Traffic Condition Is More Than Colored Lines on a Map: Characterization of Waze Alerts

Thiago Henrique Silva, Antonio A. F. Loureiro, Juliana Salles, Aline Carneiro Viana, Pedro O.S. Vaz de Melo, Jussara Marques Almeida

Traffic Condition Is More Than Colored Lines on a Map: Characterization of Waze Alerts

Thiago H. Silva[1,3], Pedro O.S. Vaz de Melo[1] Aline Carneiro Viana[3],
Jussara M. Almeida[1], Juliana Salles[2], and Antonio A.F. Loureiro[1]

[1] Federal University of Minas Gerais, Belo Horizonte, Brazil
[2] Microsoft Research, Redmond, USA
[3] INRIA, France

Abstract. Participatory sensor network (PSN) enables the understanding of city dynamics and the urban behavioral patterns of their inhabitants. In this work, we focus our analysis on a specific PSN, derived from Waze, for sensing traffic conditions. Our objective is to characterize the properties of this PSN, its broad and global spatial coverage as well as its limitations. We also bring discussions on different opportunities for application design using this network. We claim that the PSN derived from Waze has the potential to help us in the better understanding of traffic problem reasons. Besides that, it could be useful for improving algorithms used in navigation services: (1) by exploiting the provided real-time traffic information or (2) by helping in the identification of valuable pieces of information that are hard to detect with traditional sensors, such as car accidents and potholes.

Keywords: Urban social behavior, city dynamics, participatory sensing, mobile social networks, social big data.

1 Introduction

Participatory Sensing Systems (PSSs) [1,2] are revolutionizing the way we see cities, societies and the interactions among people. PSSs provide a mobile interface that allows people carrying smartphones to share data about the environment (or context) they are inserted in at any time and place. These systems certainly have the power to contribute in the process of making ubiquitous computing a reality. Consider the large variety of PSSs already deployed and functioning at global scale, such as Foursquare[1], Instagram[2], Weddar[3], and Waze[4]. Each of these systems can provide valuable information about an aspect of a given city or society in almost real-time, such as its traffic and weather conditions, local parties and festivals, riots, among others. More importantly, the cost for obtaining this data is almost negligible, since it is distributed among all the people who are sharing it.

From participatory sensing systems we can derive participatory sensor networks (PSNs), where each node in the network consists of a user equipped with a mobile device, sending data to web services. In this direction, we can view PSNs as sensing layers

---

[1] http://www.foursquare.com
[2] http://www.instagram.com
[3] http://www.weddar.com
[4] http://www.waze.com

of a global scale sensor network that uses humans in the sensing process. For example, from Waze we can obtain a layer about the traffic, from Instagram a layer containing pictures of places, from Foursquare, we can obtain a layer about the category of locations, and from Weddar, a layer about weather conditions. Each layer is responsible for sensing data related to a certain aspect, for instance traffic or weather conditions, of a specific area in the globe, such as countries, cities, or neighborhoods. In this work, we focus our analysis on a specific sensing layer, the one responsible for sensing traffic conditions. Data collected from this layer, as well as from others as mentioned above, have the potential to transform society. They enable the understanding of city dynamics and the urban behavioral patterns of their inhabitants, supporting smarter decision making. In fact, real-time traffic maps could inform more than the traffic flow's conditions (usually represented by colored lines in the map), for example, it could provide routes that cause less pollution to the city, dangerous areas to avoid, among others.

In order to evaluate the potential of the traffic sensing layer, we here analyze participatory data coming from Waze – the most popular traffic report application. Waze was created in 2008 and recently, had approximately 50 million users [3]. Waze periodically collects sensor data from mobile phones, and uses it to compute the speed of their devices to infer traffic conditions. The system also offers to its users predefined alerts stating incidents such as traffic jams and police traps, extending the information about traffic conditions. One of Waze's main features is the user engagement to contribute to the common good, i.e., Waze is not just crowdsourcing, but personal participation [3].

