Social Media as a Source of Sensing to Study City Dynamics and Urban Social Behavior: Approaches, Models, and Opportunities

Thiago H. Silva, Pedro Olmo S. Vaz de Melo, Jussara M. Almeida, and Antonio A.F. Loureiro

Federal University of Minas Gerais
Department of Computer Science
Belo Horizonte, MG, Brazil
{thiagohs,olmo,jussara,loureiro}@dcc.ufmg.br

Abstract. In order to achieve the concept of ubiquitous computing, popularized by Mark Weiser, is necessary to sense the environment. One alternative is use traditional wireless sensor networks (WSNs). However, WSNs have their limitations, for instance in the sensing of large areas, such as metropolises, because it incurs in high costs to build and maintain such networks. The ubiquity of smart phones associated with the adoption of social media websites, forming what is called participatory sensing systems (PSSs), enables unprecedented opportunities to sense the environment. Particularly, the data sensed by PSSs is very interesting to study city dynamics and urban social behavior. The goal of this work is to survey approaches and models applied to PSSs data aiming the study city dynamics and urban social behavior. Besides that it is also an objective of this work discuss some of the challenges and opportunities when using social media as a source of sensing.

1 Introduction

At the beginning, there were mainframes, shared by a lot of people. Then came the personal computing era, when a person and a machine have a close relationship with each other. Nowadays we are witnessing the beginning of the ubiquitous computing (ubicomp) era, when technology recedes into the background of our lives [1, 2].

Mark Weiser, in his classical article entitled "The computer for the 21st century", that appeared in the Scientific American magazine [3], popularized the concept of ubiquitous computing, which envisions the availability of a computing environment for anyone, anywhere, and at any time. It may involve many wirelessly interconnected devices, not just traditional computers, such as desktops or laptops, but may also include all sorts of objects and entities such as pens, mugs, phones, shoes, and many others.

Although this is not the reality yet, much has been done in this direction in the past 20 years after the publication of Weiser's seminal paper, and the key ingredients are evolving in a favorable direction for it. Observe, for example, the increasing number and popularization of numerous types of portable devices.

A fundamental step to achieve Weiser's vision is to sense the environment. The research in wireless sensor networks (WSNs) has provided several tools, techniques and

M. Atzmueller et al. (Eds.): MUSE/MSM 2012, LNAI 8329, pp. 63–87, 2013.

algorithms to solve the problem of sensing in limited size areas, such as forests or factories [4, 5]. However, traditional WSNs have their limitations, such as the high costs related to achieve very large coverage spaces, such as metropolises size areas. Consider the challenges to build and maintain such networks.

In this direction, smart phones are taking center stage as the most widely adopted and ubiquitous computing device [2]. It is also worth noting that smart phones are increasingly coming with a rich set of embedded sensors, such as GPS, accelerometer, microphone, camera, gyroscope and digital compass [6].

Social media websites such as Foursquare[1], Instagram[2], Flickr[3], Twitter[4], Waze[5], and Weddar[6] have started to create new virtual environments that integrates the user interactions and, probably because of that, are becoming very popular. Figure 1 illustrates the popularity of social media use by showing what happens on the Internet at every sixty seconds. For instance, we can see that more than 6,600 pictures are uploaded on Flickr and 320 new accounts and 98,000 tweets are generated on Twitter every minute. Besides that, Foursquare, created in 2009, registered 5 million users in December 2010, 10 million users in June 2011, and 20 million users in April 2012 [7].

The ubiquity of smartphones, associated with the adoption of social media websites, enables unprecedented opportunities to study city dynamics and urban social behavior by analyzing the data generated by users. In this way, we can consider social media as a source of sensing, each one providing different types of data. In Waze, users report traffic conditions, in Weddar, users report weather condition. Location sharing services, such as Foursquare, allows users to share their actual location associated with a specific category of place (e.g., restaurant). This enables the study of human behavioral patterns, such as mobility, and also the study of the semantic meaning of places in the city. Social media systems that allow people connected to the Internet to provide useful information about the context in which they are inserted at any given moment, as those cited above, are called participatory sensing systems (PSSs) [8, 9].

Indeed, PSSs have the potential to complement WSNs in many aspects. As WSNs are typically designed to sense areas of limited size, such as forests and factories, PSSs can reach areas of varying size and scale, such as large cities, countries or even the entire planet [9]. Furthermore, WSNs are subject to failure, since their operations depend on proper coordination of actions of their sensor nodes, which have severe hardware and software restrictions. On the other hand, PSSs are formed by independent and autonomous entities, i.e., humans, which make the task of sensing highly resilient to individual failures. The success of PSSs is directly connected to the popularization of the *smartphones* and social media.

The goal of this work is to present the state of the art of the use of participatory sensing systems to study city dynamics and urban social behavior. It surveys approaches and models applied to generate context (see Section 2.3 for the definition) from big raw

1 `http://www.foursquare.com`
2 `http://www.instagram.com`
3 `http://www.flickr.com`
4 `http://www.twitter.com`
5 `http://www.waze.com`
6 `http://www.weddar.com`

Fig. 1. Things that happen on Internet every sixty seconds. Infographic by - Shanghai Web Designers (http://www.go-globe.com/web-design-shanghai.php).

data obtained by participatory sensing systems. It is worth mentioning that it is not our objective to make an exhaustive survey in this subject. Instead, we discuss a compilation of studies that represent recurrent themes addressed by researchers nowadays. For that, we identified five classes of studies and we named them as: (1) mobility patterns; (2) understanding cities; (3) social patterns; (4) event detection; and (5) human behavior. For each class we highlight the approaches and models applied to create new knowledge and semantic meaning from the big raw data. Besides that we also discuss some of the challenges and opportunities when using social media as a source of sensing.

The rest of this chapter is organized as follows. Section 2 discusses the concept of ubiquitous computing, presenting its definition (Section 2.1), discussing its current state (Section 2.2) and also presenting the concept of context aware computing (Section 2.3), which is a central piece of ubicomp. Section 3 discusses the participation of humans in the sensing process, covering particularities of participatory sensing systems and participatory sensor networks. Section 4 presents the approaches and models used to deal with social media as a sensor, for each one of the five classes of studies considered. Section 5 and Section 6 discuss the challenges and opportunities that emerge when dealing with social media as a source of sensing, respectively. Finally, Section 7 presents the final remarks.

2 Ubiquitous Computing

Modern computing can be divided in three eras. The first is characterized by one single computer (mainframe) owned by an organization and used by many people concurrently. In the second era, a personal computer (PC) is usually owned and used by a single person. In the third era, ubiquitous computing (ubicomp), each person owns and uses many computers, especially small networked portable devices such as smart phones and tablets[1, 2].

Ubiquitous computing is related to mobile computing, although they are not the same thing, neither a superset nor a subset [10] of each other. Mobile computing devices are not mere personal organizers. They are devices (computers with processing power) that contemplate a new paradigm: mobility. Mobility has some constraints, such as finite energy sources. This paradigm is changing the way we work, communicate, have fun, study and do other activities while we are moving [11]. The fact is that ubiquitous computing must support mobility, since motion is an integral part of everyday life. Hence, ubiquitous computing relay on mobile computing, but goes much further.

2.1 Mark Weiser's Visions

To talk about ubiquitous computing we have first to mention Mark Weiser, which has been recognized as the "father" of ubiquitous computing. Weiser, called by many "Visionary", was head of the Computer Science Laboratory at Xerox Palo Alto Research Center (PARC) when he coined the term ubiquitous computing in 1988. When the ubiquitous computing program emerged at PARC, it was at first envisioned only as an answer to what was wrong with personal computing, because they were too complex, too demanding of attention, among others things [12]. During the implementation of the first ubicomp system, Weiser's group realized they were, in fact, starting a post-PC era, in other words, ubicomp was emerging [12].

