International Journal of Spatial Data Infrastructures Research, 2009, Vol. 4, 175-200

Quantifying urban attractiveness from the distribution and density of digital footprints*

Fabien Girardin[1,3], Andrea Vaccari[1], Alexandre Gerber[2], Assaf Biderman[1], Carlo Ratti[1]

[1]Massachusetts Institute of Technology, Cambridge, MA, USA
{fabieng, avaccari, abider, ratti} @ mit.edu
[2]AT&T Labs - Research, Florham Park, NJ, USA
gerber@research.att.com
[3]Barcelona Media, Barcelona, Spain

Abstract
In the past, sensors networks in cities have been limited to fixed sensors, embedded in particular locations, under centralised control. Today, new applications can leverage wireless devices and use them as sensors to create aggregated information. In this paper, we show that the emerging patterns unveiled through the analysis of large sets of aggregated digital footprints can provide novel insights into how people experience the city and into some of the drivers behind these emerging patterns. We particularly explore the capacity to quantify the evolution of the attractiveness of urban space with a case study of in the area of the New York City Waterfalls, a public art project of four man-made waterfalls rising from the New York Harbor. Methods to study the impact of an event of this nature are traditionally based on the collection of static information such as surveys and ticket-based people counts, which allow to generate estimates about visitors' presence in specific areas over time. In contrast, our contribution makes use of the dynamic data that visitors generate, such as the density and distribution of aggregate phone calls and photos taken in different areas of interest and over time. Our analysis provides novel ways to quantify the impact of a public event on the distribution of visitors and on the evolution of the attractiveness of the points of interest in proximity. This information has potential uses for local authorities, researchers, as well as service providers such as mobile network operators.

Keywords: digital earth, urban studies, urban indicators, reality mining, digital footprints, pervasive data mining.

DOI: 10.2902/1725-0463.2009.04.art10

1 INTRODUCTION

The recent large deployment of mobile devices has led to a massive increase in the volume of data regarding where people have been and when they were there. Access to aggregate measures of people's interactions and communications in the urban environment allow the generation of digital footprints that could have significant impact on urban and social studies (Microsoft, 2008). Arguably, analysis of these digital footprints can provide novel insights into how people experience the city, revealing different aspects of mobility, travel, and tourism, and allowing to study different attractors in the urban environment. This information can be useful for local authorities, researchers, as well as service providers, such as mobile networks operators. For instance, information about the distribution of population in different parts of the city at different times can lead to the development of customized services for citizens, allows accurate timing of service provision that can be based on demand, and more synchronous management of service infrastructure. In addition, in this paper, we analyze these new types of dynamic urban data to estimate the attractiveness and economic impact of points of interests in the city. In order to ensure that the social advantages of these applications are not in conflict with important privacy requirements, researchers and developers in this field must take conscientious, principled, and evident measures to protect people's privacy. In this paper, we consider two types of digital data: (i) highly aggregated, non-personally identifiable records generated by mobile phone usage on the AT&T wireless network; (ii) photos posted publicly on the photo-sharing website Flickr[1].

We present the methodologies and results of a case study in which we analyze the distribution and density of digital footprints in the area surrounding the New York City Waterfalls to quantify urban attractiveness. The Waterfalls was a public art project of four man-made waterfalls rising from the New York Harbor (in the East River) which were on display from June 26 to October 13, 2008.

Due to the large investments required for the temporary installation, its organizers also wished to study the economic impact of the event. The New York City Economic Development Corporation estimated that nearly 1.4 million people viewed The New York City Waterfalls, whether from an official vantage point, a ferry, or a tour boat (New York City Economic Development Corporation, 2008). Of the 1.4 million Waterfalls viewers, about 79,200 were incremental visitors to the City – people who visited the city or extended their ongoing visit particularly to visit Waterfalls. According to the study, the Waterfalls were viewed by New Yorkers, and by visitors from across the United States as well as from at least 55 other countries. These visitors generated – directly and indirectly – about $69 million of total economic impact on New York City. The Waterfalls were intended

[1] http://www.flickr.com

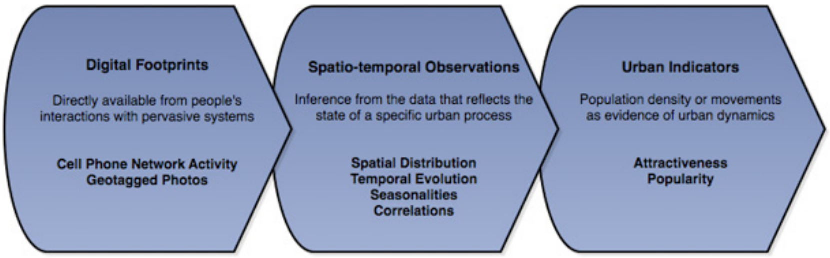
to bring visitors and New Yorkers to the waterfront who otherwise may not have done so, raising awareness to New York City's waterfront.

While traditional methods, such as people count and surveys, employed for studying the economic impact generate rather precise estimates of visitors to specific areas, our case study provides novel methods to quantify the influence of the public art exhibition on the distribution of visitors and on the attractiveness of various points on interest in the proximity of the event. Our analysis of cellular network traffic and georeferenced photos provides evidences as to where the more than one million New Yorkers and visitors were attracted in Lower Manhattan and whether the attractiveness and popularity of points of interests at the waterfronts increased during the event.

In more general terms, our contributions are twofold. First, we explore how locals and visitors share the space: our previous work has illustrated the capacity of aggregate digital footprints to uncover the presence and movements of tourists in cities such as Florence and Rome (Girardin et al., 2008). We were able to map their spatial presence, their temporal presence, and their movements within and outside the city. In addition, the spatial and temporal distribution of tourists of different nationalities seems to be characterized by different types of digital footprints. In this project, we aim to provide additional empirical evidences and more detailed analysis of the mapping of visitors and locals, revealing where they take photos and communicate through their mobile phones. Second, we develop indicators to quantify the evolution of the level of attractiveness of various points of interest. We argue here that the spatial distribution of visitors and the density of the digital footprints they leave behind reflect the attractiveness of a place. Our indicators are inspired by economics and network theory, and are used to compare the attractiveness of the main points of interests around the Waterfalls based on their relative strength, and the evolution of the centrality of the waterfront throughout the network of points of interest.

In order to explore these questions, our research process follows several steps starting with the collection of digital footprints (Figure 1). One type of digital footprints is generated by people's implicit interaction with wireless infrastructures. This results, for instance, in location logs produced when carrying a mobile phone which is in dialogue with a wireless network. In addition to these automatically generated and implicit data, another type is generated explicitly by a mobile user when generating messages and by Web users when they post content such as photos on public Web sites. Through the analysis of these data we seek to extract spatio-temporal characteristics such as seasonality, usage patterns, and spatial distribution. In subsequent phases, these parameters inform the formulation of urban indicators that help us quantify the attractiveness of popular areas.

Figure 1. Our study starts with the collection of digital footprints, followed by spatio-temporal analysis of their characteristics, which allow us to define and create applications for indicators of urban attractiveness.



