Revealing the City that We Cannot See

THIAGO H. SILVA, Universidade Federal de Minas Gerais
PEDRO O. S. VAZ DE MELO, Universidade Federal de Minas Gerais
JUSSARA M. ALMEIDA, Universidade Federal de Minas Gerais
JULIANA SALLES, Microsoft Research
ANTONIO A. F. LOUREIRO, Universidade Federal de Minas Gerais

We here investigate the potential of participatory sensor networks derived from location sharing systems, such as Foursquare, to understand the human dynamics of cities. We propose the City Image visualization technique, which builds a transition graph mapping people's movements between location categories, and demonstrate its use to identify similarities and differences of human dynamics across cities by clustering cities according to their citizens' routines. We also analyze centrality metrics of the transition graphs built for different cities, considering transitions between specific venues. We show that these metrics complement the City Image technique, contributing to a deeper understanding of city dynamics.

Categories and Subject Descriptors: J.4 [**Computer Applications**]: Social and Behavioral Sciences; H.4 [**Information Systems Applications**]: Miscellaneous

General Terms: Measurement, Design

Additional Key Words and Phrases: Urban-computing, city dynamics, participatory sensing, location based social media

ACM Reference Format:
Thiago H. Silva, Pedro O. S. Vaz de Melo, Jussara M. Almeida, Juliana Salles, and Antonio A. F. Loureiro, 2013. Revealing the city that we cannot see. *ACM Trans. Inter. Tech.* 9, 4, Article 1 (October 2013), 22 pages.
DOI: http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Smart phones [Miller 2012] are taking center stage as the most widely adopted and ubiquitous computing device [Lane et al. 2010]. Besides their computing power, smart phones are currently available with an increasingly rich set of embedded sensors, such as GPS, accelerometer, microphone, camera, and gyroscope [Lane et al. 2010], which enable the sensing of vast areas, as people carrying their portable devices share data about their locations and opinions, and collaborate among themselves. Systems that enable data sensing in this way are named participatory sensing systems (PSSs) [Silva et al. 2013b; Burke et al. 2006].

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.
Author's addresses: T. H. Silva, P. O. S. Vaz de Melo, J. M. Almeira and A. A. F. Loureiro are with the Computer Science Department, Universidade Federal de Minas Gerais, Belo Hozizonte, MG, Brazil; Juliana Salles is with Microsoft Research, Redmond, USA.

DOI: http://dx.doi.org/10.1145/0000000.0000000

Many PSSs have already been deployed [Reddy et al. 2010; Silva et al. 2013b]. Examples are location sharing services, such as Foursquare[1], which are becoming increasingly popular. Indeed, Foursquare, created in 2009, has reportedly already passed the mark of 40 million registered users worldwide. On Foursquare, users may share their current locations by checking in at different venues, which are grouped into a set of pre-defined venue categories (e.g., restaurants). PSSs are often built on top of participatory sensor networks (PSNs). In a PSN, nodes are autonomous mobile entities (users) capable of sensing the environment they are inserted in[2].

PSNs derived from systems like Foursquare offer invaluable opportunities for the study of human social networks and human behavior "in vivo", in a natural context outside laboratories, as users have their daily routines tracked down and digitally recorded. In particular, the combination of web-scale distributed sensing and ubiquitous connectivity associated with this type of system offers an unprecedented opportunity to measure large scale city dynamics at reasonably low cost, since it enables an interface of the digital world with the physical world. For example, data shared on Foursquare allows the inference of where the currently popular restaurant areas are located in New York City and in Tokyo, as well as where people usually go to after having dinner in these two cities.

Most importantly, we note that people dynamics may change unpredictably in response to events that affect the urban areas where they live, such as an extreme weather conditions, construction work on major roads, or the opening of a new trendy bar/restaurant in a previously quiet region. Tracking such changes and their implications on the city dynamics might be time consuming and quite costly if one relies on traditional methods applied in social sciences (i.e., surveys). Instead, data shared in PSNs may reflect these changes in human behavior patterns in near real time, at a much lower cost. More broadly, PSNs enable cross-city social studies that investigate similarities and differences across cities, at a much lower cost.

In this article, we aim at investigating how we can use participatory sensor networks, specifically one derived from Foursquare, to better understand the human dynamics of different cities. In this direction, our main contributions are:

— A visualization technique, called City Image, that provides a summary of the dynamics of a city and the routines of its citizens. The City Image is based on transition graphs, which map people's movements between different location categories. We apply this technique to different cities around the world to show that this compact representation is able to capture striking features of these cities.
— We demonstrate how the City Image technique can be used to identify similarities and differences of human dynamics across cities. Specifically, we propose a methodology to cluster cities according to the routines of their citizens. The core of this methodology is to measure the distance between pairs of cities using their City Image representations.
— Finally, we analyze structural properties, notably centrality metrics, of the transition graphs built for different cities, considering transitions at the granularity of specific venues. We show that these metrics complement the City Image technique, contributing to provide a deeper understanding of the city dynamics and the routines of its citizens.

The rest of this article is organized as follows. Section 2 presents the related work. Section 3 briefly discusses the concept of PSN, whereas Section 4 presents some fundamental characteristics of the PSN derived from Foursquare. Section 5 introduces our

[1]http://www.foursquare.com
[2]The sensing activity depends on whether they want to participate in the sensing process [Silva et al. 2013b].

City Image visualization technique. Section 6 shows how the City Image technique can be used for a quantitative comparison of different cities. Section 7 discusses how the City Image technique can be complemented by the analysis of centrality metrics of transition graphs. Finally, Section 8 presents the concluding remarks and future work.

2. RELATED WORK

Several studies have analyzed the spatial properties of data shared on location sharing services such as Foursquare. For example, Cheng et al. [Cheng et al. 2011] analyzed 22 million check-ins posted on more than 1,200 applications (Foursquare is responsible for 53.5% of the total). They observed that users follow simple and reproducible patterns, and also that social status as well as geographic and economic factors are coupled with mobility. Scellato et al. [Scellato et al. 2011] studied the spatial properties of the social networks connecting users of Foursquare, Gowalla, and Brightkite. Among the results, the authors showed that 40% of the social links happen within 100 km of a location assigned as the "home" of the user.

In the same direction, Cho et al. [Cho et al. 2011] investigated patterns of human mobility in three datasets, including check-ins collected from location sharing services and cellphone location data. They were particularly interested in determining how often and how far users travel, as well as how social ties may impact such movements. They observed that short-ranged travel is spatially and temporally periodic and is not affected by the social network structure, while long-distance travel is more influenced by social network ties. Concerning the places people go, Noulas et al. [Noulas et al. 2011a] showed that the distribution of check-ins at venues presents a heavy-tailed and power-law behavior. They also observed the presence of spatio-temporal patterns in Foursquare, showing considerable distinct patterns between weekdays and weekends. Using data collected from Foursquare, Chlo et al. [Brown et al. 2013] investigated social and spatial properties of social networks in cities, and proposed a model for a place-based social network.

More closely related to our work, Doytsher et al. [Doytsher et al. 2012] proposed an application that handles a social-spatial network consisting of a social network, a spatial network, and life patterns that connect users of the social network to the locations where they regularly go. In the application, users can create queries such as "friends of Marge who buy at the same grocery store that she does", using a new query language. This work highlights the fact that data recorded by location sharing services contains an extensive amount of information about people's routines, which can be explored to build valuable applications.

Concerning the understanding of the dynamics of cities and the routines of their inhabitants, Cranshaw et al. [Cranshaw et al. 2012] presented a model to extract distinct regions of a city that reflect collective activity patterns. The idea is to expose the dynamic nature of local urban areas considering spatial and social proximities of venues, derived from the geographic coordinates and the distribution of user check-ins, respectively. Similarly, Zhang et al. [Zhang et al. 2013] also investigated the identification of neighborhood boundaries using a dataset from Foursquare, while Noulas et al. [Noulas et al. 2011b] proposed an approach to classify areas and users of a city by using venues' categories of Foursquare.

Public transportation data has also been used to study city dynamics. For example, Lathia et al. [Lathia et al. 2012] used a dataset from public transportation to show that urban mobility is a viable way to better understand dynamics of urban life. The authors correlated the mobility of London inhabitants using the public transportation system with the census-based indexes of welfare in different areas of the city. The authors obtained interesting results, such as that socially-deprived communities in London

tend to be visited more than wealthy ones. Moreover, Froehlich et al. [Froehlich et al. 2009] used a dataset of a shared bicycling system to show the underlying temporal and spatial dynamics of a city. They demonstrated that simple predictive models are able to predict bicycle station usage with high accuracy.

The studies by Santi and Oliver [Phithakkitnukoon and Oliver 2011] and Noulas et al. [Noulas et al. 2011a] are very related to our present effort. In both cases, the authors analyzed check-ins extracted from Foursquare grouped by the venue categories to better understand urban social behavior. Santi and Oliver [Phithakkitnukoon and Oliver 2011] analyzed the social activity in London, Paris, and New York. Among other findings, they verified that places from Food and Nightlife categories are the strongest social hubs across the three cities. Noulas et al. [Noulas et al. 2011a] analyzed the most common transitions between venue categories around the world, aiming at identifying sequential activity transitions.

