Understanding Active Travel Networks Using GPS Data from an Outdoor Mapping App

Marcus A. Young
Transportation Research Group, University of Southampton, UK

Abstract

To support a shift to active travel there is a vital need for better data to understand active travel networks: their extent, attributes and current utilisation. Using a big dataset of volunteered geographic information from an outdoor mapping smartphone app, a methodology has been developed to analyse recorded routes to identify missing links in a routable street and path network and to visualise the relative importance of different links of the active travel network. This methodology has then been used to analyse the network for a case study area around Winchester, UK, with new pathways equivalent to 8% of the existing network dataset identified. The automated method developed can be readily applied to other locations and the outputs used to augment existing network datasets and to inform the planning and development of active travel infrastructure.

2012 ACM Subject Classification Applied computing → Transportation

Keywords and phrases active travel, map construction, GPS, volunteered geographic information

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.88

Category Short Paper

1 Introduction

To meet the UK's target of net zero carbon emissions by 2050, tackle local air pollution, and address the obesity public health crisis, there needs to be a step-change in the use of active travel modes[1]. However, the data needed by users of the active travel network, and those who maintain and develop it, is lacking. There is no single dataset that encompasses the entirety of the active travel network, and limited information on its utilisation and attributes. This gap in data provision poses a significant challenge to the successful delivery of active transport policies.

Existing routable network datasets are primarily focussed on meeting the needs of motorised transport and are unsuited to the accurate analysis and planning of active travel.[2] They mainly consist of street centrelines with attributes that are inadequate for reliable and safe pedestrian and cyclist routing (for example, pavement presence information is lacking). Relatively little attention has been given to how well these data describe the actual active travel network, in terms of both scope and real-world usability. Data from other tracking apps have been used to understand active travel, most notably Strava Metro (for example, see [5] for a review of its use to monitor cycling). However, the Strava dataset is mapped to OpenStreetMap ways so cannot be used to identify unknown parts of the network.

The research described in this paper is part of a larger on-going project - Routable Active Travel Infrastructure Network (RATIN) - being carried out by researchers at the University of Southampton and funded by Ordnance Survey. Phase one of the project was a scoping study to identify data and methods which could help provide a comprehensive routable active

[1] Active travel is defined as journeys made by transport modes that are fully or partially people-powered, irrespective of journey purpose, for example: walking, using wheelchairs, and cycling (including e-bikes).
[2] The two main options currently available for generating routable transport networks in Great Britain (GB), are OpenStreetMap (OSM) and the MasterMap Highways Network from Ordnance Survey.

12th International Conference on Geographic Information Science (GIScience 2023).
Editors: Roger Beecham, Jed A. Long, Dianna Smith, Qunshan Zhao, and Sarah Wise; Article No. 88; pp. 88:1–88:6
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

travel network dataset. This included a work package to develop methods for identifying and understanding real world active travel networks from volunteered geographic information in the form of GPS traces derived from smartphone apps. In particular, the research sought to analyse a large dataset from OSMaps.[3] The aim of the research described in this paper was to process GPS recorded routes to identify parts of the active travel network that are not currently incorporated in the Ordnance Survey (OS) Mastermap (MM) Highways and Urban Paths products and to assess the suitability of the data to understand where active travel is taking place. The only previous research to use this dataset developed a classification to describe and group walking routes based on environmental characteristics [2].[4]

2 Data and Methods

2.1 Data and study area

The OS smartphone app, OSMaps, is a popular outdoor mapping product in GB that enables users to plan future routes (by manually plotting them) or to record routes (via a smartphone built-in GPS device) as they walk or cycle (or undertake other outdoor activities). Data was provided for all routes intersecting the Hampshire and Southampton Unitary Authority areas that were generated from 2019–2021. Following data cleaning (see section 2.2) and methodological development (see section 2.3), the routes within the Winchester area, NGR squares SU43SE and SU42NE, were analysed.

