

Assessing methods for generating route networks from origin-destination data: jittering, routing, and visualisation

Robin Lovelace^{*1}, Rosa Félix², Dustin Carlino³ and Roger Beecham¹

¹University of Leeds

²CERIS, Instituto Superior Técnico, University of Lisbon

³Alan Turing Institute, London

January 17, 2022

Summary

Origin-destination (OD) datasets are widely available but transport interventions require network level data. OD-‘desire line’-route-‘route network’ conversion techniques are typically based on lines between zone centroids. This approach fails to show the diffuse nature of travel patterns. This paper presents ‘jittering’ methods to overcome these limitations and seeks to assess them. We find that jittered OD datasets result in a closer fit between observed and estimated flow in a reproducible case study using open data in Edinburgh. We conclude that jittering can add value to OD but work is needed to parameterise them in the context of route network generation techniques.

KEYWORDS: origin-destination data, mobility data, visualisation, transport planning.

1 Introduction

Origin-destination (OD) datasets are used to represent movement through geographic space, from an origin (O) to a destination (D). Also referred to as ‘flow data’¹ OD datasets usually contain not only information about where they start or end, but also about the amount of movement between zones (which are often represented by a zone centroid) or other geographic entities. Because of their ability to encode a large amount of information about millions of trips in a relatively small amount of storage space, with the maximum number of rows in an aggregate OD dataset equal to square of the number of zones squared, including intra-zonal OD pairs. Thus, the entire transport system of city the size of Edinburgh, with a population just over 500,000 people, can be represented at the level of desire lines between the city’s 111 enumeration districts (EDs) with 111^2 (12,321) rows and a number of columns depending on the number of trip types.²

^{*}r.lovelace@leeds.ac.uk

¹<https://www.ons.gov.uk/census/2011census/2011censusdata/originanddestinationdata>

²In practice, not all combinations of OD pairs have trips between them, so the square of the number of zones is an upper limit. The number of rows of data in the input OD dataset we use in this paper has 10,394 rows, 16% fewer than the maximum that could be represented by trips between every combination of zones. The .csv file associated

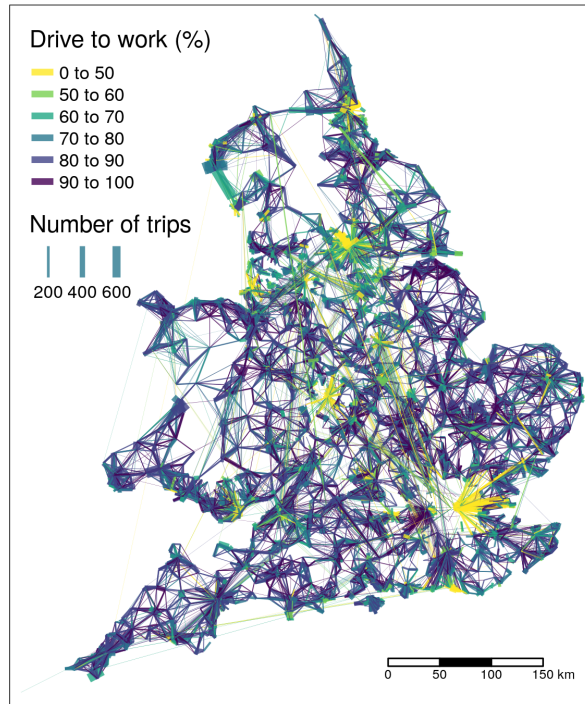


Figure 1: Illustration of typical representation of OD dataset, illustrating travel to work patterns in England. Source: author’s [analysis](<https://github.com/creds2/od-data>) of open access data from the 2011 Census.

Because of their small size and simplicity, OD datasets have long been used to describe aggregate urban mobility patterns (Carey, Hendrickson, and Siddharthan 1981). Typically, OD datasets are represented *geographically* as straight ‘desire lines’ between zone centroids, with all trips shown as departing from and arriving to a single centroid per zone, for convenience, simplicity and (historically) to reduce computational requirements. This approach, illustrated in Figure 1 below, has formed the basis of many visualisations and research projects using OD data (e.g. Rae 2009; Calabrese et al. 2011). Connecting zones with straight lines in this way has advantages: it can provide a quick summary of the movement patterns between zones, particularly when attributes such as total number of trips and the percentage of trips made by a particular mode are represented by variable aesthetics such as line width and colour, as illustrated in Figure 1.

