Reproducible methods for network simplification for data visualisation and transport planning

Robin Lovelace[1], Zhao Wang[1], Will Deakin[2] and Josiah Parry[3]

[1]Leeds Institute for Transport Studies, University of Leeds, UK
[2]Digital, Data and Technology Services, Network Rail, UK
[3]Environmental Systems Research Institute (Esri), Redlands, CA, USA

GISRUK 2024

Summary

Route network datasets, crucial to transport models, have grown complex, leading to visualization issues and potential misinterpretations. We address this by presenting two methods for simplifying these datasets: image skeletonization and Voronoi diagram-centreline identification. We have developed two packages, the 'parenx' Python package (available on pip) and the 'rnetmatch' R package (available on GitHub) to implement these methods. The approach has applications in transportation, demonstrated by their use in the publicly available Network Planning Tool funded by Transport for Scotland.

KEYWORDS: network simplification, transport planning, urban analytics, geocomputation, reproducible research

1 Introduction

Datasets representing route networks are important in every stage of modern data-driven transport planning. Geographically, the same physical network can be represented in many different ways, ranging from simple 'centreline' representations to complex representations with multiple parallel ways. For some use cases, including strategic network planning, it is important to have a simple representation of the network.

Vector geometry simplification methods include Douglas-Peucker and Visvalingam-Whyatt algorithms (de Magalhaes et al., 2014). These methods reduce the number of vertices in a line or polygon features, but do not remove parallel ways. More sophisticated methods to help simplify complex networks include the automatic detection of 'face artifacts' (Fleischmann & Vybornova, n.d.) and removal of 'slivers' to generate simplified representations of 'street blocks' (Grippa et al., 2018). However, these methods tend to be 'all or nothing' and do not provide flexibility in terms of the level of simplification or which features are removed.

We note the simplification and interpolation for with linear (one-dimensional) geometries is less mature than Polygon or MultiPolygon (two-dimensional) geometry. For example robust implementation for the areal interpolation of intensive and extensive variables are available in `R` areal-weighted-interpolation package(Prener et al., 2019) or in the `python` PySAL Tobler library (knaap eli et al., 2023).

The aim of this paper is to present new ways to simplify transport networks, with implementations in open source software for reproducible research. The code underlying the results presented in this paper are available from the following repositories:

- The `nptscot/networkmerge` repository contains the reproducible paper.
- The `parenx` Python for image skeletonization and Voronoi diagram-centreline identification is available on PyPI in the GitHub repo `anisotropi4/parenx`.
- The `rnetmatch` R package for network simplification is available on GitHub in the repo `nptscot/rnetmatch`.

2 Problem definition

Morgan and Lovelace (2020) presented methods for combining multiple overlapping routes into a single route network with non-overlapping linestrings for visualisation, implemented in the function `overline()` in the R package `stplanr`. The approach has been used to visualise large transport networks, informing investment decisions in transport planning internationally. However, the 'overline' approach does not merge parallel ways that are part of the same corridor, resulting in outputs that are difficult to interpret, as shown in Figure 1 from the Propensity to Cycle Tool for England (PCT), with segment values representing daily commuter cycling potential flows (Lovelace et al., 2017). The left panel shows Otley Road with a flow value of 818 (Figure 1a). The right panel, by contrast, shows three parallel ways parallel to Armley Road with flow values of 515 (shown), 288 and 47 (values not shown) (Figure 1b). Although this section of Armley road has a higher cycling potential than the section of Otley Road shown ($515 + 288 + 47 > 818$), this is not clear from the visualisation.

