Challenges and opportunities on the large scale study of city dynamics using participatory sensing

Thiago H. Silva, Pedro O. S. Vaz de Melo, Jussara M. Almeida, Antonio A. F. Loureiro
Computer Science, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil
{thiagohs, olmo, jussara, loureiro}@dcc.ufmg.br

***Abstract*—Cities are not identical and evolve over time, and sensing in large scale can be used to capture these differences. Research in Wireless Sensor Networks has provided several tools, techniques and algorithms to solve the problem of sensing in restricted scenarios (e.g., factory). However, sensing large scale areas, such as big cities, brings many challenges and incurs high costs related to system building and management. Thus, sensing those areas becomes more feasible when people collaborate among themselves using their portable devices, and building what has been named *participatory sensing systems*. This work analyzes an emerging type of network derived from this type of system, the Participatory Sensor Network (PSN), where nodes are autonomous mobile entities and the sensing depends on whether they *want* to participate in the sensing process. Based on four datasets of participatory sensing systems (27 million of records), we show that this type of network brings many challenges related to structural problems, e.g. instant coverage very limited, and also because of big data issues, which may restrict the use of this emerging type of network. However, it presents also, as shown here, many advantages and open opportunities, mainly related to large scale study of cities dynamics.**

***Keywords*-Participatory sensing; big data; location sharing services; social networks; smart cities**

I. Introduction

It is known that cities are not identical and evolve over time, and also that habits and routines of inhabitants are typically distinct. Given this, how can we measure those differences in large scale? Certainly, it is possible use sensing to achieve this goal.

Research in wireless sensor networks (WSN) has provided several tools, techniques and algorithms to solve the problem of sensing in restricted scenarios, such as forests or volcanoes [1]. However, sensing extensive areas (e.g., large cities) brings many challenges and incurs high costs associated with system development and management.

Sensing vast areas becomes more feasible when people carrying their portable devices (e.g., smartphones) collect data and collaborate among themselves. Smart phones are taking center stage as the most widely adopted and ubiquitous computing device. They are also increasingly coming with various embedded sensors, such as GPS and accelerometer. Systems that enable sensed data in this way are named participatory sensing systems (PSSs) [2]. In such systems, the shared data is not limited to sensor readings passively generated by the device, but also includes proactive user observations. There are several examples of PSSs already deployed, such as Waze[1], for reporting real-time traffic conditions, and Weddar[2], for reporting weather conditions. Moreover, location sharing services, such as Foursquare[3], can be seen as location categorizing applications, which allow users to share their actual location associated with a specific category of place (e.g., restaurant).

We use the concept of a Participatory Sensor Network (PSN) [2], a network derived from participatory sensing systems, where nodes are autonomous mobile entities (typically users) and the sensing activity depends on whether they want to participate in the sensing process. Although the term PSN has been previously defined [2], an analysis and discussion of this emerging type of network as performed in this paper remains untackled, as far as we know. Many applications and services to support smart cities can potentially benefit from participatory sensor networks. For that it is crucial the understating of their limitations and potential.

The main contributions of this paper are: (1) a characterization and analysis of participatory sensor networks that emerge from four real-world web-scale datasets of location sharing services, encompassing more than 27 million user check-ins; (2) a discussion of challenges when considering PSNs to support applications; (3) the presentation of promising opportunities in the use of PSNs for the large scale study of city dynamics, that could be base for tools for city planners to provide a new means to see the city, or for end users who are looking for new ways to explore the city.

The rest of this paper is organized as follows. Section II discusses the related work. Section III discusses the participation of humans in the sensing process, covering particularities of participatory sensing systems and participatory sensor networks. Section IV analyzes the main characteristics of PSNs derived from popular online sharing services, and discusses the challenges that emerge from them. We then present concrete opportunities that explore PSNs for the large scale study of city dynamics in Section V. Section VI concludes the paper.

II. Related work

Often, humans are the target in the sensing process [3], or responsible for data sharing [4]. We here focus on the latter case. More specifically, we consider systems that utilize everyday mobile devices, for instance smartphones, to form

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

interactive participatory sensor networks (see Section III-B). Participatory sensing is related to crowdsourced projects [4], but differs from studies such as [5] since they do not require participation of individuals carrying mobile devices to sense the environment and make observations at a personal level.

The term *participatory sensing systems (PSS)* has been used to define systems that enable the contribution of sensed data by users, including traffic monitoring [6] and noise level monitoring [7] applications. Other studies focus on issues of participatory sensing systems, such as the sustainability of contribution [8].