The following section gives a condensed overview of related research work that exploits digital footprints in the domains of urbanism and tourism. Then, we present the types of data collected, and the methodology we use to extract valuable indicators. Finally we conclude with a discussion on the limitations and future work.

2 RELATED WORK

People counts, surveys, and other traditional methods to identify the presence of visitors and tourists in a city are often expensive and result in limited empirical data. Methods and models that study space-usage on the basis of urban configuration such as Space Syntax (Hillier, 2007) rely heavily on such limited estimates. Similarly, the exploitation of land-use (e.g. density of hotels) and census data (e.g. museum revenues) provides only a static perspective on city dynamics. The lack of data presents particular difficulties given that most cities – though they may aim at providing advanced services – have limited human, technical, and financial resources. Today, thanks to the emergence of ubiquitous technologies, new data sources are available. Indeed, the logs produced by users' interaction with wireless and online services can allow us to develop new methods of observing, recording, and analyzing aggregate human dynamics in the city (O'Neil et al., 2006). New metrics for describing the social and spatial characteristics of the space can be developed (Kostakos et al., 2008). The distributed presence of personal devices creates a vast, geographically - referenced, sensor web (Goodchild, 2007; Zook et al., 2004; Elwood, 2008; Budhathoki et al., 2008).The accumulation of records on this network can reveal collective social behaviours with unprecedented details (Eagle and Pentland, 2006). Digital footprints present an opportunity in urban and tourism studies to build more efficient ways of collecting aggregate information about visitors' activities. For instance, tourists have many ways of leaving electronic trails: prior to their visits they generate server log entries when they consult digital maps (Fisher, 2007) or travel web sites (Wöber, 2007). During their visit, they leave traces on wireless networks (Ahas et al., 2007), whenever they use their mobile phones or through their credit card payments (Houée and Barbier, 2008). After their visit they sometimes add annotations (Mummidi and Krumm, 2008) and photos (Girardin et al., 2008; Crandall et al., 2009) to digital maps. For example, examination of such data could complement statistical analysis in areas of collective residential accommodation and optimize the provision of customized services. It can also contribute to accurate timing of service provision based on demand (e.g. rescheduling monument opening times based on the presence of tourists), estimation of the economic impact of areas of interests, and, in general, it can facilitate more synchronous management of service infrastructures.

However, privacy and ethical issues related to collecting personal data become key issues when collecting information on urban dynamics (Gutman and Stern, 2007). Other critical issues include the longevity of the studies (e.g. to prevent fatigue effects); the timeliness of data collection; and the challenge of collecting data in a scalable manner. For instance, there is great variation of standards between cellular network operators. Table 1 compares different types of data capturing methods in terms of their strength and weaknesses for applications in tourism and urban studies.

Table 1. Data capture techniques with their main strength and weakness in the context of tourism and urbanism studies.

Data capture	Strength	Weakness	Example of application
Land use and census data	Applicable to many scales and over long time period	Infrastructure and service based, static view of urban dynamics	Estimate the tourism intensity of an area
Manual surveys	Capture high-level information such as motivations and reasons for stay in specific areas	Very costly and applies to limited time periods	Capture the motivation for visiting and length of stay
Near-field communication	Precise real-time mobility data	Costly Infrastructure deployment	Describe the social and spatial characteristics of space (Kostakos et al., 2008)
GPS logs	Precise mobility data	Does not scale well if deployed for the purpose of a survey alone. limited in time and participants	Cluster tourist routes (Asakura and Iryob, 2007)
Cellphone (device-based)	Timely mobility data, potentially augmented with in-situ survey	Does not scale well if deployed for the purpose of a survey alone. limited in time and participants	Context-Aware Experience Sampling to capture the experience in-situ. (Froehlich et al. 2006)
Cellphone (aggregated network-based)	Use existing infrastructure to provide real-time density and mobility data, covering multiple geographic scales (neighbourhood, city, country)	Reveal large-scale phenomena but do not explain the reasons	Real-time traffic detection (Yim, 2007)
User-generated content	Exploit publicly available data with no need for deployment or pre-existing infrastructure	Credibility of information and no systematic coverage	Reveal flows of photographers (Girardin et al. 2008)

In this research work, we make use of georeferenced photos made publicly available by individuals on the Web. We also use aggregate, non-personally identifiable, records generated by mobile phone users who make calls and send text messages on the network. These two data sets are independent of each other.

Previous research has also shown that the widespread usage of mobile phones and the pervasive coverage of cellular networks in urban areas make these technologies efficient tools through which to study both crowds (Ratti et al., 2006) and individuals (González et al., 2008), For example, analysis of mobile data for vehicle traffic analysis (see Yim, 2007 for a review) can generate information about traffic conditions in real-time. In some other efforts, cellular network signals were correlated – with limited success – with the actual presence of vehicles and pedestrians in the city (Sevtsuk and Ratti, 2007). In a case study of tourism dynamics in Estonia, Ahas et al. proved that the sampling and analysis of passive mobile positioning data is a promising resource for tourism research and management. They show that this type of aggregated data is highly correlated with accommodation statistics in urban touristic areas. While this work described the dynamics of tourism in an entire country (Estonia) for purposes such as obtaining more detailed geographical information about urban versus rural tourism, our study takes place at the scale of a city-neighbourhood and leverages the estimated density and movement of visitors to study the attractiveness of various urban spaces over time.

We particularly focus on the attractiveness of urban centres because they have been the focal point of citizen's urban life. Therefore, the current limited abilities to monitor of their attractiveness could be regarded as an immediate threat to the liveliness of a city's economy. As we have seen, the advantage of our approach over traditional tourism and urban statistics is the longevity, scalability and timeliness the analyzed datasets offer.

3 DATASETS

We analyzed two types of digital footprints generated by phones or mobile devices that were in physical proximity to the New York City Waterfalls: cellular network activity and photo activity (Table 2). Cellular network activity was measured by analyzing aggregate statistical data about number of calls, text messages, and overall amount of network traffic generated at each AT&T antenna every hour. Photo activity was measured by adding up the number of photographers present in different areas of the city, and the number of photos they took in each location. We acquired this data by analyzing photo taken from the photo sharing website, Flickr (see Girardin et al., 2008 for a more detailed description of the data collection process).

Table 2. Sources and spatio-temporal coverage of the datasets

	Cellular Network Activity	Photo Activity
Provider	AT&T	Flickr (publicly available)
Coverage	Lower Manhattan and West Brooklyn	New York City
Time period	Aug 2007 to Aug 2008, every hour	Jan 2006 to Aug 2008, every minute

3.1 Cellular network activity

As part of our research collaboration with the mobile operator, AT&T, we were granted access to anonymous records of network activity, aggregated hourly, and generated by mobile phone users who made calls on the AT&T network between August 2007 and August 2008. The section of the network under study is serviced by a Base Transceiver Station (BTS) which covers the lower part of Manhattan and the west part of Brooklyn. A BTS represents the elementary unit of the infrastructure that is used to connect users' devices to the mobile phone network wirelessly. It provides connectivity to specific geographic regions called sectors. Each observation constitutes the number of calls that originated and terminated in each of the sectors. Table 3 provides a detailed description of the meaning of each data type. It should be noted that the data can be biased and contain spatial noise. In our case study some of the network connectivity is provided by infrastructure on the other side of the East River, thus affecting the ability to capture fine-grained information about the location of mobile phone users that were actually present in the area we aimed to study. Moreover, when an ongoing call is handed over to another antenna, the location of the call activity is still associated to the place where the call started, which creates additional noise. Finally, several additional factors directly affect the capacity to estimate network activity in specific areas. For instance, the dimension of a wireless sector can vary greatly, depending on the built environment it services and the type of phone used. Also, some sectors can partially overlap and provide signal to the same areas, thus making it difficult to determine the location of calls made from these sectors.