Mark's vision influenced a countless number of researchers. Almost one quarter of all the papers published in the Ubicomp conference between 2001 and 2005 cite Weiser's foundational articles [13]. Among the Weiser's 'foundational papers' of ubiquitous computing, perhaps the most impacting work is the one entitled "The Computer in the 21st Century", publish in *Scientific American* in 1991. In this paper, Weiser describes the ideal ubicomp future, its purposes, concerns and analogies. To illustrate its ideas he told the story of "Sal", a tale about a single mother and how the world evolves around her needs.

> *"The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it"* [3, p. 1].

Weiser believed that the most powerful things are those that are effectively invisible in use. The ideal is to make a computer so embedded, so fitting, so natural, that we use it without even thinking about it. The essence of this vision is making everything easier to do, with fewer mental gymnastics [3, 14].

Second Weiser, the style of computing that has been imposed on users in the first and second modern computing eras (mainframes and PCs, respectively) is too attention consuming, and divorce the users of what is happening around them. In the ubicomp world, as Weiser believed, computation could be integrated with common objects that you might already be using for everyday work practices, rather than considering computation to be a separate activity. If the integration is done well, the envisioned invisibility could be achieved [15, 2].

In order to clarify this concept of invisibility, consider the example based on the familiar printed page (inspired in [2]). To perform a printing it is necessary deposit ink on thin sheets of paper, and a consolidated technology is necessary for that. For a good result it is necessary to ensure that: it must be durable in use; not wick in the paper if wet; among other things. However, we rarely pay attention on the ink technologies when we read printed pages. Instead, we read pages and comprehend ideas, not necessary focusing on the technology, the characteristics of the ink, or the manufacturing process of the paper to be able to use it. In this example, the printing technology got invisible for the user, allowing the higher-level goal of reading a story, or acquiring knowledge. This kind of thinking rarely happens with traditional PCs, which demand the users continuously focus attention on the system, maintaining it and configuring it to complete a task.

Good technology is invisible, staying out of the way of the task, like a good car stays out of the way of driving. Bad technology draws attention to itself, not the task, like a car that needs a tune-up. Computers are mostly not invisible. Ubiquitous computing is about enabling invisibility in computers [16].

2.2 Ubicomp Today

As a promising research area, ubiquitous computing gave us more questions than answers [12], and many of them are still open [15]. There are many ubicomp projects around the world working on ubicomp challenges. Those projects range from prestigious computer science Schools, such as MIT (see several projects from Media Lab[7] for some examples), to mainstream computer companies, such as Microsoft (see the website http://research.microsoft.com/en-us/groups/ubicomp/ with some projects).

Since the early days of ubicomp, one of the main concerns was that computer too often remain the focus of attention, rather than being a tool through which we work, disappearing from our awareness [15]. We may have not achieved the original Weiser's vision about Ubicomp yet. But we can say that the key ingredients are evolving in a favorable direction for it. Many critical items that were rare in early 1900s are now commercially viable. Each year more possibilities for the mainstream application of ubiquitous computing open up.

The future envisioned by Weiser, ubiquitous computing, considers a computing environment in which each person is continually interacting with many wirelessly interconnected devices [15]. Today it is easy to find several microprocessors at home, available, for instance, in alarm clocks, the microwave ovens and in the TV remote controls. They do not qualify as ubicomp devices mainly because they do not communicate with each other, but if we network them together they are an enabling technology for ubicomp [1]. It soon may become a reality. For example, Google has announced in the event Google IO'11[8] an initiative called Android@Home, which allows Android[9] applications to discover, connect and communicate with appliances and devices inside the house. After connecting together several information sources with many information delivery systems we will start to have things, such as, clocks that find out the correct time after a power failure, and microwave ovens that download new recipes.

---

[7] `http://www.media.mit.edu`
[8] `http://www.google.com/events/io/2011`
[9] `http://code.google.com/android/`

Besides that, some of our computing technology are becoming ubiquitous, for instance smart phones, which are taking center stage as the most widely adopted and ubiquitous computer [2]. When we get used to the possibility of accessing a GPS-connected map, social networks and the Internet anywhere at anytime, we will realize the value of this and it will become part of our lives.

> *"Applications are of course the whole point of ubiquitous computing."* [15, p. 80]

We have to keep in mind that is not just one service that will make computing a disappearing technology, but the combination of many. Those services have to be available as needed without extraordinary human intervention [17]. The challenge is to create a new kind of relationship between people and computers, where computers do not demand too much attention, letting people live their lives [18]. Application will go beyond the big problems like corporate finance, to the little annoyances such as: where are the car-keys? Can I get a parking place? What is the best route to take now? Which pub should I go in a certain area of the city? [1].

Since ubiquitous computing has intersections with many areas of computing, several research fields can contribute to its development, including distributed computing, mobile computing, sensor networks, and machine learning. In particular context-aware computing is a key area of research that can help us to meet the original design goals of ubicomp [19].

2.3 Context-Aware Computing

Several context definitions have been proposed. Among them, those presented by Schilit et al. [20], Dey et al. [21], and Pascoe [22] are close to the definition considered by most people as the ideal one. The problem is that those definitions lack for generality. Dey and Abowd [23] proposed the following definition of context:

> *"Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves."* [23]

This is one of the most accepted and accurate definitions currently used by researchers. It can be observed that the definition is very general when considering what types of data is context, being wide enough to accept the different needs of each application. In addition, it is interesting to note that the definition is precise, not requiring a list of specific types or classes of contexts.

In this work we consider social media as a source of sensing. In this case, humans are responsible for sharing data. The data shared by the "sensors" (humans plus his/her portable device) can be then transformed in a context used to study city dynamics and urban social behavior. In the next section we discuss the participation of humans in the sensing process. After that, in Section 4, we discuss the model and approaches used to transform raw data shared by users into context information.

3 Humans in the Sensing Process

3.1 Participatory Sensing Systems

Social media systems that allow people connected to the Internet to provide useful data about the context in which they are at any given moment, such as Waze, Weddar, and Foursquare, are called participatory sensing systems (PSSs). PSS is a concept that originally considers that the shared data is generated automatically, or passively, by sensor readings from portable devices [8]. However, here it is also considered manually, or proactively, user-generated data. Systems with those characteristics have been called ubiquitous crowdsourcing [24]. 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.

To illustrate this type of system, consider an application for transit monitoring, like Waze. Users can share observations about accidents or potholes manually. Additionally, an application can calculate and share automatically a car speed based on GPS data. Since in this specific case users operate an application that was designed for a specific purpose, the sensed data is structured. If, instead, users use a microblog (e.g., Twitter), the sensed data would be unstructured. For instance, user "Smith" sends the message "traffic now is too slow near the main entrance of the university campus".

3.2 Participatory Sensor Networks

Participatory sensor networks have their users with their portable devices as the 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 participatory sensor network consists of the user plus his/her mobile device, with the goal of sending data to the systems. After that, the data usually can be collected throughout services APIs.

Similar to WSNs, the sensed data is sent to the server, or "sink node". But unlike WSNs, PSNs have the following characteristics: (i) nodes are autonomous mobile entities, i.e., a person with a mobile device; (ii) the cost of the network is distributed among the nodes, providing a global scale; (iii) sensing depends on the willingness of people to participate in the sensing process; (iv) nodes transmit the sensed data directly to the sink; (v) nodes do not suffer from severe power limitations; and (vi) the sink node only receives data and does not have direct control over the nodes. More details about PSNs can be found in [9, 25].

Figure 2 shows the components of a participatory sensor network. In particular, this figure highlights the three most important components, namely)(i) the social media as a source of sensing, (ii) the big raw data, and (iii) the context information.

The component "Social media as a sensor" encompasses users sharing data through social media systems. The component "Big raw data" is responsible for data management. As we can see in the Figure 2, the collection process may be repeated, for example, to get redundant or complementary data from other social media systems.