The objective of this work is to characterize the properties of the PSN derived from Waze, its broad and global spatial coverage as well as its limitations. Moreover, we discuss different opportunities for application design using data collected from Waze. For example, such data could be exploited to drive improvements in algorithms for navigation services and to support quicker identification of information about car accidents, potholes, and slippery roads, which are valuable information that are hard to detect with traditional sensors.

The rest of this paper is organized as follows. In Section 2 we present the related work. In Section 3 we discuss the participation of human in the process of sensing. In Section 4 we present the characterization of a PSN derived from Waze. In Section 5 we illustrate some of the possible applications based on the data shared in Waze. Finally, in Section 6 we present the conclusions and future work.

2   Related Work

Data obtained from participatory sensor networks (PSNs) may be very complex and, therefore, a fundamental step in any investigation is to analyze the collected data to understand its characteristics and usefulness. There are several proposals devoted to the study of specific characteristics of PSNs. For example, in location sharing services like Foursquare, Cranshaw et al. [4] presented a model to extract distinct regions of a city that reflect current collective activity patterns. In a previous work [1], we have characterized data collected from three distinct PSNs derived from location sharing services, such as Foursquare. Among the results, we showed the planetary scale of those networks, as well as the highly unequal frequency of data sharing, both spatially and temporally, which is highly correlated with the typical routine of people. In another

previous work [5], we performed the first characterization of Instagram using photos shared by users, analyzing them from a sensor network point of view. We showed that photo-sharing systems, particularly the Instagram, can also be used to map the characteristics of urban locations at a low cost.

Quercia et al. [6] studied how social media communities resemble real-life ones. They tested whether established sociological theories of real-life social networks hold in Twitter. They found, for example, that social brokers in Twitter are opinion leaders who take the risk of tweeting about different topics. Frias-Martinez et al. [7] proposed a technique to determine the most common activities in a city by studying tweeting patterns. Sakaki et al. [8] studied the real-time interaction of events (e.g., earthquakes) in Twitter and proposed an algorithm to monitor tweets to detect a target event.

To the best of our knowledge, Fire et al. [9] is the only prior work to analyze Waze. They showed that it might be possible to identify areas where accidents are more likely to occur by analyzing user accident reports in Waze.

Our work differs from all previous studies as is the first characterization of Waze from a crowdsensing point of view. Moreover, continuing our recent studies [1,5], we show that traffic alert sharing systems, particularly Waze, can also be exploited for mapping the characteristics of urban locations at a low cost, providing complementary data in relation to those obtained from other types of systems, such as location or photo sharing system. As previously mentioned, we believe that the personal involvement of users in such system can allow inferring much richer conclusions about traffic conditions than the usual colored information about traffic jam provided by on-line traffic websites. This work also discuss possible ways towards this goal.

3   Social Media as a Source Sensing

Social media websites based on location (also known as Location-based Social Networks – LBSNs), such as Foursquare, Instagram and Waze, build new virtual environments that integrate user interactions. Such systems have been extensively used for different applications. The ubiquity of smartphones, associated with the adoption of social media websites, enable unprecedented opportunities to study city dynamics and urban social behavior by mining the social big data shared by users of these applications.

Social media websites based on location allow people sharing useful data about the city area where they are located at any given moment, acting as a source of sensing and, thus, leading to the so-called participatory sensing systems [1,2]. Participatory sensing networks (PSNs) can be derived from such systems, where nodes represent users equipped with mobile devices sending data to web services. For example, in a Waze derived PSN, the sensed data consists of reports on traffic conditions of a specific road/street. Data from PSNs can be usually collected through the API of the specific system (e.g., Foursquare, Instagram or Waze APIs). More details about PSNs can be found in [5,1], in [1] is also discussed in more details the challenges faced by this emerging type of network. Different systems, e.g. Foursquare, Waze, and Instagram, lead to different PSNs. Each PSN may enables the access to data related to a certain city aspect, being considered a sensing layer.

4   Characterization of Waze

This section investigates the participatory sensor network (PSN) derived from Waze.