Table 3. Descriptions of the types of aggregated cellular network traffic data

Data type	Description
Aggregated calls	Number and duration of calls originated and terminated in an area
Aggregated Call Detail Records (CDR)	A CDR is a detailed record produced by a telephone exchange that provides information about each call or text. For this study, these records are aggregated over time (measured by the hour), user's home location, and the BTS it is connected to. The home location of a user is determined by the area code of a mobile telephone number for U.S.-registered phones and the country code associated with the mobile telephone number of foreign-registered phones.

Our collaborators granted us limited access to CDR data. It is important to note that many U.S. mobile phone customers may own telephone numbers with area codes that do not reflect the state in which they reside. Also, foreign visitors may acquire U.S. SIM cards and mobile telephone numbers for the duration of their stay in the U.S. Thus, there can be some inaccuracies in our inferences about the home locations of mobile phone users.

3.2 Georeferenced photos on Flickr

It's common for visitors and tourists to take photos during their trips and use a web photo sharing platform such as Flickr to share and organize photos and provide geographical attributes. Each time a photo is anchored to a physical location, Flickr assigns longitude and latitude values together with an accuracy attribute derived from the zoom level on a map that is made available for users to position their photos. Photos positioned when the user zooms in at the street level receive a higher accuracy estimate than ones positioned when the user had pulled back in the online map view. The system also adds metadata embedded by the camera in the image through the Exchangeable Image File Format (EXIF) information, enriching the spatiotemporal information (Table 4).

Table 4. Description of the Flickr dataset

Data type	Description	Size
Photographers	60% of the cases - nationality is disclosed. Activity is profiled to infer whether users are New York residents or visitors	29.235 photographers
Photos	Geographical coordinates disclosed by the photographers; date and time extracted from the digital camera metadata; semantic description of the photo is provided by the photographer.	1.197.287 photos

We pre-processed the data by classifying local photographers and visiting photographers based on their presence in the area over time as the discriminating factor with the aim of capturing the one-time tourists. We divided the time period into 30-day intervals and calculated the number of periods in which each photographer was active in an area. If a photographer took all his/her photos within one period of 30 days, we classified him/her as a visitor, while if there was an interval greater than 30 days between two photos taken, we classified him/her as resident.

We used the Flickr API[2] to retrieve the coordinates of photos and their accuracy, the time at which they were taken, and we also obfuscated the identifiers of their

[2] http://www.flickr.com/services/api/

owners. The mapping of these data allowed us to detect the main areas of photographic activities in New York because the accumulation of georeferenced photos over a period of time reveals the boundaries of areas of interest in a neighbourhood. In addition, the processing of a chronologically ordered set of photos revealed the traces of photographers; their spatiotemporal movements in the urban space.

4 METHOD

Our method aims at quantifying the distribution of digital footprints over the area of study as precisely as possible. This is more difficult in the case of cellular activity than in the case of photo activity because the former is measured only in discrete locations (i.e. at the level of the BTS) while the latter is continuous over space. For the cellular activity, we computed a radio map of partitions generated by overlapping BTS sectors. Activity inside each partition was considered uniform and was computed by accounting for the activity at each overlapping sector (see details below). For the photo activity, we employed a similar process that stores photos in equal size partitions, forming a matrix. The resulting data structures allowed us to map the density and flows of footprints over space and time: to classify them, we defined probes as spaces to analyze and compare the dynamics of particular areas of interest. The remainder of this section describes this method in more detail.

4.1 Radio map

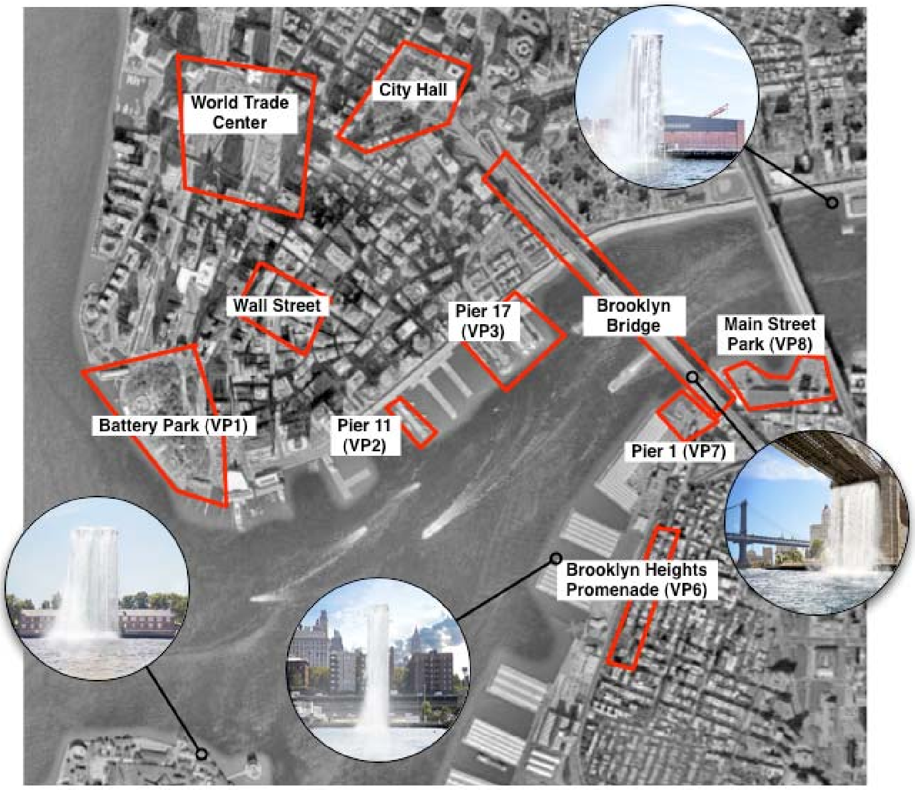
Our first step was to estimate the density of the traffic over space according to the topology of the wireless network. This was done by pre-processing wireless propagation data and generating a radio map that aims at overcoming the inherent problems of the low resolution and reliability of the cellular network statistics at relatively small urban areas.

