

TRANSPORT FINDINGS

Jittering: A Computationally Efficient Method for Generating Realistic Route Networks from Origin-Destination Data

Robin Lovelace¹ , Rosa Félix² , Dustin Carlino³ ¹ Institute for Transport Studies, University of Leeds, ² CERIS, Instituto Superior Técnico, Universidade de Lisboa, ³ Alan Turing Institute

Keywords: origin-destination data, centroid connectors, routing, route networks, modeling, open source, reproducible

<https://doi.org/10.32866/001c.33873>

Findings

Origin-destination (OD) datasets are often represented as ‘desire lines’ between zone centroids. This paper presents a ‘jittering’ approach to pre-processing and conversion of OD data into geographic desire lines that (1) samples unique origin and destination locations for each OD pair, and (2) splits ‘large’ OD pairs into ‘sub-OD’ pairs. Reproducible findings, based on the open source *odjitter* Rust crate, show that route networks generated from jittered desire lines are more geographically diffuse than route networks generated by ‘unjittered’ data. We conclude that the approach is a computationally efficient and flexible way to simulate transport patterns, particularly relevant for modelling active modes. Further work is needed to validate the approach and to find optimal settings for sampling and disaggregation.

1. Questions

Origin-destination (OD) datasets are widely used in transport planning to efficiently represent aggregate travel behavior. Despite the emergence of ‘big data’ sources such as massive GPS datasets, OD data continues to play an established — if not central — role in 21st century transport planning and modelling. Recent applications range from analysis of the evolution of urban activity and shared mobility services over time (e.g. Shi et al. 2019; Li et al. 2019) to inference of congestion and mode split (Bachir et al. 2019; Gao et al. 2021).

There has been much written on optimal zoning systems for and geographic representations of OD data (e.g. Openshaw 1977; Boyce and Williams 2015). Recent papers have presented new methods for OD dataset validation (Alexander et al. 2015), aggregation (He et al. 2018; Liu et al. 2021), disaggregation (Katranchi et al. 2016) and location of ‘connectors’ joining zone center points (centroids) with the surrounding network (Jafari et al. 2015). Broadly, there are two approaches to converting OD data into geographic representations for transport modelling:

1. Centroid to centroid representations, a common approach involving the simplifying assumption that all trip destinations and origins can be represented by (sometimes population weighted or aggregated) zone centroids (Guo and Zhu 2014; Martin et al. 2018).
2. Subdividing zones (also referred to as transport analysis zones, TAZ) at which data is available to subzones centroids (Opie, Rowinski, and Spasovic 2009) or ‘centroid connectors’ or simply ‘connectors’

“between trip ends and zonal anchors” using stochastic or deterministic approaches (Leurent, Benezech, and Samadzad 2011; Friedrich and Galster 2009).

In this paper we present a new approach which, unlike established approaches which convert centroid based desire lines to routes and then route networks (Morgan and Lovelace 2020), allows the user to adjust start and end locations based on variables such as transport network density, residential density or size of commercial buildings acting as trip attractors. This ‘jittering’ approach is flexible, enabling the user to adjust the level of disaggregation, the location of start and end points from which disaggregate OD pairs are sampled, and weights representing the importance of different trip ‘originators’ and ‘attractors’.

OD data jittering can be a simple, transparent and flexible pre-processing stage that aims to represent the diffuse nature of travel patterns. This is particularly important when designing for active travel (Buehler and Dill 2016), explaining the choice of input data to illustrate the technique in this paper: it was developed in response to feedback from Edinburgh City Council (who funded a project based on the research) that route networks based on the Propensity to Cycle Tool approach (Lovelace et al. 2017) were too sparse. Jittered OD data can be used with existing transport modelling workflows developed around centroid-based methods, as the basis of route network assignment, uptake modelling, and route network generation workflows (Morgan and Lovelace 2020). We refer to the approach as jittering, which means adding random noise for data visualization (Wickham 2016).

The jittering approach presented in this paper was motivated by the following question:

How can OD data representing trips between large geographic zones be used more effectively, to generate diffuse route networks of current or potential flow to inform local interventions?