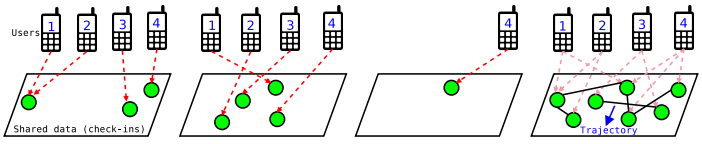
We also have previously investigated the dynamics across cities [Silva et al. 2012]. This prior work differs from those other previous efforts by two key aspects. First, we proposed a compact technique to visualize and represent city dynamics that captures people's habits and routines. Second, we assumed that social behavior highly depends on cultural and geographical factors and, unlike [Noulas et al. 2011a], we characterize the dynamics of a city based on the movement patterns (i.e., transitions) of their citizens. Thus, each city has transitions that are more and less likely to occur. The present study greatly builds upon our previous work [Silva et al. 2012] in three directions. First, we here provide a much more detailed explanation of the City Image technique, and validate its application on a much larger number of cities (30 in total). We also show how to use the City Image technique to perform a quantitative comparison of multiple cities, which is illustrated by clustering cities based on their similarities in terms of transitions. To that end, we propose a city clustering methodology, and apply it to the 30 analyzed cities, considering different periods of time. Finally, we complement our City Image technique, which is based on transitions between location categories, by proposing the analysis of centrality metrics of the transition network built by mapping people's movement between specific locations in a city. Our present effort also builds and complements other prior studies [Silva et al. 2013b; 2013a]. Whereas in these prior studies, our goal was to analyze the challenges and opportunies of studying city dynamics using participatory sensing by characterizing the spatial and temporal coverage of PSNs derived from different location-based applications, we here aim at proposing and exploring a new technique to visualize the dynamics of a city based on PSN data.

The study performed by Karamshuk et al. [Karamshuk et al. 2013] is somewhat related to the last extension. The authors investigated the problem of finding the most promising areas to open a store in New York city, using a dataset from Foursquare. To that end, they evaluated the predictive power of several machine learning features. Unlike in [Karamshuk et al. 2013], our focus is not on predicting the best spot to place a store, but rather on studying people's movements across existing locations, aiming at, for example, identifying strategic partners to make an advertising in order to direct the flow of users between two independent stores.

3. PARTICIPATORY SENSOR NETWORKS

Participatory sensing is the process where humans actively use mobile devices and cloud computing services to share local environmental data, such as their current locations and pictures they take [Burke et al. 2006]. It differs from opportunistic sensing [Lane et al. 2010] mainly by the user participation, which is key to the former and marginal to the latter. In this work, we consider that a fundamental aspect of partici-

patory sensing is the user's desire to share data, regardless of how this data is actually generated.



(a) Time 1 (b) Time 2 (c) Time 3 (d) Data Analysis

Fig. 1. Illustration of a PSN Derived from a Location Sharing Service.

Location sharing services, such as Foursquare, are examples of participatory sensing systems. The sensed data is an observation (check-in) of a user at a particular place that indicates, for instance, his/her presence at a restaurant at a certain point in time. In this work, we will use the word "check-in" to refer to an event when time and location of a particular user is recorded or, in the PSN context, sensed.

From a participatory sensing system one can derive a participatory sensor network (PSN) [Silva et al. 2013b] where the users portable devices are the fundamental building block. Individuals carrying their devices are able to sense the environment they are inserted in and make relevant observations at a personal level. Thus, each node in a PSN consists of the user plus his/her mobile device. The sensed data is sent to a server, or the "sink node". PSNs have several inherent characteristics that distinguish them from other sensor networks (e.g., wireless sensor networks), for instance, nodes are autonomous mobile entities (users), and nodes do not face severe energy constraints.

Figure 1 shows an example of a PSN derived from a location sharing service (e.g. Foursquare). The figure illustrates the locations shared by four users at three different points in time (Figures 1a, 1b and 1c), with each check-in represented by a dashed arrow. Note that users do not necessarily participate in the system all the time. Collectively, the data shared after a period of time can be analyzed in very different ways. For instance, as illustrated in Figure 1d, one can use this data to study regular user trajectories by building a network where nodes represent shared locations and edges connect shared locations by the same user. Moreover, given the ubiquity of smart phones and the increasing popularity of location sharing services worldwide, PSNs derived from such systems include people from different parts of the world, thus providing, at reasonably low cost, global scalability, as we further discuss next.

4. MAIN CHARACTERISTICS OF THE FOURSQUARE PSN

In this section, we present some characteristics of a participatory sensor network (PSN) derived from Foursquare (Section 4.1), a popular location sharing service, aiming at illustrating its potential for large scale sensing, in terms of spatial (Section 4.2) and temporal coverage (Section 4.3). A more detailed characterization of this PSN can be found in our previous work [Silva et al. 2013b].

4.1. Data Description

Foursquare is a location sharing service (also known as location-based social network) created in 2009. It is currently one of the most popular systems of its kind, registering more than 40 million users[3]. Foursquare users may share their current locations with their friends by checking in at specific virtual places (or *venues*), which, in turn,

[3]Statistic provided in Foursquare's main webpage in November 2013.

Table I. Foursquare categories

Name	Abbreviation	Sub-categories examples
Arts & Entertainment	A&E	Comedy Club, Movie Theater, Museum, Casino
College & Education	Edu	College Lab, Fraternity House, Student Center
Food	Food	Bakery, Restaurant, Coffee Shop, Pizza Place
Home	Home	Home, Residential Building
Office	Offi	Factory, Conference Room
Great Outdoors	Outd	Baseball Field, Surf Spot, Park, Cemetery
Nightlife Spot	NL	Bar, Rock Club, Nightclub, Strip Club
Shop & Service	Shop	Shoe Store, Nail Salon, Deli or Bodega, Music Store
Travel Spot	Trvl	Airport, Subway, Embassy or Consulate, Hotel

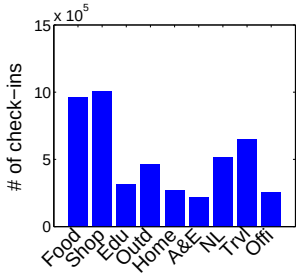


Fig. 2. Distribution of number of check-ins at all categories of venues.

represent places of the real world (e.g., a restaurant, a shop, etc). Foursquare venues are grouped into a set of pre-defined categories, and each category is further broken down into several sub-categories, as shown in Table I. Note that, in this work, we chose to separate the "Home, Work and Others" category (a single category according to Foursquare) into two distinct categories, grouping subcategories related to home as a new category named "Home", and leaving the other sub-categories under the "Office" category (abbreviation "Offi"). We did so to be able to better distinguish human dynamics using our City Image technique.

The dataset we use in this work consists of check-ins performed by Foursquare users. Since Foursquare check-ins are not publicly available by default, our data crawling was done via Twitter[4]. Specifically, we collected roughly 4.7 million tweets containing check-ins, each one providing a URL to the Foursquare website where information about the geographic location of the associated venue was acquired. For each check-in, our dataset contains the identifier of the user who did it, the latitude, longitude, identifier and category of the venue where the check-in was done, as well as the time when it was done. Our dataset comprises, in total, 4,672,841 check-ins, done in 1,929,237 different venues, during one week of April 2012.

In order to better understand our dataset, Figure 2 shows the distribution of the number of check-ins in venues of each category. As we can see, the most popular category is Shop, containing around 1 million check-ins, and the least popular is Arts & Entertainment, with 222,052 check-ins. Figure 3 shows the complementary cumulative distribution function (CCDF) of the number check-ins shared by each user. The distribution has a clear heavy tail, implying that user participation may vary widely. A heavy tail in the distribution of the number of shared check-ins and photos has also been previously observed in PSNs derived from multiple systems [Noulas et al. 2011a; Silva et al. 2013a]. Finally, Figure 4 shows the cumulative distribution function (CDF) of the time interval between consecutive check-ins by the same user regardless of the

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

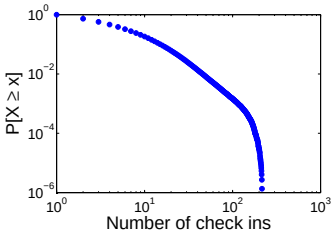


Fig. 3. Complementary cumulative distribution of total number of check-ins given by each user.

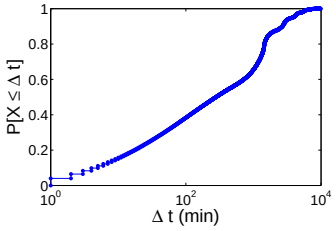


Fig. 4. Cumulative distribution of time interval between consecutive check-ins by the same user regardless of the location.

Table II. Distribution of check-ins across the selected cities.