4.1   Overview of the Dataset

Waze is a popular navigation system that uses crowdsensing to offer near real-time traffic information and routing. The system was created in 2008 and registered approximately 50 million users in 2013. Waze periodically collects data from the built-in GPS typically found in smart phones, and uses it to compute the speed of the device. With that, Waze can provide useful information about traffic conditions in different areas. The system also offers to its users predefined alerts stating incidents such as traffic jams and police traps, which extends the information about traffic conditions. It is also possible to use subcategories of incidents to better specify them, for example, "heavy traffic jam" instead of just "traffic jam".

Here, we are interested in characterizing user participation in the dissemination of alerts about traffic. To that end, we collected a dataset of Waze alerts directly from Twitter, since Waze traffic information is not publicly accessible by an API. Our dataset covers the period from December 21st, 2012 to June 28th, 2013, and consists of 212,814 tweets containing alerts about traffic shared by Waze users, each one providing the user id, type of incident (e.g., traffic jam), and the address of the incident. In order to obtain the latitude and longitude of the provided address, we performed a geocoding process using the Bing Maps API[5], which provides the confidence of the result's quality: low, medium, and high. We excluded all results classified as low. After this filtering process, we extracted 162,212 tweets containing alerts, shared by 21,852 users.

In Figure 1, we provide an overview of types of alerts reported by users of our dataset, using word clouds to represent the relative frequency[6]. Alerts were translated into English using a manually created dictionary of translation. As we can see, the most common type of reported alert is traffic jam[7], though police and hazard are also very popular.

Fig. 1. Overview of reported alerts
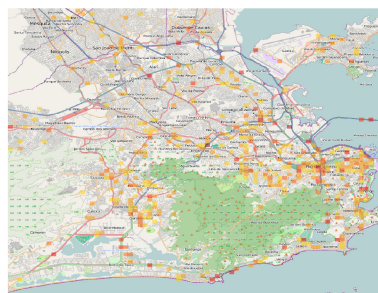
Fig. 2. Spatial coverage of Waze in Rio de Janeiro

---

[5] http://www.microsoft.com/maps/developers/web.aspx

[6] The size of the word indicates its popularity.

[7] Alerts containing a subcategory of an incident were unified to its main category, for example, "heavy traffic jam" was associated to the word "traffic jam".

4.2   Sensing Layer Coverage

In this section, we discuss the spatial coverage of the PSN derived from Waze. Towards a global view of this coverage, we first built a heatmap with all alerts shared by users in our dataset[8], and then we selected the most popular cities for further analysis.

A popular city from our dataset is shown in Figure 2. In this figure we show the number of alerts in different regions of Rio de Janeiro by a heat map, where the scale varies from yellow to red (more intense activity)[9]. The spatial coverage is not as proliferated as the one observed in location and photo sharing systems [1,5]. A factor that might help to explain it is the user population of our dataset, which is smaller than those reported in the mentioned studies. Another factor is that users might have fewer opportunities to share traffic alerts, compared to opportunities to share photos or check-ins.

In order to evaluate user participation across different regions at a finer granularity, we propose to divide the geographical area of each city into smaller rectangular spaces (or quadrants), as in a grid, and analyze the distribution of the number of alerts across quadrants. We here consider a quadrant delimited by steps of $0.0001°$ in both latitude and longitude. This scheme produces quadrants of different spatial areas, depending on the geographic location of the city, but this does not affect our analysis. For instance, it represents an area of $\approx 8 \times 11$ meters in New York and 10×11 meters in Rio de Janeiro.
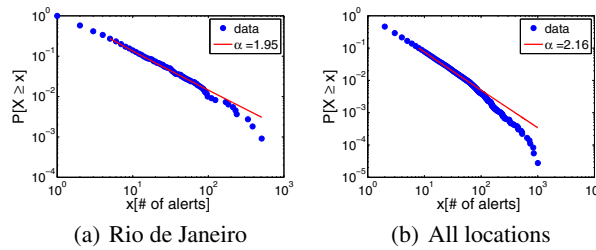