However, the approach has limitations, including:

1. people do not travel in straight lines!
2. centroid-based desire lines obscure two-way flows (Tennekes and Chen 2021)

with this dataset representing the transport system in Edinburgh (albeit only for work and representing only single stage trips in one direction) is only 0.3 MB, a compact way of storing information on travel behaviour compared with alternatives such as large GPS datasets.

Table 1: Sample of three rows from the OD dataset used in this paper (from home and other modes not shown).

geo_code1	geo_code2	all	train	bus	car_driver	car_passenger	bicycle	foot
S02001576	S02001576	151	0	6	61	7	5	70
S02001576	S02001577	132	0	11	84	10	11	15
S02001576	S02001578	40	0	5	32	2	0	1

3. unrealistic concentration of travel around certain points

One way to add richness and realism to OD data is to convert the geographic desire lines into routes on the network and then aggregate the associated data to the route segment level to create route network outputs (Morgan and Lovelace 2020). Route network generation from centroid-based desire lines addresses limitations 1 and 2 outlined above, but not 3. Recently proposed ‘jittering’ methods seek to overcome all three limitations inherent to the centroid-based conversion of OD datasets to geographic desire lines (Lovelace, Félix, and Carlinio 2022).

The aim of this paper is to quantify, for the first time to the best of our knowledge, the impact of jittering — and different input parameters used in the jittering process — on the quality of route networks derived from OD data.

2 Data and methods

The geographic input datasets on which the analysis presented in this paper build are cleaned versions of open datasets representing the transport system in Edinburgh (see Figure 2):

- Open zones data, stored in `iz_zones11.ed`
- Open road network data from OSM, stored as `road_network.ed.geojson`

A non-geographic OD dataset representing trips between the zones was also generated from the UK National travel survey 2011 data and saved as a .csv file, the first three elements of which are presented in the table below.

To test the performance of different parameters and settings for the OD-to-route-network conversion process, we focussed only on cycling trips as these were measured in the counter dataset mentioned. Furthermore, we focussed on only the desire lines representing 4 or more cycle trips, to reduce the computational requirements of the study in the time available: this reduced the number of desire lines from over 10,000 to 685.

The jittering process was undertaken with the Rust crate `odjitter`, which can be replicated using the following reproducible code run from a system terminal such as Bash on Linux, PowerShell on Windows or the Mac Terminal (Cargo must be installed for this to work).

First install the `odjitter` Rust crate and command line tool:

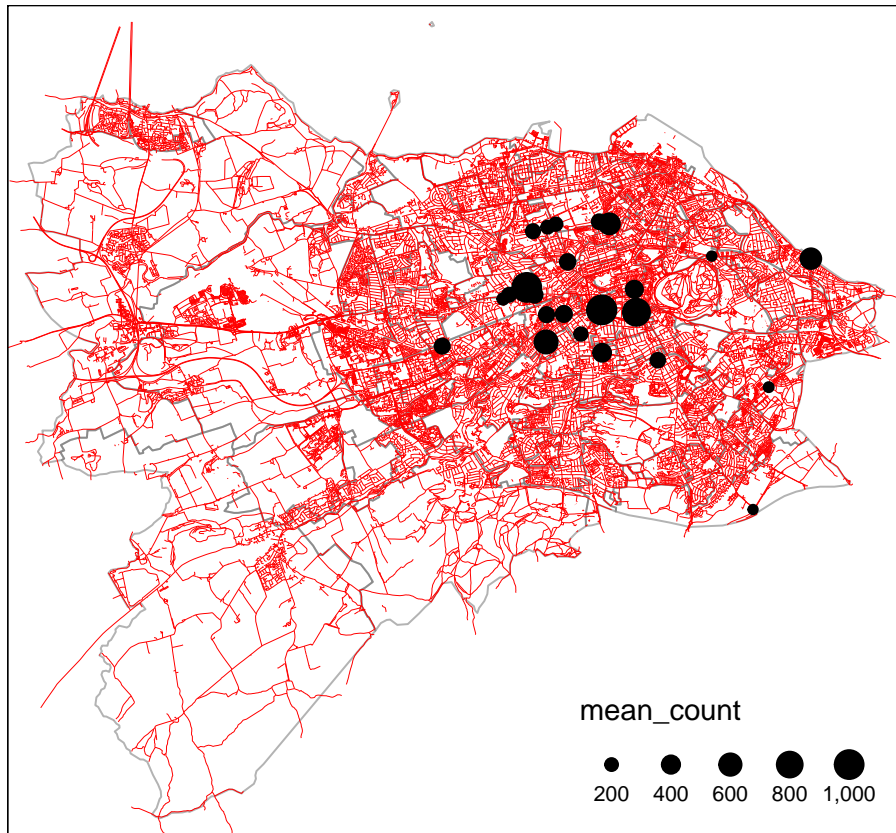


Figure 2: Overview of the study area and the input geographic datasets. Dot size is proportional to mean cycle count at counter locations.


```
cargo install --git https://github.com/dabreegster/odjitter
```