City	# of check-ins	City	# of check-ins
Bandung/Indonesia	59.332	Mexico City/Mexico	85.721
Bangkok/Thailand	67.075	Moscow/Russia	59.654
Barcelona/Spain	9.083	New York/USA	86.867
Belo Horizonte/Brazil	18.280	Osaka/Japan	27.396
Buenos Aires/Argentina	17.762	Paris/France	11.746
Chicago/USA	27.446	Rio/Brazil	27.222
Istanbul/Turkey	103.456	San Francisco/USA	17.840
Jakarta/Indonesia	158.732	Santiago/Chile	79.733
Kuala Lumpur/Malaysia	109.048	Sao Paulo/Brazil	85.640
Kuwait City/Kuwait	34.195	Semarang/Indonesia	10.518
London/UK	15.671	Seoul/Korea	26.073
Los Angeles/USA	21.961	Singapore/Singapore	65.534
Madrid/Spain	13.004	Surabaya/Indonesia	38.021
Manila/Philippines	47.343	Sydney/Australia	6.390
Melbourne/Australia	6.182	Tokyo/Japan	118.788

location (i.e., the two check-ins may be at the same venue). As we can see, a significant fraction of all pairs of consecutive check-ins by the same user occur within a reasonably short time interval. For example, 40% of the pairs of consecutive check-ins happen within at most 100 minutes from each other. This was also observed in another Foursquare dataset [Noulas et al. 2011a] as well as in a dataset collected from Instagram [Silva et al. 2013a]. The Foursquare dataset used in this present work was also explored in one of our previous work [Silva et al. 2013b], where additional descriptive statistics can be found.

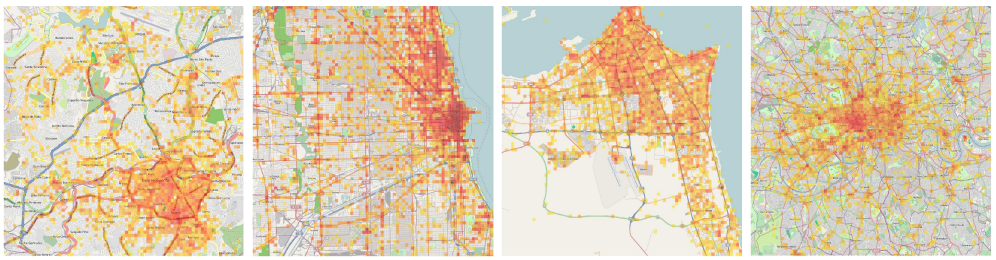
In this work, we selected 30 cities around the world to analyze. The cities and the number of check-ins available in our dataset in each of them are presented in Table II.

4.2. Network Spatial Coverage

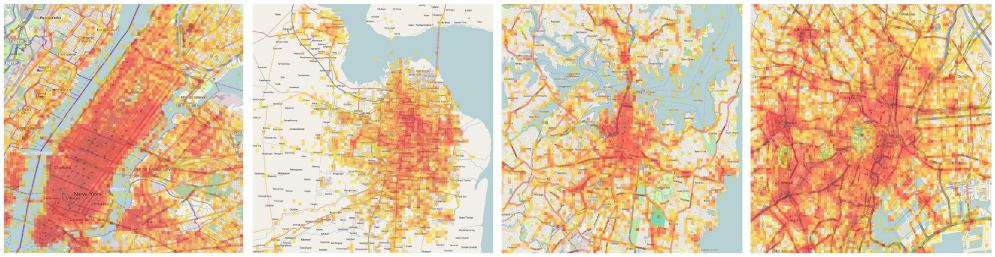
The spatial coverage of the PSN built from our Foursquare dataset is very comprehensive across the globe (figure omitted). Despite the more intense sensing activity in some continents (e.g., North America and Europe) and a higher sparsity in others (e.g., Oceania and Africa), this PSN still offers a global scale coverage at reasonably small costs. We here analyze the spatial coverage of the Foursquare PSN by focusing on eight of the selected cities, located in different continents. These cities are: Belo Horizonte (BH), Chicago, Kuwait City, London, New York (NY), Surabaya, Sydney, and Tokyo. Figure 5 shows heatmaps of the sensing activity in these cities: the darker the red color in a particular area, the larger the number of check-ins in that area. The figure shows that the PSNs of some of the cities, such as Chicago (Figure 5b), New York (Figure 5e), and Tokyo (Figure 5h), present a high coverage. On the other hand, there are some cities with very low sensing coverage, such as Sydney (Figure 5g). This was

also observed and discussed for a photo sharing system, namely Instagram [Silva et al. 2013a].

Many factors may influence the sensing coverage in a particular area. For example, economic factors might impact the usage of mobile devices by the local population, ultimately impacting sensing coverage. If most people living a given area cannot afford to buy a smartphone (or any other mobile device), the local coverage may be low. Similarly, the number of people living in a given area is also another aspect that should be taken into consideration. Since a central element of a PSN sensor is a human being, areas with low population density, such as rural areas, or areas with difficult access (e.g., high hills) are expected to have fewer data sharings (and thus lower coverage). Cultural differences are also an important aspect that must be considered. People from certain cultures might be more aware of (and worried about) privacy issues than others, and this might impact their contributions to the PSN in terms of data sharings.



(a) Belo Horizonte (b) Chicago (c) Kuwait (d) London



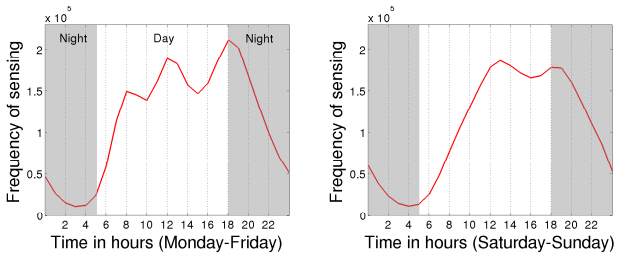
(e) New York (f) Surabaya (g) Sydney (h) Tokyo

Fig. 5. Sensed locations in eight cities. The number of check-ins in each area is represented by a heatmap. The color varies from yellow to red (higher intensity).

4.3. Seasonality

We now analyze how the seasonal behavior of humans affects the data sharing by observing the times of location sharings in our Foursquare dataset[5]. Figure 6a shows the average number of check-ins during each hour of the day, from Monday to Friday, while Figure 6b shows the same information for Saturday and Sunday. As expected, the sensing activity presents a diurnal pattern, being very low at dawn and peaking up later in the day. Considering weekdays (Figure 6a), it is possible to observe three clear peaks during the day, one around 8:00 AM (breakfast), another around 1:00 PM (lunch), and the last one at around 6:00 PM (dinner). In contrast, on weekends, there is no peak activity in the morning, the lunch peak happens around 1:00 PM, and the dinner peak is almost flat (from 6:00 PM to 7:00 PM).

[5]Timestamp is normalized according to the timezone where the check-in was performed.



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

Fig. 6. Weekdays and weekend location sharing patterns.

Thus, we note four different patterns of sensing activity depending on the time of day (day or night) and the day of week (weekdays or weekends), as highlighted in Figure 6. The patterns of check-in activity are in general distinct (except for night patterns, which are somewhat similar on weekdays and weekends), reflecting the fact that people tend to perform distinct activities on different time periods. Thus, we conduct our analysis in the next section separately for each time period.

5. THE CITY IMAGE

Similarly to Kostakos et al. [Kostakos et al. 2009], we believe that cities present distinct characteristics and evolve over time. Thus, we propose the City Image visualization technique, which exploits the movements of the city inhabitants. In summary, the City Image is a square matrix that displays a visualization of the dynamics of a city. We start by describing, in Section 5.1, a transition graph used to build the City Image. We then describe, in Section 5.2, a technique to identify and quantify the most preferred and rejected transitions (i.e., movement patterns) in a city. Finally, in Section 5.3, we show, analyze and compare the City Image for several cities.

5.1. Transition Graph

As we mentioned before, the sensing activity in a PSN is performed by mobile individuals who choose to share their information. Unlike traditional mobile wireless sensor networks, the nodes in a PSN move according to their routines or local preferences, which are dictated by the city dynamics. Thus, we propose a transition graph to map the movements of individuals in a PSN, and thus represent the city dynamics.

The proposed transition graph is a directed weighted graph $G(V, E)$, where the nodes $v_i \in V$ are the **main categories** of locations, and a direct edge (i, j) exists from node v_i to node v_j if at some point in time an individual performed a check-in at a location categorized by v_j just after performing a check-in at a location categorized by v_i. Thus, an edge represents a transition between two location categories. The weight $w(i, j)$ of an edge is the total number of transitions that occurred from node v_i to node v_j.

A transition between location categories is configured according to three requirements. First, the check-ins must be performed consecutively and by the same individual. Second, the check-ins should be performed at different venues[6]. Third, the check-ins must occur in the same "social day", which we define as the 24-hour interval starting at 5:00 am (instead of 12:00 am, since we are interested in capturing the nightlife transitions as well). Transitions that cross two different "social days" are considered only if the time interval between them is under four hours. We experimented

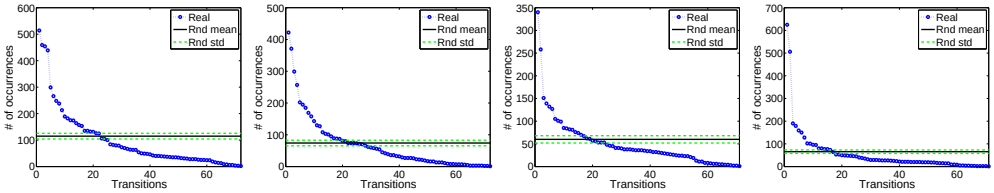
[6]The number of pairs of consecutive check-ins performed at the same venue is very small, representing at most 1.8% of the total transitions in any analyzed city.

with alternative strategies by varying the aforementioned threshold from one to five hours. The results were similar, the variability of the number of transitions that are discarded as we vary the policy is small: from approximately 2.4% to 5.2% (considering different time periods). Thus, we chose the threshold of fours hours, which is equal to the average time interval between consecutive check-ins by the same user (as shown in Figure 4).

