

BasqueRoads: A Benchmark for Road Network Selection

Guillaume Touya^{a,*}, Azelle Courtial^a

^a Univ Gustave Eiffel, ENSG, IGN, F-77420 Champs-sur-Marne, France, firstname.lastname@ign.fr

* Corresponding author

Keywords: Map generalisation, Benchmark, Road network, Selection

Abstract:

Road network selection is one of the major issues of map generalisation, as new papers are proposed every year since the first attempts of automation in the 1990's (Thomson & Richardson, 1995). New methods are regularly proposed because selecting roads for maps at smaller scales is a complex problem. Roads are at the same time present in maps to enable car navigation tasks, and because they are structuring elements that reveal the nature of the landscape (urban, rural, mountainous...). So road selection is not only about retaining the most important roads of the network, but the preservation of topology and connectivity is essential, as well as the preservation, or the typification of road patterns (e.g. a ring road), and the preservation of local density differences (between urban and rural areas for instance). It is rare to see comparisons of road selection techniques in the literature, because of the lack of open source in map generalisation, but also because of the lack of a common dataset to benchmark these techniques; new propositions on road selection are most of the time tied to their own dataset and use case. This is why we think that this BasqueRoads dataset could be useful to advance on this topic of road network selection.

This dataset (Figure 1) was created for an experiment on multi-scale cartography (Dumont et al., 2017). The target road network was included in a 1:80k scale topographic map. First, the roads were processed with an algorithm that differentiates urban and rural areas (Touya, 2010). In urban areas, blocks or meshes are iteratively aggregated, while rural roads are selected by importance considering strokes (Thomson & Richardson, 1999) and shortest paths. Roundabouts and branching crossroads are also collapsed into simple crossroads (Touya, 2010). Then, the results were manually edited to correct some connections errors and omitted important roads. As the goal was multi-scale cartography, manual edition also checked for inconsistencies between this output scale and other scales (1:50k and 1:100k) that are not included in the first version of this dataset. The shapefiles are projected in Lambert93 (2154). There is a field in both shapefiles, named *real_world_id* that links road lines that represent the same feature at different scales. The roads also include a *nature* field with values ranging from "unpaved" to "highway". This field can be used by road selection techniques to assess the importance of the road. An additional field called *lanes_nb* contains the number of lanes of the road section.

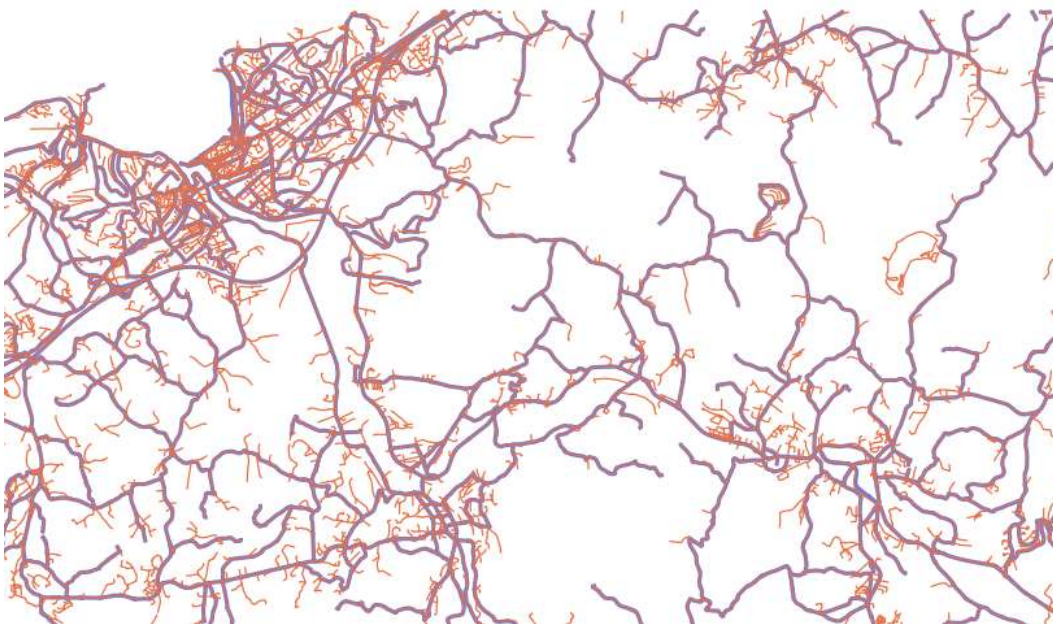


Figure 1. Extract of the BasqueRoads dataset: the final roads in blue and the initial roads in red.