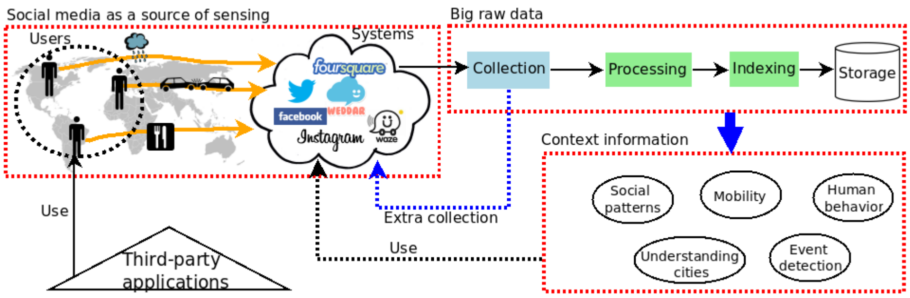


Fig. 2. Overview of a participatory sensor network

After that, the collected data needs to be processed in order to be stored. Since the amount of data coming from participatory sensing systems may be very large, all the components need to be carefully designed if the goal is to get real-time information. A more detailed discussion of some of the challenges is presented in Section 5.

After the data management stage, the data is ready to be analyzed. The component "Context information" represents five type of analysis that could be performed: Social patterns; mobility; understanding cities; human behavior; and event detection. All these classes of analysis are discussed on Section 4.

4 Approaches Used to Deal with Social Media as a Sensor

In this section we discuss the approaches and models used to extract and generate context information from participatory sensing systems data in order to study city dynamics and urban social behavior. This section will concentrate in the component named "Context information", shown in Figure 2. In this particular component is represented different classes of studies, and they will be discussed here. Section 4.1 discusses studies related to the analysis of mobility patterns. Section 4.2 considers studies that focus on the better understanding of city dynamics. Section 4.3 discusses the study of social patterns. Section 4.4 discusses studies concerned in event detection. And finally, Section 4.5 presents studies related to human behavior.

It is worth mentioning that each class of study is not necessarily mutually exclusive. For example, Long et al. [26] used a Foursquare dataset to classify venues based on users' trajectories. This work has intersections with the class "Mobility patterns" (Section 4.1), but instead of being classified in that class, it was assigned to the class "Understanding cities" (Section 4.2), since it is more concerned in the analysis of city dynamics. The main goal of this section is to present hot research topics, and present what have been done to address some of the challenges.

4.1 Mobility Patterns

This class of work focuses on studying mobility patterns of users from their logs generated from social media websites. These logs usually include spatio-temporal information, e.g., check-ins and geolocated photos. The study of mobility is useful for many purposes. For example, it is possible to understand how human allocate time to different activities, thus being a fundamental and traditional question in social science [27]. As another example, one could design new tools to help traffic engineers to understand the flow of people.

The modeling of mobility patterns has been attracting the attention of researchers in different fields, such as physics and ubiquitous computing [28–30]. For example, Song et al. [31] analyzed 50,000 cellphone users and showed that user mobility presents high predictability. It is important to point out that data derived from social media is different from GPS tracking or cellphone usage data, such as phone calls, and present special features and varied contexts. For example, check-ins in location sharing services or photos shared in a photo sharing service bring extra information of a particular place. For instance, a check-in is associated with a type of venue, e.g. pub, and a photo may bring the information about the current situation inside this venue. Again, throughout this work our focus is on studies that analyze data from social media.

Cheng et al. [32] analyzed 22 million check-ins posted from several location sharing services (Foursquare is responsible for 53.5% of the total). They found that users follow simple and reproducible patterns, and also that social status, in addition to geographic and economic factors, are coupled with mobility. **Approach:** to make their analysis they used three statistical properties to study and model human mobility patterns: *displacement*; *radius of gyration*; and *returning probability*. The *displacement* of check-ins is the distance between consecutive check-ins, measuring how far a user has moved. The *radius of gyration* measures the standard deviation of distances between the users' check-ins and the users' center mass. This measure indicates how far and frequently a user has traveled. *Returning probability* is a measure of periodic behavior in human mobility, since periodic behavior tends to happen frequently due to human routines. Besides that, the authors also studied factors that could influence mobility, such as social status and geographic and economic constraints.

Cho et al. [33] investigated patterns of human mobility in three datasets: check-ins in two location sharing services and cellphone location data. They were particularly interested in determining how often users move and where they go to, as well as how social ties may impact their movements. They observed that short-ranged travel is periodic both spatially and temporally and is not affected by the social network structure, while long-distance travel is more inuenced by social network ties. **Approach:** based on their empirical findings they built a model named Periodic & Social Mobility Model to predict mobility of users. This model is composed by three parts: (1) a model of spatial locations that a user usually visits based in a two-state mixture of Gaussians with a time-dependent state prior; (2) a model of temporal movement between these locations based on a truncated Gaussian distribution parameterized by the time of the day; (3) a model of movement that is influenced by the ties of the social network, e.g. encountering friends. In this specific model, if a user performs a check-in, then it will more likely be close in space and time to one of his/her friend's check-ins. Their model is able to predict the exact user location at any time with 40% accuracy.

Nguyen and Szymanski [34] used Gowalla, a location-based social network, to create and validate models of human mobility and relationships. In that work, the authors proposed a friendship-based mobility model (FMM) that take into account social links in order to provide a more accurate and complex model of human mobility. With this model the authors were able to study how frequently friends travel together. This model may improve the accuracy of a varied number of applications, such as traffic engineering in communication networks, transportation systems, and urban planning. **Approach:** the proposed mobility model uses a Markov Model where the states represent locations of check-ins and the links represent the probability of going from one place to another. For example, the probability of going from work to pub is defined as the ratio between the number of times a given user performs a check-in in a pub right after a check-in at work, and the number of times that user performs a check-in at work.

Zheng et al. [35] studied tourist mobility and travel patterns from geotagged photos shared on Flickr. In order to extract the travel patterns, the authors focused the analysis on tourist movement according to regions of attraction and topological characteristics of travel routes by different tourists. The authors demonstrated its potential by testing the approach on four cities. **Approach:** first it is built a database of touristic travel paths based on the concept of mobility entropy (considering Shannon's entropy), used to discriminate the touristic and non-touristic movement. Then, a significance test is applied to ensure that the resulting path is statistically reliable. For that, they devised two methods, one based on a Poisson distribution and the other on a normal distribution. Next, it is proposed a method to discover regions of attraction in a city, using for this the DBSCAN clustering algorithm. To study the touristic movement the authors considered a Markov chain model created from the visiting sequence of regions of attraction discovered by the proposed method. With that, they can estimate statistics of visitors traveling from one region to another. In order to study the topological characteristics of tour routes, the authors perform sequence clustering on travel routes, applying a modified version of the longest common subsequence as a similarity metric to minimize noise.

4.2 Understanding Cities

Information obtained from participatory sensing systems have the power to change our perceived physical boundaries and notions of space [36], as well as to understand city dynamics better. This section focuses in presenting studies in these directions. Many potential applications can benefit from these types of studies, such as tools for city planners to provide new manners to see the city, or for end users who are looking for new ways to explore the city.

Cranshaw et al. [37] presented a model to extract distinct regions of a city that reflect current collective activity patterns. The idea is to expose the dynamic nature of local urban areas considering spatial proximity (derived from geographic coordinates) and social proximity (derived from the distribution of check-ins) of venues. **Approach:** in their study the authors considered data from Foursquare. In order to explore this data, the authors developed a model based on spectral clustering. One of the main contributions is the design of an affinity matrix between venues that effectively blends spatial proximity and social proximity. The similarity of venues is then obtained by comparing pairs of these dimensions. After that, this is used to compute the clusters that may represent different geographical boundaries of neighborhoods. The clustering method is a variation of the spectral clustering proposed by Ng et al. [38], introducing a post processing step to clean up any degenerated cluster.