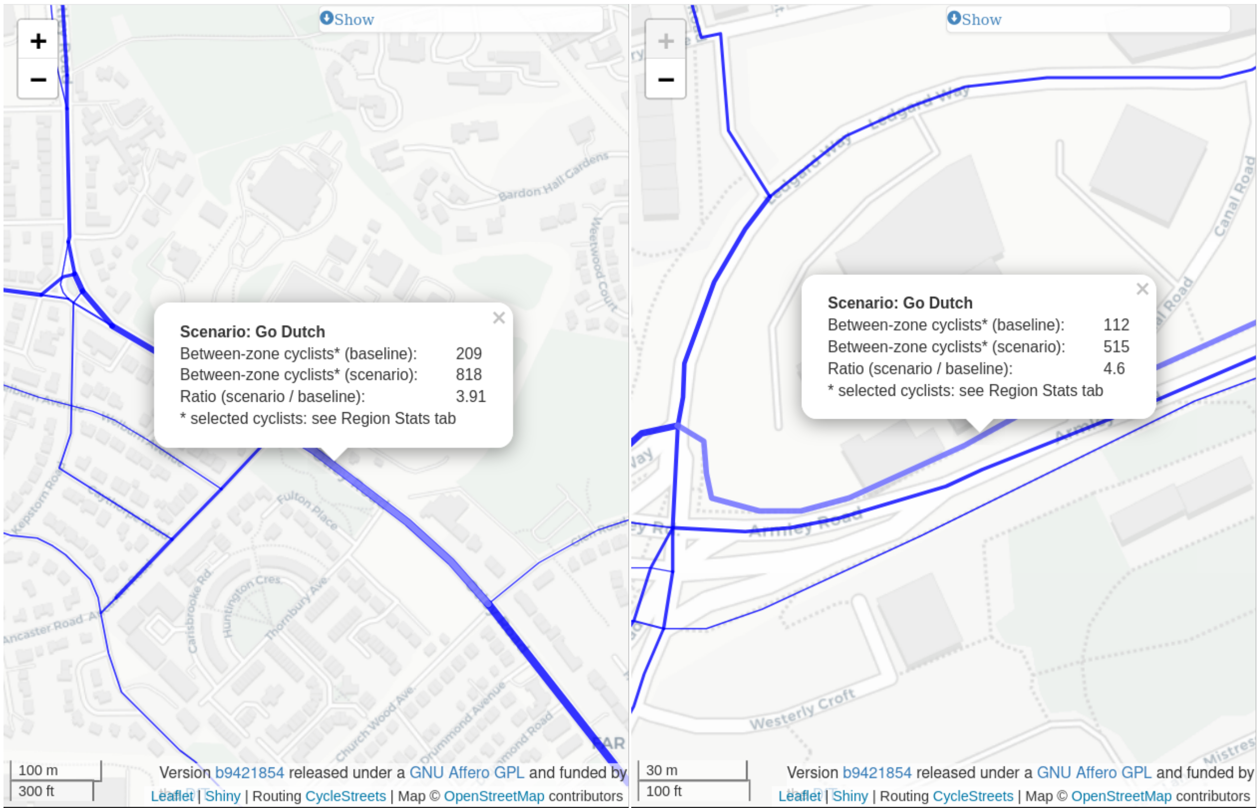
3 Methods

The key contributions of the paper are the novel methods of image skeletonization(Lee et al., 1994; van der Walt et al., 2014; Zhang & Suen, 1984), presented in Section 3.1, and simplification with Voronoi diagrams to identify central lines, covered in Section 3.2.

3.1 Simplification via skeletonization

The skeletonization approach is based on the idea of creating a buffer around the network and then skeletonizing the buffer. The buffered lines of the network are first transformed into a raster image. Subsequently, this raster image is processed through a thinning algorithm to produce a skeletal representation of the original network.

The rasterized skeletal image is then converted back into point geometry, completing the vector-to-raster-to-vector geometry transformation process, as illustrated in Figure 2.



(a) (b)

Figure 1: Illustration of issues associated with route network-level results containing multiple parallel ways on the same corridor: it is not clear from the visualisation that the corridor shown in the right hand figure has greater flow than the corridor shown in the left. Source: open access Propensity to Cycle Tool results available at www.pct.bike.

