Modelling Spatial Patterns of Outdoor Physical Activities using Mobile Sports Tracking Application Data

Rusne Sileryte and Pirouz Nourian

Abstract The paper presents a workflow for collecting, structuring and processing geo-referenced recreational mobility data from a sports tracking application as to monitor recreational usage of urban spaces. The data collected include GPS trajectories of people walking, jogging, and running for recreational purposes in European cities. The presented workflow includes systematic steps for aggregating the trajectories and attributing them to a spatial network model called Urban Space Network. The nodes of this network are the navigable spaces or streets and its links are the connections between them. A method is proposed to find a fuzzy notion of recreational space usage, using the number of distinct application users whose trajectories have been accounted for the space in question. The fuzzified space usage values are then attributed to the nodes of the network. This model can be used primarily used to observe actual patterns of space usage and has the potential to be used as ground truth data for validating and calibrating network-based models of recreational mobility. Patterns revealed by the workflow can be used to study where outdoor physically active mobility happens and where it is absent. Thus the proposed workflow can provide spatial and objective insight useful in planning, management and governance of cities in promoting active mobility that is already a rather global trend in urbanism.

Key words: Mobile Sports Tracking Application Data, Mobility Data, Urban Space Network, Recreational Usage

Rusne Sileryte · Pirouz Nourian
Faculty of Architecture and the Built Environment,
Delft University of Technology, Delft, the Netherlands
e-mail: `r.sileryte@tudelft.nl,p.nourian@tudelft.nl`

1 Introduction

Planners and legislators have already begun creating guidelines, policies, and regulations to encourage physical activities in cities (NYC DDC, 2010). A better understanding of where physical activities are conducted enables effective policy interventions to promote physically active lifestyles in different built-environment contexts. Probabilistic models on networks could bring the benefit of predicting the likely effects of changes on facilitating or hindering the phenomena.

A substantial amount of data related to outdoor physical activities can be obtained from mobile sports tracking applications. This data is constantly generated worldwide by smart device users willing to record their spatio-temporal activity. The available data is extremely big, provided voluntarily and in large numbers, public, and therefore not raising privacy issues, always available up-to-date, features the same world-wide method of collection and finally, it is constantly growing.

However, mobile sports tracking application data is collected for personal motives and the motivation of application providers is rather satisfaction of a user, than aggregate data collection. In fact, nor the data is available in a single click; neither application providers supply an interface for ready-made free access due to likely privacy issues. Furthermore, the available data is raw and so vast in size, that structuring, filtering and aggregation procedures need to be applied. The GPS (Global Positioning System) data is just a sequence of points in Euclidian space, which need to be mapped and analysed in a network space, while constructing an appropriate space network for any of the cities is a non-trivial task itself.

Three cities have been chosen as case studies based on the availability of data provided by Eurostat, similarity of rate between citys population and the chosen sports tracking application users. The chosen ones are namely Vilnius (Lithuania), Valencia (Spain) and Gothenburg (Sweden) with a ratio of 2-3 spotted application users per 1000 inhabitants. This paper describes a workflow for observing usage of urban spaces for running and walking activities in various European cities based on the mobile sports tracking application data. The developed workflow exceeds the scope of a single project in that it aims to simplify and standardize the preprocessing of mobility data.

The initial section delineates related work and clarifies how this research is distinguished from the similar works. The third one explains the procedure for acquisition of required data, which is followed by constructing an Urban Space Network. Next section explains a method to interpret the relative space usage and delivers the results. Finally, the conclusions are drawn preceded by the discussion and recommendations.

2 Related research

Previous studies of physically active human mobility rely on manual data collection by directly observing chosen locations during certain short periods (Floyd et al,

2008), asking residents in surrounding areas of a park to complete 7-day physical activity logs that include the location of their activities (Kaczynski et al, 2008), comparing recipients places of residence with their physical activity registered by accelerometers (Cohen et al, 2006) or even using a telephone survey (Lopez, 2004).

All the previously mentioned methods are performed by intensive human labour and can only be applied on relatively small-scale measurements. In addition, walkability has been mostly studied as a property of the entire neighbourhood rather than particular urban space or their networks. Yet a number of researches have explored definite relation between street walkability and the configuration of urban street network and its attributes such as transport nodes, land use, infrastructural elements, major attractors, aesthetic features, etc. (Hillier and Iida, 2005; Gauvin et al, 2005; Gebel et al, 2007).

Mobile GPS data has been already used by various researches in order to investigate spatial mobility patterns in urban settings. Van der Spek et al (2009, 2013) have carried out a research which aims to explain pedestrians behaviour in various cities by deploying GPS tracking system supplemented with questionnaires. Piorkowski (2009) has pioneered in using mobile sports tracking application data for analytic purposes. He aimed on enhancing location privacy and designing better context-aware services. Ferrari and Mamei (2011, 2013) have used Nokia Sports Tracker application data to identify the areas and temporal routines of a city most used for a given sports activity, highlight cultural and climate-related differences among cities and show differences in the routine behaviour of various demographic and social communities. Oksanen et al (2015) aim to extract frequently used routes from massive public workout data in order to define the most popular routes as a suggestion for application users.