2. Methods

The approach was developed to support public sector transport planning in Edinburgh, UK. The original study area was Edinburgh City Council, a major economic hub with ambitious [plans](#) for investment in active travel, making evidence to support investment where it will be most beneficial key. For the purposes of this study we focus on a comparatively small area around central Edinburgh. We focus in this paper on walking trips in this central area because much research into route networks has focused on cycling and, because walking trips tend to be short, they create a need to convert aggregated OD datasets into diffuse route network representations of travel. Input datasets developed for this paper can be downloaded using reproducible code that accompanies the paper; see code at <https://github.com/robinlovelace/odjitter> to fully reproduce the findings.

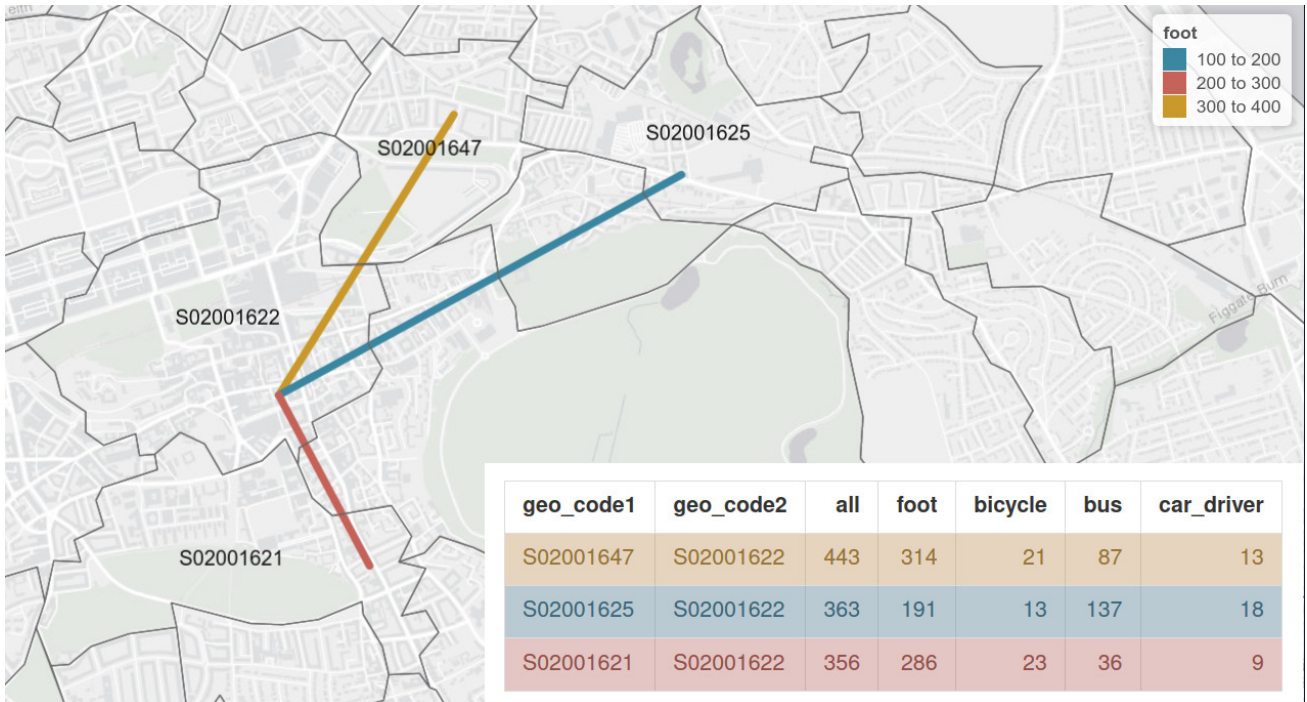


Figure 1. Illustration of input data in tabular (bottom right, inset) and geographic form (in the map).