Location sharing services have been used to study human mobility pattern as well as social relations [9]–[11]. Other studies use location sharing services to understand better cities dynamics. In this direction, Cranshaw et al. [12] present a model to extract distinct regions of a city according to current collective activity patterns. Similarly, Noulas et al. [13] propose an approach to classify areas and users of a city by using venues' categories of Foursquare. In [14] we propose a new technique to visualize the dynamics of cities based on habits and routines of people collected from check-ins on Foursquare. This present work differs from our previous work [14], because the focus of that work was the presentation of an specific technique. In another previous work [15], we consider location sharing services as a participatory sensing system, presenting initial characteristics of a participatory sensor network derived from Brightkite and Gowalla. This present work also differs from [15], because besides presenting more characteristics of Brightkite and Gowalla as a PSN, we also present characteristics from two different PSNs derived from Foursquare. This enabled, for example, the verification of similar properties among all those networks. Unlike other previous studies, we discuss the challenges and implications when dealing with participatory sensor networks. The present work also envision the presentation of opportunities of research in this direction.

III. Humans in the Sensing Process

We here focus on systems where humans are responsible for data sharing, particularly participatory sensing systems.

A. Participatory Sensing Systems

The concept of participatory sensing systems originally considers that the shared data is generated automatically, or passively, by sensor readings from portable devices [2]. However, in this work we also consider manually, or proactively, user-generated data, that has been called ubiquitous crowdsourcing [16]. The popularity of participatory sensing systems grew rapidly with the widespread adoption of sensor-embedded and Internet-enabled cell phones. These devices have become a powerful platform that encompasses sensing, computing and communication capabilities, being able to generate both manual and pre-programmed data.

Location sharing services, such as Gowalla and Foursquare, are examples of participatory sensing systems. The sensed data is a check-in of a particular place that indicates, for instance, a restaurant in a specific location, and also a signal from a user expressing his/her preference. In the rest of this work we will use the term "check-in" to refer to an event when time and location of a particular user is recorded or, in a participatory sensor network context, sensed.

TABLE I
DATASET INFORMATION

System	# of check-ins	Interval	# of Venues	Cat.
Foursquare-Year	11,743,781	2/2010 - 1/2011	490,079	no
Foursquare-Crawled	4,672,841	4/2012 (1 week)	1,929,237	yes
Gowalla	6,442,890	2/2009 - 10/2010	1,280,969	no
Brightkite	4,491,143	4/2008 - 10/2010	772,966	no
Total	**27,350,655**			

B. Participatory Sensor Networks

PSN is a simple concept, which has user's portable device as a fundamental building block. Individuals carrying these devices are able to sense the environment and to make relevant observations at a personal level. Thus, each node in a PSN consists of the user plus his/her mobile device. Nodes send data/information to participatory sensing systems, which can be crawled throughout services APIs. In this present work we consider a participatory sensor network from location sharing services. More details about PSNs can be found in [15]. Despite PSNs be a simple concept it has many challenges, and they are presented in Section IV.

IV. Characterization of PSNs derived from online sharing services

In this section we analyze the characteristics of participatory sensor networks (PSNs) derived from three location sharing services, namely Foursquare, Gowalla and Brightkite.

A. Data Description

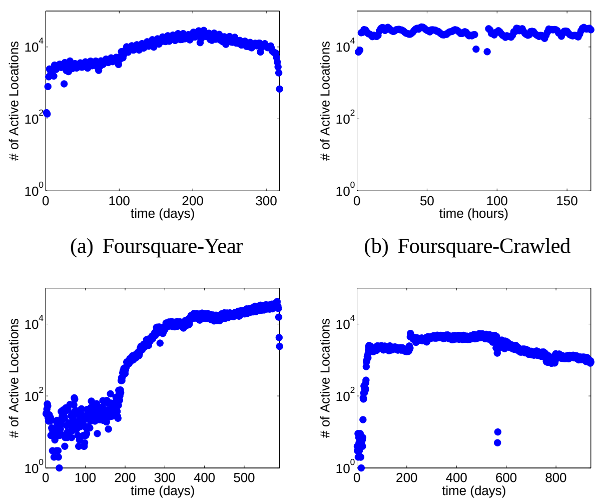
We analyze data from four real-world web-scale datasets collected from three systems. Three of these datasets, one for each system, are publicly available [9], [10]. The fourth was collected by us because we are interested in an extra piece of information: the categories of the venues people have checked-in. We collected this dataset directly from Twitter[4], since Foursquare check-ins are not publicly available by default. Approximately 4.7 million tweets containing check-ins were extracted from Twitter, each one providing a URL to the Foursquare website, where information about the geographic location of the venue was acquired. This new dataset is useful for some of the opportunities presented in Section V. To differentiate the two datasets from Foursquare we refer to the one obtained from [9] as **Foursquare-Year**, and to the one we crawled as **Foursquare-Crawled**.

