Space and time-dependant bus accessibility: a case study in Rome

Citation: Pinelli, F. et al. "Space and time-dependant bus accessibility: A case study in Rome." Intelligent Transportation Systems, 2009. ITSC '09. 12th International IEEE Conference on. 2009. 1-6. ©2009 IEEE.

As Published: http://dx.doi.org/10.1109/ITSC.2009.5309679

Publisher: Institute of Electrical and Electronics Engineers

Persistent URL: http://hdl.handle.net/1721.1/60234

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Proceedings of the 12th International IEEE Conference
on Intelligent Transportation Systems, St. Louis, MO,
USA, October 3-7, 2009

Space and time-dependant bus accessibility: a case study in Rome

Fabio Pinelli [1,2] Anyang Hou[1] Francesco Calabrese[1] Mirco Nanni[2] Chris Zegras[1] Carlo Ratti[1]
[1]MIT Senseable City Lab [2]KDD-Lab ISTI-CNR
Massachusetts avenue , 77 Cambridge, USA via G. Moruzzi, 1 Pisa, Italy
email: {pinelli, cnhanya, fcalabre, ratti} @mit.edu
mirco.nanni@isti.cnr.it

***Abstract*—The study of the accessibility in the city has always been appealing for urban planners and public trasportation companies. Nowadays, thanks to the availability of tracking devices on public transportation devices, it is possible to evaluate such accessibility very accurately, and derive useful performance measures. In this paper, we propose a complete methodological process to analysize and measure the accessibility of a city using bus GPS traces. Firstly, the process involves the application of recent results in spatio-temporal data mining in order to detect bus lines and bus stops from the traces dataset. Then an agent-based algorithm is used to simulate human mobility in the network, in order to study how the accessibility of the city changes over time, and starting from different locations in the city. Finally, the methodology is applied to bus traces collected for the city of Rome and both the detection process and the accessibility study are deeply investigate.**

I. INTRODUCTION

Measures of accessibility in the city can greatly help people assess the necessary time from one place to the other place, or search some useful geographic information such as the appropriate housing around working places within given commute time. However, due to different definitions of accessibility and research purposes, the geographical accessibility is studied in different ways. In general, accessibility measures can be categorized as infrastructure-based, location-based, person-based, and utility-based approach [14]. Among these, the location-based accessibility can be easily understood and is frequently used in public transport system. In this paper, the accessibility is defined as the accessible area that can be reached using a given transportation network.

The connectivity of network greatly affects the accessibility rather than the Euclidian distance. Normally, a travel using bus transit system includes the access to system and access of system. The former one is normally presented as the walking time that heads to or leave from the transport system and normally they varies among different bus stops. On the other hand, the in-system accessibility measurement represents the efficiency of the transportation system, which also is the efficiency of linkage among different areas in the city via the connection among bus stops.

Traditionally, accessibility research focuses on existing built environment and transportation facilities. These researches can be classified as static methods because the result will not change under different time-dependent factors, such as traffic condition and operator performances, and concern mainly on how to access nearest facilities such as hospitals, stores and office buildings. In these cases, the transportation network or timetable can be utilized to examine the static accessibility. The popularity of location-based devices, such as Global Position System and location-based service system, provides an opportunity to research on the dynamic accessibility which is seldom seen in the previous research.

This paper concentrates on the "real-time" accessibility in different areas by analyzing the real-time data that are collected by GPS devices. Our research can be traced to Real Time Rome Project, which was once exhibited at the 10th International Architecture Exhibition of Venice Biennale during 2006 [4]. The project was conducted by MIT Senseable City Lab in corporation with Telecom Italia, ATAC [13] and other organizations. It was the first time that a large urban area was covered using a variety of sensing systems, including the mobile phones and GPS devices in the buses and taxis [12], [1]. In particular, ATAC, the public transportation authority in Rome, provided the spato-temporal bus traces which were collected by the GPS devices located on the buses and sent to an ATAC server instantaneously. This paper aims at analyzing this latter dataset to provide a full understanding of dynamic performance of the public transport network.