Note how the ID codes in the first two columns of the table correspond with IDs in the zone data and how the cells in the ‘foot’ column are represented geographically on the map

Beyond the zone data illustrated in [Figure 1](#), the input dataset consisted of open access OD data from the 2011 census. The OD data can be represented as both tabular and, when start and end points are assigned to centroids within each zone, as geographic entities, as illustrated in a sample of three OD pairs presented in [Figure 1](#). To generate the route networks presented in [Figure 4](#) we used the OpenStreetMap Routing Machine (OSRM) with the profile set to ‘foot’.

The key elements of the jittering approach outlined in this paper are described in the following sub-sections, and are perhaps best understood visually, as illustrated in each of the faceted maps in [Figure 2](#). The subfigures show the flexibility of approach, with C) and D) demonstrating the use of vertices on the road network as start and end points, building on the observation from spatial network analysis that the density of the transport network is a reasonable proxy for travel demand (Cooper 2018). Other refinements including weighted subpoints could be used when data sources (e.g. building footprint areas) are available.

2.1. Sampling origin and destination points

Key to jittering is ensuring that each trip starts and ends in a different place. To do this, there must be ‘sub-points’ within each zone, one for each trip originating and departing.

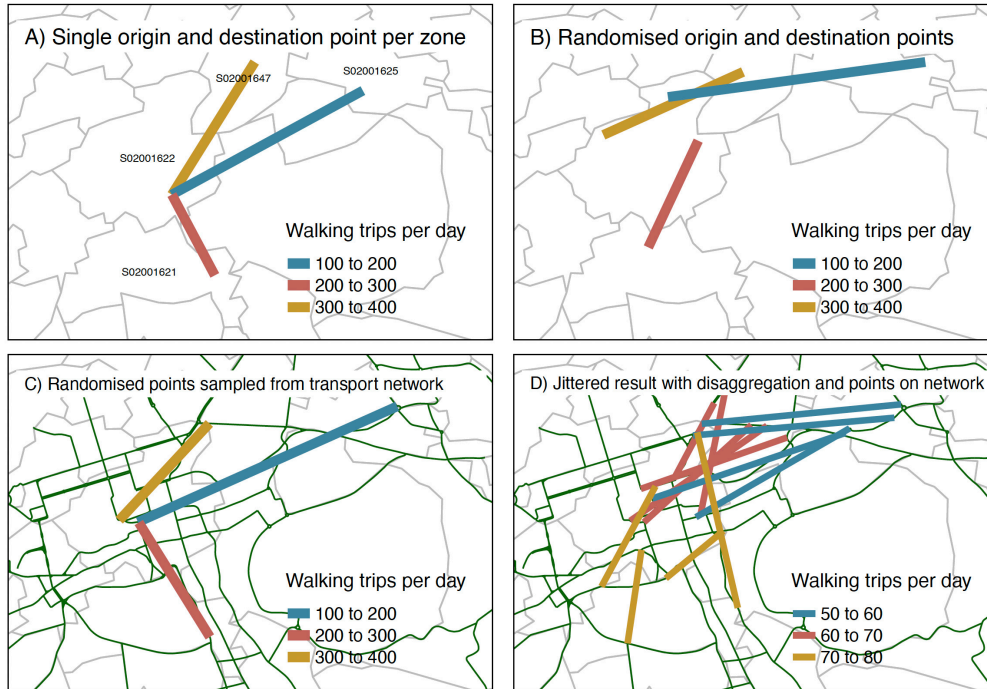


Figure 2. Illustration of jittering and disaggregation of origin-destination (OD) data with a minimal input dataset.

Subfigure A) shows the conventional way of representing OD data as desire lines between zone centroids. Subfigures B) and C) show the same desire lines but with jittered origin and destination locations based on simple random sampling of points and sampling locations on the road network. Subfigure D) shows the combined impact of disaggregation and jittering. Zone limits are represented in grey, while road network is in green.

The simplest approach is simple random spatial sampling, as illustrated in [Figure 2 \(B\)](#), which involves generating random coordinate pairs. This approach has the advantages of simplicity, requiring no additional datasets, but has the disadvantage that it may lead to unrealistic start and end points, e.g. with trips being simulated to start in rivers and in uninhabited wilderness areas.