The goal of this research, in respect to the previously described ones, is to use mobility data in tandem with other open data sources, for modelling recreational usage in a network rather than Euclidean space through an automated procedure, which later allows utilising the model as a ground truth for the further investigations of the desired phenomena.

3 Required Data

3.1 Mobile Sports Tracking Application Data

Sports tracking applications cannot provide direct access to their databases due to privacy issues; however, some of them display public workouts on dedicated websites. In that case, users are aware and content with publicly displayed (but not distributed) personal data. After considering a number of applications, Endomondo was chosen due to its popularity rate and relatively convenient data access. Workouts are available to be viewed on `www.endomondo.com/workouts/+workoutID` unless specified to be private by a user.

Fig. 1 Workflow for the
acquisition of mobile sports
tracking application data
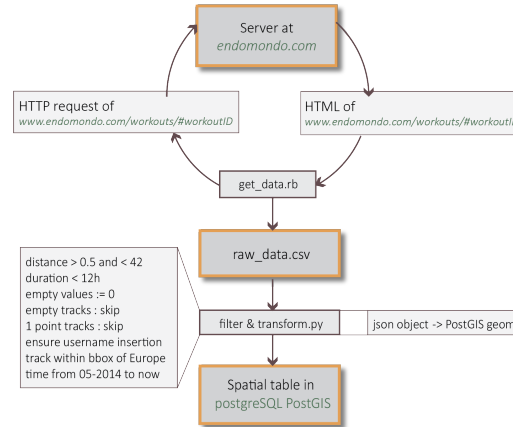
Every workout is a JSON (Java Script Object Notation) object embedded into an HTML (Hyper Text Mark-up Language) code of a page. Additional to the GPS trajectory, other available attributes include type of the workout (running, walking, etc.), date and time, user name (id), distance, duration, average and maximum speed, burnt calories, hydration, altitude and weather data. A user can choose to make any of these attributes private, edit the values or delete the workout permanently at any time (Endomondo, 2015). Tracking is based on a GPS receiver and therefore is dependent on the characteristics of each individual device.

A tutorial in Barsukov (2014) has been used as a basis for the data acquisition framework. The adapted scheme of data acquisition is shown in Fig. 1. A local script sends an HTTP (Hyper Text Transfer Protocol) request to the server for a workout with a chosen ID and either gets a negative response (in case the workout is listed as private or it has been deleted) or a positive response and an HTML code of a page, in which case the algorithm continues exploring the data. If GPS trajectory is available and listed as Running or Walking, the required fields are output into a text file, which is later filtered based on multiple criteria and transformed from a JSON object into a PostGIS geometry.

Data samples were timed every 8 days in a period of May 2014 - May 2015, aiming to have sufficient data throughout the full year and a variety of weekdays as well as occasional public holidays. Data acquisition process took approximately 1248h and resulted in more than 3.5 million valid GPS tracks of almost a million distinct users within the territory of Europe. The collected data is evenly distributed throughout a day, all seasons and weekdays.

Fig. 2 Original OSM street network expressed in polylines (left) and actual perceived urban space needed for the active recreational travel analysis (right)
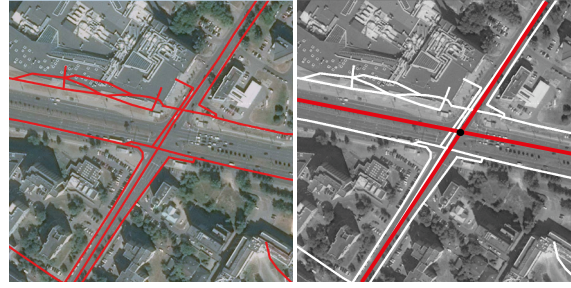
3.2 OpenStreetMap

Lately the road network provided by the OpenStreetMap (OSM) is often chosen to form the backbone of urban networks because of its universal coverage and standard defined for all modes of transport. Besides, due to its open access nature and volunteered contribution, OSM can have a very good level of completeness (Mooney, 2015) and it includes representation of paths for non-motorised means of transport, which is essential for the analysis of the jogging and walking movement patterns.

A single polyline in OSM dataset usually describes a single path; however, in some cases it can also form a boundary polygon and represent an area, which stands for various parks, squares and even wider boulevards within which no further paths are drawn. The elimination of one of the entity types would result into missing network connections, which would result into misleading snapping of GPS tracks and wrong evaluation of network configuration. Thus, the inconsistency of entity types needs further attention while processing the dataset.

Girres and Touya (2010) have listed a number of possible problems regarding the OSM street segment geometry and topology, including duplicate overlapping or missing segments, intersection nodes, etc. In addition to these, the high level of detail presented in the street network is redundant and even confusing for the later applied algorithms. While a single street in OSM can be represented by multiple lines, which stand for different car lanes, bicycle lanes, footpaths and sidewalks, all these lines are still perceived as a single space by a person engaged into an active recreational activity (Fig. 2). Therefore, the OSM street network has been later processed and coupled with additional datasets in order to overcome the identified problems and provide a neat Urban Space Network.

