8th Symposium on Location-Based Services
Vienna, 21–23 November 2011
LBS 2011

Amendment to the
Proceedings
of the 8th International Symposium on Location-Based Services

Georg Gartner and Felix Ortag (Editors)

ICA ACI
International Cartographic Association
Association Cartographique Internationale

TU WIEN
FORSCHUNGSGRUPPE KARTOGRAPHIE
Research Group Cartography

This Amendment contains papers that did not find their way to the proceedings due to an error by the editors.

The 8th International Symposium on Location-Based Services took place at Vienna University of Technology from November 21 to 23, 2011 and was organized by Vienna University of Technology's Research Group Cartography with the endorsement of the International Cartographic Association (ICA).

Additional papers from this conference were published in: "Advances in Location-Based Services", edited by Georg Gartner and Felix Ortag, Lecture Notes in Geoinformation and Cartography, Springer-Verlag, Berlin, Heidelberg, 2012.

Smartphone illustration on cover by Giovanni Meroni
LBS 2011 logo by Manuela Schmidt

Edited by Georg Gartner and Felix Ortag
Type setting of the chapters by the authors
Processed by Felix Ortag

Think before you print!

Table of Contents

Finding Interesting Places and Characteristic Patterns in Spatio-Temporal Trajectories

Udo Feuerhake, Colin Kuntzsch, Monika Sester

Institute of Cartography and Geoinformatics,
Leibniz Universität Hannover, Germany

Abstract. Trajectory data often includes knowledge about the movement and the behavior of individuals, which is useful for analyzing problems in domains like animal migration or security. In this paper we present an approach to identify interesting places and determine unusual behavior of individuals from large amounts of trajectory data.

Keywords. Spatial, Data Mining, Pattern, Places, Segmentation, Extraction, Algorithms, Incremental

1. Introduction

Understanding of space-time-patterns is relevant for many problems, e.g. animal migration, traffic analysis, security. This paper describes first stages of a framework for the interpretation of trajectory data with respect to specific patterns or dominant structures. There are several requirements concerning the interpretation of such space-time-trajectories, which are also application dependent: The trajectories can have different sampling rates, they can contain additional attributes (besides position); they can be generated by the same individual or several individuals, which may or not be known beforehand. The goal can be to identify certain patterns, which can be discerned into individual patterns of the trajectory itself and group patterns of several trajectories. A classification of space-time patterns is described by Dodge, et al. (2008). Concerning the first issue, the way the space is traversed is relevant, e.g. in terms of straight lines, circles, zigzag; then methods described by Buchin et al. (2009) can be applied. Concerning the group patterns, dynamic patterns like flocks, and also static patterns like convergence or encounter (see e.g. Laube et al. (2008)) can be identified. Furthermore, also the issue of identifying regions of common space usage can be identified, a sort of linearly extended encounter. This latter

LBS 2011 Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

pattern can also be used to identify the underlying path (network) of the moving objects (e.g. Krumm, et al. (2009)).

In the paper, we will concentrate on a scenario with different individuals traveling through space and time. We are first of all looking for encounters in the sense of places, where the individuals recurrently appear. To this end, an incremental approach is presented. After the places of interest are identified, a graph can be constructed, consisting of the interesting places as nodes and the connecting track segments as edges. Further analysis is conducted on these segments, with respect to movement type, as well as the possibility of aggregating nearby traces. Finally, general information about the connections in the graph is derived, which also include probabilities of paths through the network.

In the following, a brief description of the algorithms is given, together with illustrative examples. The paper concludes with a summary and outlook on the next steps.

2. Motivation

2.1. Context of Problem

This paper deals with the evaluation of trajectory data in an observation scenario. There the main focus lies on the identification and evaluation of movement patterns to detect critical behavior of observed individuals. Considering the fact that the developed technique shall operate on a low-performance system, it has to work efficiently.

2.2. Problem Definition and Description

A major challenge of the project is to define critical behavior. Due to the fact the evaluation is based on trajectory data, the behavior itself is represented by the movement of an individual. Therefore critical actions are directly related to critical movement. Certainly, there are various criteria characterizing movement as critical. These criteria heavily depend on the spatio-temporal context, the individual is moving in. Thus, when looking for uncommon behavior, we are looking for uncommon trajectories in the data. Common trajectories are identified by looking for clusters of similar locations and paths. Deviations from these clusters are considered as abnormal. However, they do not have to be necessarily critical. After the detection of the requested movements further decisions have to be made to classify an abnormal one as critical. Critical has then to be defined within the context of an application.

In summary we are looking for a method that separates common from special movements according to their spatio-temporal context.

2.3. Related Work

This is not the first approach structuring and evaluating trajectory data in a spatial context. Ashbrook and Starner (2003) describe a way to learn significant locations and make predictions from GPS data. They structure the existing data by finding common places, where people stay for certain time. After that they use a k-mean-similar clustering to merge the found places, to reduce their data to an essential minimum. For each of those resulting clusters, called locations, a Markov model is created, which allows predicting the people's next target.

Makris and Ellis (2001) have worked on identifying frequently used paths from video scenes. Thereby, they handle the spatial relationships among the trajectories by generating a graph from people's appearing- (entry nodes) and disappearing-points (exit-nodes) and trajectory junctions the scene. Node-usage statistics provide a measure about the most probable exit-node, so the point, the individual leaves the scene.

Another approach is presented by Baiget and Sommerlade (2008). They want to find trajectory prototypes to estimate subsequent trajectory shapes. To this end, they cluster the already obtained trajectories by their first and last points. After that, they create a trajectory prototype for each of those clusters by combining the contained segments. Those prototypes are used for predicting the most probable shape of following trajectories.

Kang et al. (2005) developed an algorithm for extracting significant places from a trace of coordinates. Instead of using GPS, they use WiFi to collect users' locations. To extract the interesting places from the location data, they suggest a time-based clustering, which relies on a distance and a time threshold.

2.4. Own Approach

Since the behavior of an individual depends on its spatial-temporal context, our problem also demands a spatial structuring of the existing trajectory data. Otherwise we would not be able to compare and interpret the movements. As in some of the related work, we have also decided to extract attractive places. But, instead of clustering any locations found by longer stays of individuals, we identify interesting places in an incremental way by counting their visits by individuals. These places will later be used to derive clustered logical segments from the trajectory data. So the segments of within each cluster will share their own spatial context.

After having clustered the trajectory segments we are able to evaluate the segments within the same environmental background. Subsequently, the aggregated trajectory segments will be clustered a second time using domain-specific parameters and analyzed with respect to their internal structure.

Thus, we present a three step-approach:

1. Extraction of attractive places
2. Segmentation and clustering of trajectory data based on the found places
3. Evaluation of segments within clusters of semantic trajectory segments

Compared to the related work we mentioned above, the main difference, next to the way of finding attractive places, is the sequence of single steps applied for reaching our goal.

3. A Three Step Approach

3.1. Requirements, Assumptions and Definitions

There are a few prerequisites and basic assumptions of our approach. The algorithm needs a sufficiently large amount of input data with a nearly constant, but quite high sampling rate. The data augmentation of our algorithm is based on the Adrienkos' basic concepts of movement data. So a trajectory $T = \{TP_0, .., TP_m\}$ is defined as a sequence of m measured tracking points $TP = (x, y, z, t)$, which contain values for space and time Andrienko et al. (2008).

We define a place to be attractive, if it has been visited several times by one or more individuals. Therefore, we define a necessary parameter n, the visit count of a place, for separating candidate places, visited n-1 times, from attractive ones. Reasonable values for n basically depend on the phenomenon to be analyzed and on the spatial density of data. The higher the density the higher n should be.

Furthermore, we define the size of a place's geometry. The smaller the size is set, the more places and clusters may be found.

Finally, we assume that individuals are getting slower or even stop at attractive places. In general, stopping can be identified by analyzing the velocity. In this work we use a fixed velocity threshold v between two consecutive tracking points. If additional knowledge about the objects or the scene is known which influence stopping, they can also be included.

LBS 2011 Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

A more detailed evaluation of the influences of the parameters will be presented together with the discussion of the results.

3.2. Step 1: Extraction of Attractive Place

In generally, candidate places are found by examining movement data and stops of individuals. Those candidates will upgrade to “attractive places” if a threshold for the number of visits (n) is reached. Since our search for candidates works on every single tracked movement, the algorithm is operating incrementally. So it works at the runtime, which is important when using it in a surveillance system as described above. In the following paragraph we explain the extraction of attractive places in detail.

For each pair of consecutive observations we decide, whether the observed individual/object has moved significantly. In this work we use the observed velocity, calculated from the travelled distance between two samples with known timestamps. This approach can be easily adapted for domains with high sampling-frequency by aggregation of more than two consecutive observations. If the calculated velocity fulfills certain stop-criteria, the observed movement M is interpreted as a stop along the trajectory. M’s center O is tested for containment within an existing place. If this check fails, i.e. there is no existing place containing the movement’s location, a new candidate place (C) is created and added to the set. In the other case the found place’s (P’s) center will be adjusted. This correction consists of moving the current center towards O. The new center will be the mean of all previous movement centers contributing to P. Further, P’s visit count is increased. If the count reaches the predefined threshold n, P will be considered an attractive place and put into the corresponding set. The following pseudo code describes the first step of the algorithm.

```
Places = Ø, AttractivePlaces = Ø, Place P, Movement M, Location L
FOREACH tracked movement
  M = new tracked movemen
  calculating parameters of M
  O = center of M
  IF M satisfying stop-criteria
      FOREACH P ϵ Places
          IF O inside P
              Increase visit count of P
              Correct center of P
              IF visit count == n
                  Upgrade P to attractive place
                  AttractivePlaces = AttractivePlaces ∩ P
              END IF
          ELSE
              Create new candidate place C
              Places = Places ∩ C
          END IF
      END FOR
  END IF
END FOR
```

3.3. Step 2: Segmentation and Clustering of Trajectory Data

In our next step the clustering requires a segmentation of the existing trajectories. Therefore the attractive places which have been found in the previous step are used. These are considered as the start and the end points of the segments, respectively.

The segmentation algorithm iterates over the tracking points of each trajectory and checks, if the current tracking point is contained inside any of the previously identified places (during the segmentation not only stops are considered, but also motion passing through a place). Whenever a trajectory point passes a place, the previous segment, if there is one, is closed and stored separately. A new segment starting at the previous trajectory segment's end place is initialized. Trajectory points not inside any of the place models are added to the current segment. Segments at the start or end of trajectories not starting or ending at place models are discarded. The result of the algorithm is a set of trajectory segments, each associated with a pair of places it starts or ends at.

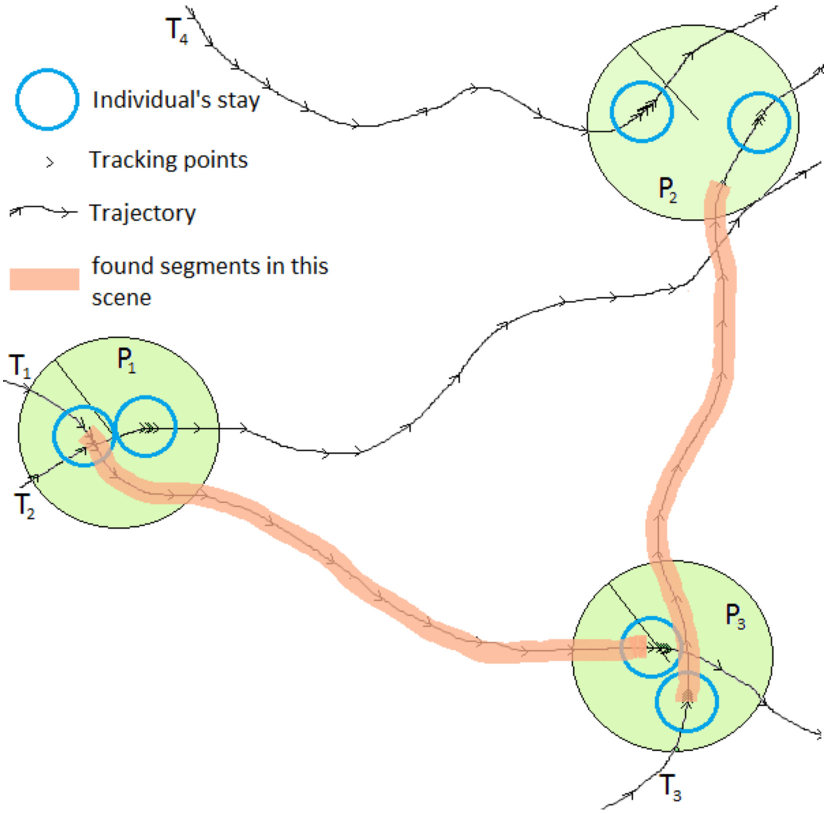


Figure 1: Illustration of trajectory segmentation

Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

```
Trajectories = {T0..TN}, Trajectory T, TrackingPoint TP
AttractivePlaces places = {P0..PM}, AttractivePlace P, AttractivePlace Pstart,
Segments = Ø, Segment S = Ø
FOR EACH T ϵ Trajectories
  FOR EACH TP ϵ T
      FOR EACH P ϵ places
          IF TP inside P
              IF S ≠ Ø
                  S = S ∩ TP
                  IF P ≠ Pstart OR T has been outside of Pstart
                      Segments = Segments ∩ S
                      S = Ø
                  END IF
              ELSE
                  Create new Segment S = { }
                  S = S ∩ TP
              END IF
          ELSE
              IF S ≠ Ø
                  S = S ∩ TP
              END IF
          END IF
      END FOR
  END FOR
  S = Ø
END FOR
```

After the segmentation, we cluster the results. To this end, we use the places again. The clustering features are the start and end places of the trajectory parts, so that every cluster is defined by a pair of places. For the example in Figure 1, there will be every combination of the places P_1, P_2 and P_3, with consideration of their order. The matrix in Figure 2 gives an impression, how the result looks like for the simple schematic example in Figure 1.

Each entry of that matrix represents the number of segments belonging to one cluster. There can be entries also on the diagonal, which means that there are loops, so segments start and end at the same place.

	P1	**P2**	**P3**
P1	0	0	1
P2	0	0	0
P3	0	1	0

Figure 2: The result of the clustering showed as a matrix

This matrix can also be regarded as a non-complete graph, where nodes are represented by the places (P_1, P_2, P_3) and edges by entries greater than 0. This fact does not really matter for the current state of work, but it may be useful for future aspects.

3.4. Step 3: Evaluation of Segments within Clusters

The segmentation presented above leads to a reduction of complexity of the collected data. Instead of one single trajectory per observed individual, we are now able to operate on trips between places. Those trips can be utilized for generating a model of typical migration behavior between places. In this work we present an approach based on clustering of spatial and geometric attributes generated by a single trajectory analysis by means of different strategies of getting from place A to place B. Depending on the domain of observation, different trajectory parameters play more or less important roles in distinguishing those place-crossing strategies and mapping them to semantic categories, e.g. spatial proximity and similar shape may implicate the use of similar routes in a street network, while the same properties would be less useful in domains without spatial restrictions on movement.

Those strategies among trips can be identified utilizing prior domain knowledge to preselect trajectory parameters used for clustering like

- spatial parameters
  - location
- trajectory geometry parameters, e.g.
  - shape
  - curvature
  - sinuosity
- temporal parameters
  - speed
  - time (of day)

4. Experiments and Parameter Evaluation

The first step of our algorithm, the places extraction step, requires three parameters. Reasonable values for the latter depend on the examined scenario. They have to be adjusted to the trajectory density and sampling rate of the data. For that purpose, we use a scenario examplarily to show the influences of the parameters to the resulting number of candidate and attractive places.

A large data set contains several trajectories of animals moving in an area of approx. 100 x 100 km, cf. Figure 3.