The rest of paper is organized as follow: In section II, we introduce the related work which includes the accessibility research; in section III, we introduce the detection process which translates the raw data into meaningful spatio-temporal information and then we design an agent-based simulation process to evaluate the bus transit accessibility of single bus stops; in section IV, we use the agent-based simulation model to examine the real-time accessibility of different parts of the cities over time; finally in section V, we summarize the whole idea and illustrate some possibilities of the proposed real-time accessibility computation method.

II. RELATED WORK

The accessibility research has a relative long track in the transportation and urban planning fields. It can be at least traced to Hansen's work [6], and is now widely accepted as an important transport issue [7]. As mentioned before, there is not a precise definition of it in academia. Nevertheless, there is a suggestion that the accessibility can be expressed as the accumulated opportunities in the general formulation, as in [8]:

978-1-4244-5521-8/09/$26.00 ©2009 IEEE

$$A_i = \sum_j O_j f(C_{ij})$$

The O_j is the opportunity that can be accessed from location i and the C_{ij} is the cost, no matter spatial distance or fare cost, between location i and j. The $f(C_{ij})$ represents the function of the cost. Obviously, it is easily to illustrate and compare the different accessibility of locations using this quantitative measure. However, if we consider an intuitive way to express and even visualize the accessibility from the perspective of citizens in this paper, such as where i can reach from a given bus stop in given time, the output, which may be the connectivity result in the whole network, should be direct and meaningful.

Thus, the popular way is to express accessibility using physical distance. Normally, some constrains are considered. A well-known approach is the space-time constrain, which was developed firstly by Hagerstrand in [5]. This approach treats accessibility as the potential locations that can be reached within the time budget. Apparently, the understandability is one of its most advantages, typically in the usage of internet-based query system for citizens. Citizens can check which areas are accessible within a given time budget or check the necessary travel time between two specific locations. For urban planners, this way is also a good way to evaluate the performance of public transport facilities.

The accessibility research in public transport often relates to its network attributes or the distribution of bus stop, which respectively be discussed above as the accessibility of the network and the accessibility to the network. A variety of data, typically in timetable and transport networks, are broadly used to build models. O'Sullivan at al., in [11], use GIS data to develop isochrones on the basis of bus route timetables and street network data . The output is an isochrone area that can be reached within a given starting point and a specific time threshold. Martin et al.'s research, in [9], is concerned with the accessibility of hospital services by public transports, by developing a software tool to analyze the bus timetable data. The result presents the travel time from bus stops to a given hospital. For the research of accessibility to the network, spatial distribution of public transport facilities are often concerned with. Chien e al. in [3] develop a mathematical model to optimize the bus stop location to improve transit accessibility. In [10], Murray et al. examine the accessibility to public transportation by focusing on the location of public transport facilities with respect to population.

No matter what kind of definition the accessibility is, the inherent common ground is to describe the spatial interaction among areas. So far, however, the traffic conditions have never been considered in the calculus of those measures, mainly because of the lack of ways to take it into account. Thus the GPS data, which records the "pulse of the city" provides a new perspective to focus on real-time accessibility.

III. EXTRACTION OF BUS SERVICE PARAMETERS FROM RAW DATA

Before describing the proposed extraction procedure, we descriptions of the used concepts: bus trajectory, bus trip, bus schedule and bus route.

Bus trajectory: The completed movement information of buses that collected from GPS devices from the start time to end time when data is collected. A trajectory should be the sets of points which include the bus spatial and temporal information in the form of raw real-time data. The start point and end point of a trajectory is subjective and determined by the data collection period. The movement between two consecutive points is assumed to be a straight-line movement. Bus trip: A bus trip should a completed travel that covers an appointed path. It includes the completed movement from the starting bus stop to the terminal bus stop. It is concerned with the real traffic condition, like congestion, operation performance and so on. The segmented trajectory is the initial form of a bus trip, which records the whole movement of a bus in its appointed route.

Bus schedule: A bus schedule is a given, or say, planned, bus timetable that covers the fixed location and time-stamp information. It is a typical example of static data because it is constant guidance to regulate the bus operator (some bus companies may have different schedule for weekday and weekend and some big events, such as Thanksgiving Day).

Bus route: A bus route is a sequence of bus stops. Generally, a bus route should have many bus trips running by several buses in different times every day. It is more likely a spatial "chain" that records the order of bus location.