Noulas et al. [39], proposed an approach to classify areas and users of a city by using venues' categories of Foursquare. This could be used to identify users' communities that visit similar categories of places, useful to recommendation systems, or in the comparison of urban areas within and across cities. **Approach:** their approach is based on spectral clustering algorithm [40, 38]. More specifically, the authors divide the area of a city to be analyzed into a number of equally sized squares, each of them will be a datapoint input for the clustering algorithm. For each area it is represented the activity performed on it based on check-ins in each existing category on that area. Then, it is calculated the similarity between two areas as the cosine similarity between their corresponding activity representation. Having the similarity information, the authors apply it in the spectral clustering algorithm.

Long et al. [26] used a Foursquare dataset to classify venues based on users' trajectories. The premise is that the venues that appear together in many users trajectories will probably be taken as geographic topics, for example representing restaurants people usually go to after shopping at a mall. The approach can be applied, for instance, to understand users' preferences to make recommendation of venues. **Approach:** the authors used the Latent Dirichlet Allocation (LDA) [41] model to discover the local geographic topics from the check-ins. With this approach, it is possible to dynamically categorize the venues in Foursquare according to the users' trajectories, what indicates the crowds preferences of venues. LDA is usually used to cluster documents based on the topics contained in a corpus of documents. For this reason, some terms used to describe the modeling make reference to this context. The authors considered that a single check-in represents a word, which is the basic unit in the LDA. A trajectory of a user consists of all the venues visited by him/her, and this represents a document in the analogy.

In our previous work [42], we propose a technique called City Image and we show its applicability using eight different cities as examples. The resulting image is a way of summarizing the city dynamics based on transition graphs, which map the movements of individuals in a PSN. This technique is a promising way to better understand the city dynamics, helping us to visualize the common routines of their citizens.. **Approach:** the proposed technique consists of a square matrix that summarizes the city dynamics. This matrix is constructed from two transition graphs. First, we construct a transition graph $G(V, E)$, where the nodes $v_i \in V$ are the main categories of the locations and an edge (i, j) exists from node v_i to node v_j if at some point in time an individual 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 weight $w(i, j)$ of an edge is the total number of transitions that occurred from node v_i to node v_j. After constructing G, we create ten random graphs $G_{Ri}(V, E_{Ri})$, where $i = 1, \ldots, 10$ and each one is constructed using the same number of transitions used to construct G. When constructing this random graph instead of considering the actual transition $v_i \rightarrow v_j$ performed by an individual, e.g., "Smith", we randomly pick a location category to replace v_j, simulating, then, a random walk for this individual. We use the distribution for the randomly generated edge weight values for $G_{R1..10}$ to build three ranges: *rejection range* (representing transitions that are not likely to happen), *favouring range* (representing transitions that are likely to happen), *indifference range* (representing transitions that neutral to be performed by users). These ranges are expressed in the visualization.

Kisilevich et al. [43] used geo-tagged photos obtained from Flickr to analyze and compare temporal events that happened in a city, and also to rank sightseeing places. More specifically, the authors presented a way to assess the attractiveness of places based on their positions in a ranking, and suggested a set of visual analytic methods that mixes computational techniques with visual interactivity in order to support analysis of the data. **Approach:** to find the attractiveness of places the authors applied the algorithm DBSCAN [44]. In order to highlight areas of people's activities within a cluster, the authors applied density maps. From the clusters obtained in the clustering part, the weight of every geotagged photo is calculated using a density function based on the relative position of photos of other users in a cluster. The calculated weight is mapped to a color, facilitating the visual inspection.

Frias-Martinez et al. [45] used a dataset from Twitter and proposed a technique to determine the type of activities that is most common in a city by studying tweeting patterns. They also proposed another technique to automatically identify landmarks in a city. **Approach:** to automatically identity urban land usage, the authors apply two methods. The first one is land segmentation. For that it is applied Self-Organizing Maps [46], which is an unsupervised neural network. After training the network, it is obtained a map that segments the urban land into geographical areas with different concentrations of tweets. Each neuron of the network represents a pointer to a region with a high density of tweets. With that, the authors apply Voronoi tessellation considering the location of the neurons to compute the land segments. Next, the authors use the segments found to detect different land usages considering the average tweet usage on them. So, for each land segment is built a unique tweet-activity vector that represents the average tweeting temporal behavior. To characterize urban land usage, it is applied the k-means algorithm, which shows common tweeting behavior across land segments. To identify the landmarks, the authors used the mean-shift clustering technique [47]. The authors considered in this algorithm that every tweet is assigned to a local maxima and a cluster represents a potential landmark. After the execution, if the resulting clusters are ranked by the number of tweets on them, then the result represents a list of the most popular landmarks.

Ji et al. [48] mine blog-based sight photos in order to discover and summarize city landmarks. Their main contribution is a generalized graph modeling framework. This study is useful, for example, for personalized tourist suggestions. **Approach:** first the authors have to extract locations of photos. For that, they collect photos with different descriptors. To identify their locations they use an application called gazetteer [49], which is able to identify location from web resources. Then they create a hierarchical visual-textual clustering scheme to organize sight photos into a "scene-view" structure for each city. For this purpose it is used the concept Bag-of-Visual-Words [50] to generate the content descriptor of photos. Bag-of-Visual-Words are clustered by their similarity measured by the cosine distance, generating then "views". After that the authors create a "scene-view", using textual clustering to aggregate "views" into "scenes". Next, they model two different graphs. The first one represents a scene, where each node is a photo and an edge exist if there is at least one word identical in the photos descriptors. For this graph they present an algorithm, PhotoRank, to discover representative views within a scene. Finally, the authors create another graph to represent the city, that encompasses a scene layer, and present an algorithm to discover city landmarks on it, which explores the PhotoRank algorithm and is inspired in [51].

4.3 Social Patterns

This class of studies concentrates in the analysis of data from social media to understand social patterns. Data from social media enables unprecedented opportunities to study human relationships in a global scale, at a relatively low cost. Examples of possibilities include community detection, products recommendation based on the discovery of similar socio-economic behavior, and new definitions of network centrality.

Scellato et al. [52] presents a study of the spatial properties of the location-based social networks arising among users. Among the results, the authors reported, for instance, that 40% of social links happens below 100 km, and that there is strong heterogeneity across users related to both social and spatial factors. **Approach:** to extract properties and verify their hypothesis, the authors analyzed datasets of three location based services: Foursquare, Gowalla, and Brightkite. In their study, the authors used two randomized models, a social model and a spatial/geo model, to assess the statistical signicance of the empirical spatial properties of the networks analyzed. The social model keeps the social connections as they are, randomizing all user locations. The geo model keeps the user locations unmodied and then assigns every social link between two users at a certain distance according to the relative probability of friendship, observed in their analysis.

Cranshaw et al. [53] introduced a new set of features of human location trail for analyzing the social context of a geographical region. They demonstrated the applicability of these features by presenting a model for predicting friendship between two users, showing significant gains over previous models for the same purpose. **Approach:** the authors used a dataset from Locaccino[10], a system that allows users to share his/her current location with other Locaccino users through Facebook[11]. For the co-location analysis, the authors split the space in grids of 0.0002 x 0.0002 latitude/longitude, which means approximately 30 meters x 30 meters. The time was considered in slots of 10 minutes. In this way, a user is co-located with another user if they are located in the same grid within a slot of time. To model the co-location of users, it is applied three diversity measures: frequency, user count, and entropy (Shannon's entropy). The frequency measure captures the raw count of users who visit a location. The user measure considers the total number of unique users in a location. The entropy measure considers the number of users observed at the location, as well as the relative proportions of observations. High entropy means that many users were observed at the location with equal proportion.