3.3 European Urban Atlas Road Land-use Data

The Urban Atlas (UA) is a joint initiative of the European Commission Directorate-General for Regional Policy and the Directorate-General for Enterprise and Industry with the support of the European Space Agency and the European Environment Agency. Its aim is to provide pan-European comparable and freely accessible land

use and land cover data for Large Urban Zones with more than 100 000 inhabitants. The resulting vector maps provide land-use classification for 21 different land-use classes with minimum overall accuracy of 85% and positional accuracy of +-5m (Urban Audit, 2007).

The 'Roads and Associated Land' class is represented by a single polygon, which comprises a city road network. The associated lands are: slopes of embankments; areas enclosed by roads, without direct access; fenced areas along roads; noise barriers; rest areas, service stations and parking areas; railway facilities; foot- or bicycle paths parallel to the traffic line; green strips and alleys (with trees or bushes). Since road lanes, cycle lanes, pedestrian paths, complicated crossroad lanes and street crossings are all covered by a single polygon, it becomes easier to determine a single space than in case of OSM dataset.

However, in order to use the polygon as a network, it has to be converted into polyline features. It also does not contain paths meant for non-motorised means of transport, and lacks most of the bridges. Due to these reasons, the UA road-land-use polygon needs to be both processed and combined with OSM data to satisfy research needs.

4 Urban Space Network

4.1 Definition of an Urban Space Network

Generally, a street network is defined as a system of interconnecting lines that represent a system of roads for a given area (Mora and Squillero, 2015). In case of this research, Urban Space Network (USN) is a network of interconnected public urban spaces, which are navigable for humans but not necessarily for vehicles. It can be defined as a network whose edges represent a single human-navigable space (i.e. street, footpath, parkway, square, etc.), and its vertices are intersections of such spaces in which there are more than two choices of moving direction. Thus, the conventional street network is merely a subset of the USN.

Specifically, the USN is a topological skeleton of the navigable urban spaces. This topological construct can be represented as a (dual) graph whose nodes and links represent spaces and connections between them respectively. From a cognitive perspective, having navigable spaces as the nodes has a number of advantages for later studies, i.e. the possibility of modelling cognitive costs of going from one space to another. However, the basic reason why recreational activities need to be modelled and analysed in a network space instead of Euclidean space is the assumption that human movement in cities is steered by the built as well as natural environment and its implied movement restrictions. E.g. while the two banks of a river might be very close to each other in an Euclidean space, they might be extremely far away in a network space and therefore the same built environment factors that determine the usage of one bank may have no influence on the other one.

Fig. 3 The framework of Urban Space Network generation, integrating OSM and UA datasets

Furthermore, the USN must be generalised, i.e. contain a single edge for a single perceived space and a single node of intersection. It has to be noted that pedestrians, in contrast to vehicles, are not compelled to use designated paths, e.g. a piece of road between 2 crossings together with all its sidewalks, bicycle and car lanes, and other associated land is considered one navigable space if there are no possibilities to navigate from it to another one.

4.2 Dataset Integration

The biggest mismatch between the UA and OSM datasets is different type of entities (polygon in case of UA and polyline and polygon in case of OSM). There are also geometrical mismatches or cases when streets in one dataset do not appear in the other. In order to overcome these issues and correct the topological errors apparent in the OSM dataset, a new polygon-based approach has been developed as in Fig. 3.
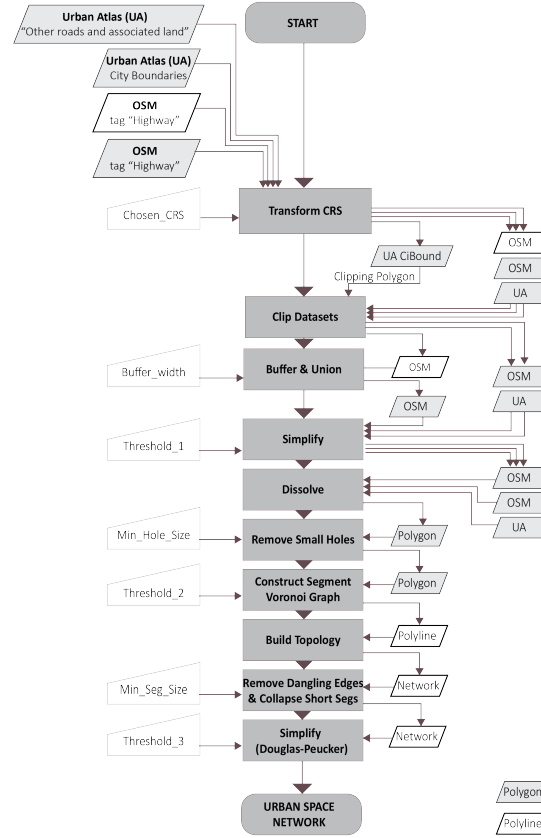
Fig. 4 Polygon-based approach for integrating UA and OSM datasets: white lines represent OSM line features, with buffered by a chosen distance (grey), white polygon represents OSM pedestrian area; green polygon comes from UA dataset of land use type 'Roads and associated land'

The first step of dataset integration is to ensure that all of them belong to the same coordinate system. For this research, along with the default WGS 84, Europe Albers Equal Area Conic (ESRI:102013) has been chosen for visualisation and calculations since it is adapted to fit Europe and uses metric unit system, thus no further recalculation from degrees to meters is needed.