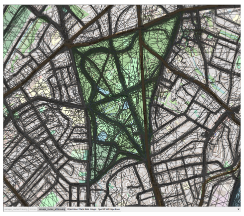
The raw data (an 80GB CSV file) was imported into a PostgreSQL/PostGIS database and contained 325,532 routes that were entirely within a bounding box around the Hampshire and Southampton boundary. The bulk of the information for each record is held as a complex JSON (not GeoJSON) formatted object in one column. The route information (without GPS timestamps) is stored in a child element that consists of one or more coordinate arrays which each contain two key:value pairs; one for latitude and one for longitude. Native JSON support within PostgreSQL (jsonb) was utilised to extract information from the JSON object making use of lateral joins to generate linestrings from the individual latitude and longitude coordinates making up route features. Information indicating whether the route was "plotted" (recorded by GPS), "routed" (manually created) or "imported" was also extracted.[5]

2.2 Data cleaning

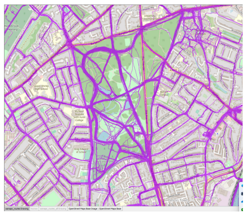
When the extracted routes were visualised it was apparent that the data contained substantial 'noise'. As shown in Figure 1a, there are many straight lines criss-crossing the area and routes with large distances between vertices making their representation of pathways followed on the ground inadequate. The following inclusion rules were applied using SQL queries to improve the dataset for subsequent analysis: entire route not a straight line; maximum segment length (distance between vertices) < 250 metres; average segment length < 125 metres; and number of segments > 2 (i.e., at least 4 vertices).

The data cleaning process reduced the number of routes from 325,532 to 66,587. The geometry of the cleaned routes is shown for an example area in Figure 1b. The median route length in the cleaned dataset was 6.4km with first and third quartiles of 3.9km and 9.9km respectively. The distribution of routes is left-skewed with a small number of routes

[3] See: `https://osmaps.com/en-GB`
[4] OS has also produced visualisations of the raw data, e.g., see: `https://bit.ly/42HvoY5`.
[5] A route can consist of a mix of "plotted" and "routed" components.



(a) Before processing.



(b) After processing.

Figure 1 OSMaps route data before and after cleaning in an example area.

extending to 75km and beyond. The user specified activity type is overwhelmingly walking (82%), with cycling and running accounting for 10% and 7% respectively, and other activities 1%. The vast majority of the routes are recorded by GPS (62,052 or 93%).

2.3 Map construction methodology

Map construction methods are used to automatically generate (or update) a street network from multiple GPS traces recorded from within vehicles. This is an active research area and many algorithms have been developed to solve the problem (see [1] for a review). In this study, the approach was used to convert the OSMaps routes into a vector map of the active travel network. A density-based map construction method was used based on the work of [3] and discussed in [1], and implemented using algorithms within the QGIS application. The method consists of line density estimation, skeletonization, conversion to vector data and a topology cleaning/refinement process. The processing steps are detailed below with example outputs shown in Figure 2. Only routes based on GPS were used for the map construction.

Line density estimation

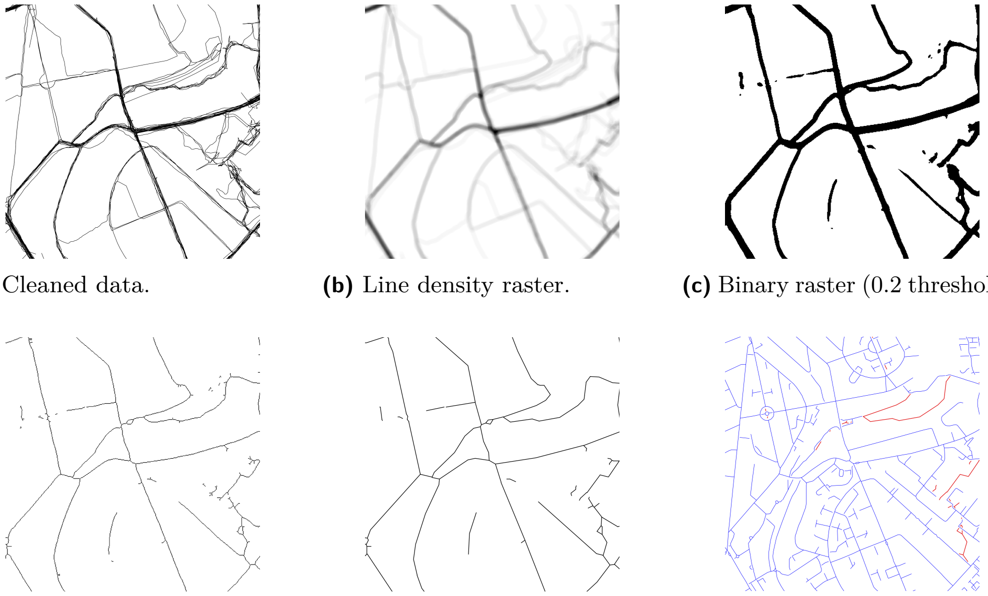
The QGIS line density interpolation tool was used to calculate a density measure of the routes within a circular neighbourhood of each cell across a raster surface. For each raster cell the length of line segments that intersect its circular neighbourhood is summed and divided by the area of the neighbourhood. A cell size of 2m was selected with a neighbourhood radius of 10m. The resultant raster is useful for visualising and analysing levels of use on the active travel network (see Figure 2b). This was followed by conversion to a binary raster, where a threshold line density is applied so that only cells with a high likelihood of a well-used pathway passing through them are assigned a value of 1 (see Figure 2c).