LBS 2011 Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

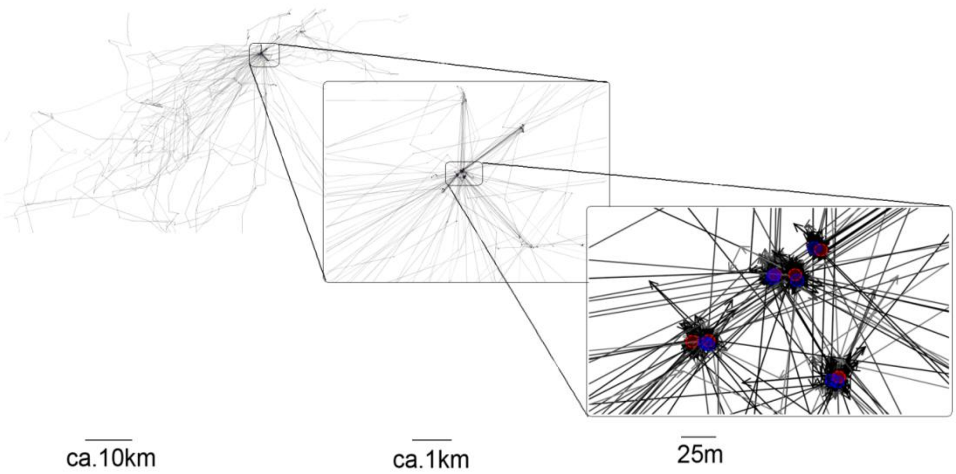


Figure 3: A part of the data set and the places we were looking for showed in three different scales

The parameter settings for the examination of 66149 tracking points are: n=2; r=5m and v=0.1m/s. The data set also contained ground truth in terms of known attractive places, which the individuals often visit. We use this list to verify our results. Our algorithm found more places than given in this list, but it did include the ground truth places as well. An example for the result is showed in Figure 4. There are small deviations, about 5-10m, between the given blue places and the red calculated ones, which mainly can be explained by the inaccuracy of GPS.

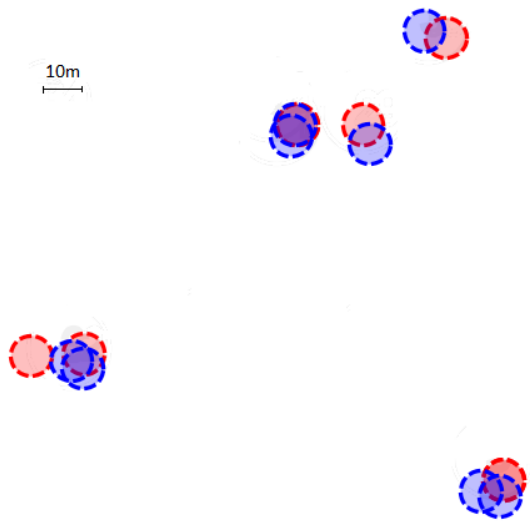


Figure 4: Comparison between blue given places and red determined places

Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

In Table 1 experiments with different parameter settings are listed. We made three series of measurements varying one parameter each time.

#	Parameter setting			Identified places		Remark
	n[-]	r[m]	v[m/s]	Candidates	Attractives	
1	2	5	0.5	5201	356	
2	2	5	1	7969	373	Changing the velocity threshold
3	2	5	2	9939	414	
4	2	5	5	16507	430	
5	2	5	10	29103	458	
6	2	5	20	36346	465	
7	2	5	30	36354	465	
8	2	5	50	36359	465	
9	2	10	0.5	3940	277	Changing the radius
10	2	20	0.5	3040	233	
11	2	50	0.5	2093	228	
z12	2	100	0.5	1532	192	
13	2	200	0.5	1017	161	
14	2	500	0.5	557	171	
15	2	1000	0.5	299	174	
16	2	2500	0.5	87	121	
17	3	5	0.5	5356	201	Changing the visit count
18	4	5	0.5	5421	136	
19	5	5	0.5	5452	105	
20	7	5	0.5	5476	81	
21	10	5	0.5	5495	62	

Table 1: Different parameter settings show the influence to the resulting places

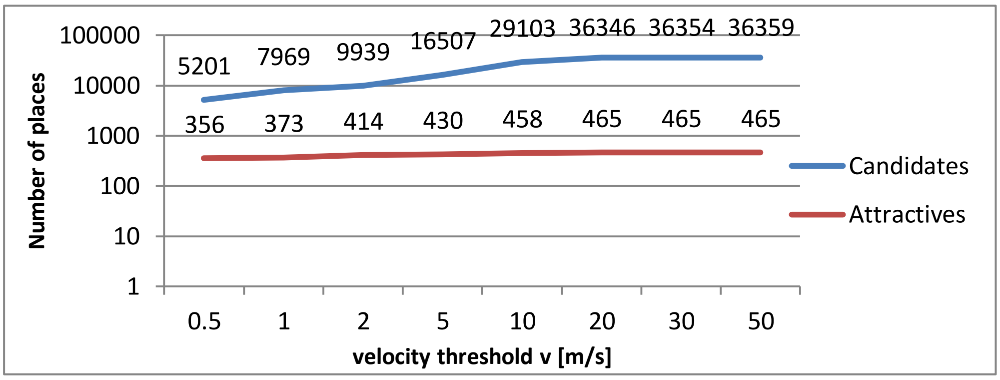


Figure 5: Varying the velocity threshold parameter

Increasing the velocity threshold v (experiments 1-8) leads to an increase in the number of places as then also locations, when paths only geometrically cross without the individual stopping, are being considered. The criteria that determine an individual stays at a place are more often fulfilled. While the upper bound for the number of candidates is equal to the number of tracking points in the dataset, the upper bound for attractive places is calculated by $N_{AttractivePlaces} = N_{TrackingPoints}/n$. Depending on the scenario and the quality of data v has to be adjusted (compare to example 1 and 2 of next chapter).

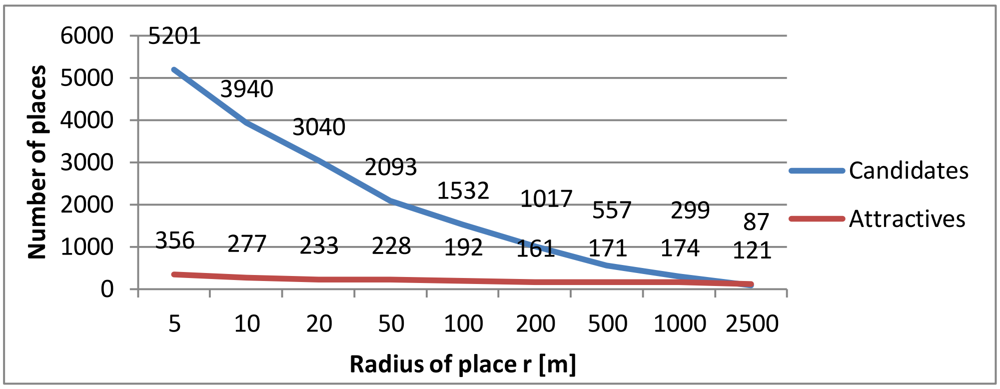


Figure 6: Varying the radius of a place

Varying the size of a place by increasing the radius r (experiments 9-16), decreases the number of candidate and attractive places. The larger the radius becomes the more movements fall into an already existing place and the less candidates are created. Small and adjacent places coincide and are treated as one, especially in areas where the concentration is high. Similar to the velocity threshold, r has to be adjusted to the data as well.

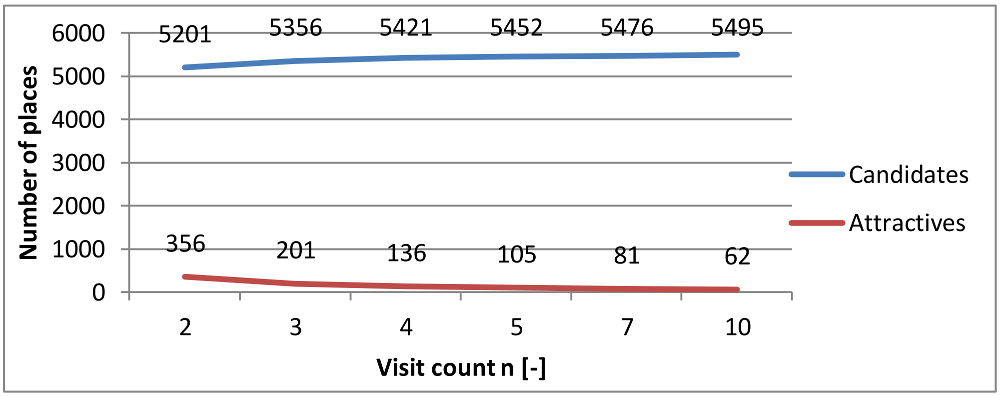
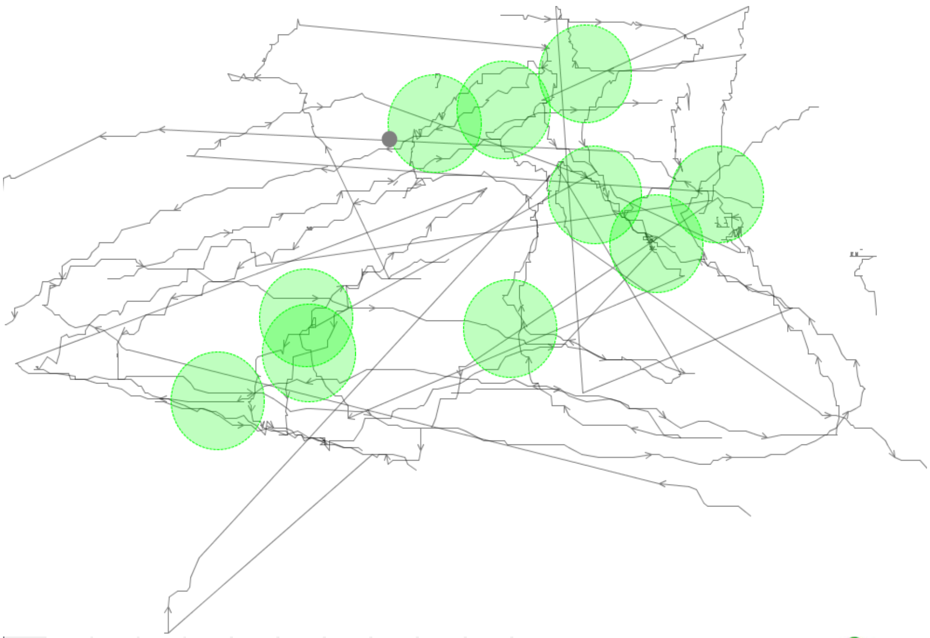


Figure 7: Varying the required visit count parameter

LBS 2011 Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

Increasing the count parameter n decreases the number of attractive places (experiments 17-21). If the density of data is low, the decrease will be higher, because the probability that several individuals visit a faraway place is quite low. This parameter has to be adjusted to the density of the dataset.

The previous result and parameter studies refer to the places-extraction step of the algorithm. The following concerns the second and third steps and shows what the results after those steps may look like. The segmentation results can easily be visualized by a graph structure. To this end, we choose another example and present the corresponding segmentation matrix and graph (cf. Figure 8).



	7	8	17	21	27	28	19	45	20	18
7	1	0	1	0	0	0	2	0	0	1
8	1	2	2	0	0	0	0	0	0	1
17	0	1	0	1	0	1	1	0	0	2
21	0	0	0	0	2	0	0	1	0	0
27	0	2	0	0	0	1	0	0	0	0
28	0	0	1	0	1	0	0	0	0	1
19	2	0	1	0	0	0	0	0	2	0
45	0	0	0	1	1	1	0	0	0	0
20	0	0	0	0	0	0	1	0	2	0
18	0	1	2	1	0	0	0	0	0	0

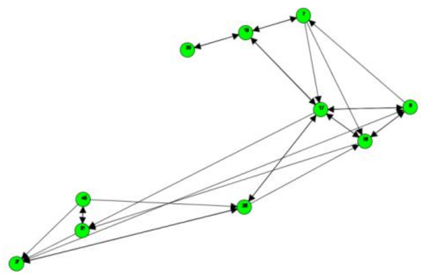


Figure 8: Example for the results after the segmentation step

Figure 9 shows simple examples for the last step of the algorithm, with and without an evaluation of segments of several clusters. There, clustering is applied using the Hausdorff-distances between each segment. The resulting clusters are symbolized by the colors red and blue. In this case some individuals (in red) have used a quite different way to traverse from one to another place. Due to this fact and this small given context, we can consider this movement to be a special one. This may change when there will be more segments within this cluster after a longer observation time.

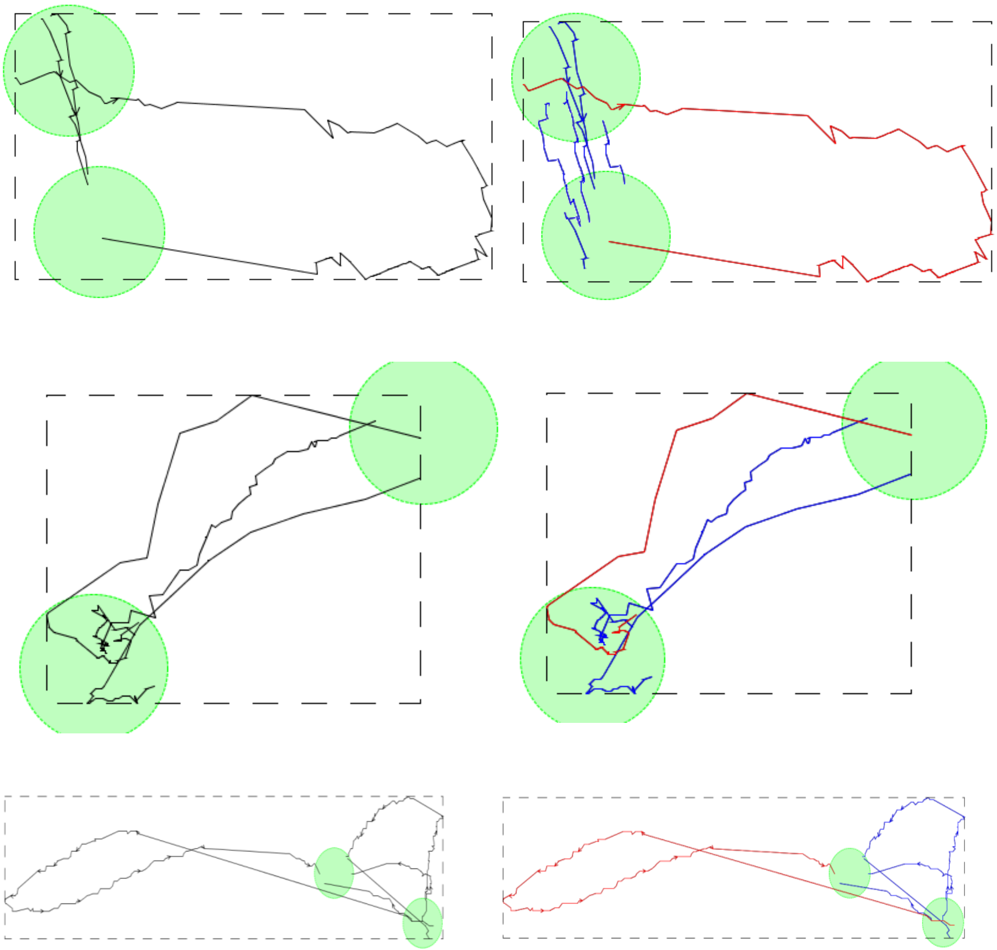


Figure 9: Examples for evaluation step with (right) and without (left) recognition of unusual movements by clustering the trajectory segments

Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

5. Transferability to Other Scenarios

The presented algorithm can be applied to trajectory data provided by different sources like GPS or video tracking. Therefore, we are going to show further examples of the examination of various data types and scenarios. While the first two examples are based on GPS data the last example uses data recorded by video cameras.