Generate jittered OD pairs with a `max-per-od` value of 50 as follows:

```
odjitter --od-csv-path od_iz_ed.csv \  
  --zones-path iz_zones11_ed.geojson \  
  --subpoints-path road_network_ed.geojson \  
  --max-per-od 50 --output-path output_max50.geojson
```

Try running it with a different `max-per-od` value (10 in the command below):

```
odjitter --od-csv-path od_iz_ed.csv \  
  --zones-path iz_zones11_ed.geojson \  
  --subpoints-path road_network_ed.geojson \  
  --max-per-od 10 --output-path output_max50.geojson
```

3 Findings

Figure 3 shows the output of the jittering process described in (Lovelace, Félix, and Carlinio 2022) and implemented with the `jitter` commands above visually. The facets a to d show the impacts of jittering with different values set for the maximum number of trips allowable per ‘jittered desire line’, and implemented with the argument `max-per-od` in the previous section. It is clear that even jittering without disaggregation has a substantial impact: Figure 3 b) represents the same number of desire lines but with widely distributed start and end points.

The maps in Figure 3 c) and d) show the impact of reducing the threshold for the number of trips allowed: although every facet represents the same number of travel, the trips are represented with around double the number of desire lines in c) and triple the number of desire lines in d) compared with the original OD dataset. To assess whether this process of disaggregation/jittering adds value to the OD data, we generated results at the route network level and evaluated them with reference to the cycle counter dataset outlined in the previous section.

The route network level results associated with the same OD pairs are shown in Figure 4. The results suggest that jittering improve network accuracy compared with a ground truth dataset, even without first disaggregating the input OD datasets. A notable finding, albeit from a small ‘ground truth’ dataset, is that jittering can improve route network results *at no no extra computational cost* in terms of routing, often the most time-consuming aspect of transport modelling workflows: networks shown in Figure 4 a) and b) both result from calculating routes associated with 685 desire lines, but b) shows a more diffuse network. Furthermore, the results suggest that disaggregation is worth extra computational costs associated with routing for more desire lines, with R-Squared results quantifying the relationship between observed daily counts and the level of flow in the nearest segment on route improving with further disaggregation.