5.2. Preferred and Rejected City Transitions

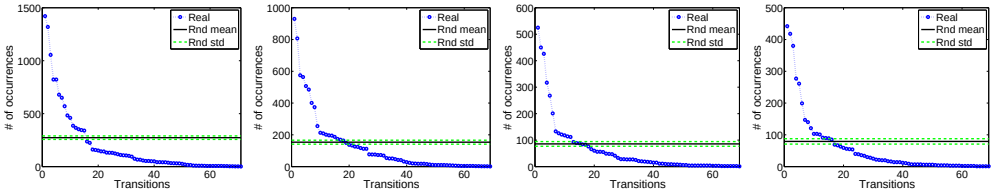
We here introduce the City Image technique, which is based on the transition graph $G(V, E)$ defined in the previous section. In summary, the City Image is a square matrix that displays a visualization of a city dynamics based on the frequency of transitions that are performed by its inhabitants.

After building the transition graph $G(V, E)$, we create ten random graphs $G_{Ri}(V, E_{Ri})$, where $i = 1, \ldots, 10$. Each such graph is built using the same number of *individual* transitions in $G(V, E)$. However, instead of considering the actual transition $v_i \rightarrow v_j$ performed by an individual (as reported in our dataset), we randomly pick a location category to replace v_j, simulating a random walk for this individual.



(a) Day - weekday (b) Day - weekend (c) Night - weekday (d) Night - weekend

Fig. 7. Observed transitions occurrences sorted in a descending order for NY city. Periods: weekday and weekend during the day and night.



(a) Day - weekday (b) Day - weekend (c) Night - weekday (d) Night - weekend

Fig. 8. Observed transitions occurrences sorted in a descending order for Tokyo. Periods: weekday and weekend during the day and night.

In Figures 7 and 8 we compare, for each pair of location categories, the number of transitions that were simulated against the number of transitions that were actually made by individuals of New York and Tokyo[7]. In these figures we consider four time periods: weekday/weekend during the day (from 5:00 am to 6:00 pm), and weekday/weekend during the night (from 6:01 pm to 4:59 am). The x-axis represents particular transitions, e.g., *Food* $\rightarrow$ *Work*, and the y-axis indicates the frequency of this particular transition. The blue curve (dotted line with a circle marker) represents the

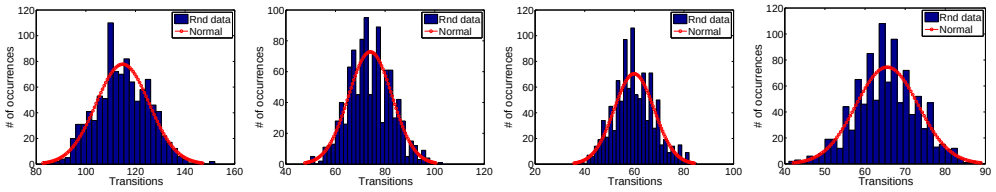
[7]These results are representative of other cities.

real transitions (i.e., represented in G), sorted in descending order of number of occurrences. The black curve (solid line) is the average number of transitions in the random graphs $G_{R1..10}$, and the two green curves (dashed lines) delimit the standard deviation. The results are shown separately for each time period. Note that, for many transitions, the number of real occurrences is significantly larger (i.e., by several standard deviations), than the expected average value in the random graphs. This implies that some transitions reflect more the preferences and habits of users from a certain city than others. There are also transitions that do not occur very often, with the number of real occurrences being much smaller than the average number in the random graphs, indicating that the inhabitants of this city strongly reject these transitions.

Based on these observations, we next identify the most and least favorable transitions to occur in a given city. To that end, we adopt one of two strategies, depending on whether the edge weights of the randomly generated graphs $G_{R1..10}$ follow a Normal distribution $N(\overline{w}, \sigma_w)$. If they are normally distributed, we compute the mean $\overline{w}$ and the standard deviation σ_w of the edge weights. We then define the *indifference range* as the interval $(\overline{w} - 3\sigma_w, \overline{w} + 3\sigma_w)$, which is expected to contain 99.73% of the randomly generated edge weight values, since the edge weights follow a Normal distribution $N(\overline{w}, \sigma_w)$. Analogously, we define the *rejection range* as the interval $[-\infty, \overline{w} - 3\sigma_w]$, and the *favouring range* as the interval $[\overline{w} + 3\sigma_w, \infty]$.

In case the edge weight distribution is not Normal, we calculate the maximum (max) and minimum (min) values of the randomly generated edge weights. We then define the *indifference range* as the interval (min, max),the *rejection range* as the interval $[-\infty, min]$, and the *favouring range* as the interval $[max, \infty]$.

For all the cities analyzed in the next section, the edge weights of the randomly generated graphs do follow a Normal distribution, as illustrated in Figures 9 and 10 for New York and Tokyo, respectively. These figures show both the histogram of the edge weights and the fitting of the Normal distribution (red curve with a circle marker). Note that, for New York city, the fitted Normal distribution has parameters $\overline{w}$ = 114.85 and σ_w = 10.712 for weekday during the day. These are the values used to delimit the *rejection range*, *indifference range* and *favouring range* for the transitions for that city in that particular time period.

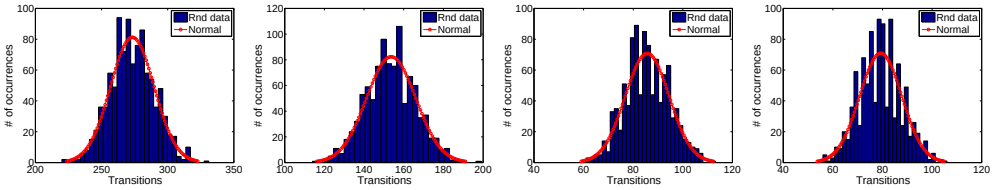


(a) Weekday - Day (b) Weekday - Night (c) Weekend - Day (d) Weekend - Night

Fig. 9. Histogram of Random Generated Transitions for NY with a Normal fitting.

5.3. Building the City Images

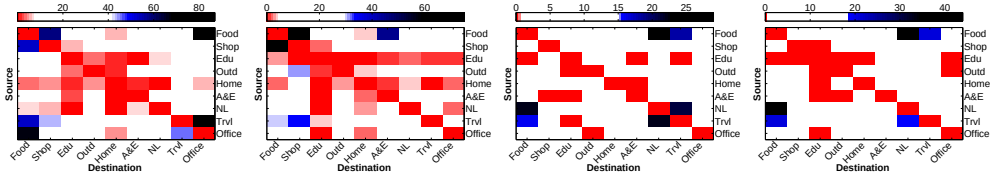
Having defined the ranges for preferred, rejected and indifferent transitions in a given city, we construct a square matrix that represents the movement patterns of the city, which is here called the City Image. In this matrix, each cell (i, j) represents the willingness of a transition from category i (line i of the matrix) to another category j (column j of the matrix). To better visualize this, we color cells that represent transitions that are *not* likely to occur in a city, i.e., transitions whose edge weight fall in the *rejection range*, in *red*. We also color transitions that are more likely to occur, i.e.,



(a) Weekday - Day (b) Weekday - Night (c) Weekend - Day (d) Weekend - Night

Fig. 10. Histogram of Random Generated Transitions for Tokyo with a Normal fitting.

transitions that fall in the *favouring range*, in *blue*. Finally, white color are used in cells that represent transitions that fall in the *indifference range*.

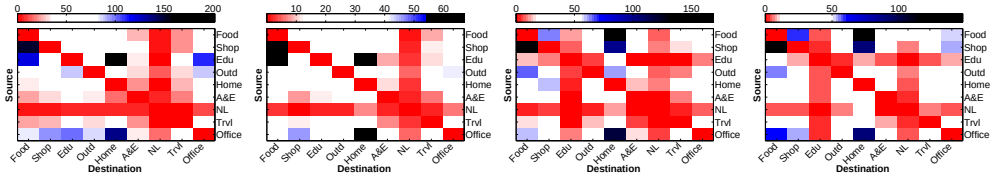


(a) Day - weekday (b) Day - weekend (c) Night - weekday (d) Night - weekend

Fig. 11. The image of London for different periods.

We built the City Image for the selected 30 cities, shown in Table II. Delving into each city, we analyze the City Image for each time period separately. Figure 11–18 present the City Images for London, Kuwait, Belo Horizonte, Chicago, Surabaya, New York, Sydney, and Tokyo. Each figure shows the City Image for one of the four time periods: weekday/weekend during the day and weekday/weekend during the night.