The BasqueRoads dataset mainly covers a rural and slightly mountainous (in the southern part) area in the French Basque region. In the rural parts, selection is mainly about removing the small dead end roads. Then, there are a number

of villages or small towns, and a town on the coast (Saint-Jean-de-Luz) where more complex road selection occurs. There is also a highway running along the Atlantic coast with interchanges. The dataset is 28 km x 19 km, and there are 12940 road sections in the initial network and 6042 road sections in the final network.

The evaluation of the benchmark is based on the comparison between the final network and the output of the network selection algorithm being tested. Both networks can be directly compared using the *real_world_id* field. The basic evaluation can be done using the standard precision, recall, and the F_1 score measures described below where *TP* means “true positive” (i.e. the roads selected that are also in the final network), *FP* means “false positive” (i.e. the roads selected that are not in the final network), and *FN* means “false negative” (i.e. the roads that are in the final network but are not selected by the evaluated road selection algorithm). F_1 score is the harmonic mean of precision and recall, so can be used as a unique score that summarises precision and recall.

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{Recall} = \frac{TP}{TP+FN} \quad F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

However, we think that this first evaluation is not enough because it only assesses how close to the reference the selection is, but not how good it is. For instance, if all *FP* and *FN* instances are very minor roads, it is not a bad result as long as the global amount of roads is the same in the selected and the final/reference networks. We could compute weighted values of precision and recall where the weight corresponds to the importance of the road. Roads are used in maps to allow the visualization of car trips, and the preservation of road connectivity is essential for this task. This connectivity can be damaged by the selection of individual road sections. Methods have been proposed to assess if connectivity has been preserved (Touya, 2010), based on the number of connected parts in the graph, before and after selection. This measure is more meaningful if we look at subsets of the network, and we can use a sliding window and compute connectivity differences in the subgraph contained in the window. This connectivity measure could be used in addition to precision and recall to evaluate the quality of a road selection algorithm. There are also specific road structures in this dataset: roundabouts, branching crossroads, dual carriageways, interchange, rest areas... (Figure 2). Specific measures to assess how these structures are typified by algorithm could be useful, but it will be for a further version of the benchmark. The final improvement for the evaluation part of this benchmark would be to introduce a measure of the preservation of local densities, to make sure towns and cities are still visible with a higher density of roads.

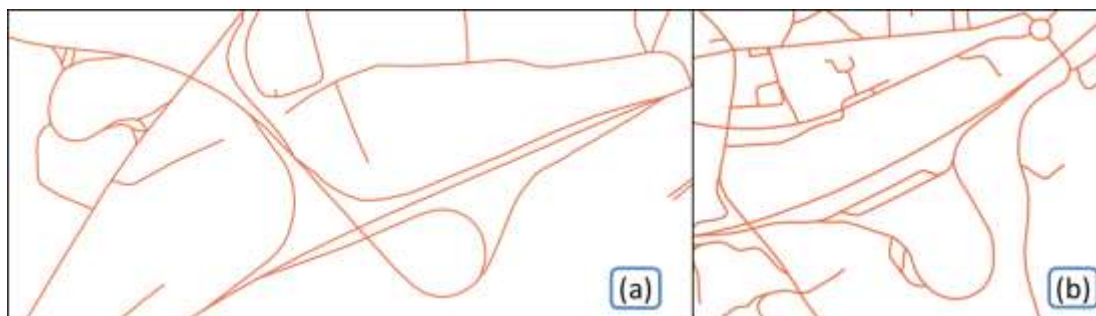


Figure 1. Road structures from the BasqueRoads dataset: (a) branching crossroads and an interchange; (b) a roundabout and a rest area.

Acknowledgement

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 101003012).

Dataset

<https://doi.org/10.5281/zenodo.5744120>

References

- Dumont, Marion, Guillaume Touya, & Cécile Duchêne. Alternative transitions between existing representations in multi-scale maps. In *Proceedings of the ICA*, 1:33. Washington, DC, USA: Copernicus Publications, 2017. <https://doi.org/10.5194/ica-proc-1-33-2018>.
- Girres, Jean-François, & Guillaume Touya. Cartographic Generalisation Aware of Multiple Representations. In *Proceedings of GIScience 2014 - Poster session*, (eds) Matt Duckham, Kathleen Stewart, Edzer Pebesma. Vienna, Austria, 2014.
- Thomson, Robert C., & Diane Richardson. A Graph Theory Approach to Road Network Generalisation. In *the Proceedings of 17th International Cartographic Conference*, 1871-80. ICA, 1995.
- Thomson, Robert C., & Diane Richardson. The “Good Continuation” principle of Perceptual Organization applied to the Generalization of Road Networks. In *the Proceedings of the 19th International Cartographic Conference*. ICA, 1999. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.202.4737>.
- Touya, Guillaume. A Road Network Selection Process Based on Data Enrichment and Structure Detection. *Transactions in GIS* 14, n° 5 (2010): 595-614. <https://doi.org/10.1111/j.1467-9671.2010.01215.x>.