Quercia et al. [54] study how social media communities resemble real-life ones. They tested whether established sociological theories of real-life social networks still hold in Twitter. They found, for example, that social brokers in Twitter are opinion

[10] `http://www.locaccino.org`
[11] `http://www.facebook.com`

leaders who take the risk of tweeting about different topics. They also discovered that most users have geographically local networks, and that social brokers express not only positive but also negative emotions. **Approach:** the authors applied network metrics about topic, geography, and emotions, regarding to parts of one's social world. These metrics include reciprocity, simmelian ties, and network constraint. Reciprocity is the proportion of edges in a network that are bidirectional. Simmelian ties are a measure that considers triadic relationships. Network constraint measure brokerage opportunities in the network, where high network constraint means less brokerage opportunities. They used Burt's formulation [55] in this specific case.

Java et al. [56] studied blog communities. For that they present a technique for clustering communities by using both the hyperlink structure of blog articles and tag information available on them. The technique was tested in a real network of blogs and tag information, as well as in a citation network. **Approach:** the authors define a community as a set of nodes in a graph that link more frequently within this set than outside it, and they also share similar tags. Their technique is based on the Normalized Cut (NCut) algorithm [57].

Sadilek et al. [58] studied the interplay between people's location, interactions, and their social ties, presenting a technique for inferring link and location information from a stream of message updates. The authors demonstrated, by analyzing users from New York City and Los Angeles, that their technique significantly outperforms other current comparable approaches. **Approach:** for link prediction their approach infers social ties by considering patterns in friendship formation, the content of people's messages, and user location. For location prediction, their technique implements a probabilistic model of human mobility, where it treats users with known GPS positions as noisy sensors of the location of their friends.

4.4 Event Detection

This class of work is focused in the identification of events through data shared in social media. This task is especially favorable due the real-time nature of certain types of social media, such as Twitter. Events might be natural ones, such as earthquakes, or not natural ones, such as the identification/prediction of stock market changes.

Bollen et al. [59] studied whether collective mood states derived from Twitter feeds are correlated to the value of the Down Jones Industrial Average (DJIA) over time. Their findings indicate that it is possible to obtain an accuracy of 86.7% in predicting the daily up and down changes in the closing values events of the DJIA. This is possible by choosing specific mood dimensions, but not all that were considered. **Approach:** to extract the sentiment expressed by the users in the tweet the authors used two tools. The first one is the OpinionFinder (OF), which extract negative or positive sentiments from the message. The second tool, Google-Profile Mood State (GPOM), extract six-dimensional daily time series of public mood. The authors use Granger causality analysis in which it is correlates DJIA values to GPOMs and OF values of n past days. The authors also trained a Self-Organizing Fuzzy Neural Network to predict DJIA values on the basis of various combinations of past DJIA values and OF and GPOMS public mood data.

Sakaki et al. [60] studied the real-time interaction of events in Twitter (e.g. earthquakes), and propose an algorithm to monitor tweets to detect a target event. To demonstrate the effectiveness of their method, the authors built an earthquake reporting system in Japan, which was capable to detect 96% of earthquakes reported by the Japan Meteorological Agency (JMA) with seismic intensity scale of 3 or more. Notification to registered users was delivered faster than the announcements that are broadcast by the JMA. **Approach:** the authors devise a classifier of tweets based on features such as the keywords in a tweet, the number of words, and their context. After that, they produced a probabilistic spatio-temporal model for the target event that can find the center and the trajectory of the event location.

Lee and Sumiya [61] present a geo-social event detection system to identity local events (e.g., local festivals) by monitoring crowd behaviors indirectly via Twitter. The system was created on the hypothesis that users probably write many posts about these local events. **Approach:** first the authors decide what the usual status of crowd behaviors is in a geographical region in terms of tweeting patterns. After that, a sudden increase in tweets in a geographical region can be an important clue. Another hint might be the increasing number of Twitter users in a geographical region in a short period of time. The authors also consider if the movements of the local users become unexpectedly elevated. The detection of unusual events in the study uses the concept of boxplot [62], which is applied to create ranges to determine the cases desired to be detected.

Becker et al. [63] analyze streams of Twitter messages to distinguish between messages about real-world events and non-event messages. They identify each event and its associated Twitter messages. **Approach:** the authors use an online clustering technique that groups together similar tweets. With that, they extract features for each cluster to help determine which clusters correspond to events. Next, the authors use these features to train a classier to distinguish between event and non-event clusters.

4.5 Human Behavior

This group of studies focus on the study of human behavior through the data shared in social media, which, as we mentioned before, can be seen as signals given by users. This type of study can be applied, for example, to the discovery of individual social proles, the discovery of collective behaviors, the analysis of sentiment and opinion evolution, and a better understanding of why individuals take certain actions.

Joseph et al. [64] analyzed a Foursquare dataset to identify groups of people and the places they go. Their model was able to identify groups of people which represent both geo-spatially close groups and people who appear to have similar interests. **Approach:** their model is based on the idea of topic modeling. For that they applied the Latent Dirichlet Allocation [41]. In the model instantiation, each check-in for a user can be thought of as a word in a document. Similar to text documents, where a “document” can have multiple words, the authors dened a multinomial distribution for the check-ins for each user by using the number of check-ins in each venue as features.

Naaman et al. [65] focused their study in the characterization of tweeting patterns in different cities located in the USA, envisioning to provide a framework for reasoning about activities performed in cities. This study might be useful to deal with challenges such as transportation or resource planning faced in urban studies. **Approach:** first the authors selected tweets from some US cities. Then, they selected the top 1000 words from the resulting dataset, and made a cleaning procedure in this dataset using the NLTK toolkit[12], removing, for example, stopwords. After that, the authors performed a study of keyword-based diurnal patterns in the considered locations. Besides that, the authors applied the concept of Shannon's entropy and Mean Absolute Percentage Error (MAPE), to measure the variability of the data within days and across days, respectively.

Poblete et al. [66] analyzed a twitter dataset aiming the discovery of insights of how tweeting behavior varies across countries, as well as the possible explanations for these differences. **Approach:** first the authors selected the top ten most active countries. Then, they extracted differences in the number of twittes per user, languages used per country, sentiment analysis (happiness), using the Affective Norms for English Words (ANEW) [67] and a Spanish version of it [68], and the content of the tweet. Moreover, they studied the social network properties for each country applying metrics, such as, clustering coefficient, diameter, and shortest paths.

Gao et al. [69] propose a model to address the "cold start" location prediction problem, by using the social network information. Results in an experiment based on a real-world location-based social network show that the approach is effective for the studied problem. **Approach:** the authors' strategy encompasses the investigation of the check-ins behavior to understand the correlations in the context of the user's social network and geographical distance. For this analysis, they considered four social cycles. With that, the authors modeled the geo-social correlations of "new check-in" behavior considering the intrinsic patterns of users' check-ins and his/her social cycles.

Yu et al. [70] used the users' behavioral patterns extracted from a Sina Weibo[13] to investigate how users' frequent activities reflect their sleeping time and living time zones. The authors showed that may be possible to detect the sleeping time of users. Their results could also be used as an alternative way to estimate time zones. **Approach:** based on the time series of the Sina Weibo usage the authors applied a simple statistical method, assuming that users keep a daily routine, going to bed and waking up on time, to detect long periods of inactivity.

5 Challenges

Considering social media as a source of sensing, constructing then a participatory sensor network imposes many challenges. Looking at Figure 2 we see that a participatory sensor network could be divided in different blocks. In Section 4 we described how researchers have been addressing challenges mainly related in the block named "Context information", which represents models and approaches to transform big raw data from social media in useful information, to be applied, for example, in applications. In this section we are concentrated in challenges related with the blocks named "Social media as a source of sensing" and "Big raw data".

Among the challenges present in these blocks we can mention data quality, data collection, data storage, data processing and indexing. The quality of the shared data is a

[12] `http://www.nltk.org`

[13] A popular Chinese micro-blogging service.

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 [24]. For instance, since users can produce sensor readings with relatively little effort, data integrity is not always guaranteed [71].