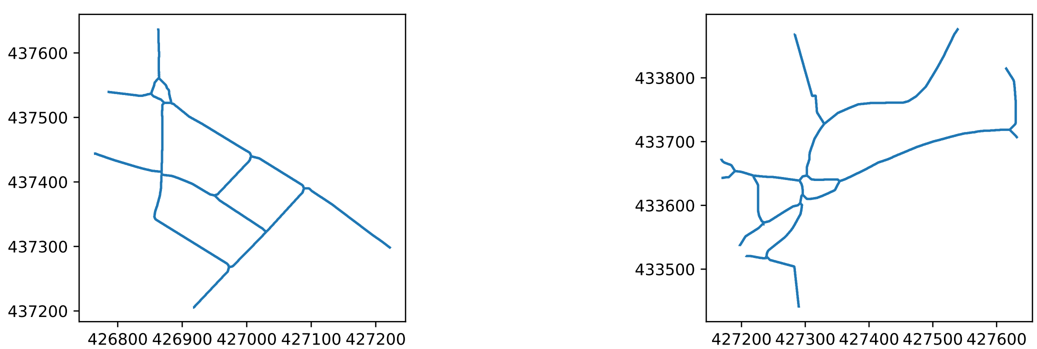


Figure 2: Simplified versions of the Otley Road (left) and Armley Road (right) networks, transformed back into line geometry.

3.2 Simplification via Voronoi polygons

In this approach, the buffers described in the previous section are converted into sequences of points. From these sequences, a centre-line is derived based on a set of Voronoi polygons with the `Shapely` library(Gillies et al., 2023) that cover these points which represents the central line of the network. This approach facilitates the creation of a simplified network representation by focusing on the central alignment of the buffered lines. The clipped Voronoi representation and the resulting central lines are illustrated in Figure 3 and Figure 4, respectively.

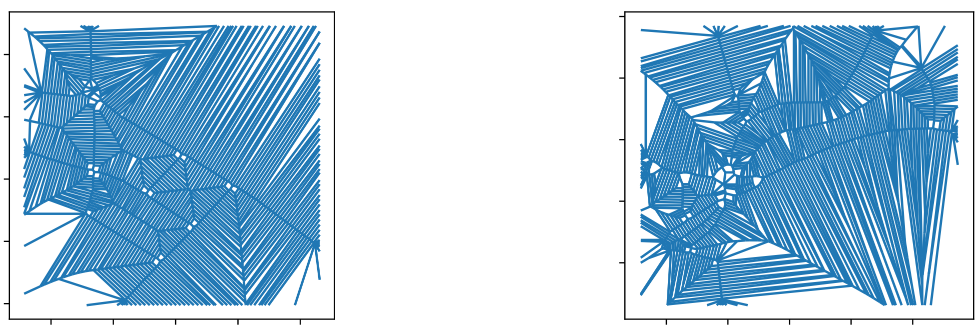


Figure 3: Clipped Voronoi diagrams of the Otley Road (left) and Armley Road (right) networks.

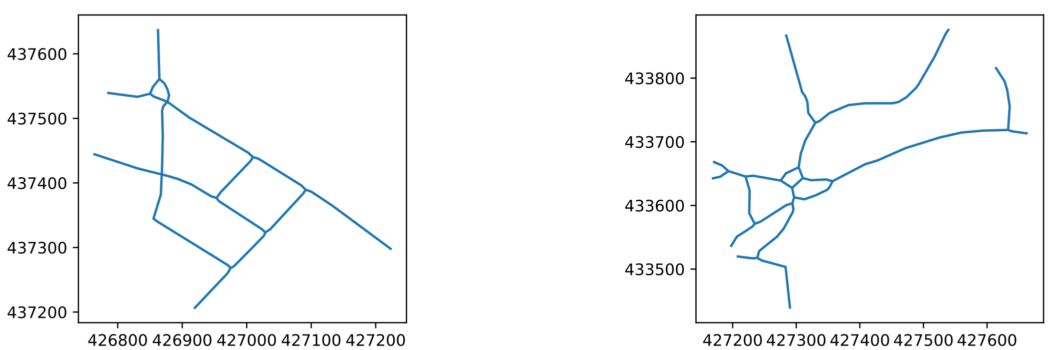


Figure 4: Simplified versions of the Otley Road (left) and Armley Road (right) networks generated by the Voronoi approach.

4 Discussion

We have demonstrated two approaches to simplify a transport network, based on 'skeletonization' and Voronoi polygons. Both approaches are based on the idea of creating a buffer around the network and then simplifying the buffer, and both offer flexibility to the user by specifying the buffer size and other parameters.