Each area of wireless coverage is divided into sectors. Each area is controlled by a transceiver that is mounted on a BTS and connects a mobile device to the rest of the network. The boundaries of each sector, which are necessary in order to compute the distribution of cellular activity, are not well defined and depend on the location of the BTS station itself. A standard approach is to compute the Voronoi diagram based on the location of the BTS stations, thus assuming that the best serving station is the closest one to any point in space. However, this approach is imprecise, because many elements of the urban landscape can interfere with the coverage of the BTS. This method also does not work well with rather the small-sized areas of interest selected for our research. Therefore, we employed a propagation model based on the Okumura-Hata model (Seybold, 2005) to estimate the boundaries of each sector based on the location, height, azimuth, Effective Isotropic Radiated Power (EIRP), type of antenna, and frequency served by each BTS, together with an analysis of the physical environment in the area which the BTS services (e.g. presence of a river). The model produces estimates of the coverage of each sector, with the overlapping areas that are used to generate a radio map of the study area formed out of around 9900 non-overlapping partitions. We estimated the activity at each partition by summing the weighted contributions of all the overlapping sectors. As a result, each coverage area can be used to describe the network activity generated by the antennas which provide it with signal.

4.2 Areas of interest

To explore the impact of the event at a neighbourhood scale and describe the evolution of the attractiveness of the surrounding areas, we defined the major attractions in the proximity of the Waterfalls such as the World Trade Center site, Wall Street, City Hall, and the Brooklyn Bridge (Figure 2).

Figure 2. Location of the four Waterfalls and definition in red, the main vantage points (VP) and attractions in proximity to the New York Waterfalls exhibit.



The probes vary widely in dimension: they could be as large as the financial district or as small as Pier 1, which was renovated with the intention of serving as a vantage point for the exhibit. The activity for each area of interest differs between the two datasets we investigated. For the Flickr dataset, we simply sum the number of georeferenced photos taken at an area, and account for the number of photographers active within each probe. For the cellular network data, the activity at each area of interest was estimated by considering the relative weights of each sector's contribution to the calling statistics measures at each element of the partition, and the overlap between areas covered by a partition and that of an area of interest. Following, we detail this process.

4.3 Interpolation model of network activity at each area of interest

Our approach is based on the following assumptions:

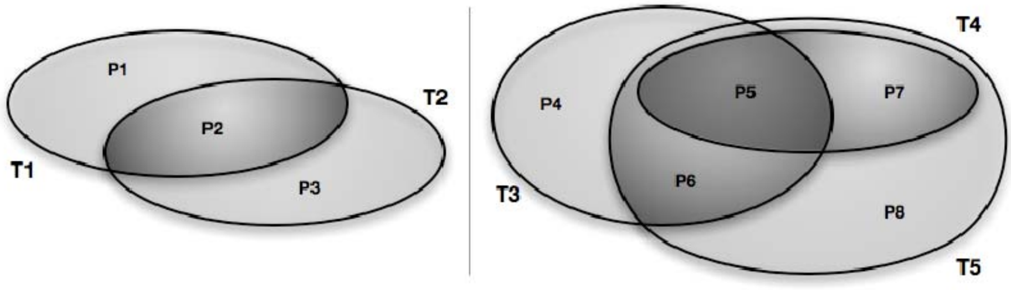
1. Phone usage is not uniformly distributed over the coverage area of each sector, but it is uniformly distributed over each partition
2. Contributions of each BTS to each partition is inversely proportional to the overall number of transceiver that service to the partition.

In the example on Figure 3, if the coverage area of transceiver T1 is divided into two partitions P1 and P2, where P1 is covered only by T1, and P2 is covered by T1 and another transceiver T2, then the density of activity contributed by T1 to P1 should be double the density of activity contributed to P2, and the same holds for T2 with respect to P2 and P3.

Furthermore, P4 is serviced by T3 only, P5 by T3 T4 and T5, P6 by T3 and T5, P7 by T4 and T5, and P8 by T5 only. Therefore:

- For T3: d(P4) = 2 x d(P6) = 3 x d(P5), and s(P4) + s(P5) + s(P6) = s(T3);
- For T4: 2 x d(P7) = 3 x d(P5), and s(P7) + s(P5) = s(T4);
- For T5: d(P8) = 2 x d(P6), and s(P8) + s(P6) = s(T5);
- Where d is the density of activity contributed

Figure 3. Interpolation model of network activity at each area of interest



The method described in this section allows us to estimate the activity at each area of interest without imposing strong a priori assumptions about the distribution of activity. The results obtained can be used to perform analysis on the data and uncover their spatio-temporal characteristics. This process and its results are described in the following section.

5 OBSERVATIONS OF PRESENCE AND ACTIVITY

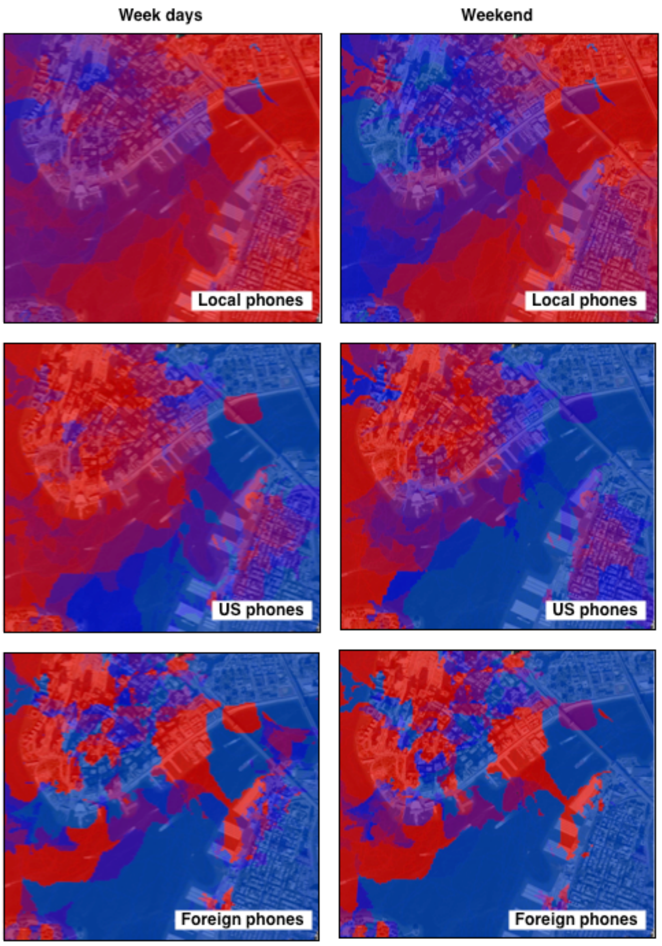
This section describes the spatio-temporal characteristics of the data. Particularly, it focuses on the spatial distribution of the presence of locals and visitors and on the evolution and seasonal patterns of their activity over time. These observations allow us to identify the major spatio-temporal activities in that lead to the production of digital footprints.

5.1 Spatial distribution of locals and visitors

For the week of August 10 to August 17, 2008, AT&T provided us with hourly aggregated data about cellular network traffic indicating the number of non-personally identifiable phone calls for each registration location of mobile phones. The registration location of locals and visitors is determined based on the area code of the handset's mobile telephone number for U.S. phones and the country code for foreign phones. This allowed us to count the activity from mobile phones registered in New York, and to quantify how much activity was associated with a mobile phone registered in the United States but outside New York, and how much activity involved a mobile phone registered outside the United States.