In the first example the data (212 trajectories) have been collected by employees of the Institute of Cartography and Geoinformatics while traveling in Hannover, Germany, during a time period of approximately two month. Using the parameters n=2, r=15m and v=0.3m/s leads to 13 attractive places found. Those can be visually inspected and assigned to existing, semantically meaningful places. For this purpose an extract from Google Maps is presented, where the attractive places are marked. Two of them represent tram stops (C, D), one is a crosswalk with traffic lights (B) and another one is the building the institute is located in (A).

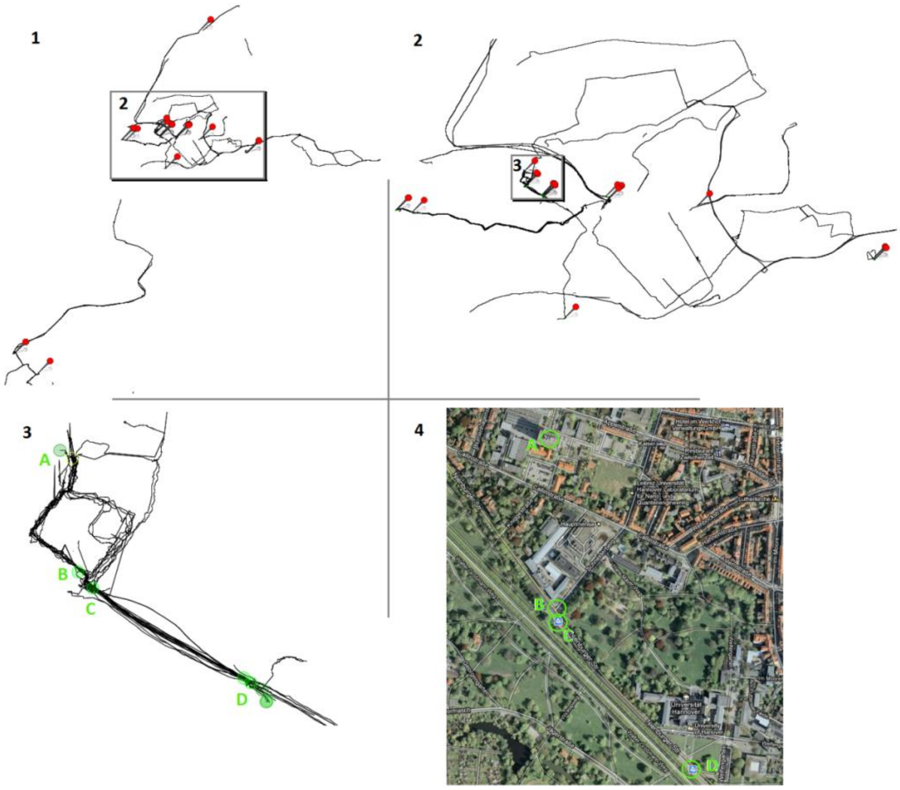


Figure 10: Extraction of interesting places in GPS data set presented in different scales (1-3). An extract of Google Maps for assigning the found places to existing ones (4)

Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

A GPS-game, in which several groups of students participated, provides the data for another example. The results shown in Figure 11 are achieved by using n=2, r=15m and v=0.1m/s as input parameters of the algorithm. Although this example also contains GPS-data, the velocity threshold can be set lower than in the example before, because it is priory known that the students were walking. The starting point of this game was in the left center. Higher densities of trajectories and of interesting places can be recognized there. Most of the places represent either meetings of different groups or road junctions, where the participants stayed for a certain time to plan where to go next.

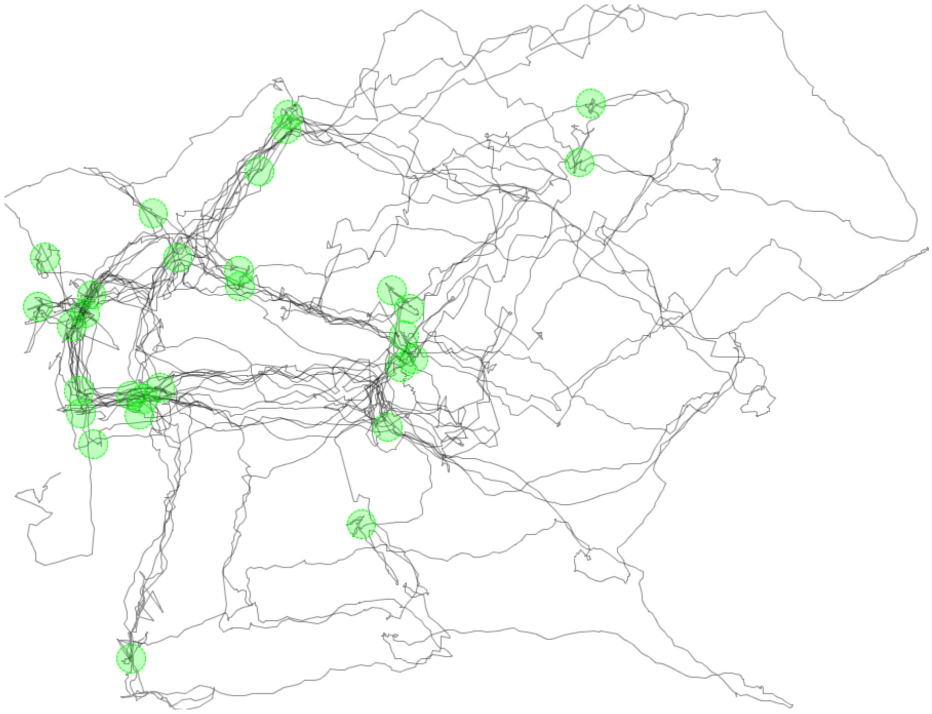


Figure 11: Interesting places found in a data set provided by a GPS-game

Another example originates from a video tracked handball match. One team has been tracked over a period of ten minutes, leading to 7 trajectories with 104993 sample points. This time the following parameters are used: visit count: 3, region radius: 0.5m, velocity threshold: 0.1m/s. Figure 12 (left) shows a snapshot of the court, the seven players of the tracked team (gray dots) and the 14 identified places (green dots) at a certain time. Considering the facts that the tracked team defends on the left and uses a specific defense formation, which is strongly kept by the players, the found places are reasonable. Those places can be interpreted and explained by visual inspection. The places 1 and 2 are places the goal keeper often stands at. The places 3 to 9 can be assigned to positions of a typical defense formation (cf. Figure 12, right). At the place in the center of the court (12) the throw-offs take place. Places 13 and 14 are offensive positions of the left wingman, at places 10 and 11 the right wingman has to wait for the throw off before entering the opposite half of the court.

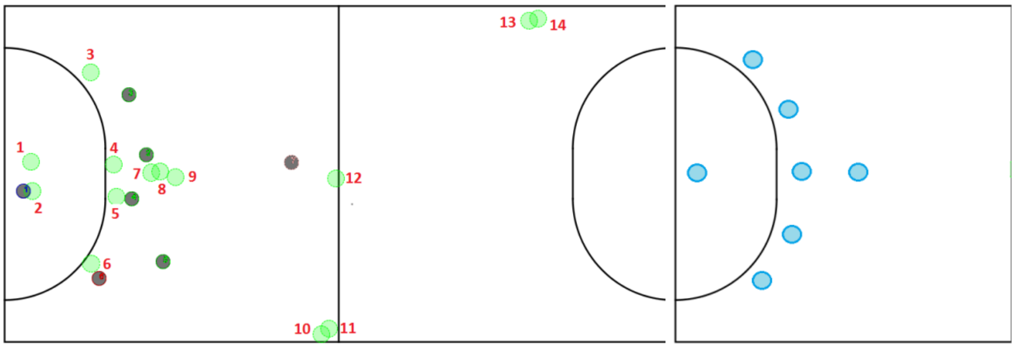


Figure 12: Examination of trajectory data provided by a video tracked handball match. Left: one snapshot with overlaid found interesting places, right: one typical defence formation during a handball match

6. Summary and Outlook

In this paper, we presented a first approach to detect abnormal movements of individuals depending on their environment. The results showed that the method we are using can be the first step to reach our overall goal. After finding typical behavior, we will be able to determine the deviations thereof, which we consider as abnormal.

This approach minimizes the data volume and computational costs by generating spatial models of movement behavior and incrementally updating with observed movements. The update mechanism does not require any reprocessing of data from previous observations. The unique processing, which consists of storing the extracted information in a more general model, reduces the amount of data. This way, for computation of a single observation we achieve a favorable runtime complexity based solely on the size of our model, making the algorithm suitable for real-time application. Given the spatial characteristics of the used model, a further speed-up of the used algorithm is possible e.g. by using spatial indexing structures.

As the interpretation of the trajectories is organized in an incremental fashion, it can also be designed in a decentralized way to achieve scalability with respect to the number of objects to track and interpret. This decentralize the algorithms is one of our main topics.

Another topic of ongoing work is to analyze and evaluate the migration graph structure. While doing this, the nodes can be classified by characteristics of entering and exiting trip edges. Next to solving typical tasks like finding the shortest or most favorite paths, many well-known concepts from graph theory can be directly applied to classify parts of the graph. We may identify places that act similar to sources and sinks (many ingoing/outgoing trips, only one outgoing/ingoing trip respectively), hubs (many ingoing and outgoing trips), loops (trips starting and ending at the same place) and so on. The graph structure can also be used for movement prediction. To this end, probabilities of possible target nodes can be calculated by including several factors like relative frequency of edge usage or target distances.

The approach is general enough to be used in several kinds of applications. In the context of LBS or pedestrian navigation, with this method popular places and frequent routes can be identified. Typical routes are also of interest for several planning purposes, e.g. city planning or traffic planning (see e.g. van der Spek, et al. (2009)).

Since in some parts of our method values have to be set a priori, it is not able to handle different scenarios automatically. It also does not adjust to varying situations. Therefore, an auto-fitting or learning technique to determine the three parameters would be very helpful.

Besides further things we are planning to deal with, the consideration of the temporal domain is important, e.g. to distinguish between trips common at typical times of day (see Makris and Ellis (2002)).

7. References

Andrienko, G., Andrienko, N., Bak, P., Keim, D., Kisilevich, S. and Wrobel, S., 2011. A conceptual framework and taxonomy of techniques for analyzing movement. Journal of Visual Languages & Computing 22(3), 213–232.

Ashbrook, D. and Starner, T., 2003. Using GPS to learn significant locations and predict movement across multiple users, Personal & Ubiquitous Computing 7: 275-286.

Baiget, P., Sommerlade, E., Reid, I., and Gonzàlez, J., 2008. Finding prototypes to estimate trajectory development in outdoor scenarios. In Proceedings of the First International Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS2008), September

Buchin, K., Buchin, M., van Kreveld, M., and Luo, J., 2009. Finding long and similar parts of trajectories. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '09). ACM, New York, NY, USA, 296-305.

Dodge, S., Weibel, R., and Lautenschütz, A. K., 2008. Towards a taxonomy of movement patterns. Information Visualization, 7, 240–252.

Kang, J.H., Welbourne, W., Stewart, B. and Borriello, G., 2005. Extracting Places from Traces of Locations, ACM Mobile Computing and Communications Review, 9(3).

Krumm J. and Cao L., 2009. From GPS Traces to a Routable Road Map // 17th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, ACM-GIS 2009, November 4-6, Seattle, Washington, USA, Proceedings. - Seattle

Laube P., Duckham M. and Wolle T. , 2008. Decentralized Movement Pattern Detection amongst Mobile Geosensor Nodes // Geographic Information Science, 5th International Conference, GIScience 2008, Park City, UT, USA, September 23-26, Proceedings. - Heidelberg : Springer

Makris, D. and Ellis, T., 2001. Finding paths in video sequences, in Proc. Brit. Machine Vision Conf., vol. 1, Manchester, U.K., pp. 263–272.

Makris, D. and Ellis, T., 2002. Spatial and probabilistic modeling of pedestrian behavior, in Proc. Brit. Machine Vision Conf., vol. 2, Cardiff, U.K., pp. 557–566.

Van der Spek S., Van Schaick J., De Bois P., De Haan R., 2009. Sensing Human Activity: GPS Tracking. Sensors. 2009; 9(4):3033-3055.

Comparative Analysis of Urban Areas through Mobile Phone Data Signatures: a Case Study in the Amsterdam Metropolitan Region

Filippo Dal Fiore[1,4], Euro Beinat[2,4], Piet Rietveld[3] and Bartosz Hawelka[4,5]

[1] Massachusetts Institute of Technology, SENSEable City Laboratory, 77 Massachusetts Ave, 20139, Cambridge (MA), USA
{dalfiore}@mit.edu
[2] University of Salzburg, Dept. of Geography and Geology, Hellbrunnerstrasse 34, 5020, Salzburg, Austria
{euro.beinat}@sbg.ac.at
[3] Vrije Universiteit Amsterdam, Dept. of Spatial Economics, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands
{p.rietveld}@feweb.vu.nl
[4] Currentcity Foundation, Keizersgracht 633, 1017s, Amsterdam, The Netherlands
{filippo, euro, bartosz}@currentcity.org
[5] University of Salzburg, GIScience Doctoral College, Hellbrunnerstrasse 34, 5020, Salzburg, Austria
{bartosz.hawelka}@sbg.ac.at

Abstract. The analysis of aggregated and anonymised data on mobile phone activities represents a novel opportunity to research human spatial behaviour on a massive scale. In this paper we present our investigation on more than 5 months worth of data provided by a Dutch telecom operator for the Amsterdam Metropolitan Region. We compared aggregated mobile phone usage behaviour in 10 functional places of the city of Amsterdam, from residential neighbourhoods to entertainment areas, from major train stations to business districts. Results reveal that the shape of weekly and daily signatures differ widely between the different urban areas, as well as the ratio between voice calls and text messages. We interpret findings by building a set of hypotheses on why each place might generate a different aggregate pattern, leveraging on the following variables: diverse contextual needs for mobile communication; socio-demographics of resident and

LBS 2011 Published in “Amendment to the Proceedings of the 8th International Symposium on Location-Based Services”, edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

transiting populations; urban time schedules; role played by major human attractors.

Keywords. mobile phone data analysis, human spatial behavior, human dynamics, voice call, text message

1. Introduction

The analysis of data on mobile phone activities represents a novel opportunity to research human spatial behaviour on a massive scale. Aggregated and anonymised data owned by mobile operators include information on where mobile phones connect to the GSM or UMTS networks, so they can reveal how individuals populate the urban space and move through it (Shoval, 2007; Mateos, 2006).

In the research presented in this paper, statistical analyses were carried out on aggregated and anonymised data provided by a Dutch mobile phone operator. This opportunity was used to characterise several functional areas in terms of their mobile phone usage patterns over time. Findings shade new light on how different urban spaces can create different conditions for human mobile communication, hence also (vice-versa) how informative mobile communication can be on how different urban spaces are used.

A few studies have examined the potential of mobile phones as sensors of human dynamics. Reades et al. (2007) analysed bandwidth consumption on the Telecom Italia network (measurements taken every 15 minutes) as a proxy of human activities in space and time for the city of Rome, Italy. They arbitrarily circumscribed a few functional locations - including the main railway station and the Olympic stadium - and reconstructed what they called weekly and daily "signatures" (Ratti et al., 2006). Peaks and lows in mobile phone traffic were revealed, mirroring expected human activities in those areas, such as morning commute or weekend lull. Normalization and cluster analysis were used to group areas with similar signatures and reconstruct functional use.

Ahas et al. (2008) used anonymised individual data owned by EMT, an Estonian mobile operator, to reconstruct the geographical location of homes and workplaces. Multiple observations of the same mobile phones enabled them to derive what they called home and work "anchor points" for most EMT subscribers. To prove the correctness of such a modelling method, the researchers showed a positive correlation between the size of Estonian municipalities (obtained from population registers) and an estimate of said size based on the number of home and work anchor points calculated. Gonzales et al. (2008) used the same dataset (i.e. a sub-sample of 100,000 EMT subscribers) to further characterise individual travel patterns, finding that such patterns tend to collapse into a single spatial probability distribution, revealing a high regularity - both temporal and spatial - in human trajectories.