Another important step is network generalization and simplification. Automated generalisation has long been a research effort of cartographers (Jiang and Claramunt, 2004; Savino, 2011; Li et al, 2014). While the previously mentioned researches mainly treat road networks formed by a single dataset and generalisation for scaling purposes, in case of this research an additional challenge is created by using multiple datasets and pedestrian routes, which do not follow such strict patterns as road lanes.

In order to unify the type of entities OSM polylines are buffered and that way transformed into a single polygon. The buffer width is decided based on the general level of detail set for the networks generalisation. When both datasets have the same type of entity, they can be dissolved into a single polygon (Fig. 4). However, beforehand they are simplified using a well-known Douglas-Peucker algorithm with the threshold of 1m in order to reduce computation time. A number of holes, which do not form a substantial gap between the paths, are cleaned by removing polygon rings smaller than a chosen threshold.

After the datasets are united into a single polygon, its centreline needs to be extracted in order to return to the polyline type of entity. The centreline of a polygon is also an approximation of all the neighbouring paths into a single network edge. The Boost library, which provides free peer-reviewed portable C++ source libraries, has been used. The Boost.Polygon.Voronoi has been used to compute a Segment Voronoi (Delaunay) Graph, which takes line segments as an input; therefore, no geometry densification is needed.

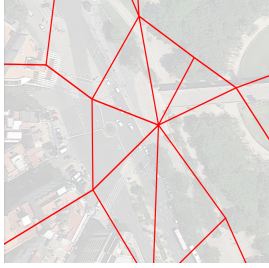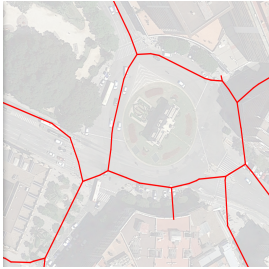**Fig. 5** Part of the resultant USN of Valencia overlaid with Google Earth image

The post processing is needed in order to decrease the complexity of the network and that way save computation time as well as to facilitate the interpretation of space usage values. This step includes building network topology, removing dangling edges, collapsing short segments and simplifying polylines. The example of resultant Urban Space Network in Valencia is shown in Fig. 5.

4.3 Validation

The benefits of the developed USN construction method are validated comparing it with a network obtained using a more commonly used approach for network generalization and simplification: that is by iteratively using topological cleaning tools followed by a vertex-snapping algorithm. The method, differently than the proposed one, is based on polylines; therefore, initially, centrelines have to be extracted from the UA road polygon and OSM pedestrian areas. Table 1 shows the differences between the two approaches considering a number of relevant aspects.

While both methods have their own benefits and drawbacks, the polygon-based method fits the purpose of this research better, since it provides a simpler outcome with less redundant connections, easily removable artefacts and a single network edge per single perceived space.

Table 1 Comparison between the polygon-based and the polyline-based dataset integration and network generalisation methods

	Polygon-based	Polyline-based
Topological validity	The outcome network is always topologically valid	The outcome needs to be cleaned from topological errors: mainly overlaps, pseudo-nodes and duplicate geometries
Junction simplicity	Junctions need to be further processed by collapsing short segments	Bigger junctions (more than 4 ways) tend to create artefacts
Redundant segments	Centerline extraction algorithm creates redundant dangles	Snapping algorithm results into redundant connections
Geometric distortions	Geometric distortions do not exceed the buffer width	Polyline geometry can get severely distorted while moving all vertices of a polyline into different directions
Attributes	No attributes preserved	Attributes are preserved
Execution time	The crucial time needed for both methods is the extraction of Segmented Voronoi (Delaunay) Graph edges, which lie inside the polygon; the buffering time in polygon-based method is comparable with the snapping time in polyline-based method	