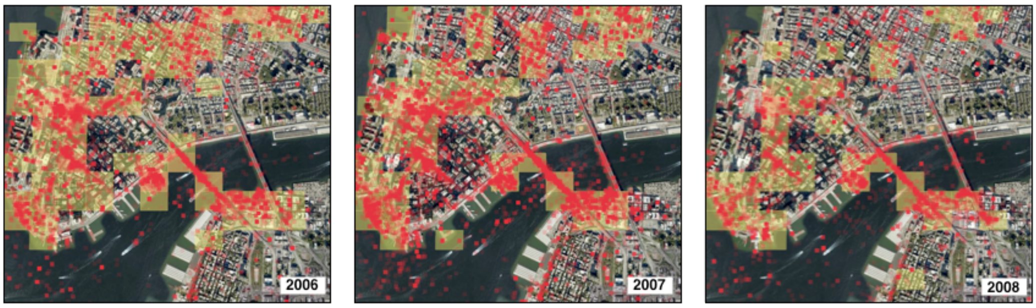
While it may not be always true, we considered it reasonable to assume that locals generate the majority of calls from mobile phone registered in New York, and that visitors generate the majority of calls that involve mobile phones registered outside New York. With this assumption, we were able to map the presence of non-identifiable locals and visitors on an average weekday and on weekend (Figure 4). The observation revealed that visitors enjoy Lower Manhattan and its main areas of interest, particularly during the weekend, much more than they visit Brooklyn, with the remarkable exceptions of Pier 1 and the waterfront. On the other hand, locals show a well-spread activity in space with stronger presence in Brooklyn over the weekend.

Figure 4. Spatial distribution of locals (New Yorkers), US visitors, and foreign visitors in the proximity of the Waterfalls. In red the areas with high density of phone calls. Locals clearly withdraw from Lower Manhattan on the weekends and the use foreign phones is limited to very precise areas of the neighbourhood.



Mapping of the density of photographers brought another perspective on the distribution of visitors. It allowed us to observe the evolution of their activity over three consecutive years (Figure 5), revealing the main areas of interest for photographers, in particular the Brooklyn Bridge, the World Trade Center site, and Battery Park.

Figure 5. Spatial distribution of photographers (in yellow) and photos (in red) in Summer 2006, 2007, and 2008 in Lower Manhattan and West Brooklyn.

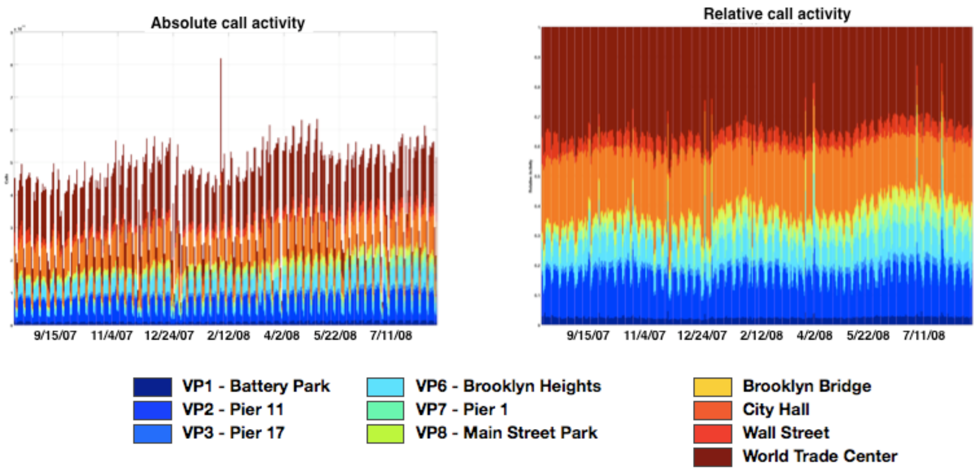


5.2 Evolution of the activity

While the previous observation allowed us to study the spatial distribution of locals and visitors, it didn't provide insights as to the evolution of their activity over time. To identify the major trends of activity, we plotted stacked bar charts that display the daily patterns of activity for each probe between August 2007 and August 2008. We generated two versions of the charts: one which represents the absolute values of phone calls – useful to understand both the behaviour of each area of interest and the overall trend, and another which represents the relative activity of each area of interest with respect to the others – useful to compare variations in behaviour between probes.

Here again, we selected phone calls as a proxy for presence of people. Figure 6 shows the evolution of the daily density of phone call activity between August 2007 and August 2008, i.e. the average number of phone calls originated or terminated in each area of interest in one day, per unit of surface. The observation shows that there is a year-long trend of positive growth in the overall number of phone calls and that most of the growth originated in an increased activity in Brooklyn (VP6 and VP8), with almost doubles in activity during that period. Moreover, it shows the presence of strong weekly seasonality, which causes the phone call activity to drop over the weekends to about half the activity of workdays with slight variations among the different areas of interest.

Figure 6. Daily absolute (left) and relative (right) density of phone calls per POI. Note: values on the vertical axes have been multiplied by a constant factor.

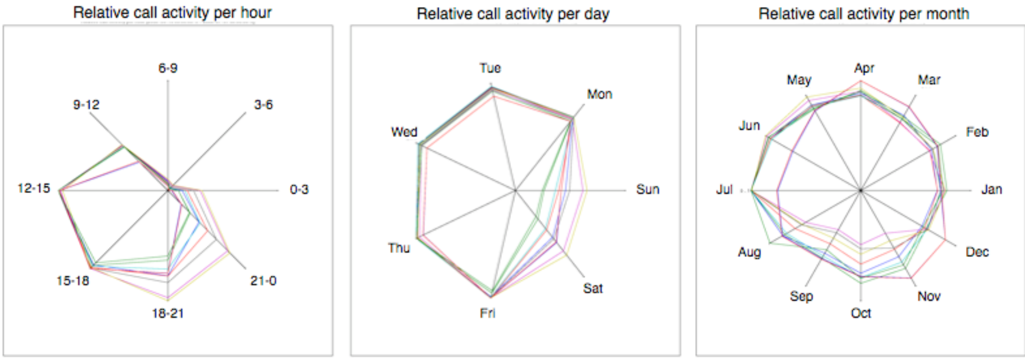


Seasonality is confirmed when observing the absolute density, which also reveals that while some areas of interest tend to have less activity over the weekends, others exhibit an increase, a divergence that can help us understand the type of traffic generated at each area of interest (e.g. work-related versus leisure-related).

5.3 Temporal patterns

Motivated by the results of our previous observations, we decided to explore temporal seasonalities further on different time horizons and with the use of spider charts. Figure 7 on the left shows the average phone call activity throughout the day (in slots of three hours each, from midnight to 3 am, from 3 am to 6 am, and so forth): all the areas of interest present the same behaviour at night and during the afternoon, while they exhibit different behaviours during the evening. Figure 7 in the middle reveals the average phone call activity per day throughout the week. It shows similar behaviour during workdays across probes, and variation during weekends, when some areas of interest drop to an activity level that is about 30% of that during workdays. Others area of interest maintain an activity of about 60% during weekends. Finally, Figure 7 on the right shows the average activity level per month throughout the year, presenting an overall higher variability of behaviours, yet still showing clearly different activity levels between August to November.

Figure 7: Temporal patterns of phone activity between POIs, per slots of three hours throughout the day (on the left), per day throughout the week (enter), per month throughout the year (on the right) - period August 2007 to August 2008.