In all four datasets, each check-in contains the venue's id, its latitude and longitude, and the time the check-in was performed. As we mentioned, our collected dataset (Foursquare-Crawled) also includes the venue category. Table I summarizes the four datasets.

[4]http://www.twitter.com

B. Coverage of the Network

Given the planetary scale coverage of the PSN in the long term observed our previous study [15], now we ask: what places are effectively covered in the short term? In Figure 1 we answer this question by showing the number of venues that were active in a given time interval. The Foursquare-Year, Foursquare-Crawled, Gowalla, and Brightkite datasets have, respectively, approximately 490 thousands, 1,9 million, 1,3 million, and 773 thousands distinct venues. Considering the total number of distinct venues in each dataset, we find that the maximum number of active venues per day, or per hour, for Foursquare-Crawled, corresponds to only 6%[5], 2%, 3.3% and 0.7% of the total for Foursquare-Year, Foursquare-Crawled, Gowalla and Brightkite, respectively. This indicates that, despite the long-term global scale coverage, the instant coverage observed is very limited considering all locations they can reach, i.e., the probability of a random location be active in a given day is very small.

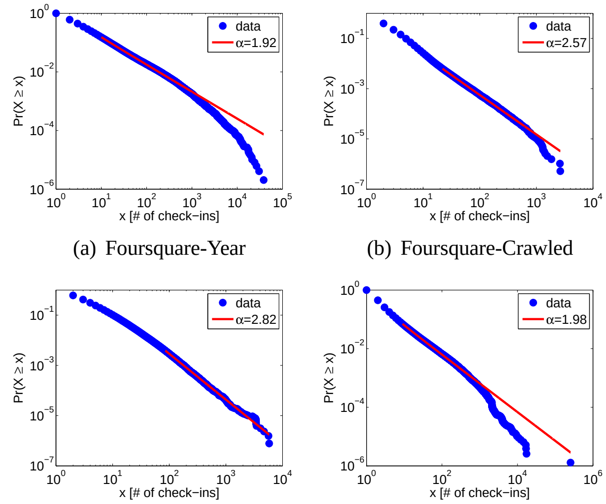


(a) Foursquare-Year (b) Foursquare-Crawled (c) Gowalla (d) Brightkite

Fig. 1. The average percentage of locations that were active in a given day

In fact, observe in Figure 2 the complementary cumulative distribution function (CCDF) of the number of check-ins per venue. It is not surprising that a power law fitting is appropriate to explain this distribution, where the majority of locations have only a handful of check-ins, while there are few locations with hundreds of them. Since we are analyzing location sharing systems, it is natural that some locations are shared more than others. For example, restaurant or pub venues are more likely to be shared than a post or bank offices, despite the fact that these places are often visited as well.

We now turn our attention to four large and populous cities located in four continents: New York City (USA), Tokyo (Japan), Sydney (Australia), and Cairo (Egypt). Figure 3 shows, for each city, the heat map of the sensing activity



(a) Foursquare-Year (b) Foursquare-Crawled (c) Gowalla (d) Brightkite

Fig. 2. The complementary cumulative distribution function of the number of check-ins per venue

in these cities, where the darker the color[6], the higher is the number of check-ins in that area. First, observe that for the cities of the first line (New York, and Tokyo) the PSN is able to practically cover the whole territory. However, for the cities of the second column (Sydney and Cairo), despite the fact they have approximately the same population of the ones of the first column, the PSN coverage is significantly lower.

These differences can be explained by several reasons. Cairo, for instance, where the coverage was the lowest (about only 10% of the whole territory), has significant cultural differences compared to the other analyzed cities. This might have a significant impact on the adoption and use of location sharing systems. Moreover, we see that the coverage in Sydney is very skewed, being not as homogeneous as in, Tokyo and New York. This is probably because the geographic aspects, i.e., large green and water areas, which limit the sensing coverage. Moreover, this city has large residential areas with few commercial venues, what also contribute for a low sensing rate. All these aspects should be carefully considered when designing participatory-based sensing applications.