The City Image captures the city dynamics in a very summarized way. Nevertheless, it can reveal striking differences in the dynamics of the same city across different time periods (weekdays and weekends, day and night), as well as across different cities. Moreover, note that the main diagonal of each matrix indicates a tendency of users not having consecutive check-ins at the same category. The City Image also provides an easy way to learn the most and least favored places and transitions of each city in a given time period.

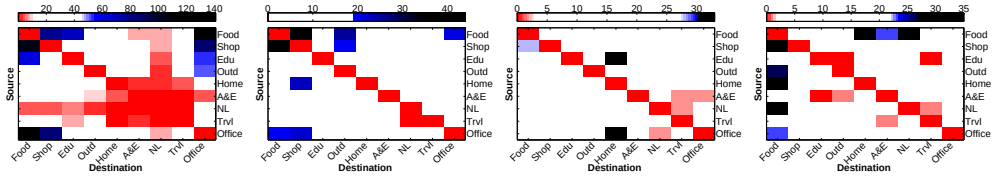


(a) Day - weekday (b) Day - weekend (c) Night - weekday (d) Night - weekend

Fig. 12. The City Image of Kuwait for different periods.

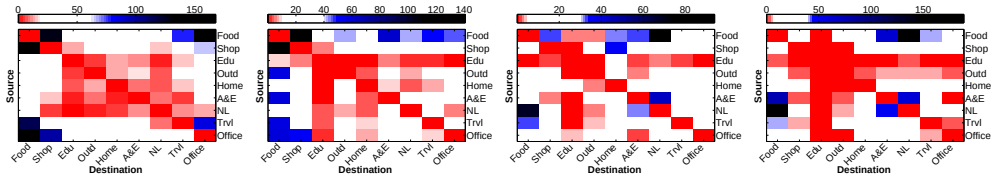
In general, using the City Image it is possible to distinguish the routines of the inhabitants of two particular cities. For instance, in Kuwait (Figure 12) and Surabaya (Figure 15) we observe the lack of favorable transitions considering the category *nightlife* for all analyzed periods. On the other hand, *nightlife* transitions are strongly favorable to happen in Chicago (Figure 14) and New York (Figure 16), not only on

weekend nights but also on weekday nights. Moreover, on weekends at night inhabitants from Kuwait and Surabaya are very favorable to perform the transitions *shop* $\rightarrow$ *food* and *food* $\rightarrow$ *home*. This might be explained by cultural differences that exist among these cities.



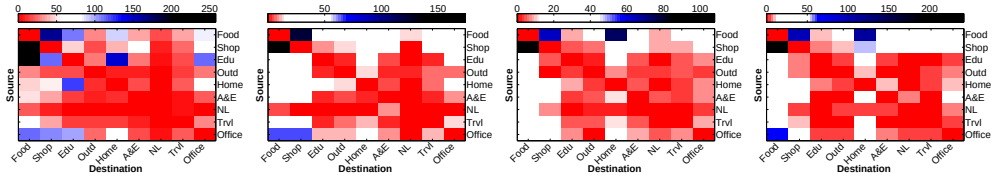
(a) Day – weekday (b) Day – weekend (c) Night – weekday (d) Night – weekend

Fig. 13. The City Image of Belo Horizonte for different periods.



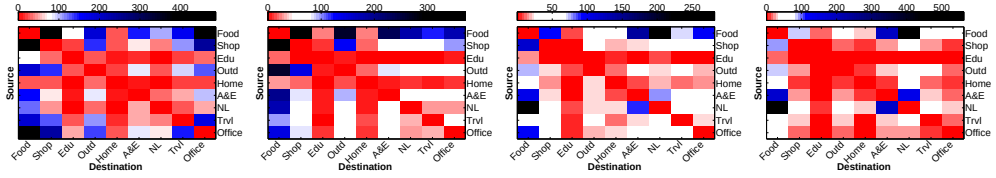
(a) Day – weekday (b) Day – weekend (c) Night – weekday (d) Night – weekend

Fig. 14. The City Image of Chicago for different periods.



(a) Day – weekday (b) Day – weekend (c) Night – weekday (d) Night – weekend

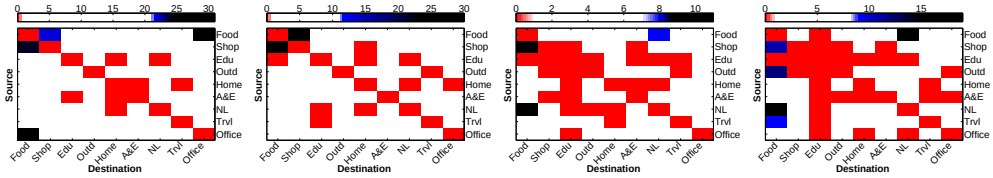
Fig. 15. The City Image of Surabaya for different periods.



(a) Day – weekday (b) Day – weekend (c) Night – weekday (d) Night – weekend

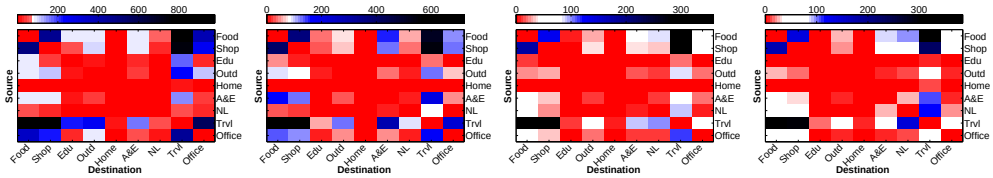
Fig. 16. The City Image of New York for different periods.

As another example, note that inhabitants of Belo Horizonte (Figure 13) are highly favorable to perform transitions containing the category *education*. This comes with no surprise since this city is an important hub of education in Brazil. In this particular City Image it is also worth noting that the transition *education* $\rightarrow$ *office* is favorable. This is because, many students in Belo Horizonte do keep a (part-time or full-time) job.



(a) Day – weekday (b) Day – weekend (c) Night – weekday (d) Night – weekend

Fig. 17. The City Image of Sydney for different periods.



(a) Day – weekday (b) Day – weekend (c) Night – weekday (d) Night – weekend

Fig. 18. The City Image of Tokyo for different periods.

This also explains the favorable transition *education* $\rightarrow$ *home* on weekdays at night, as many students who have a full-time job go to school at night. In contrast, we find that Chicago residents tend to reject any transition involving the category *education* for all analyzed periods. This is surprising, since Chicago has been a world center of higher education and research, with several universities located in the city.

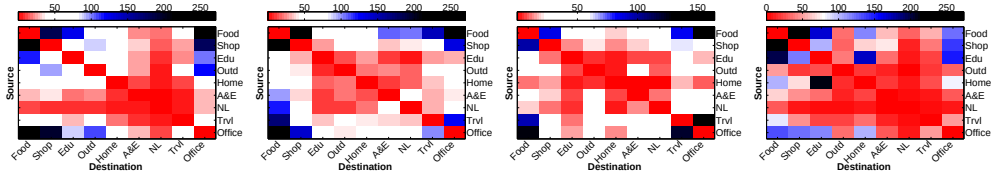
We also note that one of the most favored transitions in London (Figure 11) on weekdays during the day is *travel* $\rightarrow$ *office*. A similar trend also happens in other cities, such as New York and Tokyo (Figure 18). On the other hand, some cities, such as Belo Horizonte, Sydney (Figure 17), Kuwait and Surabaya, do not present favorable transitions containing the category *travel* on weekdays during the day. This could be associated with a larger number of people who choose to drive to get to their destinations, instead of taking public transportation.

The City Image technique, as illustrated above, is an interesting way to better understand the invisible image of a city. It provides a useful tool in various contexts, ranging from helping city planners to better understand the actual dynamics of a city, to providing tourists another source of information that might help them make their travel choices. The transition tendencies further serve as a source of fundamental information for social behavior study.

It is important to note some limitations of our dataset. First, it reflects the behavior of a fraction of the city citizens (those who actively use Foursquare). Second, since we only have a sample of the activities that occurred, external factors, such as bad weather conditions, might have affected the total number of check-ins we collected for some places, especially those at locations of the outdoor category. Nevertheless, although these limitations do prevent us from making some general assertions, they do *not* invalidate our City Image technique.

Another possible limitation of our dataset is the covered time interval, one week, which might be considered short. In order to assess to which extent this might impact the conclusions drawn from the City Images, we collected the check-ins performed on the cities of Belo Horizonte, Chicago, London, and Surabaya in the week following the period covered by our original dataset. We then recalculated the City Images for each of these cities using all the data available, thus covering a time interval of two weeks. We show the results for weekdays during the day, which is the the period where most of

the routines are performed, in Figure 19. We can observe that the new City Images are very similar to the corresponding ones produced using our original one-week dataset (Figures 13a, 14a, 11a, and 15a for Belo Horizonte, Chicago, London, and Surabaya, respectively). The strong favorable or rejection transitions remain basically the same, whereas the changes, if observed, occur in some transitions classified in the indifference range. These particular changes are expected because the larger dataset enables a clearer image of the analyzed city. The same strong similarities were observed for the City Images produced for the other periods of time (e.g., weekend night). Thus, even with a single week of data, the City Image technique is able to reveal remarkable and consistent patterns of each analyzed city.