The observations presented in this section showed that Lower Manhattan attracts more visitors than West Brooklyn, which exhibits the typical characteristics of a residential area with fewer features of tourist activity. The observations also show that there are clear seasonal differences between areas of interest, in particular at different hours of the day and during different days of the week. These results raise new questions: which areas of interest attract more tourists during the weekend? How can we quantify the growth in attractiveness at different areas over time, and which areas grew more in activity from one summer to another? In order to answer these questions, we defined indicators of urban attractiveness based on the relative density of digital footprints. We ran a comparison between the areas of interest around the Waterfalls as well as with respect to the overall New York City metropolitan area. These indicators are described in the following section.

6 URBAN INDICATORS OF ATTRACTIVENESS AND POPULARITY

In Arabic the word for indicator means pointer, and describes how an indicator is intended to point towards some desirable state or course of action. Urban indicators are about the interface between city management and data. An indicator will often be the benchmark against which policymakers and the public can assess change. Indeed, indicators are not data, rather they are models simplifying a complex subject to a few numbers which can be easily grasped and understood (de Villa and Westfall, 2001). There is a wide range of urban indicator applications from macro (i.e. global and national) to micro levels (neighbourhoods) developed by international indicator initiatives or cities authorities themselves. To our knowledge, the notion of attractiveness indicators has only been applied at a macro scale to compare the quality of life among major cities.

In contrast, our observations of the cellular network aggregate traffic and photographers activity described the density and movements of locals and visitors over a very defined space and time. This analysis was performed particularly with respect to the main areas of interest on the waterfront in Lower Manhattan and Brooklyn, where people could find vantage points to observe the waterfalls. In this section, we explore the use of indicators to measure these observations and quantify the evolution of attractiveness and the popularity of areas of interest in proximity to the Waterfalls public exhibit.

6.1 Attractiveness

We hypothesized that the density of the digital footprints that visitors leave behind gives an indication of the attractiveness of places; that is that these places of interests have beneficial features for work, social interaction, or sightseeing purposes. In addition, we conceptualize urban attractiveness as a property of a well-defined place that can have a variable size. Therefore, we developed several indicators to compare the evolution of the attractiveness of different areas around the waterfront and of areas of interest in proximity of the exhibit, based on their Comparative Relative Strength (CRS).

The CRS indicator compares the activity of one area of interest with respect to the overall activity of the city. The values are normalized over time and space therefore not relying on the absolute number of photographers but comparing their relative activity. When the CRS indicator grows in time, it shows that the area of interest is performing better than the overall city because it is attracting more or losing less activity than the rest of the city. We used three different proxies to measure the attractiveness of areas of interest: the presence of photographers, the number of calls, and the ratio of calls made by locals versus visitors.

6.1.1 Attractiveness based on the presence of photographers

Table 5 shows the variations of the CRS indicator based on the presence of photographers during the summers (June to October) of 2006, 2007, and 2008. It reveals a positive growth in the waterfront's attractiveness of 8.2% in summer 2007 and 20.7% in summer 2008 with respect to that of other areas of interest in New York City, such as Time Square and Central Park. It should be noted that the maximum growth in attractiveness (+29.9%), observed during summer 2008 was recorded in DUMBO in Brooklyn. This was probably supported by the increased presence of photographers at the proximate vantage points of Pier 1 and Main Street Park, elicited by the Waterfalls.

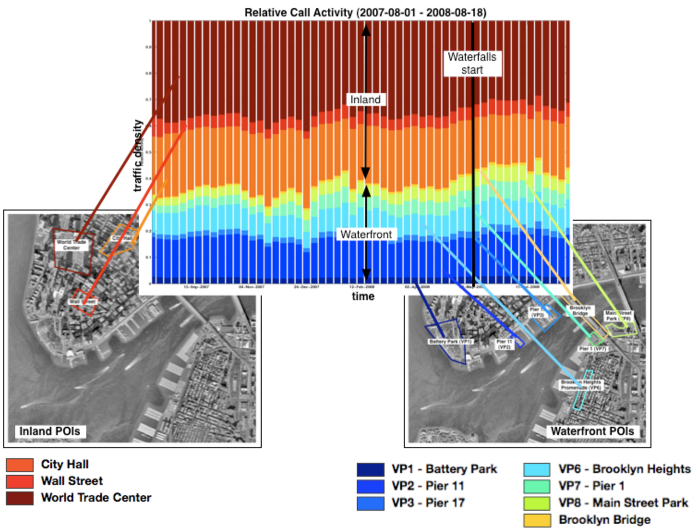
Table 5. Variation of the CRS indicators from 2006 to 2008

	Photographers 2006	Photographers 2007	Photographers 2008	CRS 2006	CRS 2007	CRS 2008	2006 to 2007	2007 to 2008
Central Park	1874	2619	1537	0.111	0.101	0.100	-0.091	-0.008
Chelsea	1146	1790	1125	0.068	0.069	0.073	0.015	0.062
East Village	652	971	606	0.038	0.037	0.039	-0.031	0.054
WTC site	564	775	374	0.032	0.029	0.024	-0.075	-0.184
Time Square	1227	1883	1026	0.073	0.072	0.067	-0.002	-0.079
Vantage Points	538	896	640	0.032	0.034	0.041	0.082	0.207

6.1.2 Attractiveness based on the number of calls

Figure 8 shows the trend of CRS of number of phone calls initiated or received every hour between August 2007 to August 2008, at each area of interest. We blurred the absolute values with a random factor to keep these data confidential. The CRS of number of phone calls reveals an increase in vantage points' attractiveness of 39.1% with respect to the attractiveness of other areas of interests in the vicinity, like the World Trade Center site, City Hall, and Wall Street. Interestingly, Pier 1 doubles its attractiveness (105%) over this period. Other areas of interest in the vicinity also experience increased attractiveness (Main Street Park, Brooklyn Bridge, and South Street Market).

Figure 8. Evolution of Comparative Relative Strength of relative phone call activity among the exhibit vantage points and other city attractions in proximity (WTC site, Wall Street, City Hall) between August 2007 and August 2008



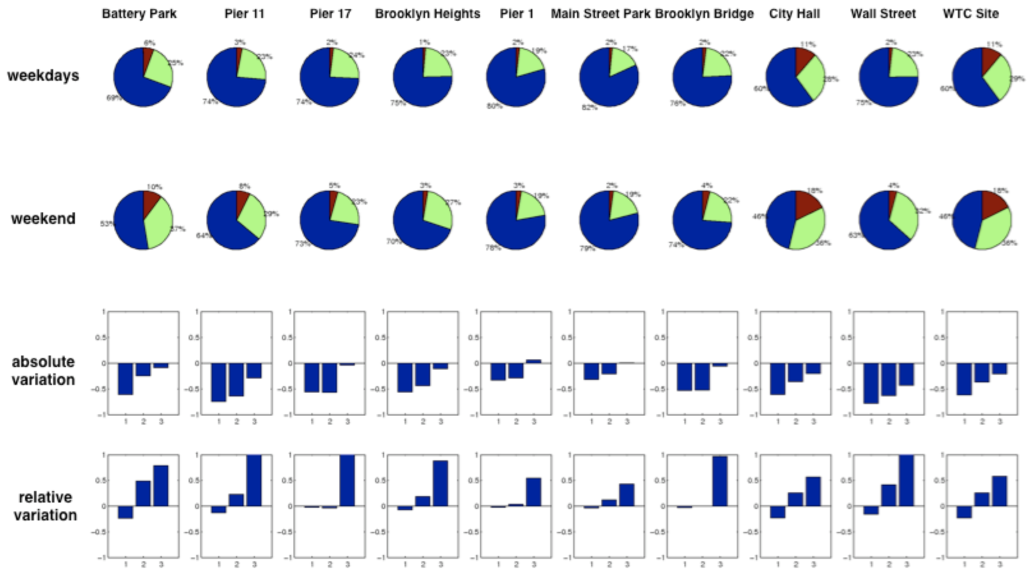
6.1.3 Comparison of attractiveness for locals and visitors

