

Large Scale Study of City Dynamics and Urban Social Behavior Using Participatory Sensor Networks

Thiago H. Silva¹, Antonio A. F. Loureiro¹ (advisor),
Jussara M. Almeida¹ (co-advisor), and Pedro O. S. Vaz de Melo¹ (co-advisor)

¹Department of Computer Science, Universidade Federal de Minas Gerais (UFMG)

1. Motivation and Objective

Smartphones play a fundamental role in today's technologically-advanced community allowing people to communicate (almost) anywhere in the world and share all kinds of contextual information (e.g., location and opinion). They are being manufactured with an increasing number of powerful embedded sensors of different categories (e.g., acoustic, sound, GPS), enabling a variety of new applications and services. Indeed, smartphones are being used for many personal sensing applications, such as for monitoring physical exercises, and for wide participatory sensing applications, which are not limited to a particular individual (e.g., traffic conditions and noise pollution) [Silva et al. 2013b].

Participatory sensing aims at monitoring large scale phenomena and require the active involvement of people to voluntarily share contextual information and/or make their sensed data available. Participatory sensing finds an effective platform for large scale reach in the increasing popularity of location-based social media applications, such as Instagram and Foursquare, which combine the features of online social networks with location-based services.

These applications enable the observations of the actions of hundreds millions of people in large scale urban areas in (near) real time and over extended periods of time. This opens an unprecedented opportunity to revolutionize the way social science is done. Unlike traditional methods that rely on survey data, new techniques can be designed to exploit participatory data, which is much cheaper, more dynamic as it reflects current situations in (near) real time, and, more important, can easily reach planetary scale. Moreover, as we argue here, such participatory sensing applications have the potential to be a fundamental tool to better understand human urban interaction, leveraging our awareness to different aspects of our lives in urban scenarios.

Participatory sensor networks (PSNs) can be derived from such systems, where each node represents a user able to sense context data with his mobile device. For example, in a PSN derived from Instagram, the sensed context data is a picture of a specific place where the user is located. Figure 1 illustrates this concept, showing sharing activities (represented by red dots) of four users at three different points in time. After a given time, we can analyze this data in different ways. For instance, the bottom rightmost portion of the figure shows, as an aggregated view, a directed graph with nodes representing locations where data was shared and edges connecting locations that were shared by the same user. Using this graph we can extract, for instance, user mobility patterns.

The main objective of this thesis is to answer the question: **Can we use PSNs to perform large scale and near real time study of city dynamics and urban social behavior?** To that end, a fundamental step is to understand the properties of PSNs. We aim to analyze PSNs derived from different systems. After that, our goal is to show how to model and extract knowledge from PSNs, individually and concurrently.

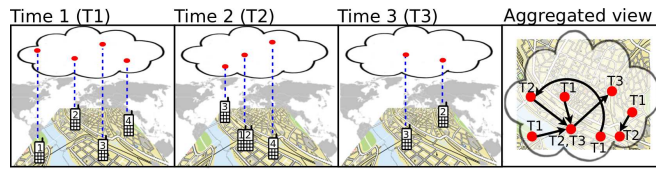


Figure 1. Illustration of participatory sensor networks.

2. Contributions

We tackle the main objective of this thesis answering three different questions: **What are the properties of PSN?** (Section 2.1); **How can we use PSNs?** (Section 2.2); **Can we combine data from different PSNs to infer new information?** (Section 2.3).

2.1. Properties of PSNs

As the data provided by PSNs may be very complex, a fundamental step in any investigation is to study the collected data in order to understand its challenges and usefulness. We have characterized and analyzed properties of three different types of PSNs: (1) photo sharing services, particularly Instagram; (2) location sharing services, particularly Gowalla, Brightkite, and Foursquare; (3) and traffic alert services, particularly Waze.

We found that PSNs share several properties in common, for instance: (i) very large scale; (ii) user participation regarding to the number of shared data, and the number of data shared per location may vary widely; (iii) and highly unequal frequency of data sharing, both spatially and temporally, which is highly correlated with the typical routine of people.

We also characterized PSNs considering the time and location where the content was shared. We aimed to understand whether the pieces of data from one system are correlated to be used for the study of city dynamics and urban social behavior. The results show the existence of correlation. This gave us insight about using different PSNs as “sensing layers” (see Section 2.3) of a predefined geographical region.

Our characterization provides a deep understanding of the properties of PSNs, revealing their potential to drive various studies on city dynamics and urban social behavior. Chapter 4 of the thesis shows more results of the characterization process.