(a) Belo Horizonte (b) Chicago (c) London (d) Surabaya

Fig. 19. The image of cities in different regions of the world during the day on weekdays.

6. QUANTITATIVE COMPARISON OF CITIES

An application that naturally emerges from the City Image technique is the numerical comparison of different cities, by exploiting the values in each square matrix. Specifically, we propose to compare two cities i and j by following the steps:

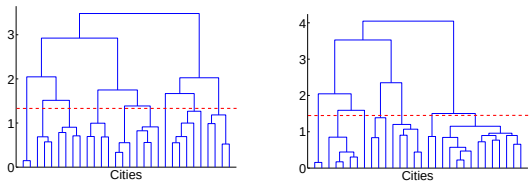
(1) For each city i, the weight of each transition t of its City Image is normalized by the maximum weight of all transitions in this particular City Image. We refer to this normalized value as t'_i. As a result, we produce a vector $T_i = (t'_{i,1}, t'_{i,2}, ...t'_{i,81})$ containing all normalized transitions (total of 81, as there are 9 location categories) for a specific City Image;
(2) We then compute the Euclidean distance $d_{i,j}$ between each pair of vectors (T_i, T_j) of cities i and j. By doing so we are calculating the distance between each considered city for all transitions.

More generally, the comparison of multiple cities produces a vector D containing the distance between each pair of cities. Vector D could then be used in several ways. For example, it could be exploited to cluster cities by similarity (in terms of movement patterns), as shown in the following steps:

(1) Build a hierarchical cluster tree for the cities based on the distances in vector D using, for example, the Ward's method [Ward Jr 1963]. This is a general agglomerative hierarchical clustering procedure, where the criterion for choosing the pair of clusters to merge at each step is based on the optimal value of an objective function. In our case, this objective function is the minimum total intracluster variance, which is computed based on the distances D;
(2) Determine the number of clusters c to be generated by visually inspecting the hierarchical cluster tree created, using, for example, a dendrogram plot of the tree;
(3) Prune the tree created in step 1 in order to have c clusters.

We applied this procedure to compare and cluster the 30 cities analyzed in Section 5.3, considering two different time periods: weekdays during the day, to study the typical time when users perform their main routines; and weekend during the night, to

study the typical period when people perform leisure activities. Figure 20 shows the dendrograms built for each period. The red lines (dashed ones) indicate the cuts used to define the number of clusters c in each case. We defined c equal to 9 clusters for weekdays during the day and 7 clusters for weekend during the night.



(a) Day – weekday (b) Night – weekend

Fig. 20. Dendrogram plots for the binary cluster tree of 30 different cities, in two different time periods.

Tables III and IV show the clustering results for weekdays during the day and weekends during the night, respectively. Note that, in general, cities from the same country or that are geographically close to each other were grouped together. The geographical proximity, which may reflect, to some extent, cultural similarity, is favorable to produce a similar behavior between the inhabitants from those cities, and might be the explanation to the clustering results. However, there are exceptions. For example, for weekdays during the day, San Francisco was grouped apart from other American cities, whereas Bangkok, far away from USA, was grouped in the same cluster as some American cities. Thus, the inhabitants of cities of the same country do not necessarily have similar behavior, reflecting heterogeneous patterns which are natural to occur in large countries, such as USA. Conversely, large geographical distances also do not necessarily imply large differences in people's habits. For instance, cities with good transportation system or many options for outdoor activities, such as beaches and parks, tend to favor transitions containing $travel$ and $outdoor$, regardless of their particular geographical location, and tend to differ from other cities, even cities in the same country, that do not have such facilities.

We note that the proposed city clustering procedure and the city distance metric could be applied to a much larger number of cities in the world, with several potential applications. One example is a personalized city recommendation system for supporting tourism-oriented applications. Such application could explore the proposed city clustering strategy to suggest new cities that the user might like, based on the user's interests (which could be inferred from prior user's interactions in the system). For example, by learning that a user liked Bandung during the day, the application might suggest Surabaya as a city to visit, as the two cities are grouped in the same cluster and thus have similarities. Location-based social media (like Foursquare) could benefit from this strategy to improve their current recommendation systems, by introducing the City Image as a new criteria.

7. VISUALIZING CITIES THROUGH NETWORK CENTRALITY METRICS

Many metrics of node centrality can be used to estimate the relative importance of a node within the graph. Although most of these metrics were first developed in social network analysis [Newman 2010], they can also be applied to a transition graph, similar to the one proposed in Section 5, enabling the study of city dynamics. Thus, in this section we build a transition graph where each node represents a specific location (and not location category, as in Section 5), and a direct edge (i, j) exists if someone performed a check-in at location j after a check-in at location i. The weight of the edge

Table III. Clustering results for weekday during the day.

Cluster	Cities
1	Bandung, Semarang, Surabaya
2	London, Paris, Madrid
3	Kuwait, Singapore, Moscow, Santiago
4	Sydney, Melbourne, Seoul, San Francisco
5	Rio, Belo Horizonte, Sao Paulo, Barcelona, Buenos Aires
6	Jakarta, Kuala Lumpur, Manila, Mexico City
7	Los Angeles, Chicago, New York, Bangkok
8	Tokyo, Osaka
9	Istanbul

Table IV. Clustering results for weekend during the night.

Cluster	Cities
1	Kuwait, Singapore, Kuala Lumpur, Manila, Bangkok
2	Tokyo, Osaka
3	Seoul, Jakarta, Bandung, Semarang, Surabaya
4	Rio, Belo Horizonte, Sao Paulo
5	Istanbul, Moscow
6	Santiago
7	Los Angeles, Chicago, San Francisco, New York, Melbourne, Sydney, Paris, Madrid, London, Barcelona, Buenos Aires, Mexico City

reflects the number of transitions between the two specific locations. These transitions are configured according to the same requirements defined in Section 5.

Table V. Centrality metrics for NY during the day and night.

Day						Night					
Degree		Closeness		Node Betweenness		Degree		Closeness		Betweenness	
Value	Venue	Value	Venue	Value	Venue	Value	Venue	Value	Venue	Value	Venue
0.04	Yankee S.	0.18	Yankee S.	0.1	Yankee S.	0.01	Yankee S.	0.06	Yankee S.	0.02	Yankee S.
0.02	Penn S.	0.18	Penn S.	0.05	Penn S.	0.007	Penn S.	0.06	Penn S.	0.01	Penn S.
0.02	Grand C.	0.18	Times S.	0.04	Grand C.	0.007	Times S.	0.06	Mad. S. G.	0.007	Tribeca F. F.
0.02	Mad. S. G.	0.17	Grand C.	0.03	Times S.	0.006	Mad. S. G.	0.06	Times S.	0.007	Mad. S. G.
0.02	Times S.	0.17	Mad. S. G.	0.03	Mad. S. G.	0.005	Tribeca F. F.	0.06	Tribeca F. F	0.007	Times S.
0.01	Bryant	0.17	Bryant	0.03	Union S.	0.005	Grand C.	0.06	Grand C.	0.006	Grand C.
0.01	Union S.	0.17	Union S.	0.03	Bryant	0.004	Webster H.	0.06	Bowery B.	0.004	Webster H.
0.01	Wash. S.	0.17	Int. Auto Show	0.02	Wash. S.	0.003	Union S.	0.06	Term. 5	0.003	Bryant
0.009	MoMa	0.17	Rockef. C.	0.02	Mad. Sq. P.	0.003	Bowery B.	0.06	Brook. Bowl	0.003	Pacha
0.008	Port A.	0.16	Port A.	0.01	Port A.	0.003	Port A.	0.06	Pacha	0.003	Radio City

Traditionally used centrality metrics are degree, closeness and betweenness centrality [Bonacich 1987]. These metrics aim to identify nodes that have central locations within the network structure. Since nodes in our networks represent locations, a central node may indicate a strategic point in the city, according to a specific metric. For example, the main idea behind the degree centrality is to identify the total number of links incident to a node, i.e., the number of incoming and outgoing edges that a node has. In our transition networks, a node with high degree indicates a location where people may arrive and depart with a high probability. Thus, degree centrality is a good measure to identify popular places in the city. These locations can be seen as city hubs.

The closeness centrality metric is related to how close a node is to all other nodes in the network, i.e., the number of edges separating a node from the others. In the context of information dissemination, the higher the closeness of a place, the higher the probability that a piece of information being disseminated from that place reaches the whole network in the least amount of time. In the perspective of a transition graph, the closeness centrality may indicate favorable locations in the network structure to start the dissemination of information to the whole network. These locations may be strategic places to install public information centers to disseminate, for example, alerts using users' portable devices in an ad hoc manner.

Finally, the main idea behind the betweenness centrality is to show how often a node is in the shortest path between any two other nodes. In our transition networks, it may indicate the most interesting locations to act as bridges to carry information among different places or regions of places (set of places). That is, the higher the betweenness of a location, the higher the chance that a user passes through that particular location.

One could explore these central nodes to sign a commercial agreement to increase their revenues by, for instance, making an advertising in order to direct flow of users to other independent business venues in the city.

