, "Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions," *Transp. Sci.*, vol. 39, no. 1, pp. 1–24, Feb. 2005, doi: 10.1287/trsc.1040.0108.

[2] M. Wirz, T. Franke, D. Roggen, E. Mitleton-Kelly, P. Lukowicz, and G. Tröster, "Inferring crowd conditions from pedestrians' location traces for real-time crowd monitoring during city-scale mass gatherings," in *Proc. IEEE 21st Int. Workshop Enabling Technol., Infrastruct. Collaborative Enterprises*, Jun. 2012, pp. 367–372.

[3] W. Daamen, Y. Yuan, D. C. Duives, and S. P. Hoogendoorn, "Comparing three types of real-time data collection techniques: Counting cameras, Wi-Fi sensors and GPS trackers," in *Proc. Pedestrian Evacuation Dyn.*, 2016, pp. 568–574.

[4] J. Cranshaw, R. Schwartz, J. I. Hong, and N. Sadeh, "The livehoods project: Utilizing social media to understand the dynamics of a city," in *Proc. Int. AAAI Conf. Weblogs Social Media*, 2012, pp. 1–8.

[5] S. Hasan, X. Zhan, and S. V. Ukkusuri, "Understanding urban human activity and mobility patterns using large-scale location-based data from online social media," in *Proc. 2nd ACM SIGKDD Int. Workshop Urban Comput.*, 2013, pp. 1–8.

[6] D. Quercia, L. M. Aiello, R. Schifanella, and A. Davies, "The digital life of walkable streets," in *Proc. 24th Int. Conf. World Wide Web*, 2015, pp. 875–884.

[7] R. Lee and K. Sumiya, "Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection," in *Proc. 2nd ACM SIGSPATIAL Int. Workshop Location Based Social Netw.*, 2010, pp. 1–10.

[8] M. Balduini, E. D. Valle, D. Dell'Aglio, M. Tsytsarau, T. Palpanas, and C. Confalonieri, "Social listening of city scale events using the streaming linked data framework," in *Proc. Int. Semantic Web Conf.* Berlin, Germany: Springer, 2013, pp. 1–16.

[9] Y. Yuan, W. Daamen, D. Duives, and S. Hoogendoorn, "Comparison of three algorithms for real-time pedestrian state estimation—Supporting a monitoring dashboard for large-scale events," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst.*, Nov. 2016, pp. 2601–2606.

[10] U. Blanke, G. Tröster, T. Franke, and P. Lukowicz, "Capturing crowd dynamics at large scale events using participatory GPS-localization," in *Proc. IEEE 9th Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process.*, Apr. 2014, pp. 1–7.

[11] J. Weppner and P. Lukowicz, "Bluetooth based collaborative crowd density estimation with mobile phones," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2013, pp. 193–200.

[12] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring millions of footprints in location sharing services," in *Proc. Int. AAAI Conf. Weblogs Social Media*, 2011, pp. 81–88.

[13] D. Quercia, R. Schifanella, L. M. Aiello, and K. McLean, "Smelly maps: The digital life of urban smellscapes," in *Proc. 9th Int. Conf. Web Soc. Media*, 2015, pp. 327–336.

[14] F. Botta, H. S. Moat, and T. Preis, "Quantifying crowd size with mobile phone and Twitter data," *Roy. Soc. Open Sci.*, vol. 2, no. 5, p. 150162, 2015.

[15] Y. Liang, J. Caverlee, Z. Cheng, and K. Y. Kamath, "How big is the crowd?: Event and location based population modeling in social media," in *Proc. 24th ACM Conf. Hypertext Social Media*, 2013, pp. 99–108.

[16] P. Georgiev, A. Noulas, and C. Mascolo, "The call of the crowd: Event participation in location-based social services," in *Proc. AAAI*, 2014, pp. 1–10.

[17] A. D. May, *Traffic Flow Fundamentals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1990.

[18] C. F. Daganzo, *Fundamentals of Transportation and Traffic Operations*, vol. 30. Oxford, U.K.: Pergamon, 1997.

[19] W. Daamen, S. Hoogendoorn, and P. Bovy, "First-order pedestrian traffic flow theory," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1934, pp. 43–52, 2005. [Online]. Available: https://trrjournalonline.trb.org/toc/trr/2005/1934/+

[20] D. C. Duives, W. Daamen, and S. P. Hoogendoorn, "Quantification of the level of crowdedness for pedestrian movements," *Phys. A, Statist. Mech. Appl.*, vol. 427, pp. 162–180, Jun. 2015.

[21] S. Chandra and A. K. Bharti, "Speed distribution curves for pedestrians during walking and crossing," *Procedia-Social Behav. Sci.*, vol. 104, pp. 660–667, Dec. 2013.