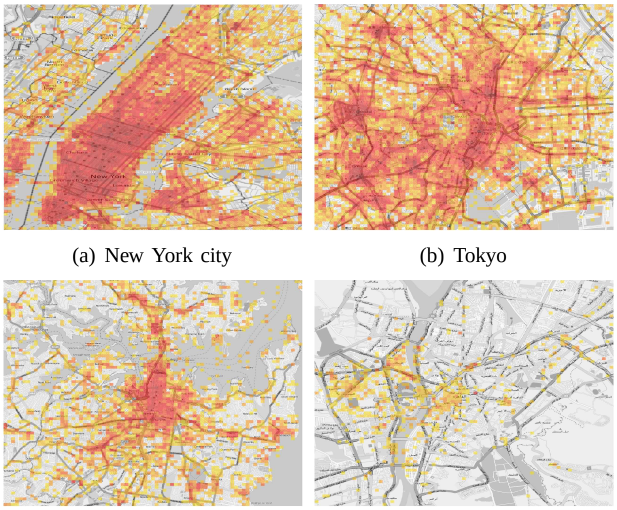
C. Seasonal Behavior of Humans

First, we investigate the frequency that users perform data sharing. Figures 4-a, 4-b, 4-c, and 4-d show the histograms of the inter-event times Δ_t between consecutive check-ins of one popular venue from the four analyzed datasets. Observe the large number of check-ins separated by a few minutes and also consecutive check-ins separated by several days. This may suggest that most of the data sharing, even in these particular popular places, happen in specific intervals of time, probably related to the human circadian rhythm, e.g., in restaurants people check-in for lunch and dinner mostly.

We now analyze how the seasonal behavior of humans affects the data sharing. Figure 5 shows the weekly location

[5] The percentage is the number of venues in the area that had at least one check-in as a fraction of total venues in the area.

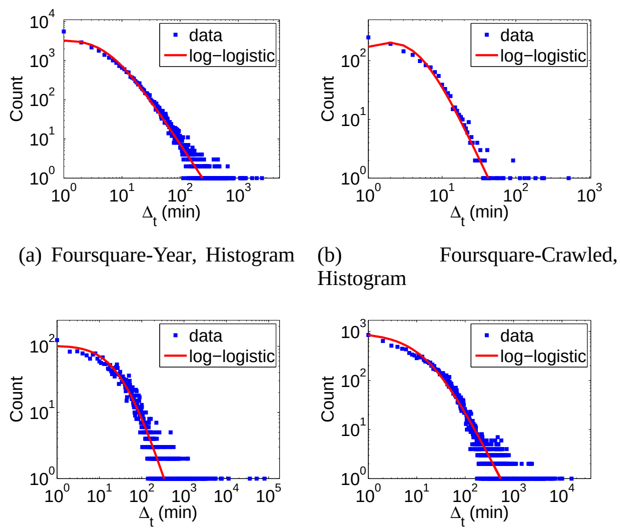
[6] Colors of the heat map for all subfigures are in the same scale.



(a) New York city (b) Tokyo

(c) Sydney (d) Cairo

Fig. 3. All sensed locations in six international cities (Foursquare datasets). The number of check-ins in each area is represented by a heat map. The color range from yellow to red (high intensity).

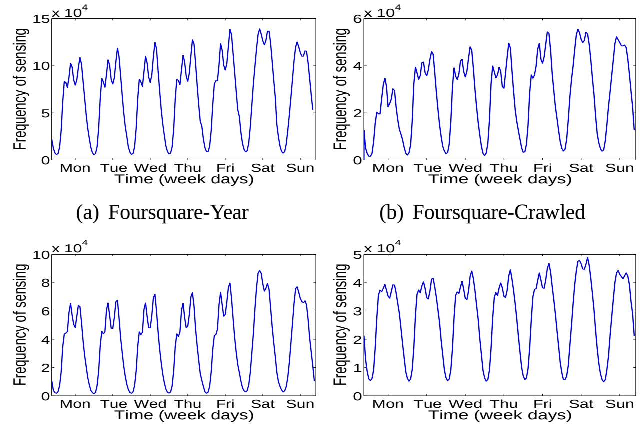


(a) Foursquare-Year, Histogram (b) Foursquare-Crawled, Histogram

(c) Gowalla, Histogram (d) Brightkite, Histogram

Fig. 4. The distribution of the inter-event times between consecutive check-ins of one popular venue of each dataset

sharing pattern for all analyzed datasets[7]. As expected, the network actuation presents a diurnal pattern, meaning that during the dawn the sensing activity is very low. We can also observe that there are two classes of behavior: weekdays and weekends. Considering weekdays, we can note, in all datasets, an increase in the activity from Monday to Friday. It is also possible to observe three peaks during the day, around breakfast, lunch, and dinner times. These peaks occur on every weekday, and are less evident on Friday during breakfast time. This might be due to specific behavior patterns, e.g., going out on Thursday night and waking up late on Friday morning. Observe the potential to study urban social behavior.