Besides that the shared data through social media in some cases is free text, not presenting structure nor codified semantics, being complex to understand and process. To better interpret such complex data, visualization techniques and tools should be developed. Another issue related to data quality is the liberty given to users in certain social media systems. Sometimes, users can post whatever, even incorrect, information in different formats. This demands mechanisms for data filtering. A reputation system may be very useful in this case.

Data collection is a challenging issue especially from third-party social media services, such as Foursquare and Waze. By default, data shared in those systems are usually private, unless users decide to make them public, for example sharing it on Twitter. This means that no public data can be available at all. Furthermore, since the data depends on the users will in contribute, there is no guarantee on the delivery of any data. This makes the use of social media as a source of sensing completely out of the control loop of system managers and application developers. Some actions can be taken to ensure that the user participation is sustained over time. An example of action could be an incentive mechanism based on micro-payments, i.e., every time a user perform a given activity, he/she receives an small payment, as proposed by Reddy et al. [72].

Another important issue is deal with a huge volume of data that social media 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 participatory sensor network. To tackle this issue we need methods to effectively store, move and process big amounts of data. New algorithmic paradigms, for example map-reduce, should be designed, as well as specific mining techniques should be created according to these new paradigms. Other methods should contemplate data engineering approaches for large networks with up to billions nodes/edges, including effective compression, search, and pattern matching methods [27].

Furthermore, participatory sensor networks are very dynamic. To illustrate the challenges that emerge with this characteristic we analyze the information flow in PSNs, which is depicted in Figure 2, particularly the two flows symbolized by the arrows labeled with the word “use”, directing from Context information component to Systems, and from Third-party applications to Users. Users rely on applications, such as Twitter or Waze, to transmit their sensed data. The sensed data is, then, transmitted to the server, or the “sink node”. The Context information component is responsible for processing the shared data and generating useful information, or contexts (Section 2.3). The systems, such as Waze, by its turn, may be fed back with the generated contexts and, from this, provides useful information to the users. Contexts can also be generated by third-party applications. For example, in Section 4, we describe an example of application that enables the identification of regions of interest in a city, which exemplifies a type context. After using this application, users may choose to change their behavior, e.g., to visit preferably popular areas, which may ultimately impact the number of potential shared data in those places. This gives an idea of how dynamic a participatory sensor network is and the challenges that emerge to deal with this dynamism.

Besides those problems there is still the problem of user's privacy. This challenge is very broad, being present in many layers of the system. Data privacy in social media systems has been currently discussed in several studies, such as: [73–75].

A wide range of novel applications opens up after dealing with the challenges of this research field. Some of the opportunities are illustrated in the next section.

6 Opportunities

In this section we present some of the promising opportunities when considering social media for the large scale study of city dynamics and urban social behavior. For that we use the Foursquare dataset analyzed in the study [42].

Nodes in a participatory sensor network move according to their routines or local preferences, and this is interesting for applications that want to capture city dynamics and urban social behavior. In this direction we present several opportunities, grouped in two categories: *Area semantics* (Section 6.1); and *Urban transitions* (Section 6.2).

6.1 Area Semantics

There are many opportunities to design semantic location services, and this sort of services will be crucial for the next wave of killer applications [76]. The opportunities pointed here exploit the information about the category of the venues present in the considered dataset. A complete list of these categories, with examples, can be found in [42].

Application accessed mostly by smartphones provides datasets that represent the social network topology and dynamics of entire cities, enabling the analysis of the social, economic, and cultural aspects of particular areas. For example, regions that provide a small amount of data compared to other regions of the same city 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 that are expensive in some countries. The preliminary results in the use of participatory sensor networks in these scenarios demonstrate good opportunities to enable the visualization of interesting facts. For example, analyzing carefully the data for the particular case of Rio de Janeiro, shown in Figure 3, 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 areas. This information may be useful, for example, 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 them in an automatic and cheaper way using a participatory sensor network. For this purpose, similar algorithms to the one proposed in [37] could be applied.

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, humans have seasonal patterns due to their routines. This seasonality has a great potential for prediction applications, since it is very likely

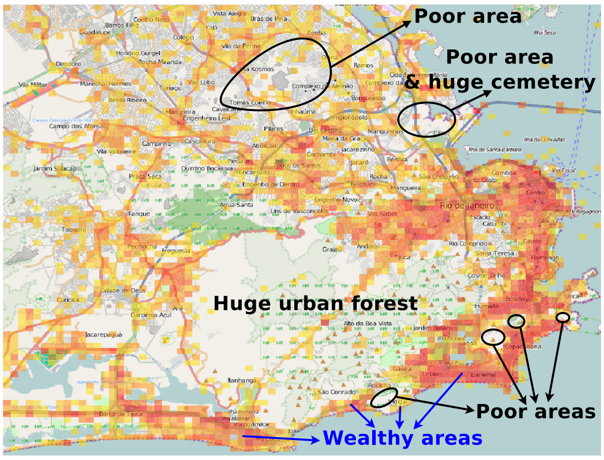


Fig. 3. Example of possible area classification by lack of sensing

that people periodically repeat their activities. We do believe that there are many opportunities for prediction given by the circadian rhythm of people, enabling the prediction, for example, of crowds. 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. For instance, Hsieh et al. [77] proposed a time-sensitive model to recommend trip routes based on the information extracted from Gowalla check-ins.

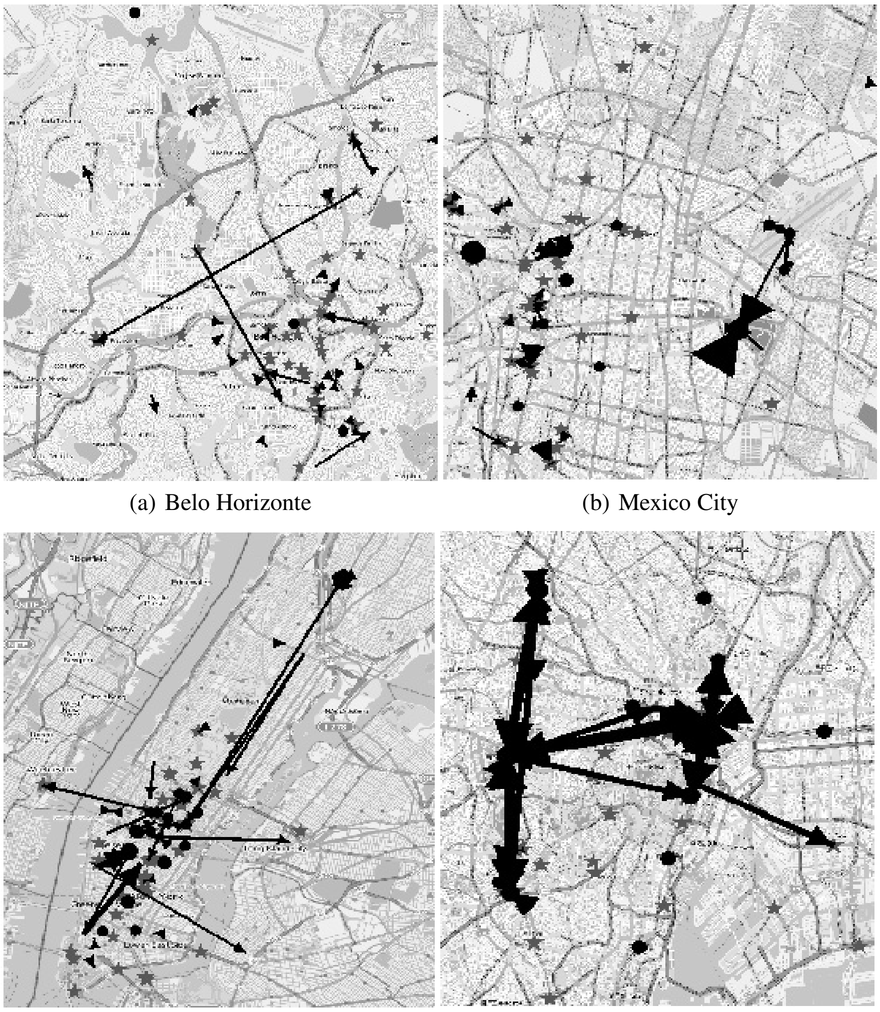
In one of our previous study [78] we present a scenario that illustrates another opportunity that exploits the same data. For that, we created a simple method to estimate the number of check-ins in certain time and space, taking into account different categories of places. We show that temporally it is possible to distinguish popular areas in different regions of the city, and this might be useful as on decision criterion when choosing an area to visit at a certain time.