With a different objective, Blondel et al. (2008) and Krings et al. (2009) examined anonymised data on who is calling whom on the GSM network of a Belgian mobile operator. As a first step, an algorithm was developed to cluster communications into “sub-communities” of individuals who frequently called each other. The algorithm’s accuracy was tested to predict and graphically cluster communities of French versus Dutch speakers in Belgium. More recently, Ratti et al. (2010) applied network analysis to telecom data in order to reconstruct regional boundaries in Great Britain (Ratti et al., 2010)

2. Description of Dataset

In Fall 2007 our research group gained access to the data owned by a Dutch mobile phone operator, whose GSM network shares the typical features of any GSM network[1] (Wilton and Charity, 2008). The area chosen for the research presented in this paper is shown in the figure 1.

From December 27th 2007 (h.08.00) through June 6th 2008 (h.06.00), hourly data on voice and SMS traffic were transferred to the servers of Currentcity.org, in Amsterdam, and the M.I.T. Senseable City Lab in Cambridge, MA: 3881 hourly measurements were received[2], corresponding to 161 days and 17 hours, for each of the 886 telecom sectors belonging to the area chosen for the investigation[3]. The data included, for each sector:

---

[1] In order to function within a GSM network, mobile phones must be located, so as to be served by the closest or most convenient cell tower, i.e. a base station with several directional radio cells to cover a 360 degree range. Some of the most common mobile phone activities – i.e. making or receiving a call, sending or receiving a SMS, entering or exiting a cell coverage area (i.e. sector) – generate data, which are saved in a database owned by the mobile operator. The spatial resolution of the network is determined by the local density of base stations: the more populated and urbanized the area, the denser the spatial distribution of base stations.

[2] Measurements refer to 00’:00” to 59’:59” of preceding hours, in such a way that - for instance - the h.16:00 measurement includes data recorded between h.15:00’:00” and h.15:59’:59”.

[3] This number is the result of a merger, i.e. a homogenous overlap by means of cross tabulation of 900 and 1800 Megahertz radio cells.

1. NC (Number of New Calls), i.e. calls initiated, including unsuccessful calls (i.e. because the receiver was not connected to a network or was busy in another conversation);

2. SMS (Number of SMS), i.e. SMS sent;

3. HV (Number of Handovers), i.e. calls transferred from/to other sectors[4]: Numbers of calls handed-over from all contiguous radio cells, i.e. Incoming Handovers (HVi) + Number of calls handed-over from the given radio cell to all contiguous ones, i.e. Outgoing Handovers (HVo); ER (Erlang), a standard unit of measure for traffic volumes, corresponding to 60 minutes of voice calls (initiated, received or handed-over from other radio cells);

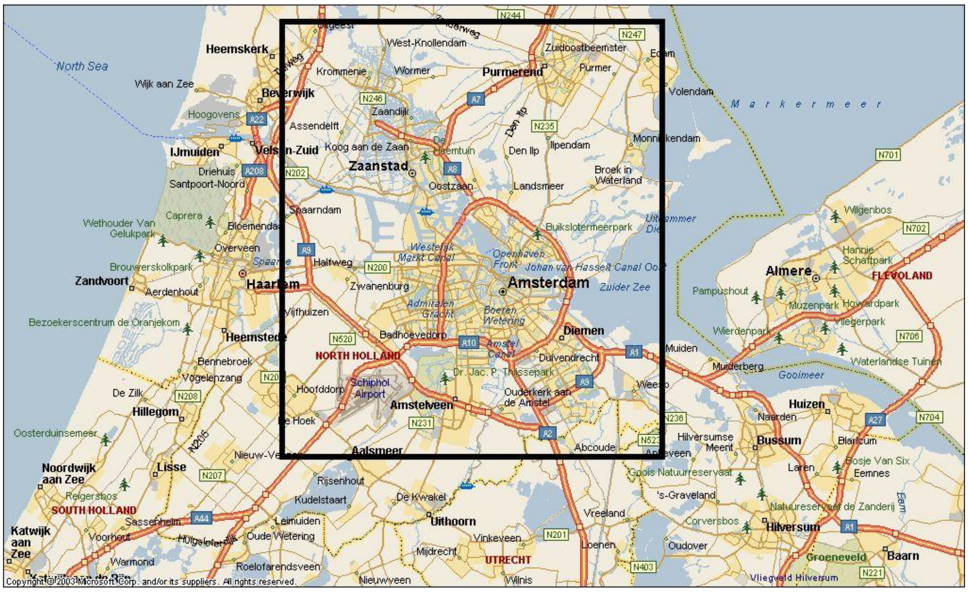


Figure 1. Within black boundaries is the area chosen for the investigation. It includes Schiphol Airport and the densely-populated urban conglomerations located north of Amsterdam, which significantly contribute to the daily commute to the city. Due to the non-linear boundaries of the sectors, the network coverage of the study area does not exactly correspond to the above box.

---

4 HV are mostly due to callers who are physically moving, but also to network operations (i.e. interference in the uplink/downlink between the mobile phone and the GSM network, or off-loading mechanisms to neighbouring cells)

3. Spatio-Temporal Patterns of Selected Areas

The objective of this section is to derive and interpret the mobile phone data signatures for several locations within the Amsterdam Metropolitan Region. Based on their different functional use and socio-economic profile the following 10 areas - representative of Amsterdam's urban diversity - were selected:

- Amsterdam Arena: the main stadium of Amsterdam;
- Amsterdam Central Station: a transport hub;
- WTC Amsterdam: a business district;
- Amsterdam Zuid Oost: a business district;
- Rembrandt Plein: an entertainment and nightlife area;
- Jordaan: a residential and tourist area;
- Schiphol Airport, including all departure and arrival terminals;
- Vondel Park: Amsterdam's central park;
- Amsterdam West: a residential neighbourhood;
- Kalverstraat: a pedestrian shopping area.

In the first part of the section (3.1), a comparative analysis is performed by means of descriptive statistics; in the second part (3.2), in-depth interpretations are provided to explain patterns of mobile behaviour in each of the 10 selected functional areas.

A certain degree of arbitrariness was used to define the physical boundaries of the 10 areas, i.e. deciding which telecom cells to include in each of them. In this respect, the following elements should be considered:

- the process of inclusion (or exclusion) of cells was based on prior knowledge of each functional area and its land use;
- in most cases only the cells which fully overlap with the functional area under analysis were included, excluding those which spill over into the surroundings;
- in most cases transit spaces (i.e. entry and exit points to and from the functional area) were included in the selection.

All graphs and signatures presented in this section were derived from the entire dataset, cleansed of two extreme events that may have skewed the results due to higher-than-normal mobile phone usage:

- the festivity of New Year's Eve (31/12/07 and 01/01/08);

- the festivity of Queen's Day (29/04/08 and 30/04/08).

Forty-eight hour observations were subtracted for each of the two events (from h.01:00 to h.00:00 of the following day), so that the resulting data set accounts for 886 sectors x 3785 hourly observations. It must also be taken into account that the actual best-serving coverage maps shown in section 3.2 do not exactly correspond to reality, for their physical boundaries have been disguised by means of Thiessen polygons, for confidentiality reasons. Nevertheless, calculations related to sectors - in particular, their physical extent and their land use distribution (see Table 1) - were performed on their real physical boundaries.

3.1. Comparative Analyses

In this section descriptive statistics are provided to characterise and compare the 10 selected areas, in terms of extent and land use distribution, as well as the mobile phone behaviour within them. In the next table each of the 10 areas is described in terms of:

- number of sectors which cover it;
- physical extent (in hectares);
- land use distribution over 14 pre-defined land use typologies.

As a first step in a comparative analysis, the following graph provides a comparative view of the 10 areas in terms of daily average number of New Calls (NC) normalized against the areas' physical extent.

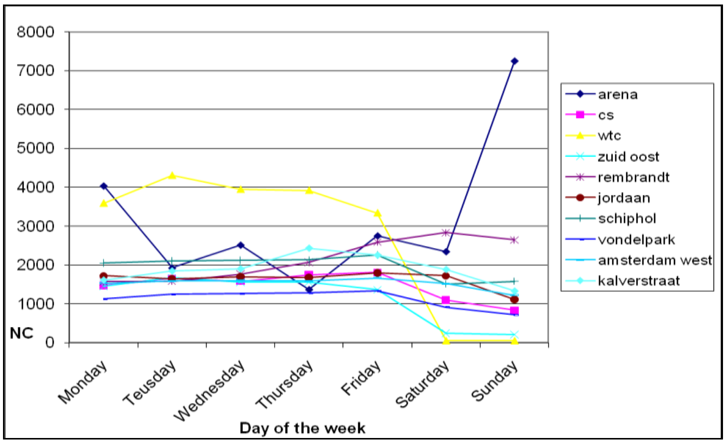


Figure 2. Weekly fluctuations of New Calls: daily averages normalized against the extent of each area (NC/km2)

Area	no. cells	area in ha	railways (%)	main roads (%)	airport area (%)	local roads (%)	residential area (%)	trade retail (%)	social-cultural facilities (%)	business area (%)	other hard surface terrain (%)	parks (%)	sport grounds (%)	day recreation terrain (%)	other agricultural terrain (%)	forests (%)	water (%)
Arena	5	2.1	0.0	0.0	0.0	10.4	0.0	0.0	0.0	0.0	0.0	0.0	89.6	0.0	0.0	0.0	0.0
Central Station	3	10.3	12.0	4.5	0.0	7.4	0.0	0.0	0.0	0.0	8.5	0.0	0.0	0.0	0.0	0.0	67.5
WTC	3	9.7	3.4	8.7	0.0	16.3	0.0	5.8	0.0	53.9	0.0	0.7	8.2	0.0	0.0	3.1	0.0
Zuid Oost	4	62.6	0.5	5.6	0.0	14.3	0.2	14.5	0.0	57.2	0.0	0.0	3.1	2.8	0.9	0.0	0.9
Rembrandt P.	1	3.9	0.0	0.0	0.0	29.8	0.0	52.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	17.8
Jordaan	2	36.6	0.0	0.0	0.0	25.6	66.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.4
Schiphol	21	156.9	0.0	1.9	78.0	7.1	0.0	1.3	0.0	0.4	0.0	2.9	0.0	0.0	8.2	0.3	0.0
Vondelpark	7	134.3	0.0	0.1	0.0	26.6	40.7	2.2	1.9	0.0	0.0	21.2	1.1	0.0	0.0	0.0	6.3
Amsterdam West	1	27.5	0.0	0.0	0.0	30.5	68.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1
Kalverstraat	2	7.4	0.0	0.0	0.0	38.1	17.3	34.2	4.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.6

Table 1. For each of 10 selected areas: number of sectors which cover it; area covered (ha); distribution of land use among given typologies (%). Source for land use data: CBS Land use map, year 2000, scale 1:25.000, juxtaposed to CBS street map, year 2000, scale 1:25.000

LBS 2011 Published in “Amendment to the Proceedings of the 8th International Symposium on Location-Based Services”, edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

WTC stands out as the area with the highest density of New Calls during the working week, possibly reflecting both a higher density of people and a higher propensity to perform new calls due to business activities. Except for Amsterdam Arena - which presents an irregular pattern with a marked peak on Sundays (the day of the soccer games) - and Rembrand Plein (a nightlife area) all other areas show a deflection of activities at weekends (in the case of WTC, almost no New Calls). In-depth interpretations for each of the 10 selected areas will be provided in the next section.

The following graph (split into two for the sake of visual clarity) depicts the evolution of New Calls over the 24 hours, normalized for each area against its average hourly value.

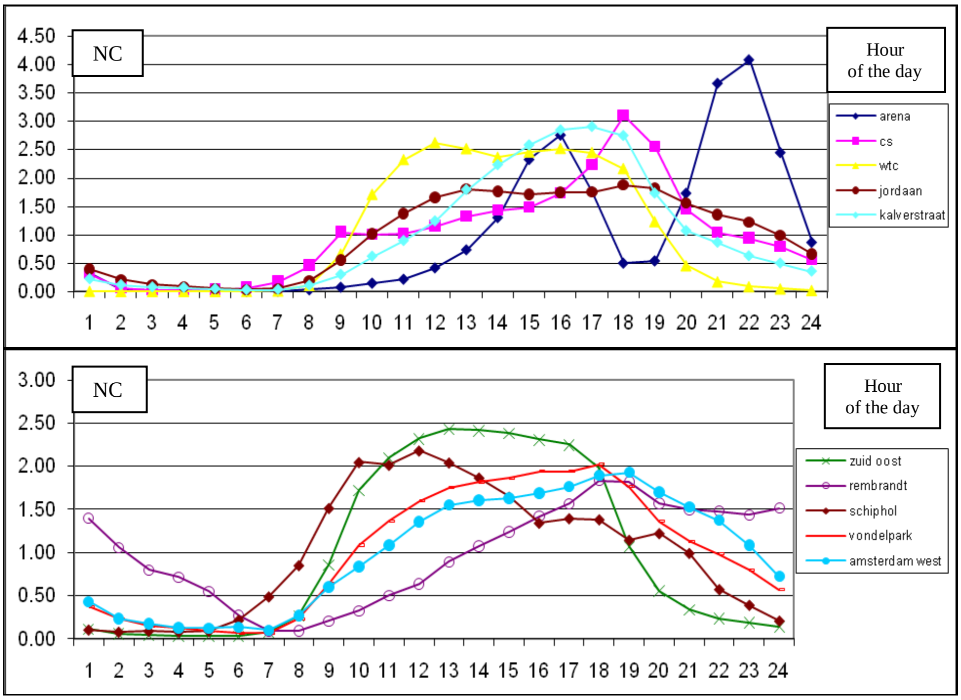


Figure 3. Daily fluctuations of New Calls: hourly averages normalized against the daily average of each area

What most remarkably distinguishes the 10 patterns is the steepness of the curve - uphill in the morning, downhill in the evening - as well as the different times when peak values are reached and how long similar values are maintained throughout the day. As further elaborated in the next section, it can be hypothesized that the predominant land use typology of

LBS 2011 Published in “Amendment to the Proceedings of the 8th International Symposium on Location-Based Services”, edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

each area - especially in its categorization as residential versus business versus entertainment - is a major explanatory factor. In the case of Arena, the two peaks in the pattern reveal the time schedule of the events which take place in the stadium.

A further line of inquiry has been the SMS share over New Calls. The scatterplot in the next figure is based on all observations for the entire monitored area and reveals a somewhat linear relationship between SMS and New Calls.