We illustrate the use of these centrality metrics by showing in Table V the top-10 locations with the largest degree, betweenness, and closeness centrality values in New York. The table presents results for two time periods, day (5:00 am to 7:00 pm) and night (6:00 pm to 6:00 am)[8], aggregating results for weekdays and weekends for the sake of avoiding hurting the presentation with excessive data. Note that most top-10 locations, according to all metrics, are widely known. Some of these locations, such as Yankee Stadium (Yankee S.), are in the top-10 according to all metrics and in both analyzed periods, whereas others appear in the top-10 list of only one metric, such as MoMa which is listed only in the degree centrality column. This demonstrates that different centrality metrics may identify different central places.

We note that the Tribeca Film Festival (Tribeca F. F.) was identified as a central place in all metrics during the night. Foursquare encouraged users to check-in in this event offering a special badge for it. This justifies the large number of check-ins and, thus, the increase of centrality. Since in the studied network nodes are venues and venues tend to be dynamic, a temporal analysis when studying centrality is desirable. In this case, it would be possible to identify that Tribeca F. F. was a temporary venue, and thus avoid considering it a central location after its expiration date.

We also note the greater diversity of central locations across metrics for the night period. In other words, there is a larger number of locations that appear among the top-10 according to only one or two metrics during the night. The type of these locations might help explain the results. Observe that nightlife places, such as Pacha and Brooklyn Bowl, are not listed in the top-10 locations with highest degrees. Yet, they are amongst the locations with highest betweenness and closeness values. This could be explained by the routine of people, who usually go to a pub or a restaurant before going to a nightlife spot. This first visited location might not be very popular, e.g. a random place close to the user's house that might be far away from the target place (nightlife spot). This could connect different regions from the network, helping to increase the betweenness of the first location. Alternatively, the first visited location could be a popular place, helping to increase the closeness.

Network visualization: The visualization of transition graphs, specially highlighting central places, is interesting because it gives fascinating insights into how people move and interact with the city. The edges in the transition graphs represent somehow a rudimentary GPS tracking. After aggregating the transitions performed by all users, the final network enables the reconstruction of typical paths that users take to move in the city. When representing the information of centrality of a place in this network we are also able to visualize and understand better how users interact with the city. Figures 21, 22, and 23[9] show such networks for Belo Horizonte and New York, during the day and night, for the degree, betweenness, and closeness centrality, respectively. Each color represents a category of place, as defined in the caption of Figure 21.

Studying the results for New York, for example, it is possible to observe that during the day there is an intense movement of people between Manhattan, New Jersey, Brooklyn, and Queens, where Manhattan is the central destination. However, during the night the movement of people between Manhattan and New Jersey is much lower,

[8]If one transition happened in the overlapped hours (5:00 am to 6:00 am, or 6:00 pm to 7:00 pm), it is considered a transition of day and night periods, respectively. NY has 49,849 check-ins during the day and 19,491 check-ins during the night.

[9]The area represented by those networks is the same as the one shown in Figure 5. Nodes disposition respects their geo-location in the city.

although the movement between Manhattan, Brooklyn and Queens is still quite intense. This might indicate that people from New Jersey tend to go to Manhattan more often to work during the day than for leisure time at night.

As another example, Figures 22a and 22b show that, during the day, both New York and Belo Horizonte have a few places that stand out with higher betweenness values than the others in the same city. This does not happen in the same proportion for the degree centrality, as shown in Figures 21a and Figure 21b. Moreover, the same discrepancies can not be observed for neither centrality metric during the night (Figures 21c, 21d, 22c, and 22d), which might be explained by the lack of peoples' routines.

Regarding the closeness metric, we can see a large number of places with high closeness during the day in both cities (Figures 23a and 23b), implying that there are many options of places to select in case one wishes to install alert dissemination schemes in the city, for example. Note also that places with high closeness are relatively well spread in both cities during the day. However, this is not the case during the night (Figures 23c and 23d). The results in this period follow the same tendency observed for the other metrics and the explanation might be the same, i.e., lack of well defined routines.

Information summarization: Tables VI, VII, and VIII show the summarization of values of each centrality metric (degree (D), betweenness (B), and closeness (C)), calculated for all places during the day and night in Belo Horizonte, New York, and Tokyo, respectively. The summarization is expressed by the percentage relative to the total of values by category of places. For example, in Table VI we can see that, during the day, all places of the category Food represent 17.7% of all degree centrality observed. These tables we help us to visualize the cities by their most important classes of places. Analyzing the top degree centrality during the day we can observe that inhabitants of Belo Horizonte concentrate a lot of activities in education, shopping and working (represented by the category Office), having the categories of places Food[10], Office, Shop and Education as the most popular. Following the same analysis, places related to working, shopping, and nightlife are quite central in New York. Studying now the centrality in Tokyo it is interesting to observe the high amount of activity in Travel places, probably related to public transportation spots. Note the high value for betweenness and the considerable lower value for closeness. This means that inhabitants of Tokyo might use public transportation to move to areas with not many central places, such as suburbs, justifying the values observed for betweenness and closeness.

Regarding to privacy issues, observe the centrality in the category of places Home. In Belo Horizonte the number of check-ins is expressive in this category. However, in NY and mainly in Tokyo people do not appear to have the same behavior. This fact might be explained to cultural differences. It is known that Japanese people are concerned with privacy issues, and apparently Brazilians are not as concerned.

Differences in the habits of inhabitants of the cities can also be captured by those tables. During the night, places related to education are still quite central in Belo Horizonte, but not in NY or Tokyo. This is explained because night courses in schools and universities are common in Belo Horizonte, since many people have to work during the day to pay their studies. In New York, as expected, the centrality of places related to nightlife and arts & entertainment is high. On the other hand, shopping places have high centrality in Tokyo for this considered period. This analysis illustrates how we can visualize characteristics of cities, and the potential of using it to differentiate them.

[10]We consider that food activities are complementary to a main activity, such as work or study, for this reason we are not mentioning it as a main activity

Table VI. Summarization of values of each centrality metric calculated for all places in BH (day and night). D=degree, B=betweenness, C=closeness.

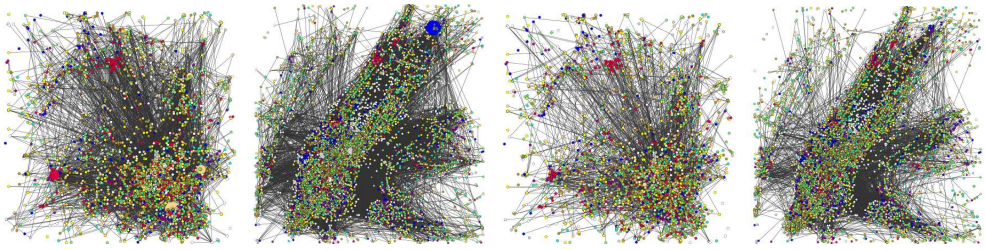
Categ.	D. (%)	B. (%)	C. (%)
		Day	
Food	17.7	14.9	21.6
Shop	13.8	22.5	12.9
Edu	14.5	15.8	10.6
Outd	9.1	15.3	6.3
Home	9.1	4.7	11.04
A&E	3.4	3.6	4.3
NL	4	3.2	5.7
Trvl	5.3	6.5	4.7
Offi	20.4	12.7	20
none	2	1	2.9
		Night	
Food	18.7	19.5	23.3
Shop	9.5	16.1	7.9
Edu	11.3	11.9	9.1
Outd	9.26	16.5	8.6
Home	15.3	6.2	14
A&E	5.5	5.6	5.2
NL	10.3	14.1	13.6
Trvl	3.9	3.4	4
Offi	14.6	6.1	13
none	1.5	0.3	1.4

Table VII. Summarization of values of each centrality metric calculated for all places in NY (day and night). D=degree, B=betweenness, C=closeness.

Categ.	D. (%)	B. (%)	C. (%)
		Day	
Food	29.5	21.8	33.3
Shop	13.5	13.2	15.1
Edu	2.5	1.9	2.5
Outd	8.8	15.2	6.1
Home	2	1	3.1
A&E	9.5	15.6	6.2
NL	10.2	8.4	10.4
Trvl	7	10.9	5.7
Offi	14.7	11	14.2
none	2	0.9	3.3
		Night	
Food	31.1	22.7	36.4
Shop	7.4	6.4	7.8
Edu	1.5	0.6	1.5
Outd	5.8	8.3	5
Home	3.2	1.1	3.6
A&E	10	17.3	7.2
NL	23.4	27.1	23.7
Trvl	6.6	9.9	6.1
Offi	9.4	6.3	6.8
none	1.6	0.5	2

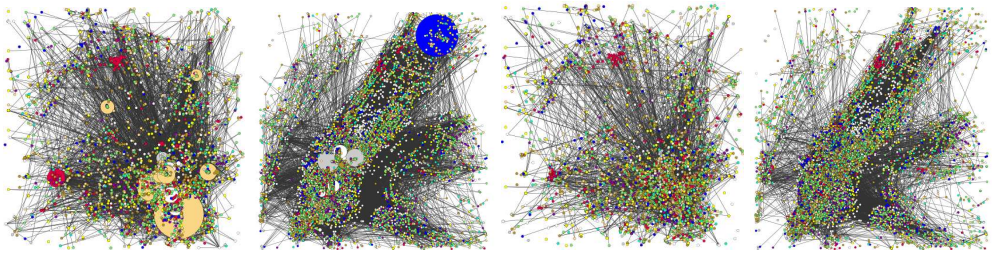
Table VIII. Summarization of values of each centrality metric calculated for all places in Tokyo (day and night). D=degree, B=betweenness, C=closeness.