|(a) Rio de Janeiro|(b) All locations|

Fig. 3. Distribution of the number of alerts

The complementary cumulative distribution functions (CCDF) of the number of alerts shared in a quadrant of the city of Rio de Janeiro, as well as across all locations in our dataset are presented in Figures 3a and 3b, respectively. Note that a power law describes well this distribution in both cases. This means that few areas have hundreds of shared alerts, while most of the quadrants have just a small number. This finding is consistent with previous results about user participation in location sharing services [1,10] and photo sharing services [5]. As in those other services, it is likely that some areas, such as large avenues in downtown, have more activity of traffic alerts. Note that the number of vehicles circulating on each region greatly impacts the local coverage of a traffic alert sharing system such as Waze, as shared alerts often refer to traffic jams and hazards, or even police traps (see Figure 1), which tend to occur more often in locations with heavier car flow. This is in contrast to location and photo sharing services, where places often visited by a large number of people are not necessarily covered by a large

---

[8] Figure omitted due to space limitations. We note, however, that user participation in Waze is low is certain regions, particularly Asia.

[9] The darkest red represents a region with 508 alerts.

amount of shared data, because the motivation of users to share data in such systems is different from Waze. For example, a large supermarket may be visited by a large number of people on a daily basis, but it is not likely that those people will share many check-ins or photos at it.
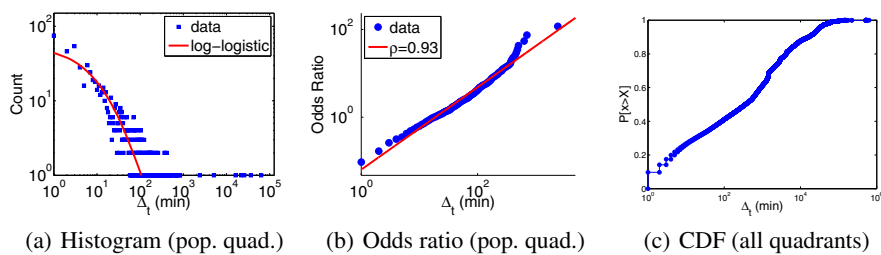
4.3   Time Intervals between Traffic Alerts

We now analyze the frequency in which users share alerts in Waze. This is important because the success of a PSN depends on the continuous participation of the users, since nodes are autonomous and responsible for their own operation and functioning.

The histogram of the inter-sharing time Δ_t between consecutive alerts (performed not necessarily by the same user), in a popular quadrant is shown in Figure 4a. Note that a log-logistic distribution[10] ($\mu = 2.931$, $\sigma = 1.065$) fits well the data, reflecting the fact that there are times when many alerts are shared within a few minutes and there are times when there is no sharing for hours. As also observed for photo [5] and location [1] sharing services, this result may indicate that the majority of alert sharing occurs at specific intervals. For instance, alerts are more likely to be common in urban areas during rush hours.

In Figure 4b, we show the odds ratio function (OR)[11] of inter-sharing time Δ_t. This function highlights the behavior of the distribution at both the head and in the tail. As also observed in previous analyses of phone SMS usage and photo sharing [5,11], the OR function of the inter-sharing time between alerts also presents a power law behavior with slope $\rho \approx 1$. This suggests that the mechanisms behind human activities can be simpler than those proposed in the literature, which depend on a larger number of parameters [12].

The CDF of all observed inter-sharing times performed by any user in the same quadrant is shown in Figure 4c. As we can observe, a considerable portion of users perform consecutive alert sharing in a short time interval. In the present case, the portion is around 25%. This is expected to happen because, for example, when an accident happens many users tend to share it in a short interval.

(a) Histogram (pop. quad.)   (b) Odds ratio (pop. quad.)   (c) CDF (all quadrants)

**Fig. 4.** Time intervals between consecutive alerts, not necessarily done by the same user

---

[10] Probability Density Function: $f(x|\mu, \sigma) = \frac{1}{\sigma} \frac{1}{x} \frac{e^z}{(1+e^z)^2}$; $x \geq 0$, where $z = \frac{log(x)-\mu}{\sigma}$.