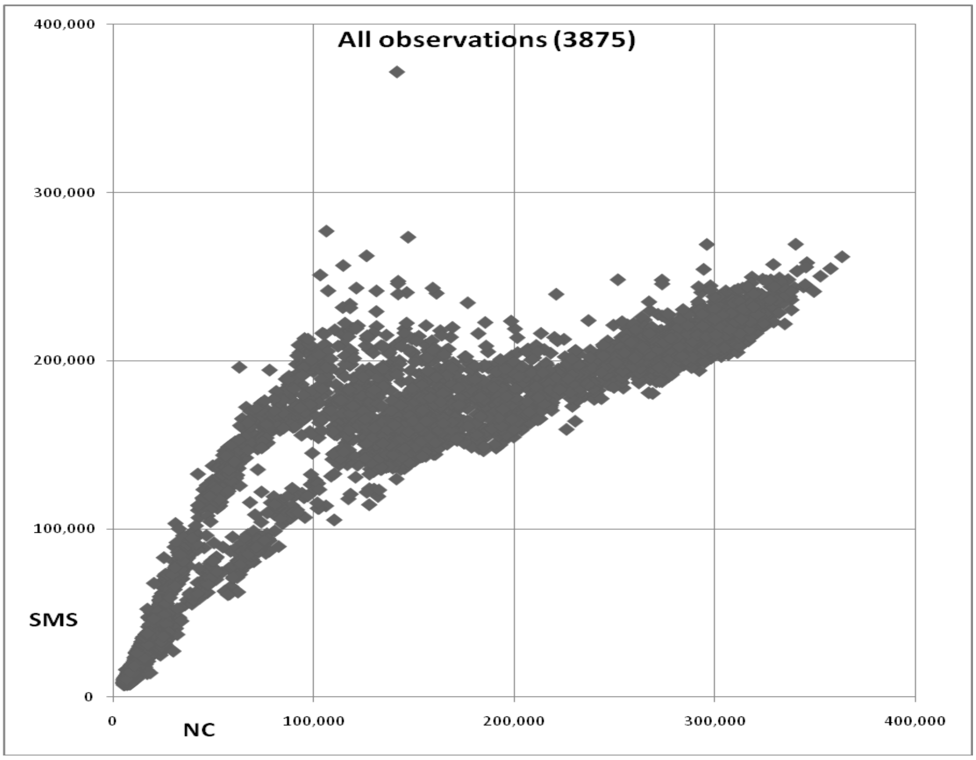


Figure 4. Scatterplot of all 3875 observations of New Calls and SMS (sum of all 886 cells in the sample). Horizontal Axis: New Calls (min.5036, ave.137430, max.363551); Vertical Axis: SMS (min.7363, ave.135245, max.371898).

The Pearson correlation equals [r = 0.895]. New Calls and SMS follow a remarkably similar pattern, with absolute values of the same order. The clearly visible outlier (highest isolated SMS value) was traced back to 22/05/08, h.22:00: remarkably, all radio cells presented anomalously high SMS values at this time, while no big event was taking place in town. The most plausible explanation is data error, in the form of temporary data miscalculation over the whole network. A striking feature of the scatterplot is its bifurcation in the middle, which enables two patterns to be distinguished, the shorter of which presents high values in the share of SMS over New Calls. In order to provide numerical evidence on the day/night and working week/weekend evolution of the share of SMS over New Calls, the next 2 figures are provided.

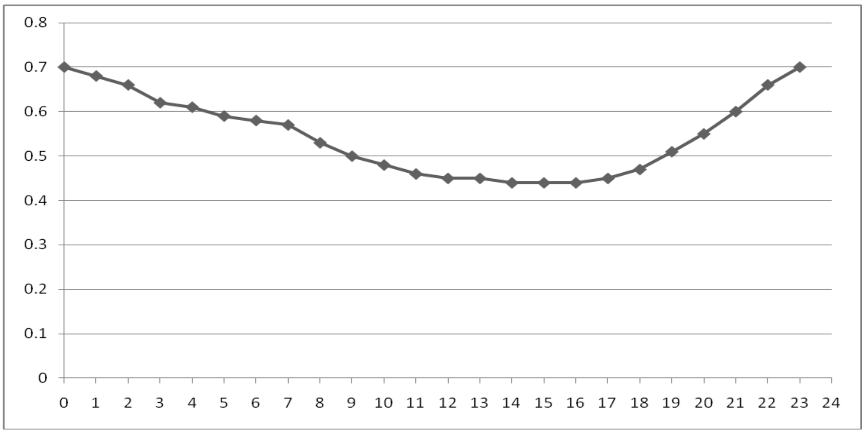


Figure 5. Daily evolution of SMS share over New Calls, i.e. [SMS/(SMS+NC)] (based on median values over the sum of 886 sectors in the sample)

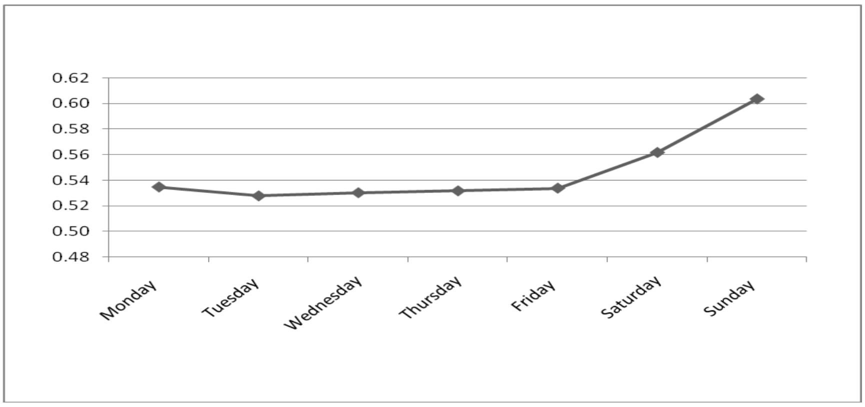


Figure 6. Weekly evolution of SMS share over New Calls, i.e. [SMS/(SMS+NC)] (based on median values over the sum of 886 radio cells in the sample)

Values for the SMS share over New Calls become higher during night hours and weekends, when users tend to send SMS rather than make new calls. The transition from day to night is more marked (2.7 decimals as the maximum difference) than the transition from the working week to the weekend (0.9 decimals as maximum difference). Two complementary hypotheses may explain this temporal difference: on the one hand, business and work-related activities may positively affect the number of new calls performed; on the other, especially night but also weekend hours may positively affect the number of SMS, since these may be more suitable for a less noisy communication with friends and family members. In this respect, while calls may be preferred to fully explain the meaning of a message and close a communication transaction, SMS may be used as a less urgent way to coordinate behaviour and keep in touch.

In the next section, individual interpretations will be provided vis-à-vis the daily evolution of the SMS share over New Calls, for each the 10 selected areas.

3.2. Interpretations by Area

This section is dedicated to the in-depth analysis of the 10 distinct patterns. Considering the purpose of interpreting them as proxies for human presence in the 10 selected areas, composite patterns were created from the algebraic sum of the indicators which have an unequivocal relationship with human presence in space and time (1 measurement equalling 1 human presence in a given sector during a given hour). Based on this criterion, New Calls, SMS and Incoming Handovers were selected, while Outgoing Handovers and Erlang were discarded[5]. The resulting composite patterns (CP) are based on the following formula:

[CP = New Calls(hourly ave.)+SMS(hourly ave.)+HVi(hourly ave.)]

Each composite patterns highlights a distinctive feature of each area, while raising questions on data interpretation.

The selection is limited to 5 sectors which cover the Amsterdam Arena Stadium (Figure 7). As expected, mobile activities peak on Sundays afternoons, when soccer games regularly take place. Smaller evening peaks characterise all 5 workdays of the week, most likely due to other events - like concerts or soccer games - which take place in the Stadium. It is worth noting that the Sunday peak spreads well beyond the two hours of the game, reflecting the arrival of spectators (more gradual) and their departure (more sudden). It should be noted that there is a car park under the Stadium. This may be used at times other than during games to serve the surrounding area, which explains the significant amount of mobile activities observed all through the week during morning and business times.

[5] HVo may be generated by users transiting and not stopping in the sector; ER does not have a linear relationship with human presence.

Amsterdam Arena

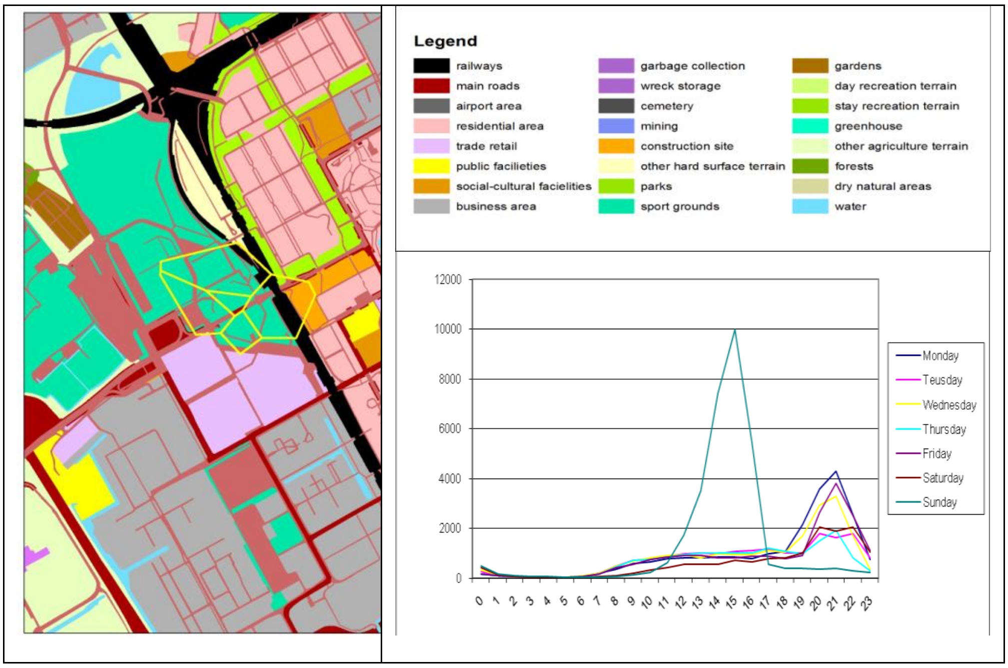


Figure 7. Selected sectors superimposed on land use (left); composite pattern of Amsterdam Arena (absolute number) (right)

In order to emphasize the impact of events over the weekly composite pattern, a soccer week (January 07-13, 2008) was extracted from our dataset and analysed separately. As shown in the following graph, on January 13th – the day of a soccer match[6] - mobile phone activities had a huge peak.

[6] Ajax versus AZ Alkmaar, h.14:30 through h.16:15, score: 6-1, attendance: 50,639 (source: ESPN Soccernet Europe, http://soccernet-assets.espn.go.com/match?id=232502&cc=null, last accessed on 25/10/2011)

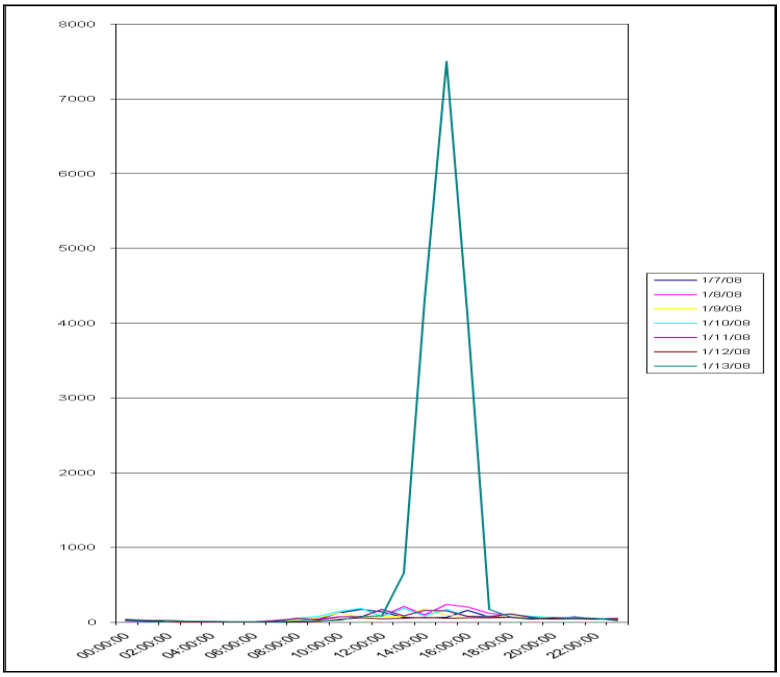


Figure 8. Composite pattern of Amsterdam Arena (absolute number) in the week between Monday 07/01/2008 and Sunday 13/01/2008

The next chart shows the daily evolution of the SMS share over New Calls for the selected sectors, as defined by the following formula: [(SMSS) = SMS/(SMS+NC)].

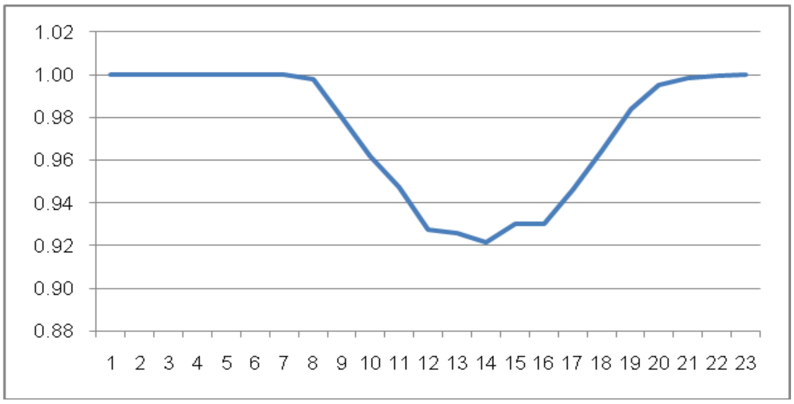


Figure 9. Daily evolution of the hourly (SMSS), at Amsterdam Arena

The balance is strikingly in favour of SMS, which account for 92-100% of the sum of SMS and New Calls. Two explanations could be advanced for this behaviour:

- the cross-sectional socio-demographics of the Arena spectators, with an over-representation of young people: they may prefer SMS over New Calls for economic convenience, as well as because its language is more suitable for the young;
- the fact that SMS may be more suitable in the noisy environment of the Stadium, as well as for short emotional communications with peers.

Amsterdam Central Station

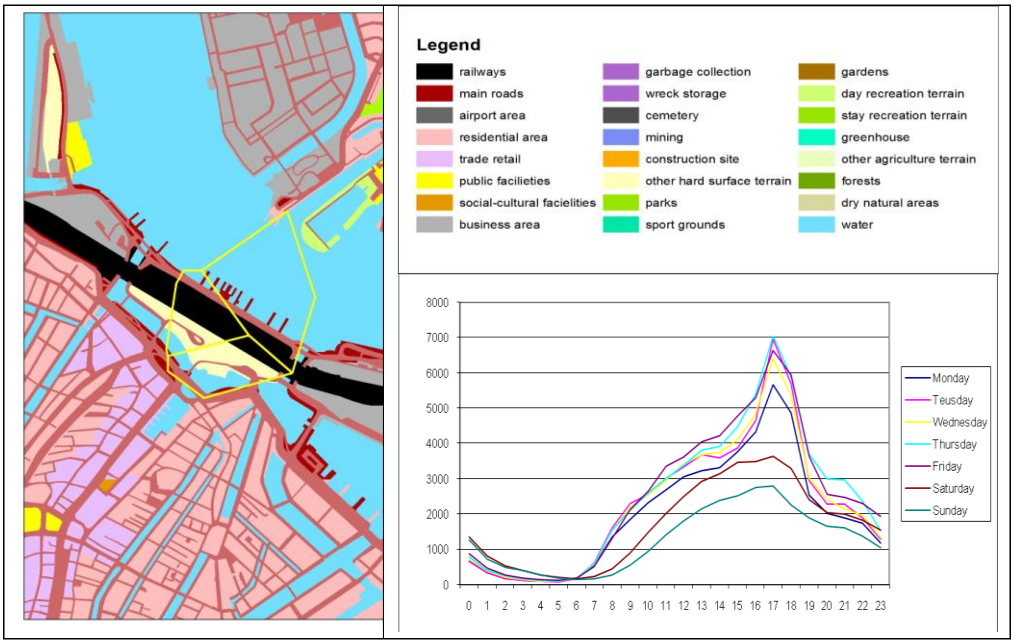


Figure 10. Selected sectors superimposed on land use; composite pattern of Amsterdam Central Station (absolute number)

The sectors selected include most of the square in front of the station, as well as a section of water behind it (due to the fact that one cell propagates through the river, since it does not find any obstacle in its way). Mobile phone activities rise throughout the day to reach their peak between h.17:00 and h.18:00, the time of the evening commute. Intuitively, one would expect an equivalent peak to account for the morning commute, but

LBS 2011 Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

activities are smaller and do not peak. This finding could be explained by the following hypotheses:

1. mobile users frequenting the Station in the afternoon are more numerous than those in the morning;
2. mobile users frequenting the Station in the afternoon use their mobile phone more than those in the morning;
3. a combination of both previous hypotheses.