The results for this contribution were reported in the following publications:

- In [Silva et al. 2012a], we perform the first analysis of PSN properties. The PSNs analyzed were derived from two location sharing services: Gowalla and Brightkite;
- In [Silva et al. 2013b], we extended the work [Silva et al. 2012a] analyzing also two different PSNs derived from Foursquare, a popular location sharing service;
- In [Silva et al. 2013c] (**2nd best paper award**) and [Silva et al. 2013d], we investigate properties of a PSN derived from a photo sharing service: Instagram;
- In [Silva et al. 2013f], we study properties of a PSN derived from Waze, a popular traffic alert system.

2.2. Understanding City Dynamics and Urban Social Behavior

This section presents our proposed methods and techniques that illustrate how PSNs can be used to enable large scale and near real time analyses of urban societies.

2.2.1. Visualizing the Invisible Image of Cities

Cities are not identical and evolve over time, while habits and routines of their inhabitants are typically distinct. Knowing that, can we use a PSN to capture the differences among cities based on common routines of their inhabitants? For that, we proposed a technique called City Image, which provides a visual summary of the city dynamics based on people movements. This technique exploits urban transition graphs to map the movements of users between locations. An urban transition graph is a directed weighted graph $G(V, E)$, where a node $v_i \in V$ is the category of a specific location (e.g., nightlife) and a directed edge $(i, j) \in E$ marks a transition between two categories. Each cell in the City Image represents the willingness of a transition from a given category at a given place to another category. Red represent rejection, blue favoring, and white indifference, as shown in Figure 2. Section 5.1 of the thesis presents further details.

City Image is a promising technique to enable a better understanding of the city dynamics, helping the visualization of common routines of their citizens. Observe in Figure 2 that *office* transitions are more likely to happen on weekdays during the day in both cities, as expected. However, note that the City Images of Sao Paulo and Kuwait city have also significant differences that reflect cultural diversities between both cities. For example, the City Image representing transitions on weekend nights show the lack of favorable transitions considering the category *nightlife* for Kuwait. This is not the case in Sao Paulo where the transition *food* \rightarrow *nightlife* is highly favorable to happen, suggesting that, in Sao Paulo, people like to eat before clubbing at night. In Kuwait city, instead, people are more favorable to perform the transitions *shop* \rightarrow *food* and *food* \rightarrow *home* on weekend nights.

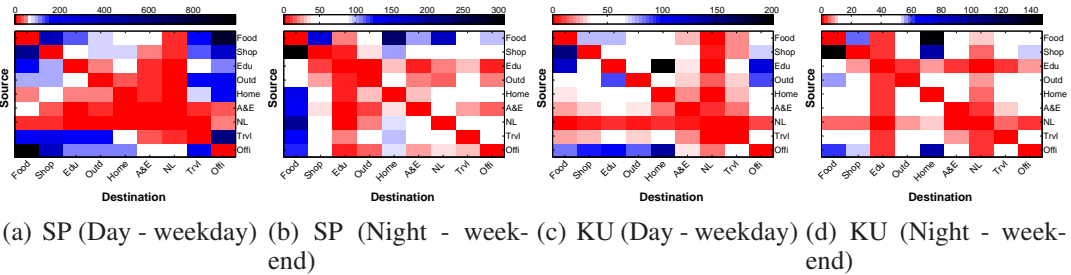


Figure 2. The City Image of Sao Paulo (SP) and Kuwait City (KU). Abbreviations: Arts & Entertainment (A&E); College & Education (Edu); Great Outdoors (Outd); Nightlife Spot (NL); Shop & Service (Shop); and Travel Spot (Trvl).

Techniques to provide easy-to-interpret visualizations of common routines of a city’s inhabitants, such as the City Image, are valuable tools to help city planners to better understand the city dynamics and make more effective decisions. They can also be used by IT designers to develop customized software services for different categories of individuals, such as recommendation of popular places to visit for tourists, and recommendation of hot spots for taxi drivers. They can also support various sorts of social studies, such as, for example, the investigation of the degree of similarity (in terms of human mobility patterns) between different cities. In particular, the City Image technique enables the clustering of cities by similarity.

2.2.2. Cultural Differences

We proposed a new methodology for identifying cultural boundaries and similarities across populations, considering eating and drinking patterns (e.g., what kind of food/drink

people prefer as well as when they often have their meals). Unlike traditional survey-based empirical studies, our methodology allows the identification of cultural dynamics much faster, capturing current cultural expressions at nearly real time, and at a much lower cost.