Figure 9 shows the CRS of three types of calls (originated from mobile phones registered in New York, the US and the rest of the World) during workdays and weekends for each area of interest. The first set of histograms shows the percentage variation of the number of calls (almost always negative because weekends are characterized by a general reduction in call activity), while the second set shows the percentage variation in CRS.

The percentage of foreign visitors' calls shows an average growth of 88% across vantage points when comparing weekdays and weekends, while the average growth of other areas of interest is 77%. However, the average CRS of US visitors' calls over the vantage points shows growth of only 15%, while the growth at the areas of interest is 31% on average. This means that while the vantage points tend to outperform the other areas of interest in attracting activity of foreign visitors, they perform worse in attracting activity of US visitors.

It should be noted that the WTC site is clearly a major area of interest for tourists in Lower Manhattan, with a locals' calling activity decreasing to less than 50% during weekends when compared to weekdays. Other areas such as City Hall and Battery Park show a similar distribution of calling activity, with New Yorkers leaving the place to visitors from the US and abroad during weekends.

Figure 9. CRS and percentage variations for locals' (blue,1), US visitors' (green,2) and foreign visitors' calls (red,3) at the vantage points and main points of interests between weekdays and the weekend.



6.2 Popularity

We assessed the popularity of an area of interest by studying its ties with other areas in the city. The stronger the ties, the more frequently an area is accessible from other places as it becomes part of a popular route. This was measured by applying network analysis techniques to study the connectivity of a network in which the nodes represent areas of interest and the arches represent flows of people between them. Flows were estimated by analyzing consecutive time stamps tagged to Flickr photos in conjunction with the reported location at which a photo was taken.

6.2.1 Popularity based on centrality

The centrality of an area of interest determines its level of membership to the popular flows of photographers. The PlaceRank indicator, inspired by the PageRank indicator developed by Google to order the importance of Web pages (see Brin and Page, 1998 for a detailed description of the PageRank calculation), determines the centrality of a location within a set of areas of interest based on the amount of digital footprints generated in each area and the traces that connect them. In particular, if more visitors visit place A than place B, then we can say that the former is more popular than the latter in the network of tourist destinations. Moreover, if the same amount of people move from place A to a new place C and from place B to a new place D, we can also say that place C is more popular than place D because place A is more popular than place B.

We computed the PlaceRank values for the areas of interest, using aggregate values during 2006, 2007, and 2008. Figure 10 helps in interpreting the data by showing a graphical representation of the connectivity of areas of interest based on the number of photographers moving within each location.

The PlaceRank indicator reveals that between 2006 and 2007 the vantage points lost their centrality by 15% while the other areas of interest increased their centrality by 10%. However, between 2007 and 2008, the vantage points gained 56% while the other areas of interest lost 30%. In 2008, the vantage points appear as central as the other areas of interest, meaning that they are on the tourist path as much as the other areas of interest in that section of the city.

Figure 10. Evolution of the flows of photographers in proximity to the exhibit based on the analysis of photos generated between June - October, 2006, June - October, 2007, and June - August 15 in 2008. At that year of the Waterfalls, VP3, VP6 and VP7 massively increase their PlaceRank.

7 DISCUSSION

The increasing deployment of wireless and mobile devices makes new types of dynamic data available. Through passive and active interactions with these ubiquitous technologies, people generate digital footprints which can be used to describe and analyze urban environments as well as human activity on a detailed level, at a large scale, in a timely fashion and without the bias of explicit survey techniques.

In this case study, we collected and analyzed aggregate and anonymous cell phone network activity data and georeferenced photos publically available online to reveal several types of digital footprints. We illustrated the use of these footprints to develop observations and define indicators that can inform urban design, planning, and management processes. Our analyses allowed to track the evolution of the attractiveness of different areas of interest in Lower Manhattan and West Brooklyn. The new information was used to supplement a study which used traditional means like manual counts and surveys to quantify the impact of the New York Waterfalls public exhibition on show around the New York waterfront from June to October 2008. Our approach relied on several indicators of urban attractiveness which we developed in inspiration by financial indicators and network theory. For instance, we compared the attractiveness of major points of interests in New York based on the relative density of digital footprints. We also analyzed the flows of visitors between several points of interest in Lower Manhattan to track the evolution of centrality of the waterfront area in comparison to the other points of interest. Mapping this new type of digital footprint analysis shows the capacity of an event to drive people to less explored parts of a city over time, information which can be highly valuable for urban design and tourism studies.

Furthermore, in the traditional urban design process, practitioners often have limited capacity to evaluate their design due to difficulties in performing reliable post-occupancy studies, which are rarely performed (Kozlowski, 2007). This case study shows that the emergence of digital footprints creates an opportunity to evaluate in detail the use of space, the impact of events, and the evolution of the built environment. This approach could not only better inform urban design and city management processes, but also enable local authorities to provide timely evidences to the public about the use of space and about the impact of interventions within the urban fabric. Indeed, the integration of our results in the official study of the economic impact of the New York Waterfalls public art project shows that the indicators proposed in this paper offer useful measures to complement traditional methods.

Our approach produced several novel research contributions for the utilization of digital footprints. The research efforts to quantify urban attractiveness using the distribution and density of digital footprints led us to explore the mapping of the spatial distribution of visitors, and to develop indicators to compare the evolution of the attractiveness of the waterfront with respect to other areas in the city.

At this stage in our research, even though the measurements and visualizations performed concur with the traditional impact study of the Waterfalls, the extent of their reliability is still unclear, as it is often the case for approaches based on user-generated content (Flanagan and Metzger, 2008). Indeed, in some cases our case study detects weak signals generated by a diffuse population over a long period of time in one of the noisiest cities in the world in terms of network activity. For instance, the location of Vantage Point 7 at Pier 1 suggests that the temporal measurements of visitors might be impaired by the noise of nearby heavy traffic on the Manhattan Bridge. Also, due to the unobstructed path available above the surface of the East River, a BTS at the waterfront can capture traffic across river, creating a bias that is hard to measure. In other words, the smaller is the area of study, the more unreliable our observations are.

Following the results of this case study, we identified three main areas for future work: the improvement of the quality of observations, the further development of urban indicators, and the implementation of visual analytics to explore and communicate the study of digital footprints.