The extraction process is to extract meaningful information from initial GPS data. Given a sequence of location and time information collected from a bus, the goal is to recognize the hidden activity pattern in which a bus performs over a period. For instance, some consecutive recorded points of GPS data may share the same (or very near) location, which maybe be the bus stop location, in different times when a bus decelerates its speed and stop for a while to pick up or drop off passengers. Also, the bus route information can be summarized from the repeated movement of buses because the sequence of location of recorded points which they send by using GPS device, if in the same path, is similar, although the exact location of recorded points may not be the same. To prepare for the simulation process, the required outputs are bus trips and bus stop location.

The general idea of simulation process is described below. First, extract the bus route information by using the repeated movements in the same path of a bus. Then, judge the location of bus stops by considering the speed and direction of a bus. Third, translate the bus trip information into a sequence point of bus stop and time. The expected output is bus trips and bus stops data, while bus routes data is also needed to generate for preparation.

A. Data preprocessing

In this section, we explain the process adopted to obtain meaningful information from the GPS bus traces. A key information for estimating the bus accessibility of the city is the activity of buses. The activity log should contain the complete list of stops of each bus during the whole analysis period, i.e. where they stops and at which bus stop:

$$schedule(bus\_id) = \langle stop_1, \ldots, stop_N \rangle$$

Second, we need the scheduled sequence of stops for each bus. The result should be a function of the following kind:

$$travel(bus\_id) = \langle (stop_1, t_1), \ldots, (stop_N, t_N) \rangle$$

Our input dataset describes the trajectories followed by each *physical* bus, while the relationship between bus trajectories and bus lines and corresponding stops is missing. Indeed, the same bus usually serves several different lines at different times. We try to discover this relationship by applying to the data some heuristics and trajectory cluster algorithms. In the following the extraction process for estimating bus stops is shown in detail:

Trajectory segmentation. We divide each bus trajectory into segments, by means of a heuristics that splits a trajectory whenever the time gap between two consecutive points exceeds a given threshold. Therefore, each resulting segment represents a trip of the bus on a line. A sample illustration of the effects of this step is shown in Figure 1, where a single bus trajectory actually corresponds to four different routes, served by the same bus at different times.

Determining bus route. The overall set of trajectory segments obtained in the previous step is processed by means of a trajectory clustering algorithm, in order to group together trajectory segments that approximately follow the same path. Each resulting cluster will represent a (tentative) bus line. The similarity between two trajectories is computed by means of the EDR distance measure — a recent, effective and noise-tolerant distance developed for multi-dimensional time series [2]. A sample cluster obtained is depicted in Figure 1(c), where a set of several different trajectories follow almost exactly the same route.

Stops detection. A multi-step heuristics is used to find the stops of each bus line (i.e., of each cluster found in the previous steps). For each cluster obtained in the previous step: (i) we detect stop points; (ii) we cluster close points intra-cluster and extra-cluster obtaining a set of bus stops; finally, (iii) we translate the trajectories as sequences of bus stops.

At end of this process, we obtain the complete schedule of the bus transportation system. Indeed, we know exactly where and when the bus is stopping. Therefore, the new dataset is the input of the simulation algorithm described in the following section. It allows us to evaluate the connectivity in Rome.

B. Connectivity estimation through simulation

The main objective of the section is to define and compute a measure of the *level of connectivity* of any point of the city at any time of the day. The analysis focuses on bus transportation, allowing to switch between bus lines by walking from one bus station to any other one at walking distance.

The analysed dataset describes the first week of September 2006 in Rome. The data are collected from 12358 different bus, recording 8330430 points. Each bus performs more than one bus lines.

The pseudo-code of the algorithm, which requires function *travel* described above, is shown in algorithm. 1.

The algorithm starts by placing an *agent* on a starting bus stop at a given time instant, both of them input parameters provided by the user. Then, all possible means of movement for the agent are considered, i.e.: all the buses that stop at its actual location, and all walks that move towards close bus stops. For each of them the agent duplicates, and all clones are put in a queue. The same process is iterated by considering next the agent in the queue that arrives first at destination. However, if the agent arrives too late w.r.t. a maximal duration (another parameter of the algorithm) or any other agent had already reached that destination earlier, the new agent is removed, since it cannot improve the route followed by its predecessor. When the queue is empty, the simulation stops.