6.2 Urban Transitions

Now we present another range of opportunities that emerges from urban transition graphs. The urban transition graph maps the movements of users between locations. This graph is a directed weighted graph $G(V, E)$, where a node $v_i \in V$ is a specific location (e.g., Times Square) 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.

Here we consider the same requirements for transitions specified in [78]. Figure 4 shows heavy weighted edges and hub nodes (top 50 edge weights and node degrees) for Belo Horizonte, Mexico City, New York and Tokyo. Stars represent the hubs, black arrows represent the edges, and black circles represent self-loops. The larger[14] the

[14] Numbers grow in logarithmic scale.



(a) Belo Horizonte (b) Mexico City

(c) New York City (d) Tokyo

Fig. 4. Top 50 edge weights and node degrees (hubs) for 4 cities. Stars represent hubs, black arrows edges, and black circles self-loops. The larger the symbol, the higher the value.

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 for Belo Horizonte and Mexico City that most of the heavy weighted edges are self-loops and low distance edges, implying that people tend to perform activities in the neighborhood of where they are while they can. On the other hand, for New York City and Tokyo, cities that are known for their fast public transportation systems, favor the existence of some long distance heavy weighted edges along the public transport links.

This scheme may be used to support various applications, for example, a public information dissemination scheme, which usually shows traditional advertising. If one knows where the city hubs are, he/she could strategically put these displays in these locations. Moreover, 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 example, advertising each other's businesses.

Urban transition graphs are useful also to display a visualization of a city based on the transitions that are likely to occur, as the demonstrated by the City Image technique presented in our previous work [42]. 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 that could be interesting for recommendation systems.

Other example in this direction is the study performed by Long et al. [26], which used a Foursquare dataset to classify venues based on users' trajectories. The study performed by Zheng et al. [35] is also one more example, since the authors show the potential in exploring transitions from geotagged photos shared on Flickr.

In this section our goal was illustrate some of the open opportunities in this field. Certainly there are many others.

7 Final Remarks

In this chapter we present the state of the art of the use of participatory sensing systems to study city dynamics and urban social behavior. This work surveys approaches and models applied to generate context from raw data obtained from PSSs. To achieve this goal we studied a compilation of studies that represent five recurrent themes addressed by researchers nowadays, namely: (1) mobility patterns; (2) understanding cities; (3) social patterns; (4) event detection; and (5) human behavior. For each class we highlight, for each study, the approaches and models applied to create new knowledge and semantic meaning from the big raw data. Besides that we also demonstrate a range of fruitful opportunities that emerge when using participatory sensing to the large scale study of city dynamics and urban social behavior.

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. Weiser, M., Brown, J.S.: The coming age of calm technology (October 1996)
2. Krumm, J.: Ubiquitous Computing Fundamentals, 1st edn. Chapman & Hall/CRC (2009)
3. Weiser, M.: The Computer in the 21st Century. Scientific American 265(3), 94–104 (1991)
4. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Computer Networks 52(12), 2292–2330 (2008)

5. Akyildiz, I., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Computer Networks 38(4), 393–422 (2002)
6. Lane, N., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.: A Survey of Mobile Phone Sensing. IEEE Communications Magazine 48(9), 140–150 (2010)
7. Molina, B.: Foursquare tops 20 million users. USA Today (April 2012), `http://content.usatoday.com/communities/technologylive/post/2012/04/foursquare-tops-20-million-users/1`
8. 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): Mobile Device Centric Sensor Networks and Applications, pp. 117–134 (2006)
9. Silva, T.H., Vaz de Melo, P.O.S., Almeida, J.M., Loureiro, A.A.F.: Uncovering Properties in Participatory Sensor Networks. In: Proc. of the 4th ACM International Workshop on Hot Topics in Planet-scale Measurement (HotPlanet 2012) (June 2012)
10. Weiser, M.: Weiser's website about ubicomp (1996), `http://www.ubiq.com/weiser/weiser.html` (Website accesed for the last time in May of 2013)
11. Satyanarayanan, M.: Fundamental challenges in mobile computing. In: Proceedings of the Fifteenth Annual ACM Symposium on Principles of Distributed Computing, PODC 1996, pp. 1–7. ACM, Philadelphia (1996)
12. Weiser, M., Gold, R., Brown, J.S.: The origins of ubiquitous computing research at parc in the late 1980s. IBM Syst. J. 38, 693–696 (1999)
13. Bell, G., Dourish, P.: Yesterday's tomorrows: notes on ubiquitous computing's dominant vision. Personal Ubiquitous Comput. 11, 133–143 (2007)
14. Weiser, M.: Ubiquitous computing. Computer 26, 71–72 (1993)
15. Weiser, M.: Some computer science issues in ubiquitous computing. Commun. ACM 36, 75–84 (1993)
16. Weiser, M.: Keynote: Building invisible interfaces (1994)
17. Abowd, G.D., Mynatt, E.D., Rodden, T.: The human experience. IEEE Pervasive Computing 1, 48–57 (2002)
18. Abowd, G.D., Mynatt, E.D.: Charting past, present, and future research in ubiquitous computing. ACM Trans. Comput.-Hum. Interact. 7, 29–58 (2000)
19. Silva, T.H., de S. Celes, C.S.F., Mota, V.F.S., Loureiro, A.A.F.: A picture of present ubicomp research exploring publications from important events in the field. Journal of Applied Computing Research 2(1), 32–49 (2012)
20. Schilit, B., Adams, N., Want, R.: Context-aware computing applications. In: Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications, pp. 85–90. IEEE Computer Society, Washington, DC (1994)
21. Dey, A.K., Abowd, G.D., Wood, A.: Cyberdesk: a framework for providing self-integrating context-aware services. In: Proceedings of the 3rd International Conference on Intelligent User Interfaces, IUI 1998, pp. 47–54. ACM, New York (1998)
22. Pascoe, M.J.: Adding generic contextual capabilities to wearable computers. In: Proceedings of the 2nd IEEE International Symposium on Wearable Computers, ISWC 1998, pp. 92–99. IEEE Computer Society, Washington, DC (1998)
23. Dey, A.K., Abowd, G.D.: Towards a Better Understanding of Context and Context-Awareness. In: CHI 2000 Workshop on the What, Who, Where, When, and How of Context-Awareness (2000)
24. Mashhadi, A.J., Capra, L.: Quality Control for Real-time Ubiquitous Crowdsourcing. In: Proc. of the 2nd Int'l Workshop on Ubiquitous Crowdsouring (UbiCrowd 2011), pp. 5–8 (2011)