We studied how these preferences change according to time of day and geographical locations. We have found that the eating and drinking choices in different countries, cities, or neighborhoods of a city reveal fascinating insights into differing habits of human beings. For instance, preferences among people in cities located in the same country tend to be very similar. We also found that the time instants when check-ins are performed in food and drink places also provide valuable insights into the cultural aspects of a particular region. For example, whereas Americans and English people tend to have their main meal at dinner time, Brazilians have it at lunch time.

Given those observations, we consider spatio-temporal dimensions of food and drink check-ins as users' cultural preferences. We then apply a simple clustering technique to show the "cultural distance" between countries, cities or even regions within a city. Comparing our results with the World Values Surveys (a very large study based many years of survey data), the similarities are striking. For more details refer to Section 5.5 of the thesis.

The investigation of cultural differences across different cities and countries are valuable in many fields and can support various applications. For example, since culture is an important aspect for economic reasons, the identification of similarities between places that are geographically separated might be valuable for companies that have businesses in one country and want to assess the compatibility of preferences across different markets.

2.2.3. Points of Interest

Every city has a set of *points of interest* (POIs), that is, areas that attract more attention of people than the others. The main touristic sights of a city are included in its set of POIs. However, POIs include more than only the city sights. For example, an area of local pubs in the suburb can be quite popular among city residents, but completely unknown and unpopular among tourists. Thus, an application that naturally emerges from PSNs is related to the identification of POIs in a city. The assumption behind this idea is that each data sharing represents, implicitly, an interest of an individual at a given moment. Thus, the data sharing by many users in a particular location at a given moment serves as evidence that this place is a POI. One advantage of using PSNs to identify POIs in a city is that such technique is robust to dynamic changes. That is, because PSNs sense dynamic data, they can automatically capture the changes in people interests over time, helping to quickly identify areas that suddenly become a POI (e.g., due to the opening of a new restaurant) or that loose popularity.

We presented a technique to identify POIs and, from them, extract the main sights of a city, as explained in detail in the thesis document (Section 5.3). We applied this technique for the Belo Horizonte city, being able to find the majority of its POIs and sights using different PSNs: Foursquare and Instagram (considering different datasets spanning different periods of time for each of them). We found that different PSNs may provide complementary data, as no single PSN found all sights. Such differences may reflect changes in the city during the time when a specific dataset was collected. For example during the collection of one particular dataset, Belo Horizonte was not receiving soccer games. This explains why the soccer stadium was not identified by the PSN derived from

that dataset. In contrast, the analysis of a more recently collected dataset from the same application correctly identified the stadium as a major sight of the city. This illustrates how our technique can automatically adapt to changes in the city dynamics, detecting unusual and trendy locations as well as uncovering possibly unexpected patterns. This is thanks to the real time nature of PSNs.

The advantage of this technique is the potential to discover trendy areas in a city and offer up-to-date views of the points of interest and sights of a city. Our results are very promising because all the eight landmarks recommended by TripAdvisor as the most important cultural and leisure areas of the city are among the identified sights.

2.2.4. Other Results

Besides those contributions, we also show in the thesis document (Sections 5.2 and 5.4), that data collected from PSNs could be used in the analysis of other social, economic, and cultural aspects of inhabitants a city.

The following publications report the results discussed in this section:

- In [Silva et al. 2012b] (**Best paper award**), we propose the City Image technique. An extended version of this study is presented in [Silva et al. 2014d];
- In [Silva et al. 2013d], we propose the technique for POI identification;
- In [Silva et al. 2014c], we propose the new methodology for identifying cultural boundaries across populations;
- In [Silva et al. 2013a], we survey models and approaches applied in PSNs to support different applications and techniques.

2.3. PSNs as Sensing Layers

A sensing layer represents data, with the corresponding attributes, from a given source of data. The data represented by sensing layers have to come from a source that can be considered a sensor. Examples of data sources are web services, such as weather condition provided by “The Weather Channel”¹; traditional wireless sensor networks; income census; and PSNs. Each layer enables the access of data related to a certain aspect of the city. A range of fruitful opportunities may emerge from this idea, because as each layer represents a partial view of the city, their aggregation can provide a deeper understanding of it.