First, improving the radio coverage maps to better distribute data non-uniformly over each sector will reduce limitations in the quality and resolution of the measured activity and increase the fidelity to the physical and social worlds. One possible solution could be to perform on-site measurements of BTS signal to refine the estimation of its coverage. Second, there is a need to estimate the reliability of the results of analyses of digital footprints. The precision of the results depends, among other factors, on the density of BTSs in proximity of the areas of interest, and on the consequent ratio between the coverage area of each BTS and the area of study. One possible solution could be to adopt fuzzy systems and fuzzy arithmetic to perform our analyses, so to keep track of the uncertainty of the data down the proposed workflow.

Third, the dynamic nature of digital footprints and their fluctuations in space and time pose new challenges. It is necessary to calibrate and validate the data using information collected through more traditional means, such as surveys and census data. From a statistical perspective, our case study focused on the estimation of the probability density function of digital footprints with respect to time. We consciously disregarded the spatial dimension by considering the activity within each probe as uniform and independent from the activity of other areas. This was done mainly because of the lack of sufficient sample data upon which to validate the spatial density of digital footprints against the spatial density of, for example, people presence. Once enough data will be available, we will leverage the recent advances in spatial kernel density estimation (G Biau, 2003) to better study the distribution of digital footprints in space. And we will develop new methods to extrapolate population densities and human activities from digital footprints.

These shortcomings currently limit the usability of digital footprints for city management and decision making. However, their visualization is already offering new tools to communicate different features of activity in urban areas. Along with interactive software, visualization is a useful tool for a multitude of actors in the city such as researchers, practitioners, service providers, and local authorities, to discuss how to interpret data and put information in context.

Finally, our analyses could be used not only to describe the attractiveness of a place, but also to further profile places of different types based on the first explorations by Reades et al, (2007). For example, an area with strong calling activity and weak photographic activity could be primarily commercial; one with weak calling activity and weak photographic activity could be rendered residential; one with strong calling activity and strong photographic activity could be friendly towards tourist and leisure-related activities. Similarly, an outdoor area which has strong attractiveness indicators during adverse weather conditions could suggest that it is critical for visitors; an indoor area with weak attractiveness indicators during adverse weather conditions could indicate that it may not be easily accessible.

REFERENCES

Ahas, R., Aasa, A., Silm, S., and Tiru, M. (2007). Mobile positioning data in tourism studies and monitoring: Case study in Tartu, Estonia. *In Marianna Sigala, L. M. and Murphy, J., (Eds.), Information and Communication Technologies in Tourism 2007*, Ljubljana, Slovenia. Springer Vienna.

Asakura, Y. and Iryob, T. (2007). Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A: Policy and Practice*, 41(7):684–690.

Biau, G. (2003). Spatial kernel density estimation. *Mathematical Methods of Statistics*, 12:371– 390.

Budhathoki, N., Bruce, B., and Nedovic-Budic, Z. (2008). Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal*, 72(3):149–160.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117.

Crandall, D. J., Backstrom, L., Huttenlocher, D., and Kleinberg, J. (2009). Mapping the world's photos. In WWW '09: *Proceedings of the 18th international conference on World wide web*, pages 761–770, New York, NY, USA. ACM.

De Villa, V. and Westfall, M. S., editors (2001). *Urban Indicators for Managing Cities: Cities Data Book*. ADB.

Eagle, N. and Pentland, A. S. (2006). Reality mining: sensing complex social systems. *Personal Ubiquitous Computing*, 10(4):255–268.

Elwood, S. (2008). Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice. *GeoJournal*, 72(3):133–135.

Fisher, D. (2007). Hotmap: Looking at geographic attention. *IEEE Trans. Vis. Comput. Graph.*, 13(6):1184–1191.

Flanagin, A. and Metzger, M. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72(3):137–148.

Froehlich, J., Chen, M. Y., Smith, I. E., and Potter, F. (2006). Voting with your feet: An investigative study of the relationship between place visit behavior and preference. In Dourish, P. and Friday, A., editors, Ubicomp, *Lecture Notes in Computer Science*, pages 333–350. Springer Berlin / Heidelberg.

Girardin, F., Calabrese, F., Dal Fiore, F., Ratti, C., and Blat, J. (2008). Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive Computing*, 7(4):36–43.

González, M. C., Hidalgo, C. A., and Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, (453):779–782.

Goodchild, M. F. (2007). Citizens as voluntary sensors: Spatial data infrastructure in the world of web 2.0. *International Journal of Spatial Data Infrastructures Research*, 2:24–32.

Gutman, M. and Stern, P., editors (2007). *Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data*. National Academies Press.

Hillier, B. (2007). Space is the Machine: A Configurational Theory of Architecture. Space Syntax, London, UK.

Houée, M. and Barbier, C. (2008). Estimating foreign visitors flows from motorways toll management system. In *9th International Forum on Tourism Statistics*.

Kostakos, V., Nicolai, T., Yoneki, E., O'neill, E., Kenn, H., and Crowcroft, J. (2008). Understanding and measuring the urban pervasive infrastructure. *Personal and Ubiquitous Computing*.

Kozlowski, M. (2007). *Urban Design: Shaping Attractiveness of the Urban Environment with the End-Users*. PhD thesis, University of Queensland, School of Geography, Planning and Architecture.

Microsoft (2008). *Being human: Human-computer interaction in the year 2020*. Technical report, Microsoft Research Ltd, Cambridge. http://research.microsoft.com/en-us/um/cambridge/projects/hci2020/downloads/BeingHuman_A4.pdf

Mummidi, L. N. and Krumm, J. (2008). Discovering points of interest from users' map annotations. *GeoJournal*, pages 215–227.

New York City Economic Development Corporation. (2008). *The new york city waterfalls: The economic impact of a public art work*. http://www.nyc.gov/html/om/pdf/2008/waterfalls_economic_impact_report.pdf. [accessed 25 January 2009].

O'Neill, E., Kostakos, V., Kindberg, T., gen. Schieck, A. F., Penn, A., Fraser, D. S., and Jones, T. (2006). Instrumenting the city: Developing methods for observing and understanding the digital cityscape. In *Ubicomp*, pages 315–332.

Ratti, C., Pulselli, R. M., Williams, S., and Frenchman, D. (2006). Mobile landscapes: Using location data from cell-phones for urban analysis. *Environment and Planning B: Planning and Design*, 33(5):727 – 748.

Reades, J., Calabrese, F., Sevtsuk, A., and Ratti, C. (2007). Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing*, 6(3):30–38.

Sevtsuk, A. and Ratti, C. (2007). Mobile surveys. In *Urbanism of Track 2007*. Delft University Press.

Seybold, J. S. (2005). *Introduction to RF Propagation*. Wiley-IEEE

Wöber, K. (2007). Similarities in information search of city break travelers — a web usage mining exercise. In *Information and Communication Technologies in Tourism 2007*, pages 77–86.

Yim, Y. (2003). *The state of cellular probes*. Technical report, California Partners for Advanced Transit and Highways (PATH).

Zook, M., Dodge, M., Aoyama, Y., and Townsend, A. (2004). New digital geographies: Information, communication, and place. *Geography and Technology*, pages 155–176.