The code is open and reproducible, hosted on the nptscot GitHub organisation. The approach demonstrates the ideal outlined by Stan Openshaw that geocomputation should be used for public benefit (Openshaw & Abrahart, 2000): the methods were developed not just because they are interesting and not previously tackled in the academic literature implemented in open source software with reproducible code. They were developed in response to a request by practitioners using early versions of the Transport Scotland funded Network Planning Tool for Scotland. The potential of the approach is illustrated with the ‘simplified network’ switch (which maps the values onto the Ordnance Survey’s simplified Open Roads dataset) in the web application www.npt.scot, as illustrated in Figure 5. A next step in terms of transport applications would be to add similar network simplification toggles to web applications in settings where simplified networks are unavailable.

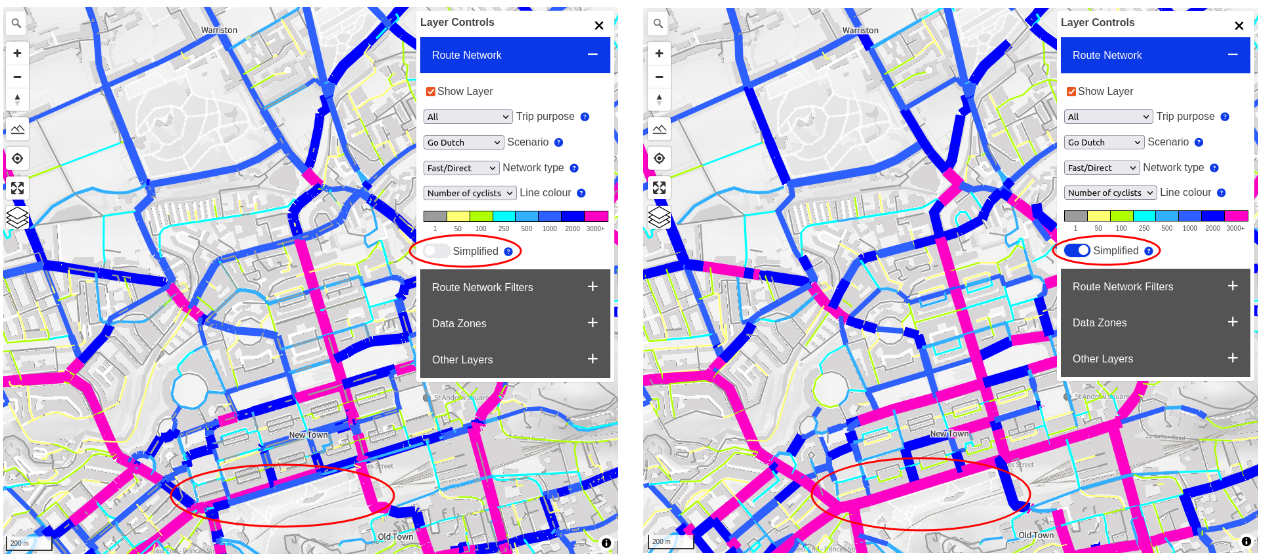


Figure 5: The Network Planning Tool for Scotland, showing the network results for central Edinburgh without simplification (left) and with simplification (right). Note that the values on Princes Street (highlighted) are hard to interpret without simplification.

A question raised by the provision of two algorithms is “which is best”. Based on our visual inspection and discussion, we sense that the Voronoi approach generates higher quality results. The skeletonization approach is more computationally efficient, but the Voronoi approach yields more accurate (in terms of true centreline) and less ‘wavy’ results. However, these are subjective judgements and we would like to test them more rigorously in future work, as outlined below.

We are confident that there will be more research and accompanying open source software to address this knotty problem. There is much more to do, including:

- Testing the impact of different parameters to find appropriate settings for different use cases
- Optimisation, noting that premature optimisation has been described as the “root of all evil”[1], but which is important for working with national or global datasets

[1]“Programmers waste enormous amounts of time thinking about, or worrying about, the speed of noncritical

- Testing the approach on other datasets, including those with different spatial scales and network types

One area where we would be particularly interested in seeing applications beyond the transport domain is river network analysis: the size of some riverine networks prohibits/slows down research. Could a network simplification pre-processing step speed-up and therefore improve results? Our hypothesis is yes, but that remains to be tested. Each of these areas represents an interesting geographical and computational challenge. More importantly, solving them has the potential to improve the quality of transport planning and policy making, demonstrating the value of geocomputation for public benefit.

5 References

de Magalhaes, S., Andrade, M., Franklin, W., & Li, W. (2014). An efficient map generalization heuristic based on the visvalingam-whyatt algorithm.