Notice that: (i) the output of the algorithm is the set of time-stamped locations reached by the simulation; (ii) switching between stations by walk might take some time, which is given by function $switch(stop_{from}, stop_{to})$, set to ∞ if the connection is not allowed (for instance, whenever the distance between the two stations is too large), and to 0 if we don't care about switching times; finally, (iii) commuting bus in the same station without changing the bus stop is implicitly performed by the algorithm, and therefore it does not require an explicit treatment, as opposed to bus stop switching. For this reason, self-switching from a station to itself is avoided (self-switching time is set to ∞).

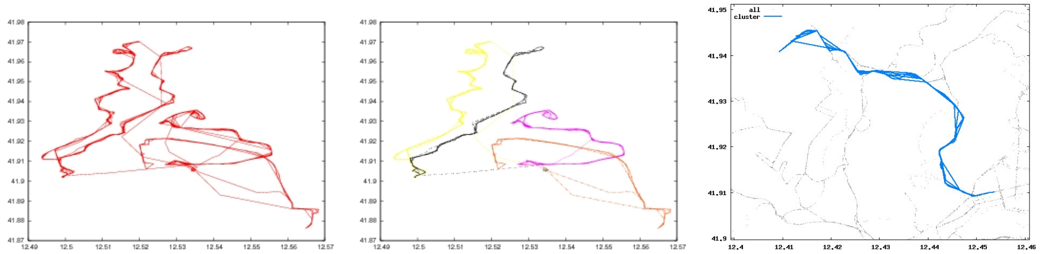
We remark that the problem solved by our simulation-based algorithm is essentially a shortest path / minimum spanning tree problem where the weight of each edge depends on time. The order of visit and filtering approach adopted (lines 4,6, and 12) ensures that almost no redundant computations are performed.

IV. EXPERIMENTS

In this section, we show the results of experiments performed adopting the above defined methodology. The first set of runs highlights possible uses of the simulation algorithm. In fact, we select a certain starting point and we measure the accessibility over time from this starting point. A second set of experiments focuses on a overall analysis of the public transportation system, that is all bus stops are starting point of the simulation algorithm.

A. Single bus stop analysis

Figure 2 reports the visual summary of a few runs of the simulation, starting at different hours of the same day (from 4 a.m. to 2 p.m., with a step of two hours) and with different maximal duration of the trip (from 1 to 3 hours). The starting location of the simulation is fixed, represented by a (red) circle. Walking is allowed only between stops not farther that 100



(a) Complete bus trajectory (b) Segmented bus trajectory (c) Sample cluster of segments

Fig. 1. Sample bus trajectory before and after splitting

```
Algorithm 1: Agent-based simulator
Require: $\mathcal{S}$, start_loc, start_t, duration.
Ensure: ReachedStops = set of bus stops reached by the simulation
for stops $s_i \in \mathcal{S}$ do
    visit($s_i$) = (INF,INF); // (t_walking, t_no_walking)
    switch($s_i$,$s_i$) = INF; // fake switching are prohibited
    Agents = (start_loc, start_t, false) ; // (l,t,walked_flag)
    while Agents is empty do
        $agent = (s_i, t, w)$ = pop from Agents the one with smallest "t";
        $(t\_w, t\_nw) = visit(s_i)$;
        if $(w = true$ AND $t \min min(t\_w, t\_nw)$ OR $(w = false$ AND $(t \min t_n w)$ then
            if (w=true) then
                t_w = t;
            else
                t_nw = t;
            MovesByBus = { ($s_i'$,t',false) — bus in Buses, $\langle(l, t''), (l', t')\rangle$ in travel(bus), $t'' >= t$ };
            if (w=false) then
                Walks = $\{(l', t', true) | t' = t + switch(l, l') < INF\}$;
            if (w=false) then
                Walks = $\{(l', t', true) | t' = t + switch(l, l') < INF\}$;
            else
                Walks = {};
            Selection = $\{(l", t", w") \in MovesByBus \bigcup Walks |$ $t" <= start_t + duration$ AND $(t_w", t_n w") = visit(l")$ AND $(w" = true$ AND $t" < min(t_w", t_n w")$ OR $w" = false$ AND $(t" < t_n w"))\}$;
            $Agents = Agents \bigcup Selection$;
Return$\{(\mathbf{l}, \mathbf{min}(\mathbf{t_w}, \mathbf{t_n w})) | \mathbf{visit(l)} = (\mathbf{t_w}, \mathbf{t_n w})\}$;
```

meters, and such movements are performed at a fixed walking speed set to 1 mt/sec. A summarization is presented in Tab. I.