Skeletonization and thinning

The GRASS r.thin tool was used to 'skeletonize' the binary raster. The tool uses the algorithm described by [4] to thin the non-null cells that represent active travel routes into linear features of single pixel width (see Figure 2d). This step is necessary so that the raster can then be converted into a vector linestring layer using the GRASS r.to.vect tool.

Topology cleaning and refinement

The vector layer was simplified by removing vertices with a tolerance of 5m and short dangles up to 15m using the GRASS v.clean tool. The resulting vector layer was compared with ground-truth (the known road and urban path network), enabling identification of parts of the active travel map that are not represented by these products. These were extracted



(a) Cleaned data. **(b)** Line density raster. **(c)** Binary raster (0.2 threshold).

(d) Skeletonized raster. **(e)** After topology cleaning. **(f)** New paths (red), MM (blue).

Figure 2 Outputs from the active travel map construction process for a small example area.

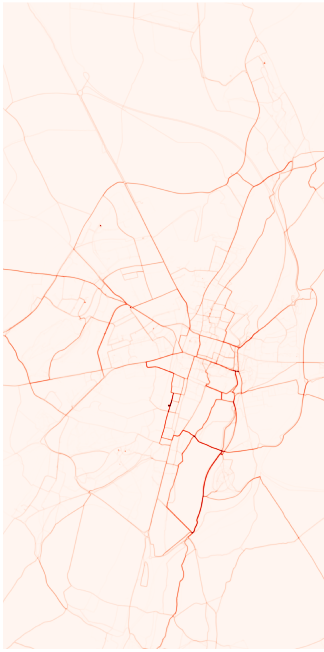
by retaining lines (or line parts) outside of a variable width buffer based on the average real-world width of the MM roadlinks or 5 metres for the urban pathlinks (see Figures 2e and 2f). The PostGIS ST_ClusterIntersection function was used to identify interconnected edge clusters which were then converted to simple line features using ST_CollectionHomogenize. This enabled the length of these features to be calculated and assessed.

3 Results and Discussion

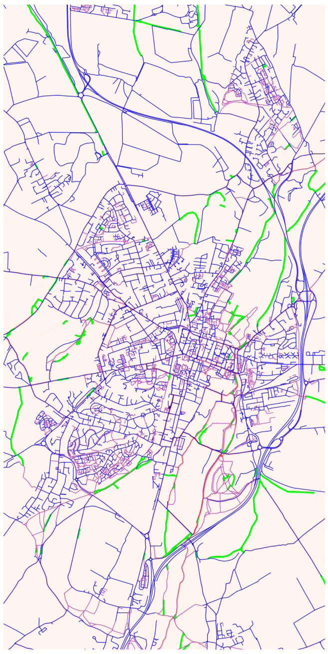
The line density raster generated for the recorded routes within the two 5km grid squares covering the case study Winchester area is shown in Figure 3a, with darker reds indicating higher line density (and therefore higher level of use). Some of the most popular active travel locations include central streets, such as Christchurch Road, St James' Lane and parts of the High Street, and off-street paths alongside the River Itchen, near Bridge Street and (via Garnier Road) near Saint Catherine's Lock (part of the Itchen Way long distance footpath).

The case study area contains approximately 375km of road links (excluding motorways that are not available for active travel) and 126km of urban path links from the OS MM network dataset. Based on a binary raster with a line density threshold of 1, a total of 38.8km of potential new links was identified, equivalent to 8% of the known network and consisting of 1049 identified line clusters. Many of these clusters are very small artefacts (789 are less than 1 metre) and the median length is 7.7m with the first and third quartiles 1.1m and 22.1m respectively. The top 10% of line clusters are 60m or more in length and account for 28.3km (73%) of the newly identified paths. These are shown in green in Figure 3b overlain on the line density raster with the MM streets and urban paths.