[11] $OR(x) = \frac{CDF(x)}{1-CDF(x)}$, where $CDF(x)$ is the cumulative density function, in this case, of the inter-sharing time Δ_t distribution.

4.4   User Activity

We now analyze the contribution of individual users in the PSN derived from Waze. In Figure 5, we show that the distribution of the number of alerts shared by each user of our dataset has a heavy tail, as observed for photo sharing [5]. This implies in a great variability of user participation. For instance, 35% of the users contributed with only one alert during approximately the six-month period covered by our dataset, while 16% and 0.006% of users contributed with more than 10 and 100 alerts, respectively, in the same period. These proportions are similar to those observed in photo sharing [5].

(a) All distances   (b) Median distance per user
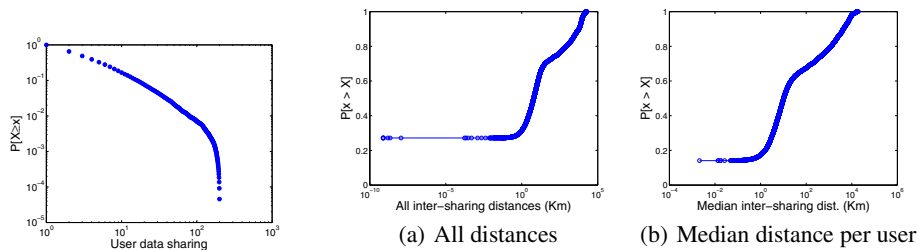
Fig. 5. CCDF of the number of shared alerts (same user)

Fig. 6. Distribution of the geographical distance between consecutive data of the same person

We now analyze the spatial distance between consecutive alerts by the same user, by taking the distance [13] between the geographic coordinates associated with both alerts. In Figure 6a, we show the CDF of the distances between consecutive alerts shared by each user, for all users. Note that a large portion of the distances are very short: for instance, around 30% are below 1 meter. Such large fraction of small distances between consecutive sharings were also observed in photo sharing [5] and, to a lesser extent, location sharing services [10]. In the latter, Noulas et al. [10] observed that 20% of the consecutive sharings by the same user were in locations that were apart from each other by up to 1 km. For photos and alerts, this fraction raises to approximately 45% and 80%, respectively. This suggests that users tend to share multiple alerts in the same location.

In Figure 6b, we show similar results for the distribution of the median distance between consecutive sharings for each user. That is, even aggregating results for each user, we still observe that alerts are shared at very short distances: around 15% of users share alerts 1 meter apart from each other.

4.5   Influence of User Routines

In this section, we study how user routines affect the temporal frequency of alert sharing. In Figure 7a, we show the temporal variations[12] of the number of alerts shared throughout the week (Monday to Sunday), for all locations of our dataset. As expected, user participation presents a diurnal pattern, and the activity during late night hours and dawn is much lower than previously observed in location and photo sharing patterns [1,5]. During that period, traffic problems are typically rare, whereas users have

---

[12] The time of sharing was normalized according to the timezone where the alert was shared.

more opportunities to share data in location and photo sharing systems (e.g., in a night club or in a concert).

Intense user activity during the weekends, as observed in location sharing services [1,14] and photo sharing services [5], is not observed here. This might indicate that the reasons motivating users to contribute alerts are distinct from the ones to perform check-ins. In Figure 7b, we show the average number of data sharings throughout the day, separately for weekdays (Monday to Friday) and weekends (Saturday and Sunday). Note the two clear peaks of activity, one around 7 to 8 AM and the other around 6 PM, coinciding with typical rush hours in urban areas. This result is different from the three clear peaks previously observed in location sharing services [1,14], around breakfast, lunch and dinner times, as well as from the two peaks during lunch and dinner times in photo sharing [5].
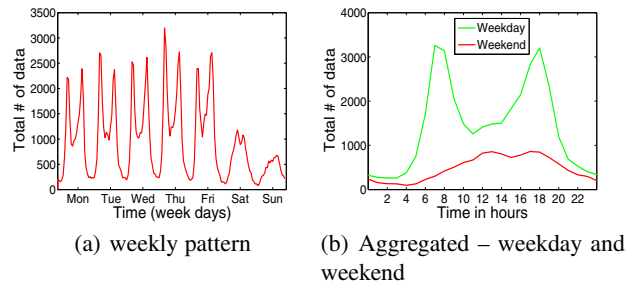