The results show, as expected, that longer trip durations yield consistently larger reachable areas. Moreover, the hour of the day has a deep impact on accessibility. For instance, at 6 a.m. and at 12 a.m. a large area can be covered 3 hours, while such are is much reduced starting at different hours.

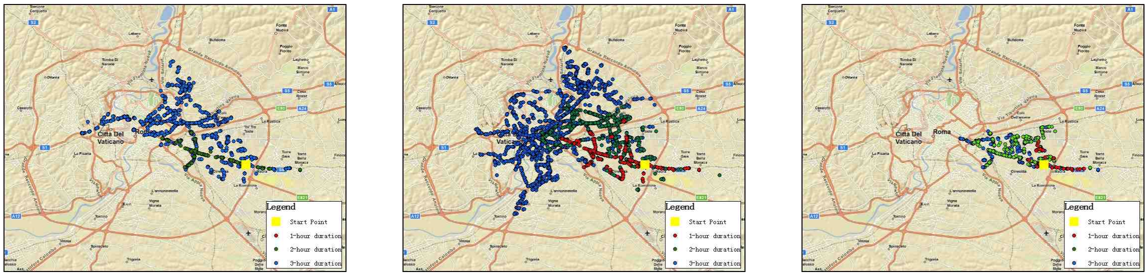
Hour of the day	N. of bus stops reached
4am + 1 hour	0.04%
4am + 2 hour	1.71%
4am + 3 hour	13.75%
6am + 1 hour	3.62%
6am + 2 hour	9.99%
6am + 3 hour	37.79%
8am + 1 hour	0.63%
8am + 2 hour	3.07%
8am + 3 hour	4.59%
10am + 1 hour	0.08%
10am + 2 hour	0.76%
10am + 3 hour	6.03%
12am + 1 hour	4.16%
12am + 2 hour	13.70%
12am + 3 hour	39.07%
2pm + 1 hour	1.44%
2pm + 2 hour	3.29%
2pm + 3 hour	7.05%

TABLE I
N. OF BUS STOPS REACHED IN DIFFERENT HOUR OF THE DAY.

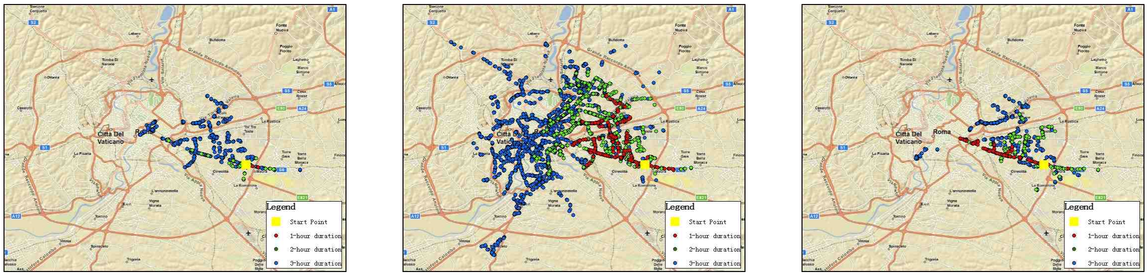
Moreover, we perform other experiments in order to evaluate the connectivity at same hour but in different days of the week. In particular, we compare the of Wednesday and Sunday. Fig. 3(a,b,c) shows how the covered area is different in these two days starting at 8am. How we expect, during the weekend day the covered area is smaller than in a week day. If the simulation starting time is set up to 12 am, we obtain similar results. Indeed, the covered area by the buses in week end days is smaller than the covered area of the week day (see Fig. 3(d,e,f)). Tab. II resumes all results.

