Generalising OpenRailwayMap to 1:10k and 1:50k

Julien Gaffuri[1]

1. Eurostat/GISCO, Luxembourg city, Luxembourg; julien.gaffuri@ec.europa.eu

Keywords: Generalisation, railway, Europe, OpenStreetMap, OpenRailwayMap, graph theory, stroke.

1. Introduction and objective

GISCO is a permanent service of the European commission in charge of coordinating GIS activities of the commission and some other European institutions. These activities include the provision of pan-European datasets to support GIS activities of different commission services on regional and urban policies, transport, energy, agriculture, environment, etc. GISCO is currently providing railway geographical data at scale 1:250k based on EuroRegionalMap[1] transport layers produced by EuroGeographics. These data are used with success for spatial analyses at European level (mainly for routing/accessibility analyses and cartographic purposes) – it shows however some limitations for analyses at more local levels due to a too coarse level of detail (Poelman et al. 2017). Complementary data sources to address this need with more detailed railway geographical data have been identified, such as OpenRailwayMap[2] (ORM). ORM describes the railway network with an impressively detailed resolution, representing each single railway track. The required level of detail is however for scales 1:10k (2m resolution) and 1:50k (10m resolution), showing train lines linking main stops, instead of single tracks and the whole railway network infrastructure elements. Generalising ORM data has already been addressed by (Touya et al. 2014; Savino et al. 2015) and encouraging results have been produced, mainly for 1:25k. Our objective is to present the results of some tests, which have confirmed the feasibility of producing pan-European (and maybe global) railway geographical data at 1:10k and 1:50k from ORM.

2. Tests and results

The input data is extracted from overpass-api.de facility selecting all OpenStreetMap features with a tag *"railway"* over Sweden. The following tags related to railways and specified in ORM schema[3] are selected: *railway, usage, service, railway:traffic_mode, name, description, gauge, railway:track_class, maxspeed, highspeed, direction, historic, bridge, bridge:name, tunnel, tunnel:name, electrified, electrified:rail, voltage, start_date, end_date, incline, operator*. The approach adopted is to progressively apply structure detection and generalisation algorithms to identify pertinent operations for 1:10k and 1:50k simplification.

Phase 1 - Filtering/selection: A quick comparison of the input ORM data with EuroRegionalMap shows the need for filtering/selecting tracks. The objective is to select railway tracks relevant for railway transport accessibility analysis, for both freight and passengers, excluding urban public transport such as subway and tram. The following selection procedure simply based on tag values is chosen: *"railway"!='construction' AND "railway"!='dismantled' AND "railway"!='elevator' AND "railway"!='funicular' AND "railway"!='historic' AND "railway"!='historic_path' AND "railway"!='historical' AND "railway" != 'miniature' AND "railway"!='planned' AND "railway" != 'platform' AND "railway" != 'platform_edge' AND "railway" != 'proposed' AND "railway" != 'razed' AND "railway" != 'turn-*

[1] http://www.eurogeographics.org/products-and-services/euroregionalmap
[2] https://www.openrailwaymap.org/
[3] https://wiki.openstreetmap.org/wiki/OpenRailwayMap

table' AND "railway" != 'abandoned' AND "railway" != 'tram' AND "railway" != 'subway' AND "service" IS NOT NULL. A limitation is due to incorrect classification of some tracks, which is sometimes due to limitations in the tagging scheme. For example, tram tracks under construction are simply tagged as *"railway"='construction'* – there is no way to retrieve the fact that these tracks are tram tracks.

Phase 2 – Connectivity analysis: The connected components of the graph structure are built and ordered by size, in order to filter isolated tracks not linked to the network. The main network component is identified, as well as other secondary networks, which are not linked to the main network physically with tracks, but only through passenger platforms. These secondary networks are connected to the main network with functional "fictive" links. For each connected component, potential nodding issues are detected with the characterisation of dangle node pairs using the ratio of their Euclidian distance over their distance through the network: Low values are obtained for potential nodding issues.