5 Space Usage

5.1 Filtering GPS Trajectories

In the initial state acquired GPS tracks are rather a set of coordinates, which are not in any way related to the USN (Fig. 6), therefore in order to define the usage measure, GPS tracks need to be processed filtered and snapped to the underlying network.
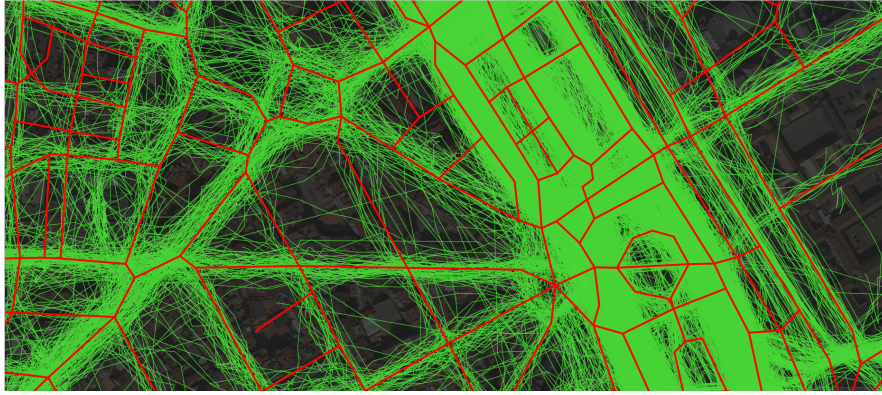
Fig. 6 GPS track (transparent green) on a USN (single line red) prior to snapping

Filtering of GPS points is needed in order to remove blundering values, which appear in GPS trajectories due to various reasons: lack of satellites in sight due to environment obstructions, cold start or signal multipath. Filtering outliers has been detached from the initial filter that takes place while writing data into the database in order to reduce total filtering time and be able to process only the relevant GPS tracks. However, this is a trade-off between filtering time and loss of individual GPS point attributes. Consequently, such methods as proposed by (Schuessler and Axhausen, 2009; Auld et al, 2013; Biljecki, 2010), which suggest removing the outliers from GPS data based on the unrealistic altitude, sudden speed and acceleration jumps or sudden changes in heading become unavailable.

In case of this research the under-filtering is less of a problem than over-filtering due to the snapping algorithm, which relies on a sequence of points. In addition, scarce data should not be lost during the outliers filtering. Therefore, the definition of an outlier has been formulated as following: it is a point that lies from both of its neighbours further than three times the median while the distance between the neighbours is less than the smaller distance between the point and each of its neighbours. Median refers to the median distance between two consecutive points calculated for each GPS track individually.

The heuristics of using three medians comes from the evaluation of a sample set of 100 randomly chosen GPS tracks from different cities, which can be visually confirmed as not having outliers. The calculation is based on the ratio between the median distance between two consecutive points and a maximum deviation from the median in each of the test tracks. The average value of 100 tested ratios appeared to be 2.2215 with standard deviation of 0.8429. Thus, if the ratio between a point and its neighbours is higher than the mean ratio plus the standard deviation, the point can be considered as suspicious.

Fig. 7 GPS snapping algo-
rithm: bright line indicates
original GPS track; dark
line indicates USN edges to
which the GPS track has been
snapped.

5.2 GPS Track Snapping on an Urban Space Network

The reason for the geometric mismatch between the GPS tracks and the space net-
work lies both in the inaccuracy of the GPS measures and the data used to construct
the USN. Most of the map-matching algorithms tend to deal with the GPS tracks of
vehicle movements, which are in many aspects different from the workout data. E.g.
runners as in contrast to vehicles, do not necessarily stay on a designated path, they
can change moving direction at junctions as well as in the middle of a path, do not
have any movement restrictions or predictable moving speed. Moreover, nothing is
known about the characteristics of a GPS device, positioning data quality, satellites
in range or the frequency of GPS fixes. Due to these reasons, most of the advanced
algorithms cannot be implemented and therefore only geometrical and topological
data is used for snapping.

The GPS snapping algorithm has been developed based on the algorithms pro-
posed by Marchal et al (2004); Yang et al (2005); Quddus and Washington (2015).
It is a topological algorithm, which relies on the multiple hypothesis technique. It
allows to keep track of several positions or paths at once and to select eventually
which candidate is the best. The first point is snapped to the two closest segments
of the extracted piece of the whole network. Later, the best-fit edge is decided by
checking the following points and choosing the best matching one. The path is aug-
mented through topological connections of the best fitting edge, always choosing
two of them based on a single point and deciding the better one based on a sequence
of points up until the last GPS point is reached. The sample results of map matching
algorithm can be seen in Fig. 7.

The accuracy of the map-matching algorithm has been computed by visually
comparing the GPS track with the assigned USN edges of 25 randomly selected
samples in Vilnius city, which all together make up almost 5000 GPS points.
Mapping accuracy has been computed as a number of correctly assigned network

edges over the number of all edges considered (assigned, over-assigned and under-assigned) and results into 85% of overall mapping accuracy, which is reasonable for a geometrical/topological, map matching algorithm. The standard deviation of GPS points to the network edge they are snapped to is 15.859m.

Moreover, over-assignment is more frequent than under-assignment. This happens often due to lack of edges in the network, i.e. recreational activities happening in spaces which are not represented by any edge in the network. This can happen because of two reasons either the lack of an existing path in the OSM or UA data or the absence of a path as such, e.g. running in out-door stadium, in meadows or private lands.

5.3 Value of Recreational Usage

After snapping GPS tracks every space in a network gets an attribute of a number of distinct application users spotted therein. While the overall goal is to model the recreational space usage, i.e. give an indication to every space of how much the particular space is used for recreation, the actual counts cannot clearly represent the measure. Moreover, literal quantification of recreational usage is impossible, since it is a rather qualitative notion.