Day	Hour of the day	N. of bus stops reached
Sunday	8am + 1 hour	0.03%
	8am + 2 hour	0.83%
	8am + 3 hour	**14.19%**
Wednesday	8am + 1 hour	1.90%
	8am + 2 hour	9.91%
	8am + 3 hour	**35.65%**
Sunday	12am + 1 hour	0.39%
	12am + 2 hour	3.34%
	12am + 3 hour	**25.39%**
Wednesday	12am + 1 hour	1.07%
	12am + 2 hour	3.38%
	12am + 3 hour	**15.67%**

TABLE II
N. OF BUS STOPS REACHED IN DIFFERENT DAYS AT THE SAME HOUR.

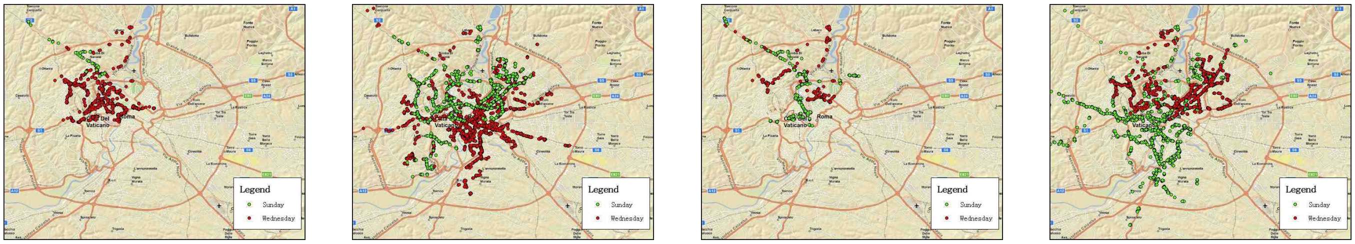


(a) accessibility at 4am (b) accessibility at 6am (c) accessibility at 8am



(d) accessibility at 10am (e) accessibility at 12am (f) accessibility at 2pm

Fig. 2. Accessibility at different hours of the same day. Red dots represent the bus stops reachable in a hour, green dots in 2 hours and blue dots in 3 hours. The starting point is indicated with a bigger yellow dot.



(a) Sunday-Wednesday 8AM+2h (b) Sunday-Wednesday 8AM+3h (c) Sunday-Wednesday 12AM+2h (d) Sunday-Wednesday 12AM+3h

Fig. 3. Accessibility at same hours in different day. Red dots indicate Wednesday results, and green dots on Sunday

B. *Overall analysis*

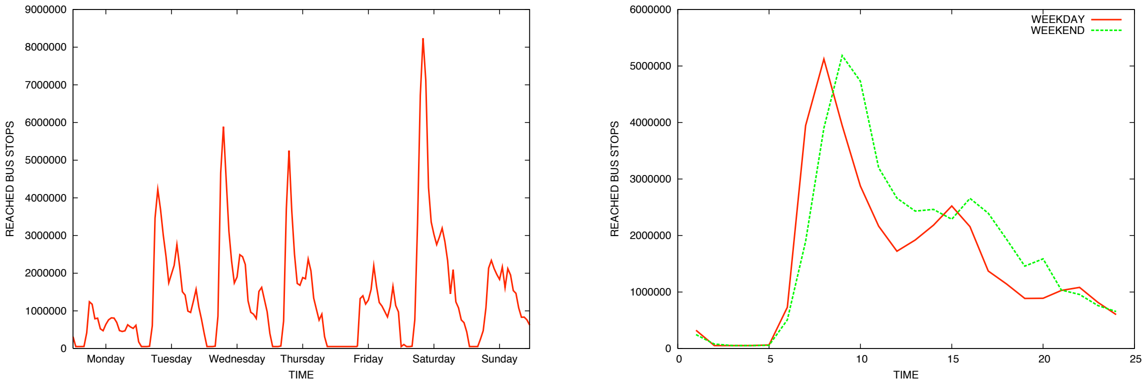
In this section, we show the results obtained using all the bus stops as starting point. In such way, we are able to evaluate the performances of the whole system during a whole week. Further experiments are performed in order to evaluate the network dynamics in time. All the bus stops are used as starting points for each hour of the day and with a hour as duration of the simulation. The results are shown in Fig. 4(a), where we note that there are several peaks that correspond to the rush hour in the morning. We obtain a higher peak on Saturday because a special event occurred in Rome: *the white night*. Another interesting results is shown in Fig. 4(b), where we show the different peak hour between week day and week-end day. The results are obtained averaging the results represented in Fig. 4(a). As we can see, the peak hour in the weekend seems one hour postponed w.r.t. the weekday.