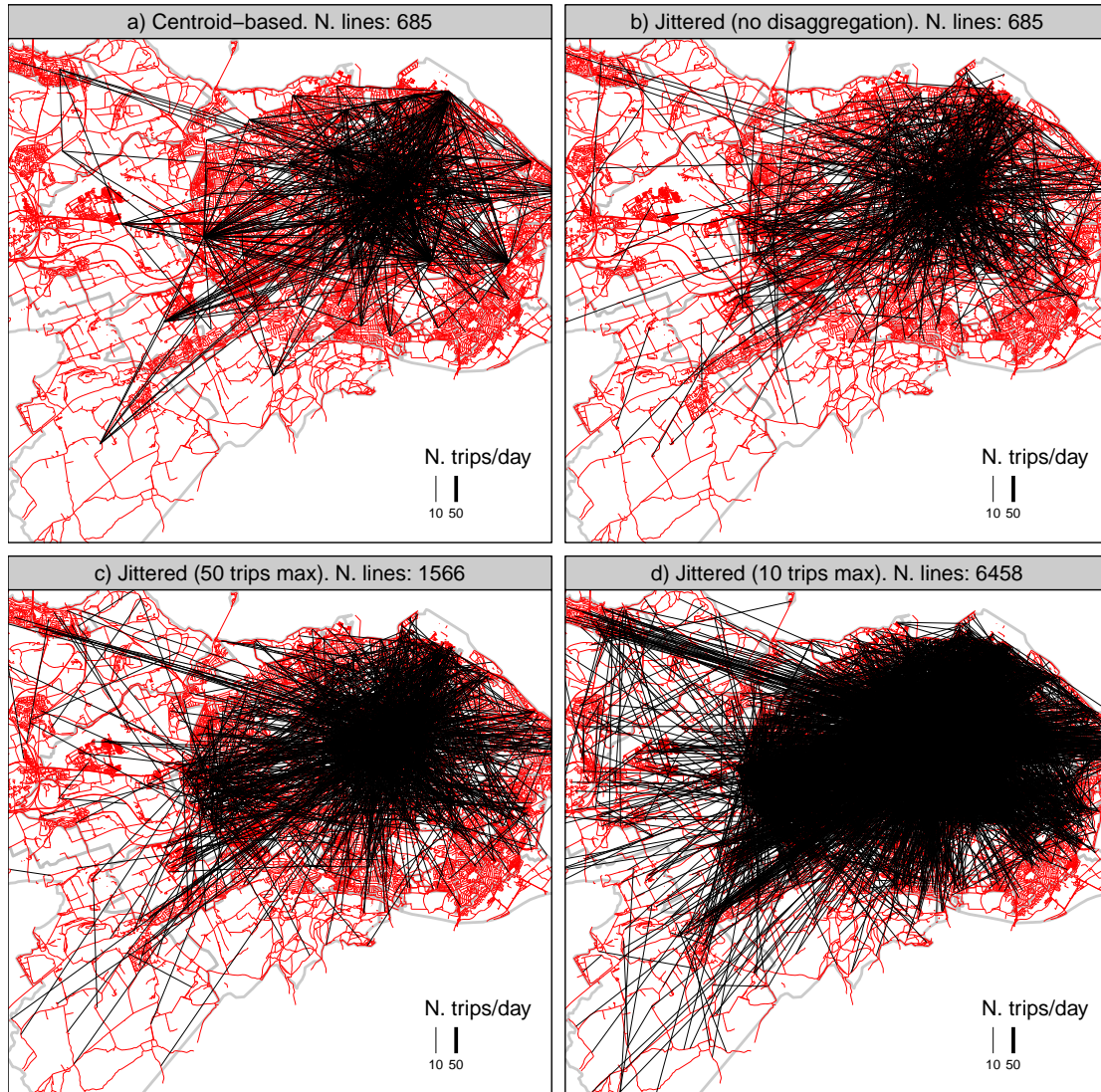


Figure 3: Results at the desire line level. The top left image showsunjittered results with origins and destinations going to zone centroids (as in many if not most visualisations of desire lines between zones). Top right: jittered results without disaggregation. Bottom left: result with a maximum number of trips per jittered OD pair of 50. Bottom right: result result with a maximum number of trips per jittered OD pair of 10.