Fleischmann, M., & Vybornova, A. (n.d.). A shape-based heuristic for the detection of urban block artifacts in street networks.

Gillies, S., van der Wel, C., Van den Bossche, J., Taves, M. W., Arnott, J., Ward, B. C., et al. (2023). Shapely. https://doi.org/10.5281/zenodo.5597138

Grippa, T., Georganos, S., Zarougui, S., Bognounou, P., Diboulo, E., Forget, Y., Lennert, M., Vanhuysse, S., Mboga, N., & Wolff, E. (2018). Mapping urban land use at street block level using openstreetmap, remote sensing data, and spatial metrics [60 citations (Crossref) [2024-01-03] Number: 7 Publisher: Multidisciplinary Digital Publishing Institute]. *ISPRS International Journal of Geo-Information*, *7*(7), 246. https://doi.org/10.3390/ijgi7070246

knaap eli, e., Cortes, R. X., Rey, S., Arribas-Bel, D., Fleischmann, M., Gaboardi, J., & Frontiera, P. (2023, September 26). *Pysal/tobler: V0.11.2* [DOI: 10.5281/zenodo.8378249]. Zenodo. https://zenodo.org/records/8378249

Knuth, D. E. (1974). "structured programming with goto statements". *Computing Surveys*, *6*(4), 261–301. https://doi:10.1145/356635.356640

Lee, T.-C., Kashyap, R., & Chu, C.-N. (1994). Building skeleton models via 3-d medial surface/axis thinning algorithms. *Computer Vision, Graphics, and Image Processing*, *56*(6), 462–478.

Lovelace, R., Goodman, A., Aldred, R., Berkoff, N., Abbas, A., & Woodcock, J. (2017). The propensity to cycle tool: An open source online system for sustainable transport planning [Citation Key: lovelace_propensity_2017]. *Journal of Transport and Land Use*, *10*(1). https://doi.org/10.5198/jtlu.2016.862

Morgan, M., & Lovelace, R. (2020). Travel flow aggregation: Nationally scalable methods for interactive and online visualisation of transport behaviour at the road network level [tex.ids: morgan_travel_2020a publisher: SAGE PublicationsSage UK: London, England Citation Key: morgan_travel_2020]. *Environment and Planning B: Planning and Design*. https://doi.org/10.1177/2399808320942779

parts of their programs, and these attempts at efficiency actually have a strong negative impact when debugging and maintenance are considered. We should forget about small efficiencies, say about 97% of the time: premature optimization is the root of all evil. Yet we should not pass up our opportunities in that critical 3%" (Knuth, 1974).

Openshaw, S., & Abrahart, R. J. (2000). *Geocomputation* [Citation Key: openshaw_geocomputation_2000]. Taylor & Francis.

Prener, Christopher, Revord, & Charles. (2019). areal: An R package for areal weighted interpolation. *Journal of Open Source Software*, *4*(37). https://doi.org/10.21105/joss.01221

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., D. Warner, J., Yager, N., Gouillart, E., Yu, T., & scikit-image contributors. (2014). Scikit-image: Image processing in python. https://doi.org/10.7717/peerj.453

Zhang, T. Y., & Suen, C. Y. (1984). A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, *27*(3).

Acknowledgements

Thanks to Transport Scotland for funding the development of the Network Planning Tool for Scotland, and to the many users who have provided feedback on the tool. Thanks to Sustrans Scotland, and Matt Davis and Angus Calder in particular, for collaborating on the project.

Biographies

Robin Lovelace is Associate Professor of Transport Data Science at the Leeds Institute for Transport Studies (ITS) and Head of Data Science at the UK government agency Active Travel England. Robin specializes in data science and geocomputation, with a focus on modeling transport systems, active travel, and decarbonisation.

Zhao Wang is a researcher at the Leeds Institute for Transport Studies (ITS). Zhao specializes in machine learning, data science and geocomputation for transport planning and engineering.

Will Deakin is the Trains Portfolio Architect in IT Delivery at Network Rail, the British national rail infrastructure manager. Will is a passionate advocate of the use of data science, visualisation and open data to help support reproducible and fact based policy making.

Josiah Parry is a Senior Product Engineer at Environmental Systems Research Institute, Inc (Esri) and an open source software developer.