(a) weekly pattern

(b) Aggregated – weekday and weekend

Fig. 7. General temporal sharing pattern (all locations)

We now analyze the hourly variations of alert sharings in six large cities: Chicago and New York (Figure 8a[13]); Belo Horizonte and Sao Paulo in Brazil (Figure 8b); and London, and Paris (Figure 8c). Note that the curve of each city follows the general trend observed for all locations (Figure 7b).
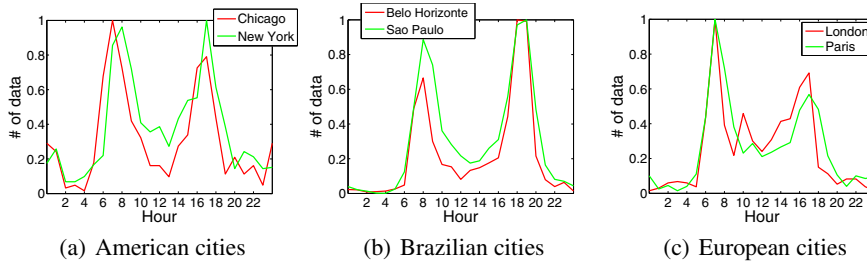

(a) American cities

(b) Brazilian cities

(c) European cities

Fig. 8. Alerts sharing throughout the day in different cities around the world

We can also observe that the peaks reflect distinct rush times that are related to the common working hours of different cities. In Chicago (Figure 8a) the morning peak is around 7 AM, as in the two European cities (Figure 8c). In contrast, in New York and in the Brazilian cities (Figure 8b), the morning peak is usually one hour later, suggesting that people tend to leave later to work in those cities. The second most expressive peak in

---

[13] Each curve is normalized by the maximum number of alerts shared in the city in question.

both American cities is around 5 PM, which is similar to the European cities. However, this is distinct from the Brazilian cities, which have a peak of activity around 6 PM.

To complement this analysis, we performed, from July 16th to July 18th, an hourly collection of traffic conditions of Paris, using Google Maps. We note that the time of the observed peaks reflects relatively well intense traffic conditions reported by Google Maps, whereas the reduced activity prior and after the peaks also reflects better traffic conditions. This suggests that this information could be used to assure the quality and improve traffic condition information services, such as those offered by Google Maps.

4.6 Discussion

We showed the planetary scale of the studied PSN, derived from Waze. We also showed the highly unequal frequency of data sharing, both spatially and temporally, which is highly correlated with the typical routine of people. Our characterization provided a deep understanding of the properties of this particular PSN, revealing its potential to drive various studies on city dynamics and urban social behavior, as discussed next.

5 Opportunities for New Services and Applications

This section discusses some possible situations where a PSN derived from Waze can be exploited to build new services and applications. As discussed in Section 4.1, the most often reported problem by Waze users is traffic jam. Since this is a common cause of complaints and many other problems may end up resulting in traffic jam, a natural question that arises is: What are the causes of traffic jam? This is not an easy question to answer, and it may vary from place to place. However, the shared alerts in Waze might help us to understand the causes.

More specifically, we note that the analysis of the traffic alerts can lead to a more detailed investigation of traffic conditions. For instance, the real-time identification of locations with potholes or animals in the road, whose detection is hard with traditional sensors, becomes more feasible when users participate in the sensing process. This is useful to discover not obvious reasons for a frequent traffic jam. Besides that, such detection opens opportunities for various services, such as, help smart cars in the correct identification of problems on the road.