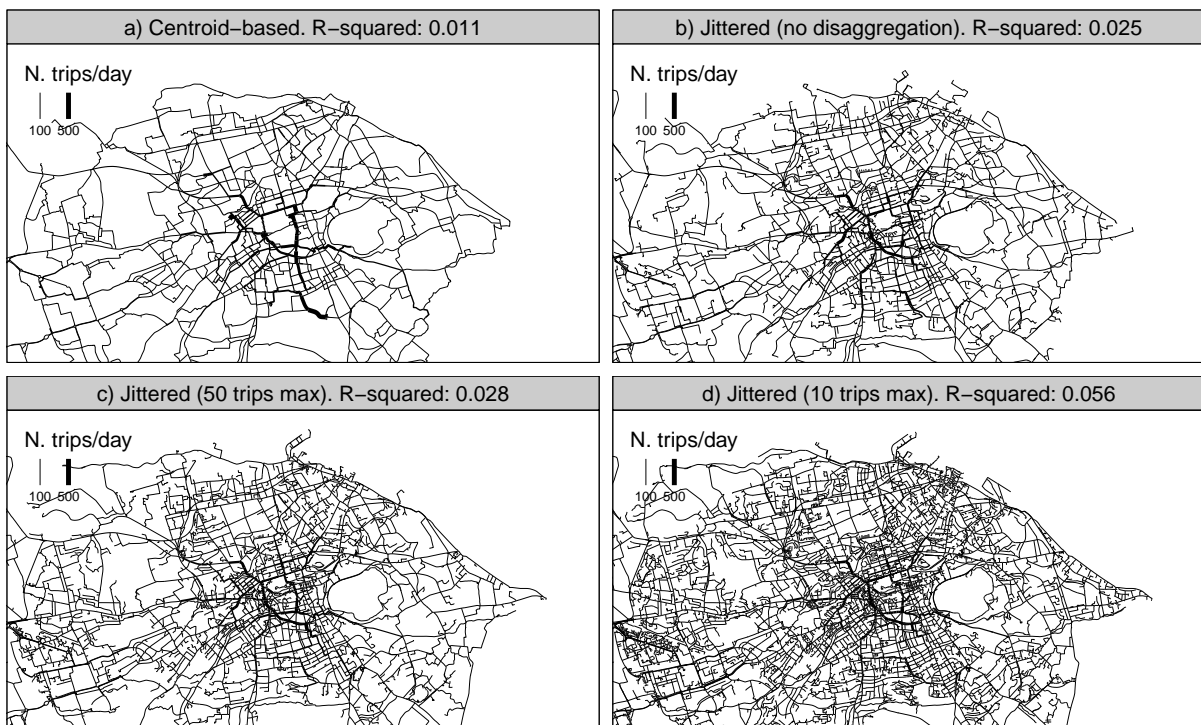


Figure 4: Route network results.

4 Discussion

The results presented above support the hypothesis that ‘jittering’ techniques outlined in (Lovelace, Félix, and Carlino 2022) can add value to OD data. The process of randomising start and end points to vertices on a transport network, combined with variable levels of disaggregation — the process of splitting single OD pairs into multiple desire lines based on a threshold value for the maximum number of trips that can be represented in a single jittered desire line — clearly leads to more diffuse networks. The progressive improvements in network-counter fit are encouraging, although we note that the best fit implies that only 6% of the variability on daily average count datasets can be explained by route network level results. Thus, the results should be seen more as a proof-of-concept highlighting the *potential* value of jittering when converting OD datasets to route networks, rather than definitive findings.

The approach is not without limitations. Despite the variability of places where the automatic bicycle counters are located, they are only 40 in number, which were used to test the method. It should also be noted that the OD data is from 2011, while the counter datasets are from 10 years later and represents cycle trips for all purposes, not just travel to work. These limitations help explain the poor counter-network results and, more importantly, suggest ways to gain further understanding of ways to improve network generation processes. Promising avenues of future research could include:

- Exploring the impact of other parameters in the OD to route network generation process, including:
 - The routing profile used, which can ‘prefer’ different route types, resulting in ‘quiet’ to ‘fast’ networks (Desjardins et al. 2021)
 - Further disaggregation levels, including full disaggregation (one desire line and route per trip) and generation of ‘centroid connectors’ (Jafari et al. 2015; Friedrich and Galster 2009)
 - Testing different jittering strategies used to sample origin and destination points within zones, such as using open building datasets to generate start and end points (Lovelace, Félix, and Carlino 2022)
- Repeating the tests outlined in this paper but with larger and richer input datasets

We believe that pursuing such lines of inquiry should be a priority for sustainable transport planning. New insights could lead to improved evidence on which to develop investment strategies, such as route networks presenting estimates of baseline levels, and potential uptake, of sustainable transport modes, improving on route network results derived from ‘centroid-based’ desire lines (e.g. Goodman et al. 2019; Lovelace et al. 2017; Biba, Curtin, and Manca 2010).³

³See the Technical Note produced by consultancy PJA for Staffordshire’s Local Cycling and Walking Infrastructure Plan for a good example of the use of centroid-based desire lines for routing in practice: <https://www.staffordshire.gov.uk/Transport/transportplanning/documents/Appendix-B-PJA-GIS-Analysis-Technical-Note.pdf>

5 Acknowledgements

This work was supported by ESRC and ADR's 10DS Fellowship funding, and the Alan Turing Institute.

This research was supported by the Portuguese Foundation for Science and Technology (FCT) with the the PARSUK Portugal-UK Bilateral Research Fund.

6 Biography

Robin is an Associate Professor of Transport Data Science working at the University of Leeds' Institute for Transport Studies (ITS) and Leeds Institute for Data Analytics (LIDA). Robin is undertaking a fellowship to support evidence-based decision making in central government in collaboration with the No. 10 Data Science Team and is an Alan Turing Fellow, specialising in transport modelling, geocomputation and data driven tools for evidence-based decision making to support policy objectives including uptake of active travel to maximise health, wellbeing and equity outcomes, and rapid decarbonisation of local, city, regional and national transport systems.