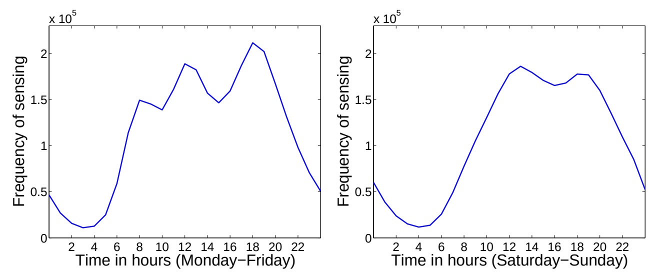


(a) Foursquare-Year (b) Foursquare-Crawled

(c) Gowalla (d) Brightkite

Fig. 5. Weekly location sharing patterns

We further analyze the different behavioral patterns on weekdays and weekends, focusing now on the Foursquare-Year dataset[8]. Figure 6-a shows the average number of check-ins of each hour from Monday to Friday. Figure 6b shows the same information for Saturday and Sunday. As we observe, the peaks during weekdays happens on 8:00AM (breakfast), 12:00PM (lunch), and 6:00PM (dinner). On weekends, there is no peak activity in the morning, the lunch peak happens around 1:00PM, and the dinner peak is flatter (comprising 6:00PM to 7:00PM). We can also observe that the activity is more intense on weekends.



(a) Monday to Friday (b) Saturday and Sunday

Fig. 6. Weekdays and weekend location sharing patterns

D. Discussion of Challenges

We now discuss the challenging issues when dealing with PSNs. By the characterization we can summarize the most important structural issues found in:

- Instant coverage is very limited;
- very unequal distribution of check-ins in venues;
- coverage is not comprehensive in all cities;
- long periods of inactivity.

All these issues suggest that not every type of applications is suitable to be built in PSNs. For example, PSN to monitor

[7] The server timestamps where converted to the local time of the check-in.

[8] Similar characteristics were observed for all the other analyzed systems.

city problems, such as noise and air pollution, or potholes on the streets might follow the same structural challenges (because a fundamental piece still a human, with his/routines and preferences), which might not meet the requirements of coverage and frequency of sensing demanded by such applications. Thus, if one wishes to design a participatory-based application with a more comprehensive contribution per area, one should incentive users to participate in places that usually they would not. A punctuation or reward system is one of many types of incentive that might work in this case.

PSNs are very scalable because their nodes are autonomous, i.e., users are fully responsible for their own functioning. Since the cost of the network infrastructure is distributed among the participants, this enormous scalability and coverage are achieved without significant costs. The key challenge to the success of this type of network is to have sustained and high quality participation. In other words, the sensing is efficient as long as users are kept motivated to share their resources, sensing data frequently.

Besides these structural problems, the construction of applications in this type of network face challenges also in data quality, data collection, data storage, data processing and indexing. The quality of the shared data is a challenge that has been relatively well tackled in the web domain, however there are unique challenges for controlling the quality of shared data when dealing with ubiquitous user contributions [16]. For instance, since users can produce sensor readings with relatively little effort, data integrity is not always guaranteed [17].

Data collection is a challenging issue especially when is used web-based services from third-parties, such as Foursquare or Waze. By default, data shared in those systems are usually private, unless users decide to make them public somehow, for example sharing it on Twitter. This means that, in this case, no public data can be available at all.

Another important issue is how to deal with a huge volume of data that location-based systems can offer, because it tend to be large and complex being difficult to process and index using traditional database management tools or data processing applications. This imposes challenging issues to offer real-time services using a PSN.

V. Opportunities for the Study of City Dynamics

In Section IV, we show that all PSNs studied share common characteristics, which may restrict some kind of application. We now present some promising opportunities for the large scale study of cities dynamics using the studied PSNs, considering all the aforementioned characteristics[9].

Unlike traditional mobile Wireless Sensor Networks, the nodes in a PSN move according to their routines or local preferences, which generate the skewed behavior in the sensing activity that we have seen so far. This may be a problem for certain kinds of applications, but it is not the case for those that want to capture city dynamics, especially the ones that do not present severe timing constraints. In this direction we present several opportunities, grouped in two categories: *Area semantics*; and *transitions in the city*.

Area Semantics: Semantic location services will be critical for the next wave of killer applications [18], and there are many opportunities to design them. The opportunities listed here exploit the information about category of the venues[10] present in the Foursquare-Crawled dataset.

Together with geographic neighborhoods, cities can be divided into semantic neighborhoods. To illustrate this idea, consider Figure 7-a. This figure shows a heat map for two categories of venues: Arts & Entertainment, ranging from yellow to red, and Great Outdoors, ranging from light to dark blue. Again, darker colors represent larger numbers of check-ins. Note that it is possible to distinguish popular areas of venues related to the Arts & Entertainment and Great Outdoors categories. Using simple clustering algorithms to classify these regions, such as the one in [12], it may be possible to offer to a tourist, for instance, an intuitive and automatic visualization of the points of interest in a given city.