In order to quantify a qualitative measure, a fuzzy notion of likeliness has been used (Klir and Yuan, 1995). It describes how likely it is that a space is used for active recreational travels and is measured in the range of 0 to 1, where 0 means no usage and 1 means that a space is definitely used. All values in between indicate how much a space is used compared to the other ones. It is important to note that values of recreational usage are not numerical but rather of an ordinal nature. Fuzzy normalisation primarily serves for visualisation purposes enabling more intuitive and comprehensible overview of space usage.

In order to perform the fuzzy normalization, first the cumulative frequency for each space usage value is calculated. The number of distinct users spotted in a space over the whole study period is denoted as u. The set of spaces with k or less users is defined as $S(k) = u|u \leq k$; as to which the cumulative frequency is $f^C(k) = |S(k)|$. Then normalized space usage and its normalized cumulative frequency are defined respectively as: $k^n = k/k_{max}$, in which k_{max} is the maximum number of spotted distinct users for a single space in the whole range of spaces; and $f_n^C(k) = (f^C(k) - f^C(0))/(f^C(k_{max}) - f^C(0))$ in which $f_n^C(k)$ denotes the normalized cumulative usage frequency of k or less users.

Any space which has a number of spotted application users above 0, is regarded as 'somewhat used for recreation'. The distribution of values differ per city due to the different proportions of network size and number of application users and because of different distributions of recreational activity, which are dependent on individual characteristics of the built environment. Higher number of attractive spaces shares out the total number of the users, while lower amount of attractive spaces concentrates the users within them.
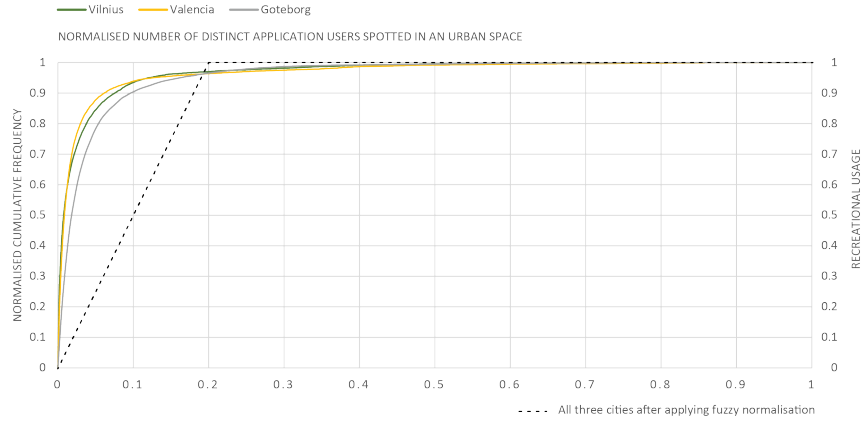
Fig. 8 The solid lines correspond to the normalised cumulative frequency of urban spaces, while the dashed line corresponds to the recreational usage of those spaces, when the normalised number of spotted distinct application users is the same. All normalisations are done according to the cities own minimum and maximum values.

However all normalized frequencies have similar distributions and approximately even out at one point corresponding to 20% of the maximum usage. Simply put, in all three cases only 3% of all the network spaces have a number of spotted users higher than 20% of the maximum registered. Therefore, this point has been used as a reference for the likeliness coefficient. For example, in case of Vilnius the maximum number of spotted users in an urban space is 592, which means that all spaces, which have 118 or more users, are regarded as 'used for recreation'. Accordingly, a space, which has 59 users, is considered to have recreational usage value of 0.5. Fig. 8 illustrates the dependency between the number of users spotted in a single space and its recreational usage.

6 Results

The resulting maps for all three case study cities can be seen in Fig. 9.

While examining the maps closer, it can be noticed that large recreational areas in all of the three cities attract the most of the recreational activities. While these results are expected, the interesting things can be noticed while closer examining non-recreational areas. For example, in case of Vilnius, some heavy traffic streets are used for recreation more than the nearby green zones, while in case of Valencia almost all recreational activities are concentrated in the parks or green alleys, leaving densely urbanised areas excluded. In case of Gothenburg, the most recreational spaces almost evenly spread throughout the city, interconnecting with each other and forming their own 'recreational network'.