Rosa is an urban cycling and active mobility researcher at Instituto Superior Técnico - University of Lisbon, and a PhD in Transportation Systems in the MIT Portugal program. Rosa is interested in GIS for transportation, and has been working on cycling uptake in low cycling maturity cities, and socioeconomic impacts of shared mobility.

Dustin is a software engineer at the Alan Turing Institute, where he's creating an ecosystem of interoperable digital twins to study urban transportation and pandemics. He's the creator of the A/B Street transportation planning platform, and a proponent of open source code and the Rust programming language.

Roger is an Associate Professor of Visual Data Science in School of Geography, University of Leeds. His research develops, applies and evaluates visualization techniques in the analysis of large geospatial datasets. A current focus is on the use of visual techniques for supporting inference in geographical analysis.

References

- Biba, S., K. M. Curtin, and G. Manca. 2010. "A New Method for Determining the Population with Walking Access to Transit." *International Journal of Geographical Information Science* 24 (3): 347–64. <https://doi.org/10.1080/13658810802646679>.
- Calabrese, Francesco, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. 2011. "Estimating Origin-Destination Flows Using Mobile Phone Location Data." *IEEE Pervasive Computing* 10 (4): 36–44. <https://doi.org/10.1109/MPRV.2011.41>.
- Carey, Malachy, Chris Hendrickson, and Krishnaswami Siddharthan. 1981. "A Method for Direct Estimation of Origin/Destination Trip Matrices." *Transportation Science* 15 (1): 32–49. <https://doi.org/10.1287/trsc.15.1.32>.
- Desjardins, Elise, Christopher D. Higgins, Darren M. Scott, Emma Apatu, and Antonio Páez. 2021. "Correlates of Bicycling Trip Flows in Hamilton, Ontario: Fastest, Quietest, or Balanced

- Routes?" *Transportation*, June. <https://doi.org/10.1007/s11116-021-10197-1>.
- Friedrich, Markus, and Manuel Galster. 2009. "Methods for Generating Connectors in Transport Planning Models." *Transportation Research Record* 2132 (1): 133–42. <https://doi.org/10.3141/2132-15>.
- Goodman, Anna, Ilan Fridman Rojas, James Woodcock, Rachel Aldred, Nikolai Berkoff, Malcolm Morgan, Ali Abbas, and Robin Lovelace. 2019. "Scenarios of Cycling to School in England, and Associated Health and Carbon Impacts: Application of the 'Propensity to Cycle Tool'." *Journal of Transport & Health* 12 (March): 263–78. <https://doi.org/10.1016/j.jth.2019.01.008>.
- Jafari, Ehsan, Mason D. Gemar, Natalia Ruiz Juri, and Jennifer Duthie. 2015. "Investigation of Centroid Connector Placement for Advanced Traffic Assignment Models with Added Network Detail." *Transportation Research Record: Journal of the Transportation Research Board* 2498 (June): 19–26. <https://doi.org/10.3141/2498-03>.
- Lovelace, Robin, Rosa Félix, and Dustin Carlino. 2022. "Jittering: A Computationally Efficient Method for Generating Realistic Route Networks from Origin-Destination Data," January. <https://doi.org/10.31219/osf.io/qux6g>.
- Lovelace, Robin, Anna Goodman, Rachel Aldred, Nikolai Berkoff, Ali Abbas, and James Woodcock. 2017. "The Propensity to Cycle Tool: An Open Source Online System for Sustainable Transport Planning." *Journal of Transport and Land Use* 10 (1). <https://doi.org/10.5198/jtlu.2016.862>.
- Morgan, Malcolm, and Robin Lovelace. 2020. "Travel Flow Aggregation: Nationally Scalable Methods for Interactive and Online Visualisation of Transport Behaviour at the Road Network Level." *Environment & Planning B: Planning & Design*, July. <https://doi.org/10.1177/2399808320942779>.
- Rae, Alasdair. 2009. "From Spatial Interaction Data to Spatial Interaction Information? Geovisualisation and Spatial Structures of Migration from the 2001 UK Census." *Computers, Environment and Urban Systems* 33 (3): 161–78. <https://doi.org/10.1016/j.compenvurbsys.2009.01.007>.
- Tennekes, Martijn, and Min Chen. 2021. "Design Space of Origin-Destination Data Visualization." *Computer Graphics Forum* 40 (3): 323–34. <https://doi.org/10.1111/cgf.14310>.