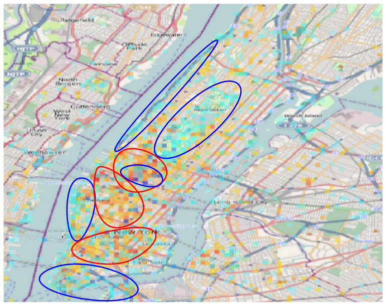
Moreover, datasets collected from application accessed mostly by smartphones represent the social network topology and dynamics of entire cities, enabling the analysis of the social, economic, and cultural aspects of particular areas. For instance, regions that provide a small amount of data to the PSN might indicate a lack of technology access by the population, since the frequent use of location sharing services often relies on smartphones and 3G or 4G data plans, which, usually, are expensive. The preliminary results in the use of PSN in these scenarios demonstrate good opportunities to enable the visualization of interesting facts, some of them discussed in Section IV-B. For instance, analyzing carefully the data for the particular case of Rio de Janeiro, illustrated in Figure 7-b, we observe that it is common to find very poor areas next to wealthy ones. Note the small sensing activity in the circle areas indicated as poor. This information may be useful to guide better public politics in those areas. The same information can be obtained using traditional methods, such as surveys, but in this new way we may be able to obtain the same results more quickly and cheaply.

Other opportunities to classify areas emerge when jointly considering the time and venue where the check-ins are performed. It may be possible to visualize crowds in a city in near real-time. Besides that, we observed in Section IV-C that the seasonal patterns may be due to the circadian rhythm present in human routines. This seasonality has a great potential for prediction applications, since it is very likely that people repeat their activities in a periodic manner. We do believe that there are fruitful opportunities for prediction given by the circadian rhythm of people, enabling the prediction, for instance, of how crowded a place will be. This type of information is valuable in many scenarios, such as services for smart cities to avoid traffic in certain areas and offer alternative routes for users.

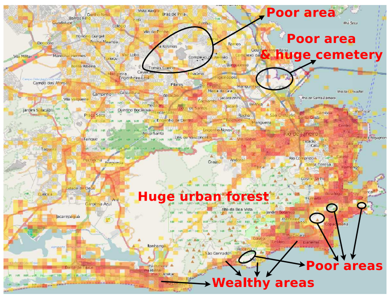
The following scenario illustrates another opportunity that exploits the same data. For that, we created a simple method

[9]For simplicity, we consider only the Foursquare-Crawled dataset.

[10]A complete list with examples is available in [14].



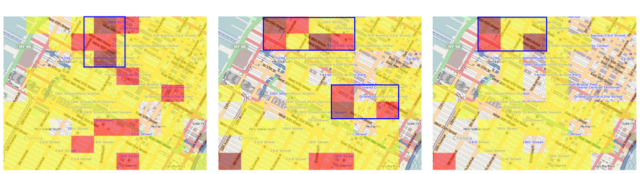
(a) Classification by categories



(b)Classification by lack of sensing

Fig. 7. Examples of possible area classifications

to estimate the number of check-ins in certain time and space. This method average the number of check-ins for the area of interest at a given time, taking into account every category separated. Figures 8-a, 8-b, and 8-c show the check-ins estimation for "Food" places at 7:00PM, "Nightlife" category at 11:00PM, and "Nightlife" category at 1:00AM for the same area respectively. Consider that Bob and Alice have tickets to watch their favorite rock band at Madison Square Garden, located in the area depicted, on Saturday at 8:30 PM to 10:30 PM. They want to have dinner in a popular place before the concert, and after that go clubbing nearby the arena. Since they do not know New York, they decide to use the information provided by an imaginary application represented on Figures 8-a, 8-b, and 8-c. A candidate area to have dinner is marked by a blue rectangle in the Figure 8-a. Regarding to where to go clubbing after the concert, the result shown in Figure 8-b indicates at least two potentially good areas (blue rectangles). Since the couple plan to club until late at night, a tiebreaker criterion could be the estimation of the number of check-ins late at night for the same category, as shown in Figure 8-c. The result indicates one of the two areas as the best choice.



(a) Food, Sat. 7PM (b) Nightlife, Sat. 11PM (c)Nightlife, Sun. 1AM

Fig. 8. Check-ins estimation for different times and type of places

Transitions in the City: We present now another range of opportunities that rise from transition graphs. The location transition graph maps the movements of users in a PSN. It is a directed weighted graph $G(V, E)$, where a node $v_i \in V$ is a specific location (e.g., Eiffel Tower) and a direct edge $(i, j) \in E$ marks a transition between locations. That is, an edge exists from node v_i to node v_j if at least one user performed a check-in in the location represented by v_j just after performing a check-in in the location represented by v_i. The weight $w(i, j)$ of an edge is the total number of transitions that occurred from v_i to v_j.