25. 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 the IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS 2013) (May 2013)
26. Long, X., Jin, L., Joshi, J.: Exploring trajectory-driven local geographic topics in foursquare. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. UbiComp 2012, pp. 927–934. ACM, New York (2012)
27. Giannotti, F., Pedreschi, D., Pentland, A., Lukowicz, P., Kossmann, D., Crowley, J., Helbing, D.: A planetary nervous system for social mining and collective awareness. The European Physical Journal Special Topics 214(1), 49–75 (2012)
28. Brockmann, D., Hufnagel, L., Geisel, T.: The scaling laws of human travel. Nature 439(7075), 462–465 (2006)
29. Zheng, Y., Zhang, L., Xie, X., Ma, W.Y.: Mining interesting locations and travel sequences from gps trajectories. In: Proceedings of the 18th International Conference on World Wide Web, pp. 791–800. ACM (2009)
30. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. Nature 453(7196), 779–782 (2008)
31. Song, C., Qu, Z., Blumm, N., Barabási, A.L.: Limits of predictability in human mobility. Science 327(5968), 1018–1021 (2010)
32. Cheng, Z., Caverlee, J., Lee, K., Sui, D.Z.: Exploring Millions of Footprints in Location Sharing Services. In: Proc. of the Fifth Int'l Conf. on Weblogs and Social Media, ICWSM 2011 (2011)
33. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2011, pp. 1082–1090. ACM, San Diego (2011)
34. Nguyen, T., Szymanski, B.K.: Using location-based social networks to validate human mobility and relationships models. arXiv preprint arXiv:1208.3653 (2012)
35. Zheng, Y.T., Zha, Z.J., Chua, T.S.: Mining travel patterns from geotagged photos. ACM Trans. Intell. Syst. Technol. 3(3), 56:1–56:18 (2012)
36. Bilandzic, M., Foth, M.: A review of locative media, mobile and embodied spatial interaction. International Journal of Human-Computer Studies 70(1), 66–71 (2012)
37. 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 the Sixth Int'l Conf. on Weblogs and Social Media (2012)
38. Ng, A.Y., Jordan, M.I., Weiss, Y., et al.: On spectral clustering: Analysis and an algorithm. Advances in Neural Information Processing Systems 2, 849–856 (2002)
39. Noulas, A., Scellato, S., Mascolo, C., Pontil, M.: 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 2011 (2011)
40. Luxburg, U.: A tutorial on spectral clustering. Statistics and Computing 17(4), 395–416 (2007)
41. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. The Journal of Machine Learning Research 3, 993–1022 (2003)
42. Silva, T.H., Vaz de Melo, P.O.S., Almeida, J.M., Salles, J., Loureiro, A.A.F.: Visualizing the invisible image of cities. In: Proc. of IEEE International Conference on Cyber, Physical and Social Computing, CPScom 2012 (November 2012)
43. Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., Andrienko, G.: Event-based analysis of people's activities and behavior using flickr and panoramio geotagged photo collections. In: IEEE 2010 14th International Conference on Information Visualisation (IV), pp. 289–296 (2010)

44. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise, Kdd (1996)
45. Frias-Martinez, V., Soto, V., Hohwald, H., Frias-Martinez, E.: Characterizing urban landscapes using geolocated tweets. In: IEEE 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom) Privacy, Security, Risk and Trust (PASSAT), pp. 239–248 (2012)
46. Kohonen, T.: The self-organizing map. Proceedings of the IEEE 78(9), 1464–1480 (1990)
47. Cheng, Y.: Mean shift, mode seeking, and clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(8), 790–799 (1995)
48. Ji, R., Xie, X., Yao, H., Ma, W.Y.: Mining city landmarks from blogs by graph modeling. In: Proceedings of the 17th ACM International Conference on Multimedia, pp. 105–114. ACM (2009)
49. Wang, C., Xie, X., Wang, L., Lu, Y., Ma, W.Y.: Detecting geographic locations from web resources. In: Proceedings of the 2005 Workshop on Geographic Information Retrieval, pp. 17–24. ACM (2005)
50. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2161–2168. IEEE (2006)
51. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. Journal of the ACM (JACM) 46(5), 604–632 (1999)
52. Scellato, S., Noulas, A., Lambiotte, R., Mascolo, C.: Socio-spatial Properties of Online Location-based Social Networks. In: Proc. of the Fifth Int'l Conf. on Weblogs and Social Media, ICWSM 2011 (2011)
53. Cranshaw, J., Toch, E., Hong, J., Kittur, A., Sadeh, N.: Bridging the gap between physical location and online social networks. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, pp. 119–128. ACM (2010)
54. Quercia, D., Capra, L., Crowcroft, J.: The social world of twitter: Topics, geography, and emotions. In: The 6th International AAAI Conference on Weblogs and Social Media, Dublin (2012)
55. Burt, R.S.: Structural holes: The social structure of competition (1992)
56. Java, A., Joshi, A., Finin, T.: Detecting commmunities via simultaneous clustering of graphs and folksonomies. In: Proceedings of WebKDD, vol. 2008 (2008)
57. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 888–905 (2000)
58. Sadilek, A., Kautz, H., Bigham, J.P.: Finding your friends and following them to where you are. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, pp. 723–732. ACM (2012)
59. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science 2(1), 1–8 (2011)
60. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, pp. 851–860. ACM, New York (2010)
61. Lee, R., Sumiya, K.: Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection. In: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, pp. 1–10. ACM (2010)
62. McGill, R., Tukey, J.W., Larsen, W.A.: Variations of box plots. The American Statistician 32(1), 12–16 (1978)
63. Becker, H., Naaman, M., Gravano, L.: Beyond trending topics: Real-world event identification on twitter. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, ICWSM 2011 (2011)

64. Joseph, K., Tan, C.H., Carley, K.M.: Beyond local, categories and friends: clustering foursquare users with latent topics. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 919–926. ACM (2012)
65. Naaman, M., Zhang, A.X., Brody, S., Lotan, G.: On the study of diurnal urban routines on twitter. In: Sixth International AAAI Conference on Weblogs and Social Media (2012)
66. Poblete, B., Garcia, R., Mendoza, M., Jaimes, A.: Do all birds tweet the same?: characterizing twitter around the world. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 1025–1030. ACM (2011)
67. Bradley, M.M., Lang, P.J.: Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Technical Report C-1, The Center for Research in Psychophysiology, University of Florida (1999)
68. Redondo, J., Fraga, I., Padrón, I., Comesaña, M.: The spanish adaptation of anew (affective norms for english words). Behavior Research Methods 39(3), 600–605 (2007)
69. Gao, H., Tang, J., Liu, H.: gscorr: modeling geo-social correlations for new check-ins on location-based social networks. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 1582–1586. ACM (2012)
70. Yu, H., Sun, G., Lv, M.: Users sleeping time analysis based on micro-blogging data. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 964–968. ACM (2012)
71. Saroiu, S., Wolman, A.: I am a sensor, and i approve this message. In: Proc. of the Eleventh Workshop on Mobile Computing Systems and Applications, HotMobile 2010, pp. 37–42. ACM, Annapolis (2010)
72. Reddy, S., Estrin, D., Hansen, M., Srivastava, M.: Examining micro-payments for participatory sensing data collections. In: Proc. of the 12th ACM International Conference on Ubiquitous Computing, Ubicomp 2010, pp. 33–36. ACM, New York (2010)
73. Pontes, T., Magno, G., Vasconcelos, M., Gupta, A., Almeida, J., Kumaraguru, P., Almeida, V.: Beware of what you share: Inferring home location in social networks. In: 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW), pp. 571–578 (2012)
74. Toch, E., Cranshaw, J., Drielsma, P.H., Tsai, J.Y., Kelley, P.G., Springfield, J., Cranor, L., Hong, J., Sadeh, N.: Empirical models of privacy in location sharing. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, Ubicomp 2010, pp. 129–138. ACM, Copenhagen (2010)
75. Brush, A.B., Krumm, J., Scott, J.: Exploring end user preferences for location obfuscation, location-based services, and the value of location. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, Ubicomp 2010, pp. 95–104. ACM, Copenhagen (2010)
76. Kim, D.H., Han, K., Estrin, D.: Employing user feedback for semantic location services. In: Proc. of the 13th International Conference on Ubiquitous Computing, UbiComp 2011, pp. 217–226. ACM, New York (2011)
77. Hsieh, H.P., Li, C.T., Lin, S.D.: Exploiting large-scale check-in data to recommend time-sensitive routes. In: Proc. of the ACM SIGKDD Int Workshop on Urban Computing, UrbComp 2012, pp. 55–62. ACM, Beijing (2012)
78. 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: 18th IEEE Symposium on Computers and Communications (ISCC 2013), Split, Croatia (July 2013)