[22] S. Bocconi, A. Bozzon, A. Psyllidis, C. T. Bolivar, and G.-J. Houben, "Social glass: A platform for urban analytics and decision-making through heterogeneous social data," in *Proc. 24th Int. Conf. World Wide Web*, 2015, pp. 175–178.

[23] A. Psyllidis, A. Bozzon, S. Bocconi, and C. T. Bolivar, "Harnessing heterogeneous social data to explore, monitor, and visualize urban dynamics," in *Proc. Planning Support Syst. Smart Cities, 14th Int. Conf. Comput. Urban Planning Urban Manage. (CUPUM)*, Cambridge, MA, USA, Jul. 2015, pp. 1–22.

[24] B. Hecht, L. Hong, B. Suh, and E. H. Chi, "Tweets from Justin Bieber's heart: The dynamics of the location field in user profiles," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2011, pp. 237–246.

[25] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

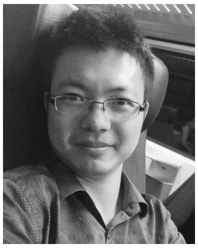
[26] J. Hauke and T. Kossowski, "Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data," *Quaestiones Geograph.*, vol. 30, no. 2, pp. 87–93, 2011.

[27] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, "Universal and nonuniversal properties of cross correlations in financial time series," *Phys. Rev. Lett.*, vol. 83, no. 7, p. 1471, 1999.

[28] J. Yang, C. Hauff, G.-J. Houben, and C. T. Bolivar, "Diversity in urban social media analytics," in *Proc. Int. Conf. Web Eng.* Springer, 2016, pp. 335–353.

V. X. GONG is currently pursuing the Ph.D. degree with the Web Information Systems Group and ALLEGRO Project Team, Delft University of Technology, The Netherlands.

His research interests include data modeling and analyzing based on various data sources in order to understand pedestrians and cyclists behavior in an urban context, which involves identifying relevant information from various social media networks, such as Twitter, Instagram, and Foursquare, developing methods to model and analyze data, and analyzing the performance in comparison with other data sources.



J. YANG received the Ph.D. degree from the Web Information Systems Group, Delft University of Technology, The Netherlands, in 2017. He is currently a Senior Researcher with the eXascale Infolab, University of Fribourg, Switzerland.

His research interests include building effective human-machine loop systems that combine human intelligence with machine scalability to solve complex tasks at scale. The topic lies at the intersection of human computation, machine learning, recommendation, and user modeling. His work finds its natural application in human computation, recommendation, question answering, and urban computing systems.

W. DAAMEN is currently an Associate professor with the Chair of Traffic Operations and Management, Department of Transport and Planning, Delft University of Technology, The Netherlands.

His research interests include theory, modeling, and simulation of traffic (pedestrians, cyclists, vehicles, and vessels), and innovative methods have been developed to collect microscopic traffic data, which are used to underpin theories and models describing traffic operations.

A. BOZZON is currently an Assistant Professor with the Web Information Systems Group, Delft University of Technology, The Netherlands. He is also a Research Fellow with the AMS Amsterdam Institute for Advanced Metropolitan Solutions, The Netherlands, and a Faculty Fellow with the IBM Benelux Center of Advanced Studies, The Netherlands.

His research interests include the intersection of crowd-sourcing, user modeling, and web information retrieval. He has studied and created novel social data science methods and tools that combine the cognitive and reasoning abilities of individuals and crowds, with the computational powers of machines, and the value of big amounts of heterogeneous data.

S. HOOGENDOORN is currently a Professor and the Head of the Chair of Traffic Operations and Management, Department of Transport and Planning, Delft University of Technology, The Netherlands. He is also a Principal Investigator with the AMS Amsterdam Institute for Advanced Metropolitan Solutions, The Netherlands, a Faculty Fellow with the IBM Benelux Center of Advanced Studies, The Netherlands, and a Strategic Advisor with ARANE, The Netherlands.

In the past five years, his research has involved theory, modeling, and simulation of traffic and transportation networks. He focused on innovative approaches to collect microscopic traffic data and the use of these data to underpin the models and theories that he have developed, using new techniques for model identification.

G. J. HOUBEN is currently a Professor of web information systems with the Software Technology Department, Delft University of Technology. He is also a Principal Investigator with the AMS Amsterdam Institute for Advanced Metropolitan Solutions, and a Faculty Fellow with the IBM Benelux Center of Advanced Studies.

His research interests include web engineering, web science, and user modeling, adaptation and personalization. He is a Managing Editor of the *Journal of Web Engineering*, and an Editorial Board Member for the *Journal of Web Science*, the *International Journal of Web Science*, *User Modeling and User-Adapted Interaction*, and the ACM *Transactions on the Web*.