Phase 3 – Stroke analysis: Chains of continuous/aligned tracks, aka "strokes" (Thomson 2006), could be built to detect main track lines as described in (Savino et al. 2015). This construction can be based not only on the deflection angle at each node, but also on the similarity of tag values of successive tracks as suggested by (Thomson 2006). This phase has not been implemented in our test, but our level of confidence in its pertinence is high.

Phase 4 – Network face analysis for train line and service/station area generalisation: This phase is based on the faces of the network graph. These faces are often very narrow and elongated polygons, whose width is estimated with the measure proposed by (Huber, 2013). This measure has not been used so far in map generalisation despite its high potential. Considering that the distance between parallel railway tracks is usually very regular and well-known, this width estimation can be used:

1. to identify main tracks detected in phase 3 to be collapsed into main train lines. Typical generalisation algorithms such as (Thom 2005; Haunert et al. 2007; Szombara 2013) can be applied to this case.
2. to detect station/service areas as clusters of narrow and parallel faces of the network graph. These areas can alternatively be built using morphological operators such as dilatation-erosion operations on the track geometries with a distance equal to the target resolution.

The objective of the generalisation procedure should be to suppress situations where faces of the network graph are too narrow by collapsing the corresponding tracks into either a train line, or a service/station area. An analysis of the too short graph edges and too small faces has also been performed. Such small network elements should also be collapsed and not appear anymore in the target dataset.

3. Conclusion: Remaining problems and plans

These tests have shown encouraging results to reach the objective. Phase 4 is the most challenging phase and is necessary only for the 1:50k generalisation – railway tracks can indeed be kept for a 1:10k dataset. Future plans are to finalise the procedure described (especially phase 4) and apply it to the whole European territory, assuming that the conclusions made on Sweden remain valid. The intension is also to include data on stations and stops, by integrating especially external information from RINF register[4] and UIC codes[5]. The methodology and source code is intented to be shared as part of the EuroGen[6] library.

[4] http://www.era.europa.eu/Core-Activities/Interoperability/Pages/RINF.aspx
[5] https://uic.org/location-codes-enee
[6] https://github.com/eurostat/EuroGen

References

Haunert, J.-H., Sester, M., 2007. Area Collapse and Road Centerlines based on Straight Skeletons. *GeoInformatica*, 12 (2), pp. 169–191.
Huber, WA, 2013, Calculating average width of polygon? GIS, URL:gis.stackexchange.com/q/20279/664 (version: 2013-08-13).
URL gis.stackexchange.com/q/20279
Poelman, H., Ackermans, L., 2017. Passenger rail accessibility in Europe's border areas. *European commission working papers on regional and urban policy*. URL: http://ec.europa.eu/regional_policy/en/information/publications/working-papers/2017/
Savino, Sandro & Touya, Guillaume. (2015). Automatic Structure Detection and Generalization of Railway Networks. *International Cartographic Conference*. Rio de Janeiro, Brazil, August 2015.
Szombara, S., 2013. Unambiguous Collapse Operator of Digital Cartographic Generalisation. *Proceedings of 16th ICA Workshop on Generalisation and Map Production*. Dresden.
Thom, S., 2005. A strategy for collapsing OS integrated transport network dual carriageways. In: *workshop in generalisation and multiple representation*. ICA commission in map generalisation and multiple representation, la Corona, Spain.
Thomson, R., 2006. The' stroke' concept in geographic network generalization and analysis. In: Riedl, A., Kainz, W., Elmes, G. (Eds.), *Progress in Spatial Data Handling*. Springer Berlin Heidelberg, pp. 681-697. URL http://dx.doi.org/10.1007/3-540-35589-8_43
Touya, Guillaume & Girres, Jean-François. (2014). Generalising Unusual Map Themes from OpenStreetMap. *17th ICA Workshop on Generalisation and Multiple Representation*. Vienna, Austria, September 2014.