It is considered here the following requirements for a transition to exist. First, the check-ins must be performed consecutively and by the same individual. Moreover, they must occur in the same "social day". We define a "social day" as the 24-hour interval starting as 6:00AM (instead of 0:00AM), since we are interested in capturing the nightlife transitions as well. This type of transition graph is powerful, because it helps to identify the major flows of people in a city (e.g., people leaving from a restaurant to a club) and hub venues, i.e., venues that receive people coming from or going to diverse areas of the city. We illustrate the potential of this opportunity by building the location transition graphs for large big cities - New York (U.S.) and Tokyo (Japan) - considering only check-ins performed on weekdays (i.e., from Monday 6:00AM to Friday 6:00PM[11]). A transition occurring between two "social days" will still be considered if the time interval between the check-ins is under 4 hours. We tested several different policies for characterizing transitions and the results are very similar, since only a small percentage of transitions are discarded/considered when we vary the policy.

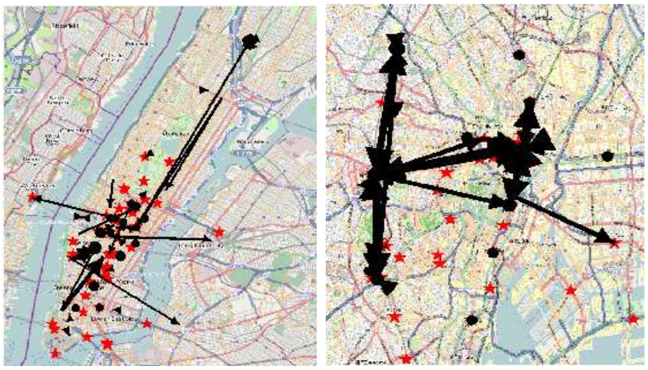
Figure 9 shows heavy weighted edges and hub nodes (top 50 edge weights and node degrees) for NY and Tokyo. Red stars represent the hubs, black arrows represent the edges, and black circles represent self-loops. The larger[12] the symbol, the larger the value. Note that the city flow is very concentrated and skewed, as expected, with a small fraction of the city areas having most of the heavy weighted edges and hubs. Note also that, cities that are known for their fast public transportation systems, as those analyzed, favor the existence of some long distance heavy weighted edges along the public transport links.

The scheme shown in Figure 9 may be used to support various other applications. Consider, for example, a public information dissemination scheme, such as the public displays, which usually show traditional advertising (e.g., short commercials or fast news). If one knows where the city hubs are, he/she could strategically put these displays in these locations. Besides that, if one verifies an unusual and constant flow of people between two independent business venues in a city, the owners could sign a commercial agreement to increase their revenues by, for instance, advertising each other's businesses.

Transition graphs are useful also to display a visualization of a city based on the transitions that are likely to occur,

[11]Friday nights are usually more similar to Saturday nights than other weekday night.

[12]Numbers grow in logarithmic scale.



(a) New York City (b) Tokyo

Fig. 9. Top 50 edge weights and node degrees (hubs) for 2 cities. Red stars represent hubs, black arrows edges, and black circles self-loops. The larger the symbol, the higher the value.

as the demonstrated by the City Image technique presented in our previous work [14]. In this technique the transition graph is slightly modified, where the graph $G(V, E)$, have nodes $v_i \in V$ as the main categories of the locations (e.g. "Food"), and an edge (i, j) exists from node v_i to node v_j if at some point in time a user performed a check-in in a location categorized by v_j just after performing a check-in in a location categorized by v_i. The City Image technique builds a square matrix that summarizes the city dynamics relying on this transiting graph, as well as in a random graph that simulates a random walk for individuals (for details see [14]). The city image can be expanded to consider sub-categories instead of main categories. Since PSN data is highly skewed, few top transitions between sub-categories should be good indicators of the city dynamics. This technique could be useful as a way to measure the distance between two cities, enabling cities comparison and clustering worldwide.

VI. Conclusions and Future Work

In this paper we characterize, discuss challenges and demonstrate opportunities of participatory sensor networks (PSNs), an emerging type of network comprised of autonomous mobile entities with sensing capability. Unlike in wireless sensor networks, the sensing process in PSNs depends on whether nodes want to participate. Using four different large scale datasets, we analyzed the main characteristics of PSNs derived from three location sharing services, namely Foursquare, Gowalla and Brightkite. Our analysis pointed out several challenges of this emerging type of network, which may restrict its use, but also showed that there are good opportunities. In particular, we demonstrate a range of fruitful opportunities that emerge when using PSNs to the large scale study of city dynamics.