Many of the additional links of the active travel network identified may be available in other datasets, for example the definitive maps of public rights of way (which are available as GIS datasets from some local authorities in GB) or recorded in OpenStreetMap (which may also be partly sourced from the definitive maps, although investigations have shown



(a) LD raster - darker lines indicate greater level of use (more recorded routes).



(b) LD raster overlain with MM roads (blue), paths (purple), and new paths >60m (green).

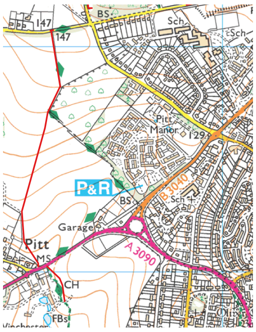
Figure 3 Outputs from the active travel map construction process for Winchester area.

that OSM does not include all the additional links). However, the actual path used may differ from that recorded in the definitive map. An example is shown in Figure 4a, where the extracted path (red) is not recorded as a legal right of way and is the preferred route to the A3090 as indicated by the line density raster (Figure 4b).

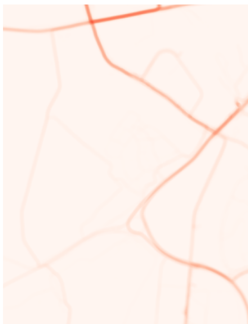
Limitations and future work

Users of OSMaps are a self-selecting group of outdoor enthusiasts who have chosen to use this app. It will be biased towards longer leisure journeys rather than shorter travel to work, school or functional journeys. This will affect the relative levels of use of different parts of the network identified by the line density raster. Some important routes, for example a path giving access to a school, may appear insignificant in the line density raster and be excluded when the line density threshold is applied when creating the binary raster. Recorded tracks that on the ground relate to different paths that are close and parallel to one another can appear to be a single path when the binary raster is created. It may be possible to limit this by using a smaller cell size and adjusting the line density neighborhood distance.

Future work will develop methods to automatically link newly identified paths to the existing MM network and seek to improve the extraction of new paths, perhaps using map-matching algorithms that are usually used to identify the road network link (centreline) that is being driven by a vehicle [6]. The use of buffers around paths and roads can be indiscriminate and more problematic in built-up areas where GPS traces can be deflected from their true position by tall buildings. Future work will also consider how this data (potentially with other big data sources) could be incorporated into a decision-making tool that would enable local authorities to understand how active travel networks are being used and thus aid future planning for maintenance, enhancement and additions to infrastructure. This will include further disaggregation of the data to analyse different types of activity.



(a) New path shown in red, existing legal right of way shown as green diamonds.



(b) line density raster with line intensity indicating relative level of use of the two paths.

Figure 4 Example of new pathway identified that has higher level of use than existing right of way.

4 Conclusion

While substantial effort is put into monitoring motorized traffic (for example, the statistics compiled by DfT[6]), much less attention, beyond some monitoring of on-street cycle use, is given to understanding the use of active travel networks. This research has shown how a big dataset of volunteered geographic information from an outdoor leisure mapping smartphone app can be used to visualise where active travel is taking place, understand the relative importance of different parts of the active travel network, and through an automated process identify pathways that are not currently contained within a motorized vehicle-oriented street and urban path based network. The automated method that has been developed is readily transferable to other locations and/or other sources of this type of data.

References

1 Mahmuda Ahmed, Sophia Karagiorgou, Dieter Pfoser, and Carola Wenk. *Map Construction Algorithms*. Springer, softcover reprint of the original 2015 edition, 2019.

2 Andrea Ballatore, Stefano Cavazzi, and Jeremy Morley. The context of outdoor walking: A classification of user-generated routes. *The Geographical Journal*, Advance online publication, 2023.

3 James Biagioni and Jakob Eriksson. Inferring Road Maps from Global Positioning System Traces: Survey and Comparative Evaluation. *Transportation Research Record: Journal of the Transportation Research Board*, 2291(1):61–71, 2012.

4 B.-K. Jang and R.T. Chin. Analysis of thinning algorithms using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(6):541–551, 1990. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

5 Kyuhyun Lee and Ipek Nese Sener. Strava Metro data for bicycle monitoring: a literature review. *Transport Reviews*, 41(1):27–47, 2021.

6 Mohammed A. Quddus, Washington Y. Ochieng, and Robert B. Noland. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation Research Part C: Emerging Technologies*, 15(5):312–328, 2007.

[6] see: `https://www.gov.uk/government/collections/road-traffic-statistics`