With this in mind, we propose a framework for integrating multiple sensing layers, which can be applied to more sophisticated services than services based on a single layer. Finally, we present applications that illustrate the use of the proposed framework and the potential of using multiple sensing layers. Refer to Chapter 6 of the thesis for more details.

The results for this contribution were reported in the following publications:

- In [Silva et al. 2013e], we perform a comparative study of different PSNs derived from Instagram and Foursquare, and verified if they can complement each other;
- In [Silva et al. 2014a], we, among other things, introduce the concept of sensing layers used in this work;
- In [Silva et al. 2014e] and [Silva et al. 2014b], we formalize the concept of sensing layers and present a framework for working with multiple sensing layers.

¹<http://www.weather.com>.

3. Final Remarks

In summary, in this work we show that the use of PSNs can help us better understand the dynamics of cities and urban social behavior, and from this we are able to offer smarter services to meet people's needs. Future work can be found in Section 7.2 of the thesis.

Besides the publications made from the contributions of this thesis, our work has been also honored with a prize for **best work** in “2013 Thesis Seminar” promoted by **DCC/UFMG** and featured by the media, for instance in: **MIT Technology Review** (<http://goo.gl/rpVrDi>), **Revista INFO** (<http://goo.gl/sMBEBK> and <http://goo.gl/F7DIAh>), **Irish Times** (<http://goo.gl/VGvF2p>), **Revista FAPESP** (<http://goo.gl/eALxkA>) and **Boletim UFMG** (cover) (<http://goo.gl/11B0kI>). This thesis is also a base of two **short courses** that are going to be presented at **SBRC'15** and **JAI'15**.

References

- Silva, T., Vaz De Melo, P., Almeida, J., and Loureiro, A. (2014a). Large-scale study of city dynamics and urban social behavior using participatory sensing. *IEEE Wireless Comm.*, 21(1):42–51.
- Silva, T. H., Vaz de Melo, P., Almeida, J., Viana, A., Salles, J., and Loureiro, A. (2014b). Participatory Sensor Networks as Sensing Layers. In *Proc. of IEEE SocialCom'14*, Sydney, Australia.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J., and Loureiro, A. A. F. (2012a). Uncovering properties in participatory sensor networks. In *Proc. of ACM HotPlanet'12*, Low Wood Bay, UK.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., and Loureiro, A. A. (2013a). Social media as a source of sensing to study city dynamics and urban social behavior: Approaches, models, and opportunities. In *Ubiquitous Social Media Analysis*, volume 8329, pages 63–87. Springer.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., and Loureiro, A. A. F. (2012b). Visualizing the invisible image of cities. In *Proc. of IEEE CPScom'12*, Besancon, France.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., and Loureiro, A. A. F. (2013b). Challenges and opportunities on the large scale study of city dynamics using participatory sensing. In *Proc. of IEEE ISCC'13*, pages 528–534, Split, Croatia.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., and Loureiro, A. A. F. (2013c). Uma Fotografia do Instagram: Caracterização e Aplicação. In *Proc. of SBRC'13*, Brasília, Brazil.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., Musolesi, M., and Loureiro, A. A. F. (2014c). You are What you Eat (and Drink): Identifying Cultural Boundaries by Analyzing Food & Drink Habits in Foursquare. In *Proc. of ICWSM'14*, Ann Arbor, USA.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., Salles, J., and Loureiro, A. A. F. (2013d). A picture of Instagram is worth more than a thousand words: Workload characterization and application. In *Proc. of IEEE DCOSS'13*, pages 123–132, Cambridge, USA.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., Salles, J., and Loureiro, A. A. F. (2013e). A comparison of foursquare and instagram to the study of city dynamics and urban social behavior. In *Proc. of ACM UrbComp'13*, pages 1–8, Chicago, USA.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., Salles, J., and Loureiro, A. A. F. (2014d). Revealing the city that we cannot see. *ACM Trans. Internet Technol.*, 14(4):26:1–26:23.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., Salles, J., Viana, A., and Loureiro, A. A. F. (2014e). Definição, Modelagem e Aplicações de Camadas de Sensoriamento Participativo. In *Proc. of SBRC'14*, Florianópolis, Brazil.
- Silva, T. H., Vaz de Melo, P. O. S., Viana, A., Almeida, J. M., Salles, J., and Loureiro, A. A. F. (2013f). Traffic Condition is more than Colored Lines on a Map: Characterization of Waze Alerts. In *Proc. of SocInfo'13*, pages 309–318, Kyoto, Japan.