Finally, we set up other experiments in order to evaluate the accessibility distribution over space. We run the simulator for each hour of the day with a duration equal to one hour. The results are shown in Fig.6. The zone where the color is more clear corresponds to a higher accessibility. This zone connects residential area to business area. Moreover, this hyper-accessibility seems to fill a gap in the public transportation system. In fact, the subway network for archeological reason is not well developed in Rome and it does not cover this zone. The results are shown in Fig. 6.

V. CONCLUSION AND FUTURE WORK

In this paper we define a methodology to first detect bus lines and bus stops from GPS data, then we apply a simulation agent-based algorithm to measure and show the accessibility in the city. Moreover, the simulation algorithm allows a deep investigation of the accessibility of the city. In this paper we provide just some preliminary results. They demonstrate how it can be applied and how it is appropriate to achieve this problem. As future work, we plan to apply the simulation algorithm in order to evaluate how the connectivity changes over the time in the city, and which are the less connected zone of the city and how they change over the time.

Acknowledgements The authors wish to thank Assaf Biderman, Fosca Giannotti and Dino Pedreschi for their technical



(a) Overall weekly accessibility (b) Comparison between peak hours

Fig. 4. Overall accessibility in different hour of the day

(a) 6:00 (b) 8:00 (c) 10:00

Fig. 5. Overall accessibility spatial distribution (morning)

(a) 12:00 (b) 14:00 (c) 16:00

Fig. 6. Overall accessibility spatial distribution (afternoon)

support, and for fruitful discussions. This work has been carried out within the M.I.T. Portugal. Moreover, the authors wish to thank At&t and Audi Volkswagen. F. Giannotti and D. Pedreschi. The project M.I.T. Portugal. At&t. Audi Volkswagen

REFERENCES

[1] Francesco Calabrese and Carlo Ratti. Real time rome. networks and communication studies. In *Official Journal of the IGU's*, volume 20, pages 247–258. Geography of Information Society Commission, 06.

[2] Lei Chen, M. Tamer Özsu, and Vincent Oria. Robust and fast similarity search for moving object trajectories. In *SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 491–502, New York, NY, USA, 2005. ACM.

[3] Steven I. Chien and Zhaoqiong Qin. Optimization of bus stop locations for improving transit accessibility. In *Transportation Planning and Technology*, pages 211–227, 2004.

[4] Biennale di Venezia. 10th International Architecture Exhibition of Venice Biennale, 2006. `www.biennale.org`.

[5] Torsten Hägerstrand. *The domain of human geography*. 1973.

[6] Walter G. Hansen. How accessibility shapes land use. In *Journal of the American Institute of Planners*, pages 73–76, 1959.

[7] Susan Hanson. *The geography of urban transportation*. The Guilford Press, 1986.

[8] J. G. Koening. Indicators of urban accessibility: theory and application. In *Transportation*, volume 9, pages 145–172, 1980.

[9] David Martin, Hannah Jordan, and Paul Roderick. Taking the bus: incorporating public transport timetable data into health care accessibility modelling. In *Transportation Research Part A*, volume 40, pages 2510–2525, 2008.

[10] Alan T. Murray, Rex Davis, Robert J. Stimson, and Luis Ferreira. Public transportation access. In *Transportation Research Part D*, pages 319–328, 1998.

[11] David O'Sullivan, Alastair Morrison, and John Shearer. Using desktop gis for the investigation of accessibility by public transport: an isochrones approach. In *International Journal of Geographical Information Science*, pages 85–104, 2000.

[12] Jonathan Reades, Francesco Calabrese, Andres Sevtsuk, and Carlo Ratti. Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing*, 6(3):30–38, 2007.

[13] Atac S.p.A. Agenzia per i trasporti autoferrotranviari del comune di roma, 2008. `http://www.atac.roma.it/`.

[14] Chris Zegras. Sustainable urban mobility: exploring the role of the built environment. In *PhD Thesis*. MIT, 2004.