While more trains may be scheduled in the afternoon (see sub-section on Schiphol Airport), it can be hypothesized that demand for communication is higher in the afternoon since people may want or need to get in touch with relatives and friends for the coordination of leisure time activities. On the other hand, in the morning such demand may not exist, since subsequent work activities may already be organized.

The next chart shows the daily evolution of the SMS share over New Calls for the selected sectors.

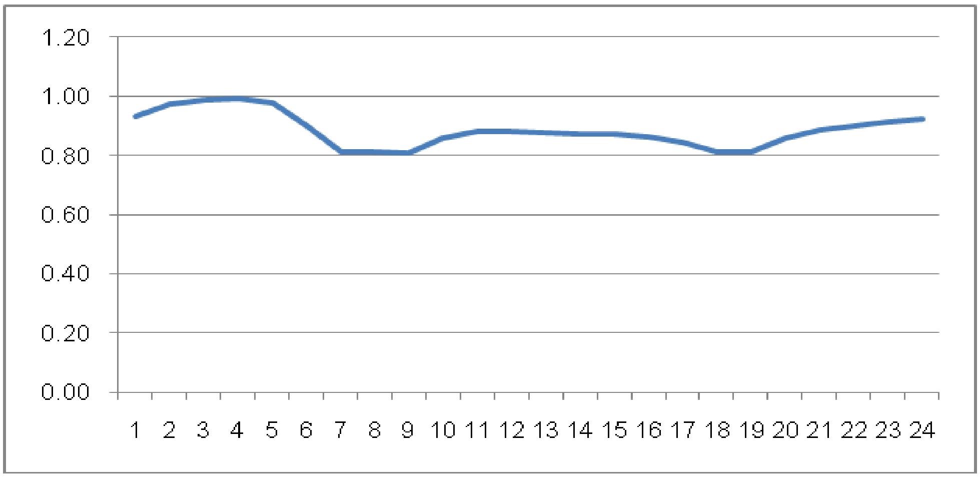


Figure 11. Daily evolution of the hourly (SMSS), at Amsterdam Central Station

SMS account for more than 80% of SMS+NC at all times of the day. This result is in line with evidence from the literature (Lasen, 2002) that less intrusive SMS may be preferred to new calls inside trains. Also, the noise of the station and on the station platforms may make people less inclined to initiate calls.

WTC Amsterdam

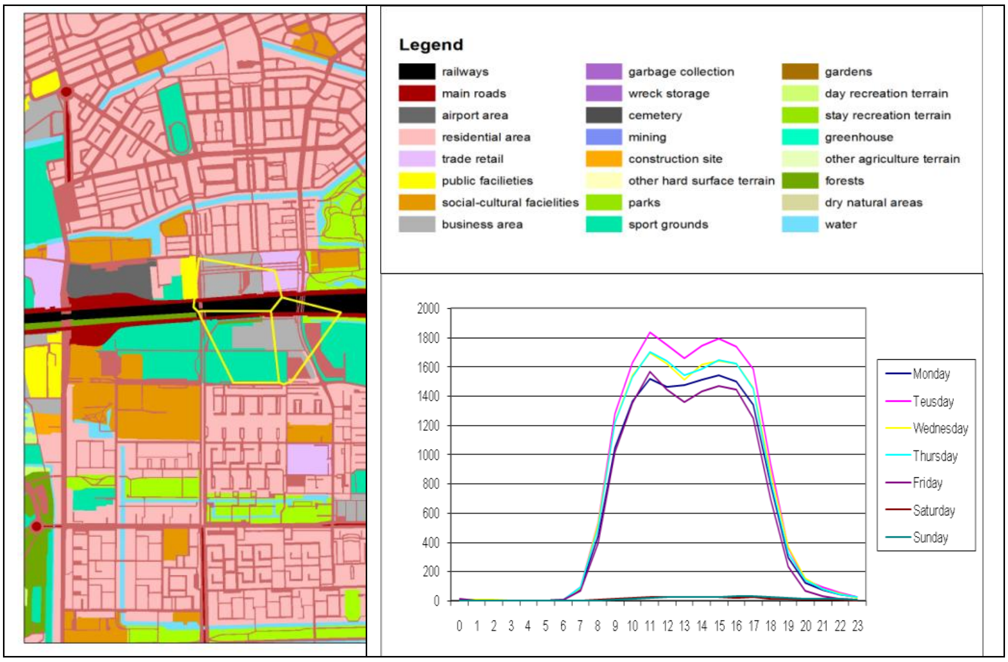


Figure 12. Selected sectors superimposed on land use; composite pattern of WTC Amsterdam (absolute number)

The 3 selected sectors include most of the WTC business area, the Amsterdam Zuid Station (railway and tram), and part of the new developments south of it. Sustained activities (> 1000) are maintained from h.09:00 to h.17:00, coinciding with business hours. Peaks are visible between h11:00 and h.12:00 and between h.15:00 and h.16:00, while between h.12:00 and h.15:00 a dip can be spotted, most likely due to lunch time and after lunch lull. To further highlight the business-related nature of the district, virtually no activities can be observed during night hours (h.21:00 through h.06:00) as well as on Saturdays and Sundays.

The next chart shows the daily evolution of the SMS share over New Calls for the selected sectors.

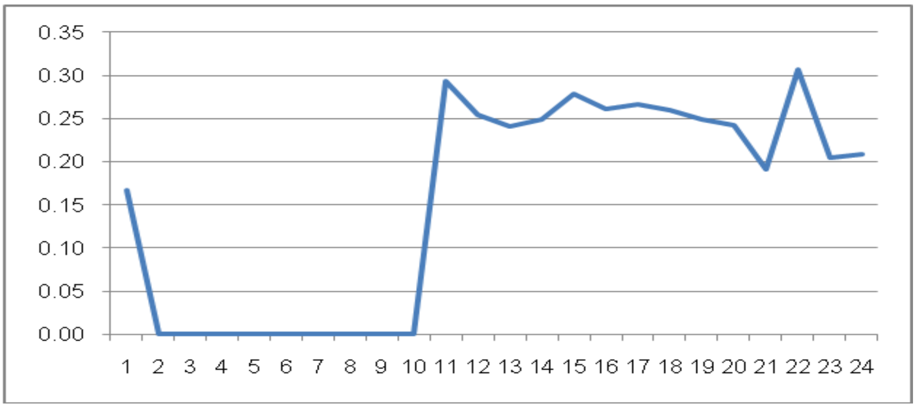


Figure 13. Daily evolution of the hourly (SMS), at WTC Amsterdam

The number of New Calls always prevails over SMS, not only during night hours (when virtually no SMS are sent and a very low number of calls are initiated) but also during business hours. This finding supports the hypothesis that New Calls may be the preferred medium for business activities, also considering that their higher cost does not burden individual callers but rather the budget of their employers, who may opt for a flat rate. Also, new calls may be easier to handle than SMS for drivers passing by on the highway.

Amsterdam Zuid Oost

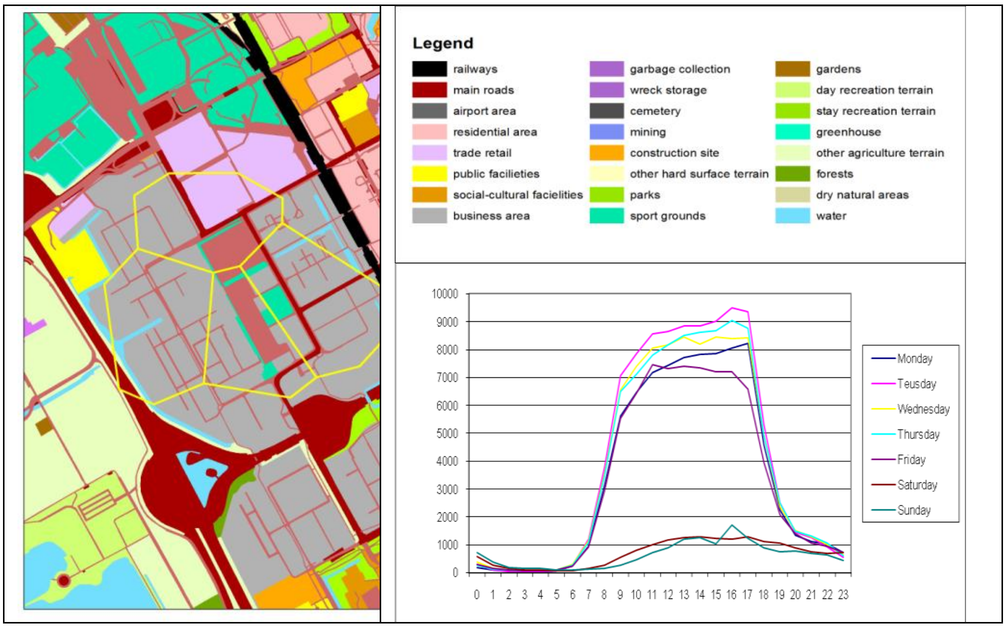


Figure 14. Selected sectors superimposed on land use; composite pattern of Amsterdam Zuid Oost (absolute number)

Most of the Amsterdam Zuid Oost business district is covered by the 4 selected sectors; the highway and railway which circumvent it are excluded. The area is more than 6 times larger than the selected area for WTC Amsterdam (see Table 1), but presents a remarkably similar land use configuration. Its pattern mimics WTC's for its sustained level of activities during business hours, but differs in that it shows the highest peak at the end of business days (Monday through Thursdays, at h.16:00 and h.17:00), as well as non-negligible activities on weekends. This may be due to its larger area and to the fact that a few retailers are located in Amsterdam Zuid Oost: most notably Ikea, which is open at weekends and which may attract more clients in the last hours of business days; there are also a few restaurants, a supermarket and several banks.

The next chart shows the daily evolution of the SMS share over New Calls for the selected sectors.

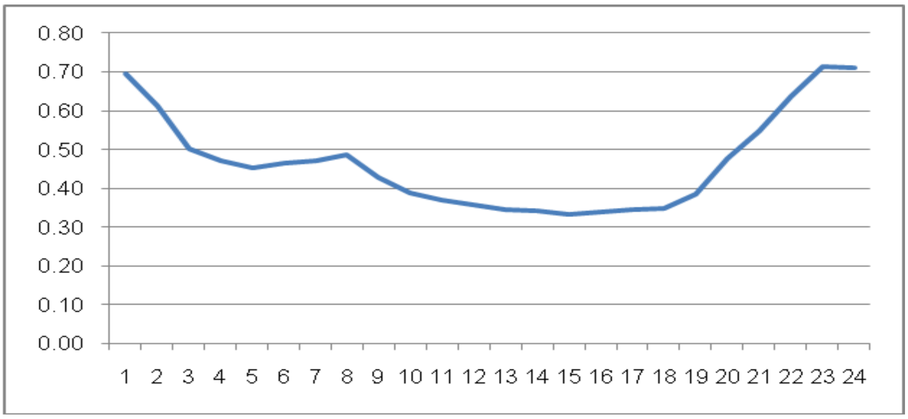


Figure 15. Daily evolution of the hourly (SMSS), at Amsterdam Zuid Oost

As in the case of Amsterdam WTC, New Calls are predominant during the day, but this time not in the evening or at night. Such evening and night SMS activities may not only be generated by those people who populate the restaurants in the area, but also by the small portion of residential area (0.2%) included in the selection.

LBS 2011 Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

Rembrandt Plein

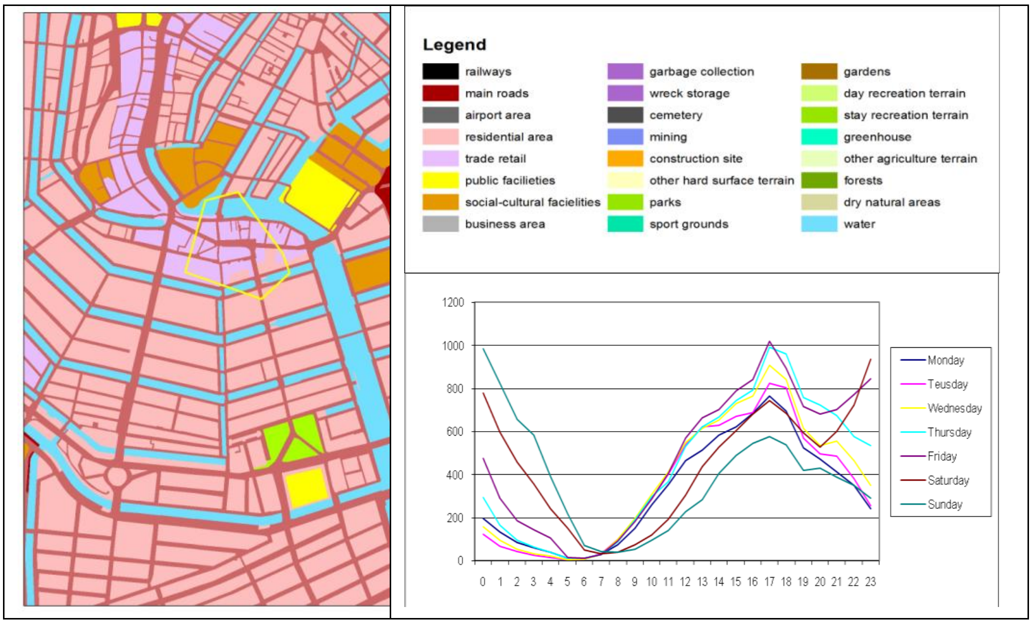


Figure 16. Selected sectors superimposed on land use; composite pattern of Rembrandt Plein (absolute number)

In this case, only 1 sector covers the Rembrandt Plein square and its surroundings. In this sector, 52.5% of the area falls under "trade/retail" land use, and includes numerous restaurants, cafés, theatres, hotels, pubs and disco clubs. The resulting pattern highlights the nightlife character of the area: the peak is reached between h.17:00 and h.18:00, but mobile activities continue until late at night especially during weekends when they have a rebound between h.20:00 and h.21:00. The place is the most "silent" between h.05:00 and h.08:00 on weekday mornings; after h.08:00 activities show a steep and constant increase up to h.18:00, while after 18:00 a mirroring decrease takes place until 5:00 of the following day.

The next chart shows the daily evolution of the SMS share over New Calls for the selected sectors.

Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

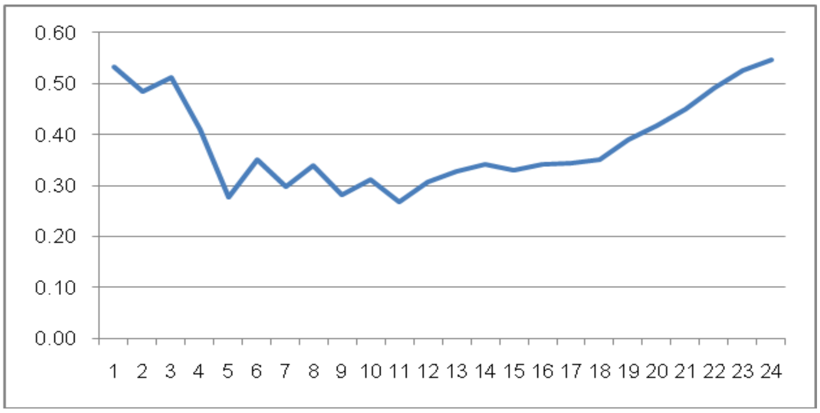


Figure 17. Daily evolution of the hourly (SMSS), at Amsterdam Zuid Oost

With the exception of a few night hours (h.23:00; h.00:00; h.01:00; h.03:00), New Calls are predominant over SMS. It may be hypothesized that the latter may be particularly suitable for the younger generations who populate this area at night, while the former may be preferred by people in need of more detailed information and quicker coordination. Also, new calls may be preferred by more affluent people who may populate the many restaurants and hotels located in the area.