We envision two main directions of future work: extend our analysis to include other types of participatory sensing systems to build a more accurate mapping of the city dynamics; and build applications and services for smart cities exploring some of the opportunities presented here, such as traffic monitoring, information dissemination and recommendation systems.

Acknowledgment

This work is partially supported by the INCT-Web (MCT/CNPq grant 57.3871/2008-6), and by the authors individual grants and scholarships from CNPq, CAPES (scholarship 7356/12-9), and FAPEMIG.

References

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292 – 2330, 2008.

[2] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," in *Workshop on World-Sensor-Web (WSW'06)*, Boulder, Colorado, USA, 2006, pp. 117–134.

[3] E. C. Larson, T. Lee, S. Liu, M. Rosenfeld, and S. N. Patel, "Accurate and privacy preserving cough sensing using a low-cost microphone," in *Proc. of the 13th int. conf. on Ubiq. comp. (UbiComp'11)*. Beijing, China: ACM, 2011, pp. 375–384.

[4] M. Goodchild, "Citizens as sensors: The world of volunteered geography," *GeoJournal*, vol. 69, no. 4, pp. 211–221, 2007.

[5] D. Hristova, A. Mashhadi, G. Quattrone, and L. Capra, "Mapping community engagement with urban crowd-sourcing," in *Proc. When the City Meets the Citizen Workshop (WCMCW)*, Dublin, Ireland, Jun. 2012.

[6] S. B. Eisenman, E. Miluzzo, N. D. Lane, R. A. Peterson, G.-S. Ahn, and A. T. Campbell, "Bikenet: A mobile sensing system for cyclist experience mapping," *ACM Trans. Sen. Netw.*, vol. 6, no. 1, pp. 6:1–6:39, Jan. 2010.

[7] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Ear-phone: an end-to-end participatory urban noise mapping system," in *Proc. of the 9th ACM/IEEE Int. Conf. on Infor. Proc. in Sensor Networks*, ser. IPSN '10. Stockholm, Sweden: ACM, 2010, pp. 105–116.

[8] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava, "Examining micro-payments for participatory sensing data collections," in *Proc. of the 12th ACM int. conf. on Ubiquitous computing (Ubicomp '10)*. Copenhagen, Denmark: ACM, 2010, pp. 33–36.

[9] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring Millions of Footprints in Location Sharing Services," in *Proc. of the Fifth Int'l Conf. on Weblogs and Social Media (ICWSM'11)*, Barcelona, Spain, 2011.

[10] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proc. of the 17th ACM Int. Conf. on Know. Disc. and Data Min. (KDD'11)*, San Diego, California, USA, 2011, pp. 1082–1090.

[11] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo, "Socio-spatial Properties of Online Location-based Social Networks," in *Proc. of the Fifth Int'l Conf. on Weblogs and Social Media (ICWSM'11)*, Barcelona, Spain, 2011.

[12] 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. of the Sixth Int. Conf. on Weblogs and Social Media*, 2012.

[13] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-based Social Networks," in *Proc. of the Fifth Int'l Conf. on Weblogs and Social Media (ICWSM'11)*, 2011.

[14] T. H. Silva, P. O. S. Vaz de Melo, J. M. Almeida, and A. A. F. Loureiro, "Visualizing the invisible image of cities ," in *Proc. of IEEE Int. Conf. on Cyber, Phy. and Social Comp. (CPScom'12)*, Besancon, France, Nov. 2012.

[15] ——, "Uncovering Properties in Participatory Sensor Networks," in *Proc. of the 4rd ACM Int. Work. on Hot Top. in Planet-scale Meas. (HotPlanet'12)*, Lake District, UK, June 2012.

[16] A. J. Mashhadi and L. Capra, "Quality Control for Real-time Ubiquitous Crowdsourcing," in *Proc. of the 2nd Int. Workshop on Ubiquitous Crowdsouring (UbiCrowd'11)*, Beijing, China, 2011, pp. 5–8.

[17] S. Saroiu and A. Wolman, "I am a sensor, and i approve this message," in *Proc. of the 11th Work. on Mobile Comp. Systems and App.*, ser. HotMobile '10. Annapolis, Maryland: ACM, 2010, pp. 37–42.

[18] D. H. Kim, K. Han, and D. Estrin, "Employing user feedback for semantic location services," in *Proc. of the 13th int. conf. on Ubiq. comp. (UbiComp'11)*. Beijing, China: ACM, 2011, pp. 217–226.