Fig. 9 Visualisation of USN coloured according to its recreational usage values as defined after the fuzzy normalisation; case studies of Vilnius, Valencia and Gothenburg
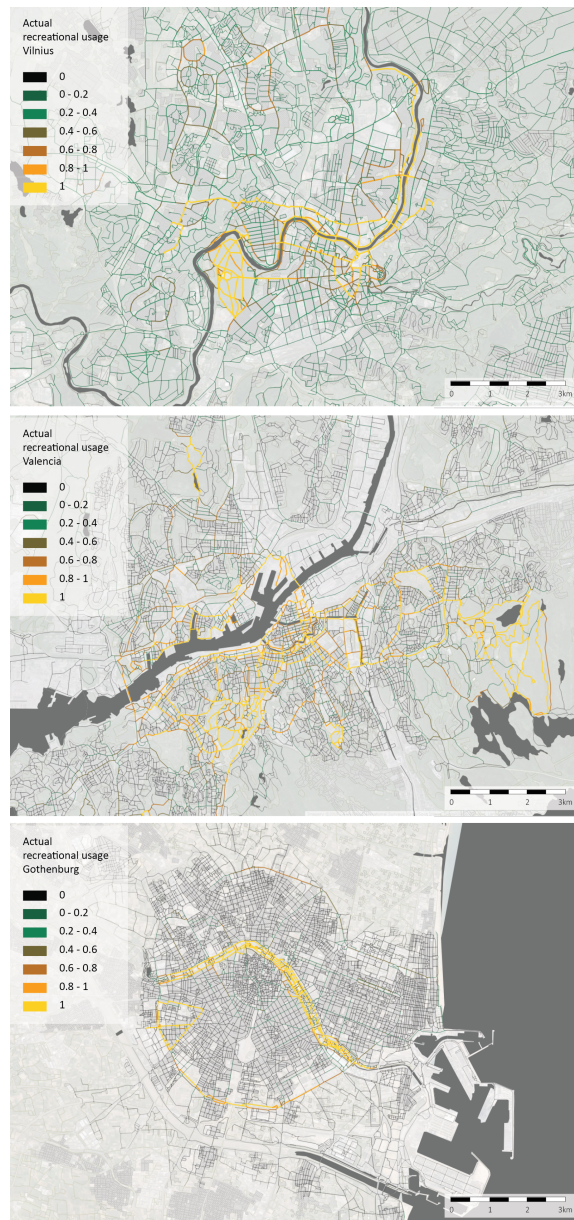
The constructed network can be overlaid with a number of related maps in order to visually inspect the relationships between different phenomena. However, more importantly, network nodes can also be attributed with a number of measures, such as space greenness, network centrality, land use, etc. in order to use quantitative methods to find associations between the values.

Finally, by looking at the maps, it can be noticed that it is not only the attractiveness of a single space that enables presence of recreational activity but also its position in a broader network of spaces; not in a sense of being in an attractive area but in a sense of being connected to other attractive spaces. In other words it is the position of a space in the network space that matters more than its position in the Euclidean Space. Therefore, these findings suggest that while analysing recreational activity patterns, it is the network-based analysis that must play bigger role than the neighbourhood-based analysis. This highlights the role of USN as an essential construct in this research.

7 Discussion and Future Work

First of all, a collaboration between sports tracking application and a researcher would significantly improve the efficiency of data acquisition. Furthermore, knowing such characteristics as user age group, occupation, education, etc. might give a better overview of data validity and allow deeper investigation of recreational travel patterns. Currently, user group analysis is not possible due to the privacy matters.

Furtermore, the running and walking activities have been considered equally, while they might also have different movement patterns. In addition, various other types of recreational travels could be added among which recreational cycling, orienteering, roller skiing, skateboarding, etc. Generally, the collected data is limited to only one sports tracking application which limits the set of tracked individuals to those who have knowledge of a foreign language, possession of a smart phone, ability to use the application and, of course, having given a consent to be tracked. Therefore, it must be acknowledged that acquired data represents only a certain subset of all recreational travels conducted in a city, which may cause related bias to the research results.

Even though the integration of UA and OSM datasets improves the completeness of a USN, a number of paths and connections remain unknown. This problem could be tackled by upgrading the GPS network-snapping algorithm. The missing paths could be added to the constructed USN based on the clusters of GPS tracks. This would also improve the mapping accuracy of the snapping algorithm itself. Furthermore, some heavy traffic roads should rather be considered as barriers as well as water bodies, so that only certain connections through them would be possible. Finally, buffering sometimes may cause connection of spaces, which actually do not reach in reality due to topography, water features, etc.

8 Conclusions

The conducted research has investigated how mobile sports tracking application data can be used to model and visualise the recreational usage of an Urban Space

Network. An automatic and non-labour intensive method has been devised for data acquisition, management and processing. Collected GPS tracks have been filtered from blundering fixes and snapped to a USN with 85% mapping accuracy. GPS tracks when aggregated per single network edge form a measure, which is later normalised using fuzzy normalisation methods, and represents how much a space is used for recreation compared to the other ones.

Before processing the mobility data, a systematic workflow has been developed for constructing an Urban Space Network using OSM data complemented with UA road land use data. The method relies on integration of datasets, generalisation and simplification through buffering linear features, combining all polygons and using Segmented Voronoi (Delaunay) Graph to extract polygon centreline, which, after minor processing and additional simplification is used as a representation of a USN. The constructed network is relevant for the desired type of analysis and differs from conventional street networks in that it includes paths for both motorised and non-motorised means of transport, which run through urban fabric as well as parks and urban forests. A particular characteristic of the USN is that it has low granularity, however, well-preserved space connectivity.