Jordaan

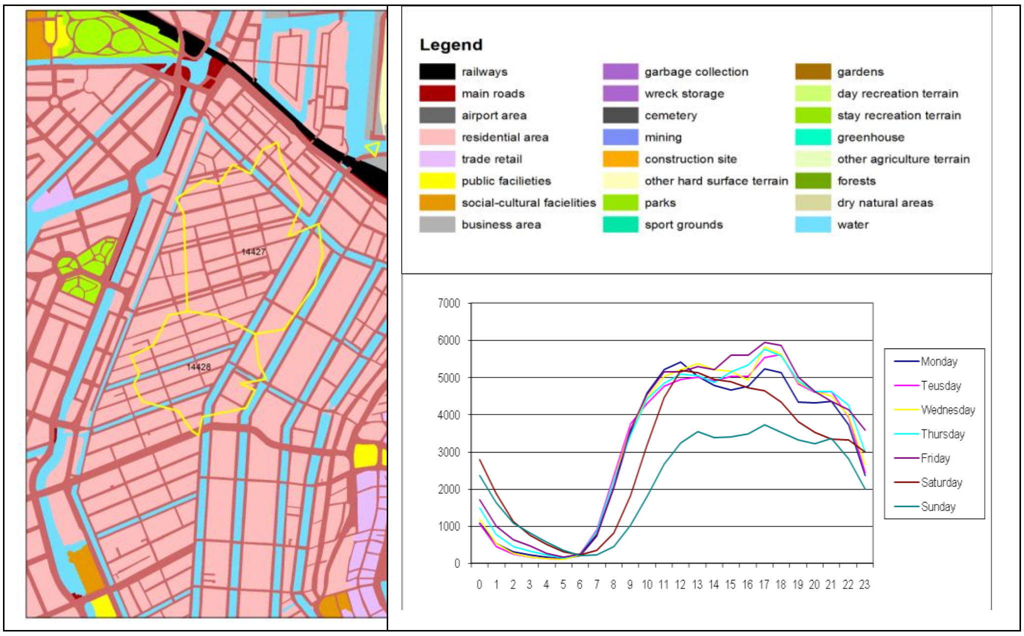


Figure 18. Selected sectors superimposed on land use; composite pattern of Jordaan (absolute number)

Only the 2 sectors that most thoroughly overlap with the historical neighbourhood of Jordaan have been included in the selection. Despite the prominence of residential land use (66% of the total), the Jordaan hosts a variety of tourist and recreational activities: several restaurants, cafes, art galleries and boutique shops, as well as a big open market ("Noordermarkt") are located in the area. The pattern of mobile activities reflects the mixed use of the neighbourhood, with high levels of activities maintained throughout the day; peaks appear between h.11:00 and h.12:00, as well as between h.17:00 and h.18:00, followed by a slow decrease until late at night (as observed for Rembrandt Plein).

The next chart shows the daily evolution of the SMS share over New Calls for the selected sectors.

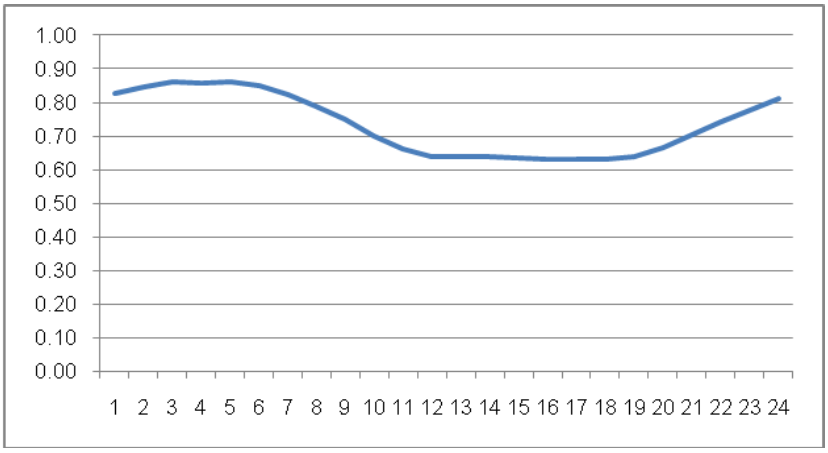


Figure 19. Daily evolution of the hourly (SMSS), at Jordaan

SMS are predominant throughout the 24 hours, possibly due to the presence of young resident and visiting populations.

LBS 2011 Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

Schiphol Airport

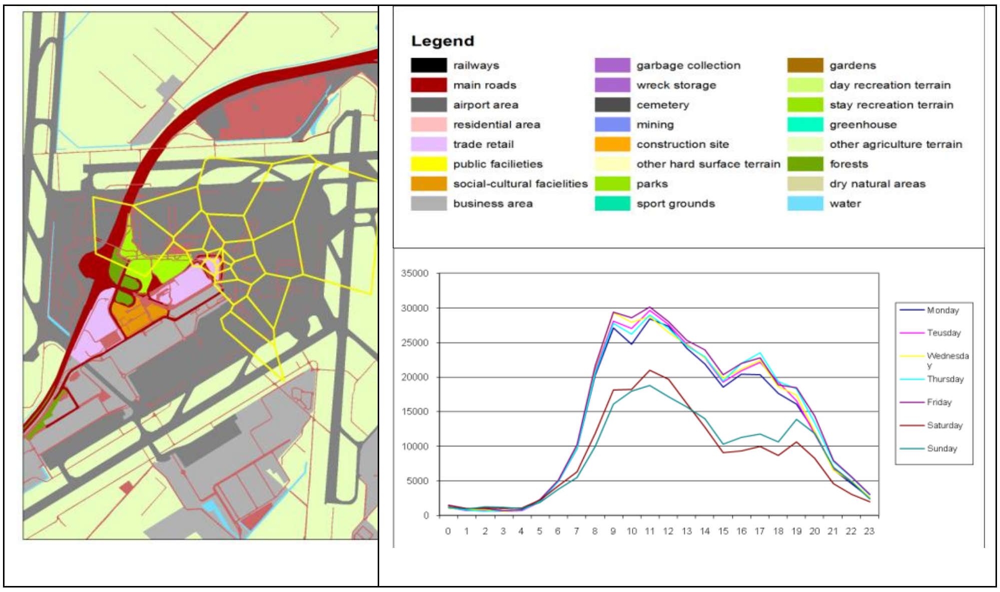


Figure 20. Selected sectors superimposed on land use; composite pattern of Schiphol Airport

The 21 selected sectors include: departures and arrivals, transit and shopping center ("Schiphol Plaza"), most of the gates, the Schiphol train station within the airport, the Food Village, the Airport's bus and taxi station, the Hilton and Sheraton hotels, and small sections of the adjacent office buildings and highways. The highest peak is reached in the morning, between h.08:00 and h.11:00. Smaller peaks appear at h.16:00 (working week) and at h.19:00 (weekends). Saturday and Sunday patterns are consistenly smaller than on weekdays. This distribution of peaks is the opposite of what was seen for Amsterdam Central Station (see related sub-section), where the highest peak is reached in the afternoon instead of the morning. None of the following hypotheses can be excluded:

1. mobile users frequenting the Airport in the morning are more numerous than those in the afternoon;
2. mobile users frequenting the Airport in the morning use their mobile phone more than those in the afternoon;
3. a combination of both previous hypotheses.

In order to disambiguate such multiple hypotheses, the number of incoming and outgoing flights to and from Schiphol was counted, for the morning (h.08:00 to h.12:00) and the afternoon (h.15:00 to h.19:00).

While 720 flights transit through Schiphol in the mornings (280 arrivals, 440 departures), only 485 flights do so in the afternoons (210 arrivals, 275 departures)[7]. When performing the same calculation on the composite pattern presented above (average for all Thursdays in the sample), there were 28% more mobile activities in the morning than in the afternoon, not fully accounting for the 48% difference observed for the number of flights. This mismatch could be explained by hypothesizing that the larger number of travellers in transit requires more personnel at the airport, as well as people accompanying them, taxi drivers, public transport and all other services related to travellers. Finally, it cannot be excluded that usage of mobile phone by mobile travellers might be more intensive in the morning, in particular accounting for business travellers and travellers who just disembarked from intercontinental night flights.

The next chart shows the daily evolution of the SMS share over New Calls for the selected sectors:

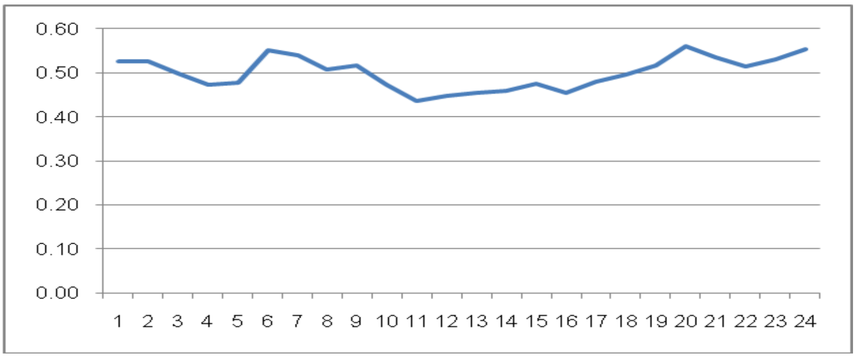


Figure 21. Daily evolution of the hourly (SMSS), at Schiphol Airport

The share fluctuates remarkably around the 50% threshold, possibly revealing a high variety of attitudes to mobile communication. In this respect, it may be that calls are preferred by business travellers, while SMS by leisure travellers. SMS may be a cheap and pragmatic solution for short communications such as “Arrived ok” or “Leaving on time”, specially for foreigners roaming on the network at higher costs.

[7] Calculations have been based on data available on (http://www.schiphol.nl/Flightinfo/) for Thursday 23/10/2009, more than 1 year later when compared to the time to which the data apply (27/12/2007 to 06/06/2008), with unknown consequences on the reliability of the comparison. Most flights appear multiple times because they are operated by multiple airlines.

Vondel Park

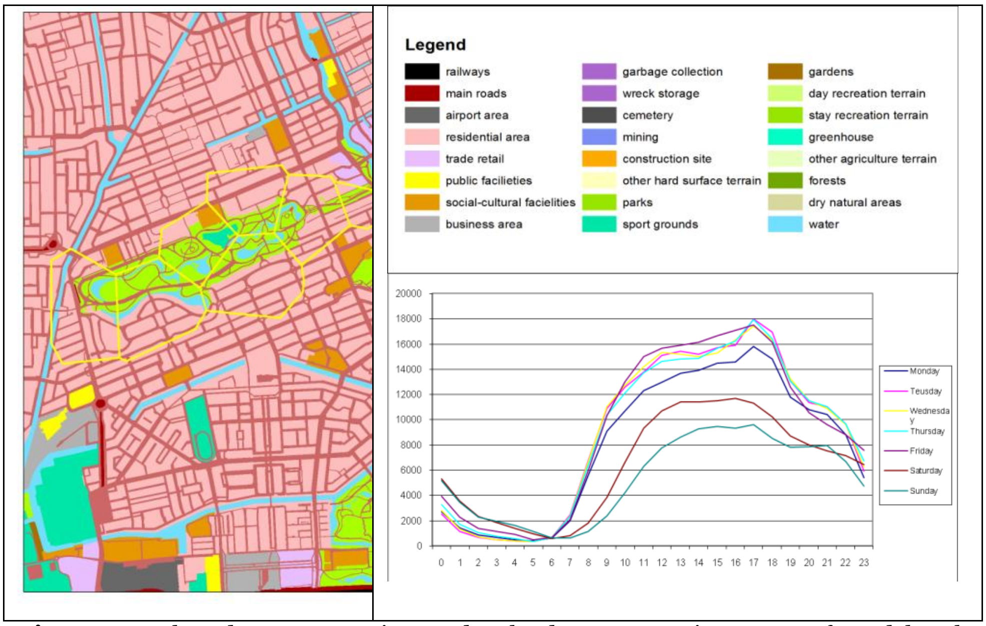


Figure 22. Selected sectors superimposed on land use; composite pattern of Vondel Park

The 7 selected sectors include the entire green area of the park (21.2% of the overall selected surface), but also large sections of the surrounding neighbourhoods. The latter are for the most part upscale residential (40.7%) and include numerous small offices as well as hotels. A considerable amount of local roads is covered in the selection (26.6%). Vondel Park itself hosts a few cafes as well as an open air theatre and a museum.

Mobile phone activities steadily increase during the day to peak between h.17:00 and h.18:00 in the afternoon. It can be assumed that human presence in the park and surrounding residential areas may be higher after business hours, when users return home or go to the park. This hypothesis is supported by the fact that on Saturdays and Sundays the late afternoon peak is less marked, and overall mobile activities are lower. Nevertheless, one might expect more users to spend time in the park and surrounding neighbourhoods during weekends and a few hypotheses may be advanced to explain such apparently counterintuitive findings:

- the mix of populations visiting the park and the surrounding area in the working week (for instance, business people and students) is substantially different from the one frequenting it at weekends (for instance, families and visitors), and may be more actively using their mobile phones;

- a considerable amount of residents and commuters may not be present in the areas on weekends;
- people transiting in the area - with public and private transportation - may be less numerous on weekends.

Finally, it is interesting to note that both on workdays and at the weekend mobile phone activities continue to be considerable late into the night, highlighting on the one hand the residential nature of the neighbourhood, on the other the fact that the park is open from dawn to dusk (which is very late during spring in Amsterdam).

The next chart shows the daily evolution of the SMS share over New Calls for the selected sectors:

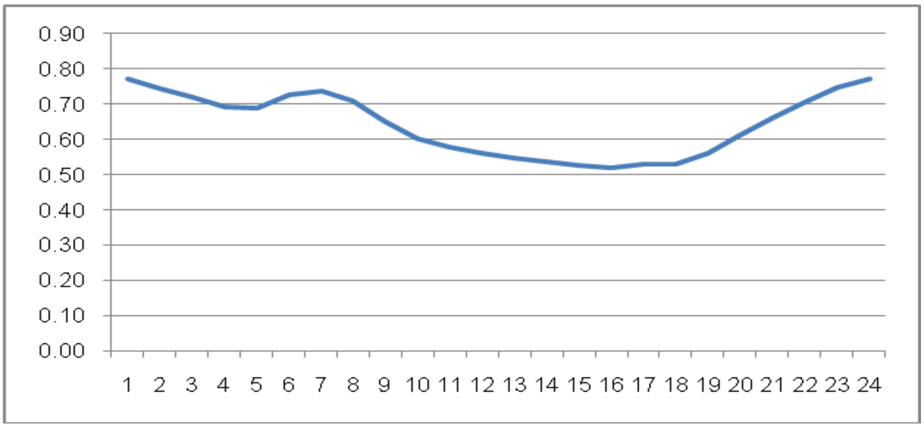


Figure 23. Daily evolution of the hourly (SMSS), at Vondel Park

SMS are predominant throughout the 24 hours, though less remarkably so during the central hours of the day. This share may reflect a high propensity to perform new calls among the affluent resident population outside business hours, paralleled by a higher propensity to send SMS by people visiting the park and transiting in the area during the afternoon.