To overcome the limitations of the simple random sampling approach, the universe of possible coordinates from which trips can originate and end can be reduced by providing another geographic input dataset. This dataset could contain known trip attractors such as city centers and work places, as well as tightly defined residential ‘subzones’. For highly disaggregated flows in cases where accurate building datasets are available, building footprints could also be used. A useful, and widely available (Barrington-Leigh and Millard-Ball 2017), input for subsampling is a transport road network, as illustrated in [Figure 2 \(C\)](#). Additional refinements to the stochastic selection of origin and destination based on weights relating to other datasets are possible, as discussed in the final section.

2.2. Disaggregation

Both of the jittering techniques outlined above generate more diffuse route networks. However, a problem with OD datasets is that they are often highly variable: one OD pair could represent 1 trip, while another could represent

Table 1. Attribute data associated with an OD pair before disaggregation.

representation	geo_code1	geo_code2	all	foot
original	S02001647	S02001622	443	314

Table 2. Attribute data associated with an OD pair after disaggregation.

representation	geo_code1	geo_code2	all	foot
disaggregated	S02001647	S02001622	88.6	62.8
disaggregated	S02001647	S02001622	88.6	62.8
disaggregated	S02001647	S02001622	88.6	62.8
disaggregated	S02001647	S02001622	88.6	62.8
disaggregated	S02001647	S02001622	88.6	62.8

1000 trips. To overcome this problem a process of disaggregation can be used, resulting in additional OD pairs within each pair of zones. The results of disaggregation are illustrated geographically in [Figure 2 \(D\)](#) and in terms of changes to attributes, in [Tables 1](#) and [2](#). As shown in those tables, updated attributes can be calculated by dividing previous trip counts by the number of OD pairs in the disaggregated representation of the data, 5 in this case. To determine how many disaggregated OD pairs each original OD pair is split into, a maximum threshold was set: an OD pairs with a total trip count exceeding this threshold (set at 100 in this case) is split into the minimum number of disaggregated OD pairs that reduce the total number of trips below the threshold.

3. Findings

We found that jittering generates desire lines, and route networks, that are more geographically diffuse than those resulting from the established centroid-based approach. [Figure 3](#) shows the use of simple random sampling and sampling nodes on transport networks with reference to a real world example. While the simple random sampling method of jittering presented in [Figure 3 \(B\)](#) may be appropriate in some specific cases, we advocate using pre-defined sub-points. Using sub-points representing vertices on the transport network, as illustrated in [Figures 3 C and D](#), is supported by ‘spatial network analysis’ (SNA) approaches to transport modelling (e.g. Cooper 2018). Weighted points representing trip origins and destinations such as houses and commercial buildings could also be used.

The results of converting the desire lines to routes and then route networks are illustrated in [Figure 4](#), which shows progressively more diffuse networks. Greater disaggregation leads to more diffuse networks as shown in [Figure 4](#).