Categ.	D. (%)	B. (%)	C. (%)
		Day	
Food	25.4	15.7	39.2
Shop	16.3	13.9	18
Edu	3.1	1.8	3
Outd	4	4.2	4.8
Home	0.2	0.1	0.4
A&E	5.1	4.2	4.8
NL	2.9	1.3	5.7
Trvl	32.8	50.8	11.8
Offi	8.8	7.4	10
none	1.3	0.6	2.4
		Night	
Food	26.9	10.8	35.4
Shop	13.3	11.1	15.7
Edu	1.1	0.6	1.1
Outd	3	3.2	3.7
Home	0.4	0.1	0.7
A&E	5	3.2	5.4
NL	7.5	3.6	10.8
Trvl	35.8	63.5	20
Offi	5.4	2.8	5.4
none	1.4	0.6	1.8



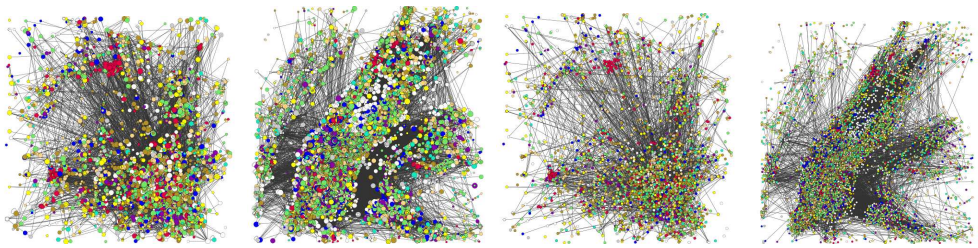
(a) BH - Day (b) NY - Day (c) BH - Night (d) NY - Night

Fig. 21. Node degree - For two cities in different countries. Each node color represents an specific category of places. Blue=Arts& Entertainment; Red = College & Education; Light Green = Food; Yellow = Home; Green Moss = Office; Purple = Nightlife Spot; White = Great Outdoors; Beige = Shop & Service; Grey = Travel spot; Cyan = no category.



(a) BH - Day (b) NY - Day (c) BH - Night (d) NY - Night

Fig. 22. Node Betweenness - For two cities in different countries. Colors legend: see caption of Figure 21.



(a) BH - Day (b) NY - Day (c) BH - Night (d) NY - Night

Fig. 23. Node Closeness - For two cities in different countries. Colors legend: see caption of Figure 21.

8. CONCLUSIONS AND FUTURE WORK

Participatory sensor networks (PSNs) have the potential to become a fundamental tool to study social behavior at large scale. A simple and important type of sensing data is human location, which clearly captures one of the key dimensions of the dynamics of a city. Currently popular location sharing services, such as Foursquare, allow users to share their actual locations, which are associated with different location categories. Thus, PSNs derived from such services provide an unprecedented opportunity to analyze large scale city dynamics.

In this article we investigated the potential of PSNs derived from Foursquare to study city dynamics. First, we presented a visualization technique called City Image, and illustrated its use in 30 different cities around the world. This technique summarizes the city dynamics based on transition graphs that map the movements of individuals between different location categories in the PSN. We also showed the use of this technique for clustering cities based on their similarities in terms of movement patterns, which can be exploited to build city recommendation systems (see below). Finally, we investigated the use of centrality metrics, computed on transition networks built at the granularity of specific venues, as a means to complement the City Image technique towards a deeper understading of the city dynamics.

The proposed City Image technique can be a valuable component in the design and improvement of various socio-technical systems, such as:

— New recommendation systems for driving and supporting tourism-oriented applications. For instance, a city recommender system could exploit the similarity between cities, captured by our City Image technique, as well as information about the profile and interests of users to provide personalized city recommendations;
— New tools to support city planners to detect (in near-real time) and react to changes in the dynamics of the city. For example, urban traffic actions could be employed to react to the appearance of new crowded areas;
— New customer recommendation systems for taxi drivers (and other classes of workers) to help them meet the current demand in different regions of the city.

Possible directions for future work include: extending our study to include other types of participatory sensing systems; exploiting the City Image technique and the proposed city clustering methodology to build new recommendation services (such as the aforementioned ones), and investigating privacy issues related to the City Image technique.

REFERENCES

P. Bonacich. 1987. Power and Centrality: A Family of Measures. *The Amer. Jour. of Sociology* 95, 5 (1987), 1170–1182.

Chlo Brown, Anastasios Noulas, Cecilia Mascolo, and Blondel Vincent. 2013. A Place-focused Model for Social Networks in Cities. In *Proc. of SocialCom'13*. Washington, USA.

J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. 2006. Participatory sensing. In *Proc. Workshop on World-Sensor-Web (WSW)*.

Zhiyuan Cheng, James Caverlee, Kyumin Lee, and Daniel Z. Sui. 2011. Exploring Millions of Footprints in Location Sharing Services. In *Proc. of ICWSM'11*. Barcelona, Spain.

Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proc. of KDD'11*. San Diego, USA, 1082–1090.

Justin Cranshaw, Raz Schwartz, Jason I. Hong, and Norman Sadeh. 2012. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In *Proc. of ICWSM'12*. Dublin, Ireland.

Yerach Doytsher, Ben Galon, and Yaron Kanza. 2012. Querying Socio-spatial Networks on the World Wide Web. In *Proc. of WWW'12*. Lyon, France, 329–332.

Jon Froehlich, Joachim Neumann, and Nuria Oliver. 2009. Sensing and Predicting the Pulse of the City through Shared Bicycling. In *Proc. of the 21st Int'l J. C. on Art. Int. (IJCAI'09)*. 1420–1426.

Dmytro Karamshuk, Anastasios Noulas, Salvatore Scellato, Vincenzo Nicosia, and Cecilia Mascolo. 2013. Geo-spotting: mining online location-based services for optimal retail store placement. In *Proc. of KDD '13 (KDD '13)*. Chicago, Illinois, USA, 793–801.

Vassilis Kostakos and others. 2009. Understanding and Measuring the Urban Pervasive infrastructure. *Personal and Ubiquitous Computing* 13, 5 (June 2009), 355–364.

Nicholas D. Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T. Campbell. 2010. A survey of mobile phone sensing. *Comm. Magazine, IEEE* 48, 9 (Sept. 2010), 140–150.

Neal Lathia, Daniele Quercia, and Jon Crowcroft. 2012. The Hidden Image of the City: Sensing Community Well-Being from Urban Mobility. In *Proc. of Pervasive'12*. Newcastle, UK.

Geoffrey Miller. 2012. The Smartphone Psychology Manifesto. *Perspectives on Psychological Science* 7, 3 (2012), 221–237.

Mark Newman. 2010. *Networks: an introduction*. Oxford University Press, Inc.

Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. 2011a. An Empirical Study of Geographic User Activity Patterns in Foursquare. In *Proc. of ICWSM'11*. Barcelona, Spain.

Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. 2011b. Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-based Social Networks. In *Proc. of ICWSM'11*. Barcelona, Spain.

Santi Phithakkitnukoon and Patrick Oliver. 2011. Sensing Urban Social Geography Using Online Social Networking Data. In *Proc. of ICWSM'11*. Barcelona, Spain.

Sasank Reddy, Deborah Estrin, and Mani Srivastava. 2010. Recruitment Framework for Participatory Sensing Data Collections. In *Proc. of Pervasive'10*. Helsinki, Finland, 138–155.

Salvatore Scellato, Anastasios Noulas, Renaud Lambiotte, and Cecilia Mascolo. 2011. Socio-spatial Properties of Online Location-based Social Networks. In *Proc. of ICWSM'11*. Barcelona, Spain.

Thiago H. Silva, Pedro O. S. Vaz de Melo, Jussara M. Almeida, and Antonio A. F. Loureiro. 2013a. A picture of Instagram is worth more than a thousand words: Workload characterization and application. In *Proc. of IEEE DCOSS'13*. Cambridge, USA, 123–132.

Thiago H. Silva, Pedro O. S. Vaz de Melo, Jussara M. Almeida, and Antonio A. F. Loureiro. 2013b. Challenges and opportunities on the large scale study of city dynamics using participatory sensing. In *Proc. of ISCC'13*. Split, Croatia, 528–534.

Thiago H. Silva, Pedro O. S. Vaz de Melo, Jussara M. Almeida, Juliana Salles, and Antonio A. F. Loureiro. 2012. Visualizing the invisible image of cities. In *Proc. of IEEE CPSCom'12*. Besancon, France.

Joe H Ward Jr. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* 58, 301 (1963), 236–244.

Amy X. Zhang, Anastasios Noulas, Salvatore Scellato, and Cecilia Mascolo. 2013. Hoodsquare: Modeling and Recommending Neighborhoods in Location-based Social Networks. In *SocialCom'13*. Washington, USA.