The visualisation of results has proved that analysing recreational usage in a network space instead of Euclidean space brings clearer insight and provides a basis for understanding and explaining the usage patterns and their associations with built environment effects. Finally, testing all processes and algorithms in parallel for three different case studies has ensured that the collected data as well as the developed methods would not be dependent on a specific urban structure and can be repeated for any of the European cities with sufficient application users.

References

Auld J, Williams C, Mohammadian A (2013) Prompted recall travel surveying with gps. In: Transport Chicago Conference, Zugegriffen, vol 15

Barsukov N (2014) Generating running route maps. URL `http://barsukov.net/programming/2014/07/26/endomondo-code.html`, cited 1 Dec 2014

Biljecki F (2010) Automatic segmentation and classification of movement trajectories for transportation modes. Master's thesis, TU Delft, Delft University of Technology

Cohen DA, Ashwood JS, Scott MM, Overton A, Evenson KR, Staten LK, Porter D, McKenzie TL, Catellier D (2006) Public parks and physical activity among adolescent girls. Pediatrics 118(5):e1381–e1389

Endomondo (2015) Endomondo sports trackers. URL `http://www.endomondo.com`

Ferrari L, Mamei M (2011) Discovering city dynamics through sports tracking applications. Computer 44(12):63–68

Ferrari L, Mamei M (2013) Identifying and understanding urban sport areas using nokia sports tracker. Pervasive and Mobile Computing 9(5):616–628

Floyd MF, Spengler JO, Maddock JE, Gobster PH, Suau LJ (2008) Park-based physical activity in diverse communities of two us cities: an observational study. American Journal of Preventive Medicine 34(4):299–305

Gauvin L, Richard L, Craig CL, Spivock M, Riva M, Forster M, Laforest S, Laberge S, Fournel MC, Gagnon H, et al (2005) From walkability to active living potential: an ecometric validation study. American journal of preventive medicine 28(2):126–133

Gebel K, Bauman AE, Petticrew M (2007) The physical environment and physical activity: a critical appraisal of review articles. American journal of preventive medicine 32(5):361–369

Girres JF, Touya G (2010) Quality assessment of the french openstreetmap dataset. Transactions in GIS 14(4):435–459

Hillier B, Iida S (2005) Network and psychological effects in urban movement. In: Spatial information theory, Springer, pp 475–490

Jiang B, Claramunt C (2004) A structural approach to the model generalization of an urban street network. GeoInformatica 8(2):157–171

Kaczynski AT, Potwarka LR, Saelens BE (2008) Association of park size, distance, and features with physical activity in neighborhood parks. American journal of public health 98(8):1451

Klir G, Yuan B (1995) Fuzzy sets and fuzzy logic, vol 4. Prentice Hall New Jersey

Li Q, Fan H, Luan X, Yang B, Liu L (2014) Polygon-based approach for extracting multilane roads from openstreetmap urban road networks. International Journal of Geographical Information Science 28(11):2200–2219

Lopez R (2004) Urban sprawl and risk for being overweight or obese. American Journal of Public Health 94(9):1574–1579

Marchal F, Hackney J, Axhausen K (2004) Efficient map-matching of large gps data sets: Tests on a speed monitoring experiment in zurich. Arbeitsbericht Verkehrs und Raumplanung 244

Mooney P (2015) An outlook for openstreetmap. In: OpenStreetMap in GIScience, Springer, pp 319–324

Mora AM, Squillero G (2015) Applications of evolutionary computation. In: 18th European Conference, EvoApplications 2015, Springer, vol 9028

NYC DDC (2010) New york city active design guidelines: Promoting physical activity and health in design. URL `http://centerforactivedesign.org/dl/guidelines.pdf`, cited 29 Nov 2014

Oksanen J, Bergman C, Sainio J, Westerholm J (2015) Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data. Journal of Transport Geography 48:135–144

Piorkowski M (2009) Sampling urban mobility through on-line repositories of gps tracks. In: 1st ACM international workshop on hot topics of planet-scale mobility measurements, ACM

Quddus M, Washington S (2015) Shortest path and vehicle trajectory aided map-matching for low frequency gps data. Transportation Research Part C: Emerging Technologies 55:328–339

Savino S (2011) A solution to the problem of the generalization of the italian geographical databases from large to medium scale: approach definition, process design and operators implementation. PhD thesis, Universita di Padova

Schuessler N, Axhausen K (2009) Processing raw data from global positioning systems without additional information. Transportation Research Record: Journal of the Transportation Research Board (2105):28–36

Urban Audit (2007) State of european cities report. Tech. rep., Study contracted by the European Commission

Van der Spek S, Van Schaick J, De Bois P, De Haan R (2009) Sensing human activity: Gps tracking. Sensors 9(4):3033–3055

Van der Spek SC, Van Langelaar CM, Kickert CC (2013) Evidence-based design: satellite positioning studies of city centre user groups. Proceedings of the ICE-Urban Design and Planning 166(4):206–216

Yang Js, Kang Sp, Chon Ks (2005) The map matching algorithm of gps data with relatively long polling time intervals. Journal of the Eastern Asia Society for Transportation Studies 6:2561–2573