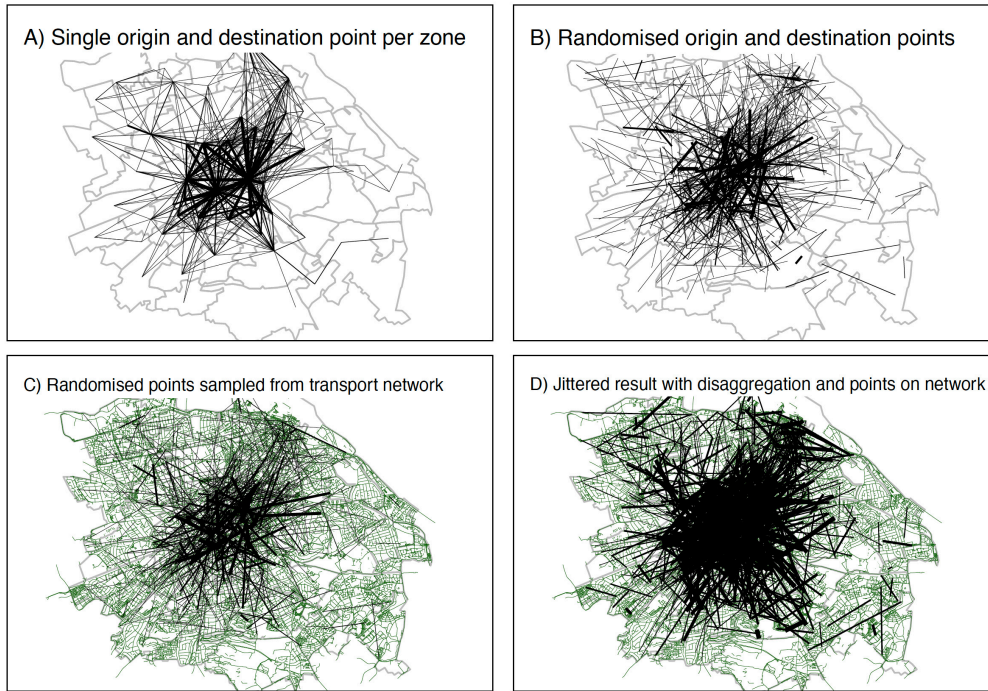


Figure 3. Results showing the conversion of OD data to geographic desire lines using population weighted centroids for origins and destinations (A) and jittered results.

The jittered results illustrate jittering with simple random sampling of origin and destination locations (B), sampling on the network (C), and sampling on the network plus disaggregation of OD pairs representing more than 100 trips (D).

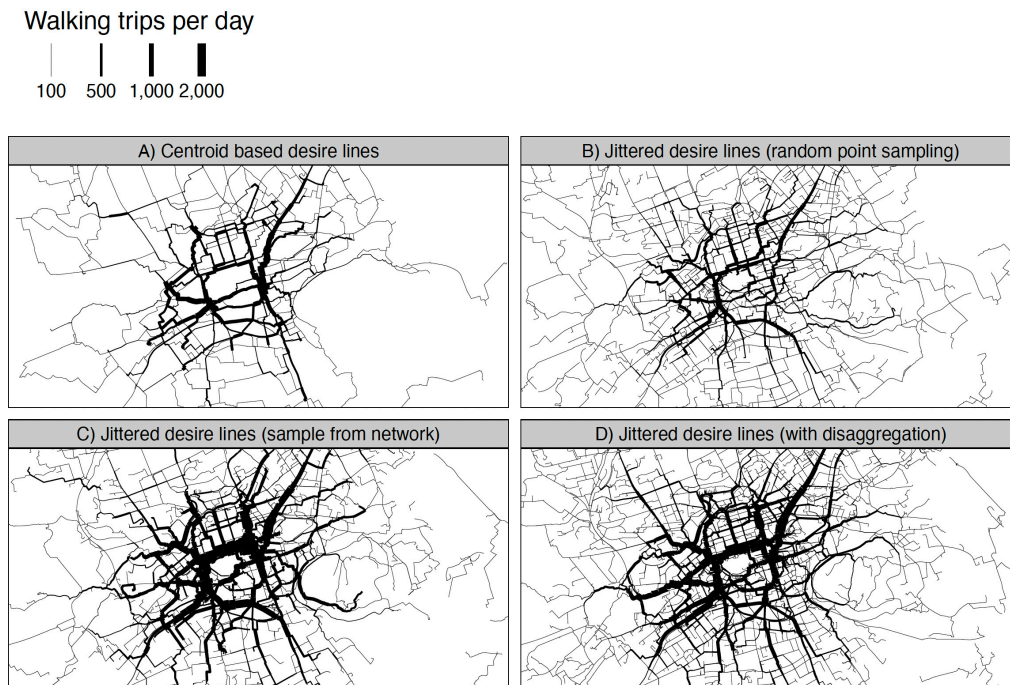


Figure 4. Route network results derived from non-jittered OD data (A) and OD data that has been jittered (B to D).

The route network results correspond to the desire lines shown in Figure 4, with start and end points sampled from: random locations in geographic space (B); nodes on the transport network network (C); and nodes on the network plus disaggregation of OD pairs representing more than 100 trips (D).