Amsterdam West

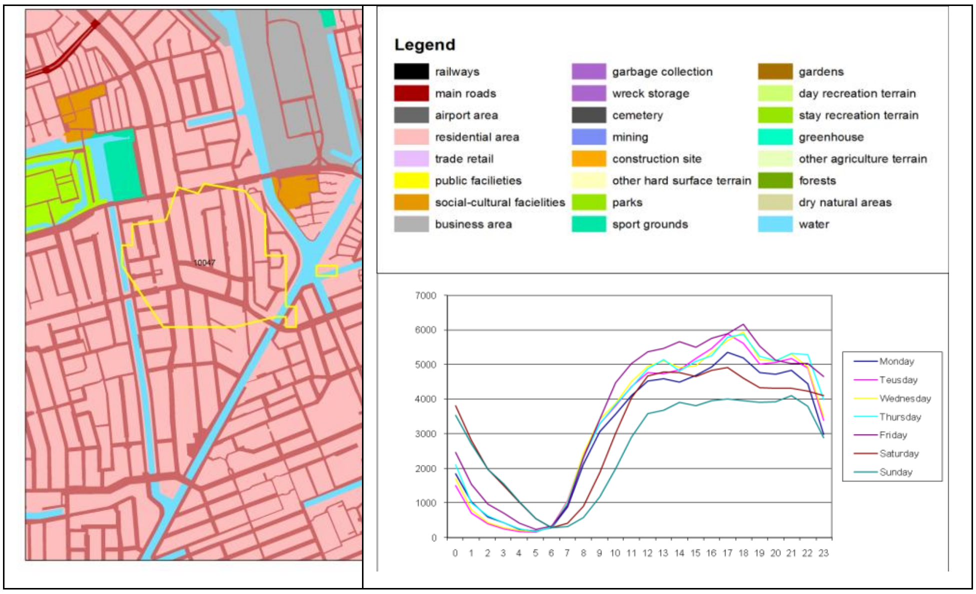


Figure 24. Selected sectors superimposed on land use; composite pattern of Amsterdam West

The selected sector covers a large section of the multi-ethnic residential neighbourhood of Amsterdam West. 68.4% of its land is classified as residential, the rest as local roads (see Table 1). The area includes several supermarkets and a few lines of public transport. On working week days mobile phone activities peak between h.17:00 and h.18:00 (Monday through Wednesday) or between h.18:00 and h.19:00 (Thursday and Friday), presumably when residents are coming back from work or after-work errands. Mobile phone activities steeply increase during morning hours (after h.06:00 in the working week, after h.07:00 at weekends), while they steeply decrease after h.21:00, except for Saturdays and Sundays. In most cases, a slight decrease can be observed between h.13:00 and h.15:00, most likely coinciding with after-lunch lull.

The next chart shows the daily evolution of the SMS share over New Calls for the selected sector:

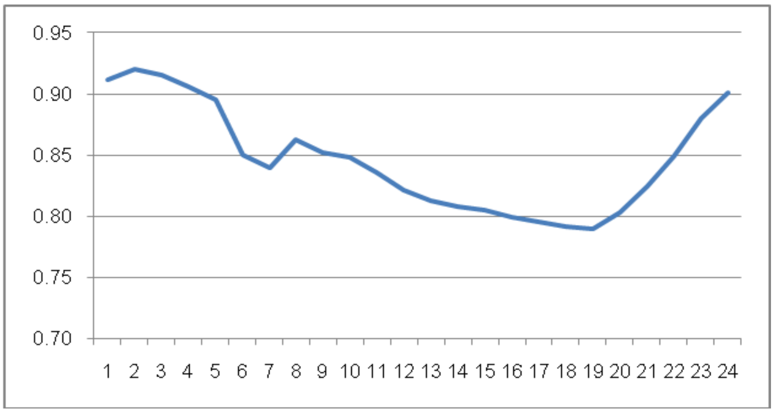


Figure 25. Daily evolution of the hourly (SMSS), at Amsterdam West

The vast majority of communication activities are in the form of SMS, throughout the day. This finding may be explained by the social composition of the neighbourhood, which hosts lower-income immigrant communities, possibly characterised by a younger population.

Kalverstraat

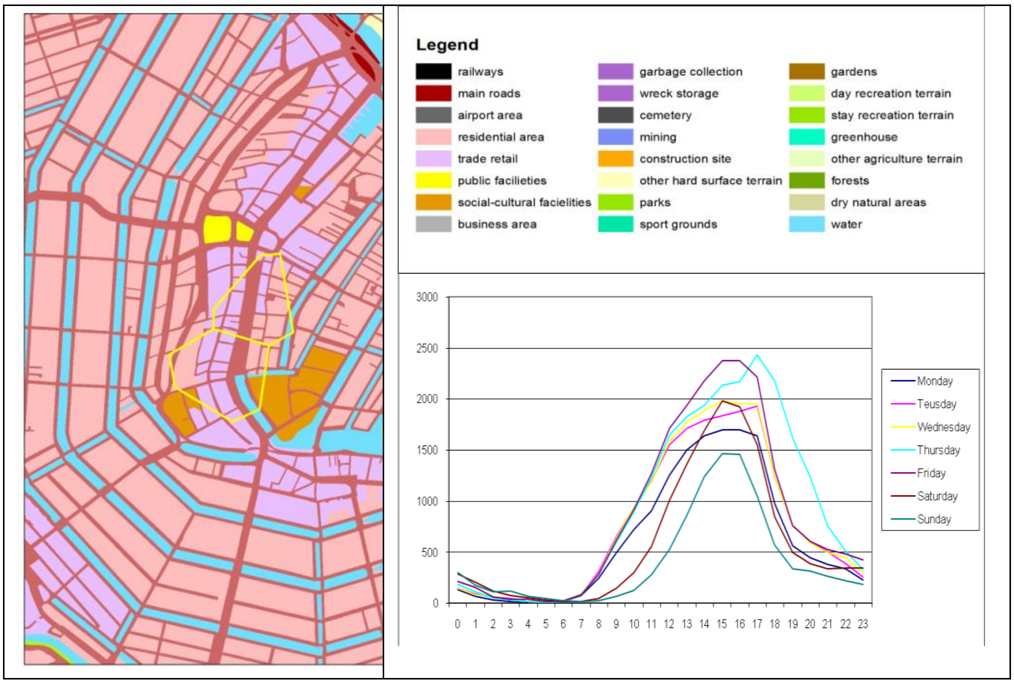


Figure 26. Selected sectors superimposed on land use; composite pattern of Kalverstraat

The two selected sectors include the northern section of Kalverstraat a pedestrian street in the centre of Amsterdam entirely dedicated to shopping and portions of the surrounding neighbourhood, a commercial and residential area, which is cut across by a major local road ("Rokin"). Mobile phone activities seem to reflect opening times of retail shops, which are the following:

- Monday: h.13:00 to h.18:00;
- Tuesday: h.9:00 or h.10:00 to h.18:00;
- Wednesday: h.9:00 or h.10:00 to h18:00;
- Thursday: h.9:00 or h.10:00 to h.18:00 or h.21:00;
- Friday: h.9:00 or 10:00 to h.18:00;
- Saturday: h.9:00 or h/10:00 to h.17:00;
- Sunday: h.12:00 to h.17:00.

Mobile phone activities are consistently more numerous on Thursdays and Fridays. Thursdays witness the highest peak (between h.17:00 and h.18:00), but this does not happen as late as one may expect considering the extended opening times of shops. All through the week, a steep decrease can be observed after shop closing times, and is therefore - consistently - an hour earlier on Saturdays and Sundays.

The next chart shows the daily evolution of the SMS share over New Calls for the selected sectors:

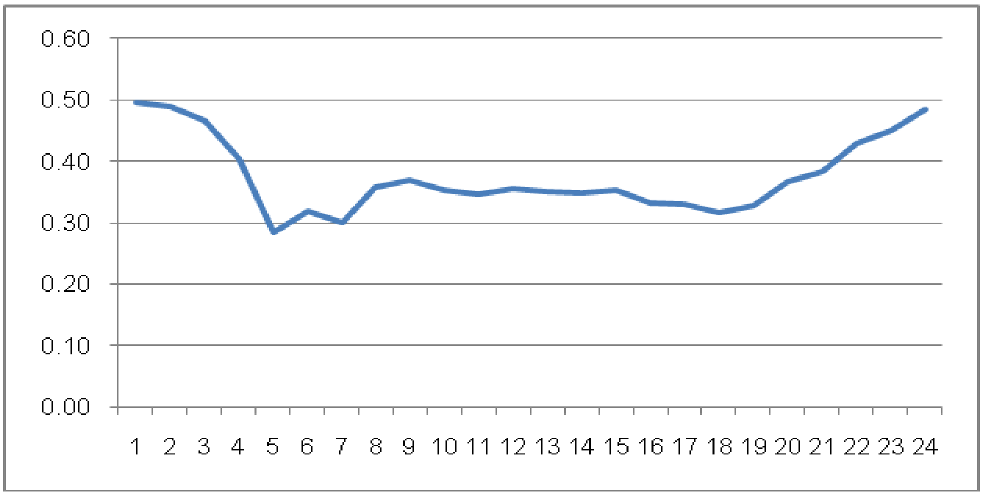


Figure 27. Daily evolution of the hourly (SMSS), at Kalverstraat

New Calls predominate throughout the day, but less remarkably at night. It may be hypothesized that shopping, strolling and tourist excursions are activities which attract relatively affluent people (especially considering that Kalverstraat lies in the heart of Amsterdam's city centre) on the one hand, while on the other they are less suitable for sending SMS, which would mean stopping and being distracted in the middle of shopping and tourist crowds.

4. Conclusions

The findings of this study highlight the high spatio-temporal variability and complexity of mobile phone behaviour. At the same time, they suggest that mobile phone data can be used as a proxy for human presence to reconstruct human behaviour in space and time, even if (given the availability of only aggregated data) it is not possible to infer how many people are generating mobile activities over time (10 calls could be performed by 1, 5 or 10 users).

Our analyses proved that the evolution of traffic volumes over time always reflects the day/night rhythm, as well as working schedules (in particular the h.09:00 to h.17:00 Dutch workday) and working week/weekend alternation. The traffic volumes enabled us to detect less apparent phenomena too, like the higher probability of having more travellers at Schiphol Airport during mornings (versus afternoons), which seems to find confirmation in a calculation of the number of flights.

While the daily and weekly evolution of mobile activities is most often explainable by knowing the type of place and its functional uses, the interpretation of the SMS share over New Calls is more problematic, also due to the relatively scarce literature on the different contexts and communicational aims of texting versus calling (Baron and Ling, 2007; Kasesniemi, 2003; Mante and Piris, 2002). Our analyses on the SMS share over New Calls reveal that these two indicators do not always follow a parallel path. In many cases, the share seems to be in favour of SMS in the evening and at night, when business calls may be missing and SMS could be preferred because they are less noisy and intrusive. Also, different areas seem to be differently characterised in terms of the propensity to send SMS versus calling, possibly reflecting both different socio-demographics - assuming that the youngest and the less affluent would be the heavier SMS users - and different contextual constraints (SMS turned out to be the preferred medium in Amsterdam Arena Stadium, which is a very noisy environment during soccer games). It is assumed that the above can only be hypotheses, verifiable only by means of an in-depth analysis of subscription and marketing data at individual user level.

Several other hypotheses can be advanced on the underpinning reasons for the observed differences between the 10 different urban areas. Nonetheless, such hypotheses could only be substantiated by collecting qualitative data on how individuals use their phone in different urban spaces, for example through participant or non-participant observations in the actual physical spaces, or by means of interviews with users (Bankers, 2004). Qualitative analysis on the actual content of mobile conversations would most likely be of great help (Cohen and Lemish, 2002), but content data are protected by privacy laws and in most cases not even collected by telecom operators. As an alternative, content analysis of public conversations over the web 2.0 - for example on popular social networks like Twitter.com or FourSquares.com - might offer new insights on the role played by different urban spaces and land uses in creating a need for human communication.

References

Ahas R, Silm S, Saluveer E, Jarv O (2008) Modelling Home and Work Locations of Populations Using Passive Mobile Positioning Data. In Gartner, G. & Rehrl, C. Location Based Services and TeleCartography II, From Sensor Fusion to Context Models, pp. 301-315. Berlin: Springer

Bankers J (2004) Understanding the usage of voice calls. An explorations of motivations, barriers and preferences. Master Thesis, Eindhoven University of Technology

Baron N, Ling, R (2007) Emerging Patterns of Mobile Phone Use: Electronically Mediated Communication in Transition. Paper presented at Mobile Media 2007, July 24 2007, Sydney, Australia

Blondel V, Guillaume J, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment, P10008

CBS Statline (2000) Gemiddelde bevolking; geslacht, burgerlijke staat, leeftijd en regio. Amsterdam: CBS

Cohen A, Lemish D (2002) Real time and recall measures of mobile phone use: some methodological concerns and empirical application, New Media and Society, 5(2), 167-183

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

Kasesniemi E (2003) Mobile Messages: Young People and a New Communication Culture. Tampere: Tampere University Press

LBS 2011 Published in “Amendment to the Proceedings of the 8th International Symposium on Location-Based Services”, edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.

Krings G, Calabrese F, Ratti C and Blondel V D (2009) Urban gravity: a model for inter-citytelephone communication networks. Journal of Statistical Mechanics: Theory and Experiment, L07003

Lasen (2002) A comparative study of mobile phone use in public places in London, Madrid and Paris, Digital World Research Centre, University of Surrey.

Mante E, Piris D (2002) SMS use by young people in the Netherlands, Revista de Estudios de Juventud, 52, 47-58

Mateos P (2006) Spatiotemporal accuracy in mobile phone location: assessing the new cellular geography. In Dynamic and Mobile GIS: Investigating Change in Space and Time, Drummond J, Billen R, Joao E, Forrest D (2006) (eds.), London: Taylor and Francis

Ratti C, Pulselli R, Williams S, 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

Ratti C, Sobolevsky S, Calabrese F, Andris C, Reades J, Martino M, Claxton R and Strogatz S (2010) Redrawing the map of Great Britain from a network of human interactions PLoS ONE, 5(12), e14248

Reades J, Calabrese F, Sevtsuk A, Ratti C (2007) Cellular Census: Explorations in Urban Data Collection, IEEE Pervasive Computing, 6(3), 30-38

Shoval N (2007) Sensing Human Society, Environment and Planning B, 34, 191-195

Wilton A, Charity T (2008) Deploying wireless networks. Cambridge: Cambridge University Press

Integrated Spatiotemporal Analysis of Mass Events

Christian E. Murphy

Department of Cartography
Technische Universität München,
Germany
christian.murphy@bv.tum.de

Abstract. Despite various existing methods for data mining and visual data exploration it remains difficult to explore large volumes of event location data adequately over space and time. This work presents a technique to statistically analyze mass events, which supplies a simplified and abstracted presentation visualized simultaneously in two dimensional space plus time. Therefore the visualization is set into a three dimensional environment with the variable time as third dimension. In aggregated area-units the mass events are generalized by performing a kernel density estimation on the time distance of occurring events. This provides density curves which are rotated around z-axes to produce solids of revolution. These figures are used as cartographic symbols and integrate the display of space and time information. The resulting solids of revolution indicate the amount of events by the local radius (distance from z-axis) and show the corresponding time on the z-axis value. This novel spatiotemporal analysis ensures that the general survey of the holistic spatiotemporal phenomenon is always allocated. This has three main benefits: (1) any geographic location at any given time or time period can be visually explored in detail. (2) All major and minor spatiotemporal hotspots (or coldspots) are clearly visible in one single statistical model and (3) the smooth rotating curve respects the continuous nature of time. The model is implemented to be visually explored in an earth viewer. The user of this visualization has the freedom to analyze either the complete phenomenon, specific time periods or to animate the phenomenon´s progression. This visual tool is introduced using a mobile phone location dataset and can also be used to support the exploratory data analysis of other spatiotemporal event data phenomena.

Keywords. visual analytics, geostatistics, interactive visualization, mass event data

Published in "Amendment to the Proceedings of the 8th International Symposium on Location-Based Services", edited by Georg Gartner and Felix Ortag, LBS 2011, 21–23 November 2011, Vienna University of Technology, Vienna, Austria, 2011.