Using OpenStreetMap in spatial network analysis

Open Street Map (OSM) is currently the world’s most prominent open mapping platform. A non-profit foundation registered in England and Wales since 2006, OSM passed the million-user mark in 2013, containing 21 million miles of road data and recruiting 1000 new contributors per day, resulting in ever increasing accuracy\(^1\). More than 100 universities have research associated with Open Street Map\(^2\).

On the other hand, the crowd-sourced nature of OSM means that some unique problems are encountered in its usage. Data quality is not consistent, with more accuracy in some regions than others; there is also a lack of consistency between regions when it comes to the recording of link attributes. The definitive OSM source data is corrected and updated daily, although with a bias towards more updates in some areas than others.

These issues notwithstanding, we expect OSM to improve over time, and we have already found it to be very useful here at sDNA headquarters. The author recently produced a model of the Cardiff city region based on OSM, that correctly predicted 90% of the variance in vehicle traffic flows, and 75% of the variance in flows of pedal cycles. In the UK, OSM is probably the most complete digital record of pedestrian and pedal cycle routes published to date.

Background reading: considerations for preparing all spatial networks
The remainder of this document assumes you have read our recent guide: Cooper (2014) “Preparing models for use in spatial network analysis”, available in the Documentation section of the sDNA website. Most of the issues encountered in using OSM are special cases of the issues described there; we describe here only the considerations unique to OSM. If you haven’t read the other guide, please read it first.

Downloading OSM
Numerous online services exist for extracting OSM data into shapefiles. For the model described above, the data source was Mapzen: https://mapzen.com/metro-extracts/. Mapzen extracts come with more attribute data attached than most (including e.g. cycle routes) but are available only for urban areas. If you need larger scale or rural OSM extracts, you may find other services suit your purposes better.

Spatial referencing in ArcGIS
OpenStreetMap is stored and downloaded on the WGS84 datum. ArcGIS fails to correctly identify the transformation necessary to project OSM data to a national grid. The easiest way to resolve this is (before loading the data) to mark it as WGS84 using ArcCatalog.

- In ArcCatalog, right-click on the feature class you have downloaded
- Choose “Properties”

\(^1\) https://www.mapbox.com/osm-data-report/
\(^2\) http://wiki.openstreetmap.org/wiki/Research
Choose “XY Coordinate System”

You should see that the data is already set to WGS84 (as described in the projection data for the shape file). Change this to ArcMap’s own definition of WGS84 so ArcMap can correctly select transforms. A suitable choice is Geographic Coordinate Systems -> Spheroid Based -> WGS 1984 Major Auxiliary Sphere.

Repeat for any other OSM feature classes you have

Once you have loaded the feature classes, you can change the spatial reference of the data frame to e.g. “British National Grid” and the data will project correctly.

Before further processing, project your input data to the grid you will be using with ArcToolbox -> Data Management -> Projections -> Feature -> Project.

Selecting subtypes from OSM

OSM contains numerous types of lines feature beside roads and paths – for example walls, fences and river banks. You will need to filter out the roads and paths only. This can be achieved using Select Layer by Attributes.

In Cardiff (as of November 2014) a road was present when the ‘highway’ field was set to one of the following: living_street, motorway, motorway_link, primary, primary_link, residential, road, secondary, secondary_link, service, services, tertiary, tertiary_link, trunk, trunk_link, unclassified. Note that other areas may differ, and these classes may change over time. We recommend examining all values which appear in the ‘highway’ field to decide for yourself which to use.

Cycle lane data was not extensive enough to be useful, but data on cycle routes was consistent. Cycle routes do not necessarily denote highways, but are a separate feature that coincides with the highways used. Some of the marked cycle routes coincided with features which were classed as paths rather than highways (a data inconsistency). Therefore it was necessary to extract all candidate highways for cycle routes using the following procedure:

- Select all lines where “route” = bicycle
- Create 5m buffers surrounding these lines
- Create a new data field on OSM, “on_cycle_route”
- Use spatial join to set on_cycle_route=1 for all OSM lines within the cycle route buffers
- Use the query "highway" <> " AND "on_cycle_route" = 1 to extract all highways coinciding with cycle routes.

With further use of the “select by attribute” tool it was possible to create two fields on the extracted highway network.

- cycle_route – set to 1 if on a cycle route, 0 otherwise
- car_net – set to 1 if on car (road) network, 0 otherwise

Reducing such data to numeric values is important if it is to be preserved during the sDNA Prepare process.
Connectivity and geometry errors
As of November 2014, the OSM data for Cardiff contained a number of connectivity and geometry errors. These were fixed by planarizing with a 1m cluster tolerance. It was first necessary to extract bridges and tunnels, to avoid planarizing these also, using the SQL query:

\[ \text{\texttt{"bridge" <> 'no' AND "bridge" <> ' ' or "tunnel" = 'yes'}} \]

Again, it is strongly recommended to inspect all values which appear in the bridge and tunnel fields in order to be sure that you are capturing all information provided by OSM. In our case the bridge field could contain the values "", "no", "suspension", "viaduct", "swing", "transporter" and "yes".

The feature class was thus split into two separate classes; one for brunnels and the other for ordinary links. The ordinary links were planarized. In order to ensure that the ordinary link layer connected correctly to the brunnels layer, topology checking was used. A point feature class was created from the endpoints of all ordinary links (using the feature vertices to points tool). A topology rule was then specified that brunnels must have endpoints covered by ordinary link endpoints.

For the entire Cardiff region, there were about 100 violations of this latter rule. The vast majority were cases where a brunnel’s endpoint lay on another brunnel. The remaining cases were points where a brunnel failed to correctly intersect an ordinary link, and were fixed by hand.

Conclusions
OSM, despite its inconsistencies, is a very useful spatial model as of November 2014, and we expect it only to improve for the foreseeable future.

OSM data is made freely available for all users, and comes without cumbersome acknowledgement requirements. Although we recommend you check the legal terms for yourself, in most cases, all that is required is a simple “© OpenStreetMap contributors” alongside any map produced from OSM.

OSM is of course a community crowd-sourced project. In the short run, those of us who use OSM for modelling purposes are likely to discover errors in the course of our work, and in the long run we can help OSM itself by reporting or fixing these. Thus we encourage users to get involved: for details, visit the link below.

\[ \text{\url{http://wiki.openstreetmap.org/wiki/Getting_Involved}} \]

3 The “oneway” field was yet more complex, with values “”, “yes”, “no”, “-1” and “yes;no”.