In the same direction, the identification of problematic roads might also be possible by looking at the number of alerts reported on a road. For example, if we take the top five reported locations in Belo Horizonte (Cristiano Machado Av; Raja Gabaglia Av; Contorno Av; Beltline Rd; Amazonas Av) and Paris (A15; N104; A6 - E15; A13 - E05; N118), we observe that all of them are roads that typically present traffic problems, especially on rush hours. This shows that it is possible to identify problematic roads using traffic alerts. However, the main advantages of using a PSN of traffic alerts do not lie on discovering common problematic roads, but on detecting unusual ones. This is possible thanks to the capability of Waze alerts in describing real-time incidents, what can help to discover particular patterns not generally known.

This information could be used to improve algorithms for navigation services. Besides that, traffic information services, such as Bing Maps, could also benefit from this

information to assure the quality about the reported traffic condition, as we observed in the Section 4.5. Moreover, an urban planner could use this information to assess the effectiveness of previous roadworks. For instance, it could be verified if roadworks in the Cristiano Machado Avenue, a very problematic road in Belo Horizonte, were effective to reduce the number of problems reported in that road.

6   Conclusion and Future Work

The better understanding of dynamics of cities and the social behavior that happens on them can be achieved with the help of participatory sensor networks (PSNs). This understanding allows system designers to offer smart services that meet people's needs. In this paper, we studied a PSN derived from Waze, the most popular traffic report application, and characterized the properties of this network. Among other results, we showed that the routines have a considerable impact on the data sharing, for example, the peaks of activity reflect distinct rush times that are related to the common working hours of the analyzed cities. A future direction we intend to pursue is the design of new applications that explore traffic alerts shared in Waze, following the opportunities presented in Section 5.

References

1. Silva, T.H., Vaz de Melo, P.O.S., Almeida, J.M., Loureiro, A.A.F.: Challenges and opportunities on the large scale study of city dynamics using participatory sensing. In: Proc. of ISCC 2013, Split, Croatia (July 2013)
2. Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M.B.: Participatory sensing. In: Workshop on World-Sensor-Web, WSW 2006 (2006)
3. Goel, V.: Maps That Live and Breathe With Data. The New York Times (June 2013)
4. Cranshaw, J., Schwartz, R., Hong, J.I., Sadeh, N.: The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In: Proc. of ICWSM 2012, Dublin, Ireland (2012)
5. Silva, T.H., Vaz de Melo, P.O.S., Almeida, J.M., Salles, J., Loureiro, A.A.F.: A picture of Instagram is worth more than a thousand words: Workload characterization and application. In: Proc. of DCOSS 2013, Cambridge, USA, pp. 123–132 (May 2013)
6. Quercia, D., Capra, L., Crowcroft, J.: The social world of twitter: Topics, geography, and emotions. In: Proc. of ICWSM 2012, Dublin, Ireland (June 2012)
7. Frias-Martinez, V., Soto, V., Hohwald, H., Frias-Martinez, E.: Characterizing urban landscapes using geolocated tweets. In: Proc. of SocialCom 2012, Washington, USA (2012)
8. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proc. of WWW 2010, Raleigh, USA, 851–860. ACM (2010)
9. Fire, M., Kagan, D., Puzis, R., Rokach, L., Elovici, Y.: Data mining opportunities in geosocial networks for improving road safety. In: Proc. of IEEEI 2012, pp. 1–4 (2012)
10. Noulas, A., Scellato, S., Mascolo, C., Pontil, M.: An Empirical Study of Geographic User Activity Patterns in Foursquare. In: Proc. of ICWSM 2011, Barcelona, Spain (2011)
11. Vaz de Melo, P.O.S., Faloutsos, C., Loureiro, A.A.: Human dynamics in large communication networks. In: Proc. of SDM 2011, Mesa, USA (2011)
12. Malmgren, R.D., Stouffer, D.B., Motter, A.E., Amaral, L.A.N.: A poissonian explanation for heavy tails in e-mail communication. PNAS 105(47), 18153–18158 (2008)
13. Sinnott, R.W.: Virtues of the Haversine. Sky and Telescope 68(2), 159+ (1984)
14. Cheng, Z., Caverlee, J., Lee, K., Sui, D.Z.: Exploring Millions of Footprints in Location Sharing Services. In: Proc. of ICWSM 2011, Barcelona, Spain (2011)