The advantages of this approach include simplicity, low computational cost and flexibility, with disaggregation (and network diffusion) levels adjusted depending on requirements. Disadvantages relate to the use of random number generators (RNG), which can reduce reproducibility (overcome this by setting a ‘seed’, which makes the findings reproducible) and influence findings (generate more than one set of results and undertake testing to mitigate this drawback). Jittering is particularly well suited to modelling walking and cycling, which require diffuse networks. Taking disaggregation further, the approach can generate one desire line per trip that could feed into agent based models (ABM) such as A/B Street and MATSim (Carlino et al. 2022; Horni, Nagel, and Axhausen 2016). Jittering has few input data requirements, enabling its use in situations where sub-zones are unavailable.

This is, to the best of our knowledge, the first time that stochastic spatial sampling and disaggregation of OD data has been described in a single approach. The approach is implemented in the open source Rust crate [odjitter](#). Implementations in R packages [od](#) and [odjitter](#), an interface to the Rust implementation, enable others to reproduce the findings, raising the possibility of interfaces to other languages.

The results also raise research questions, including:

- Are the jittered results measurably better when compared with counter datasets on the network?
- How would results from jittering OD data compare in other situations, e.g. to model motor traffic?
- Which jittering settings (including sampling strategies and levels of disaggregation) represent the best ‘boom for buck’ in terms of network accuracy relative to computational requirements?
- And can further refinements, for example sampling with weights to increase the proportion of trips associated with large buildings and commercial centers, or modifying disaggregation threshold values depending on variables such as zone size, improve results?

Before further refinements are made, we advocate empirical research to validate the jittering approach outlined in this paper as a foundation for further work on OD data pre-processing and disaggregation. Such research requires case studies that have both good open OD data and good observed travel behavior data, for example from manual and automatic counters at point locations on the network (Lindsey et al. 2013) and other sources of data such as trajectory datasets from GPS devices (Zheng et al. 2016).

.....

Acknowledgements

This work is funded by the Economic and Social Research Council (ESRC) & ADR UK (grant number ES/W004305/1) as part of the ESRC-ADR UK No.10 Data Science (10DS) fellowship in collaboration with 10DS and ONS (who funded Robin Lovelace).

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

We would like to acknowledge the GeoRust community (<https://georust.org>) for developing core dependencies and implementation advice.

Submitted: December 23, 2021 AEST, Accepted: March 26, 2022 AEST



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-SA-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-sa/4.0> and legal code at <https://creativecommons.org/licenses/by-sa/4.0/legalcode> for more information.

REFERENCES

- Alexander, Lauren, Shan Jiang, Mikel Murga, and Marta C. González. 2015. "Validation of Origin-Destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data." *Transportation Research Part B: Methodological*, 1–20. <https://doi.org/10.1016/j.trc.2015.02.018>.
- Bachir, Danya, Ghazaleh Khodabandelou, Vincent Gauthier, Mounim El Yacoubi, and Jakob Puchinger. 2019. "Inferring Dynamic Origin-Destination Flows by Transport Mode Using Mobile Phone Data." *Transportation Research Part C: Emerging Technologies* 101 (April): 254–75. <https://doi.org/10.1016/j.trc.2019.02.013>.
- Barrington-Leigh, Christopher, and Adam Millard-Ball. 2017. "The World's User-Generated Road Map Is More Than 80% Complete." *PLOS ONE* 12 (8): e0180698. <https://doi.org/10.1371/journal.pone.0180698>.
- Boyce, David E., and Huw C.W.L. Williams. 2015. *Forecasting Urban Travel: Past, Present and Future*. Edward Elgar Publishing. <https://doi.org/10.4337/9781784713591>.
- Buehler, Ralph, and Jennifer Dill. 2016. "Bikeway Networks: A Review of Effects on Cycling." *Transport Reviews* 36 (1): 9–27. <https://doi.org/10.1080/01441647.2015.1069908>.
- Carlino, Dustin, Michael Kirk, Dcarlino, Mateusz Konieczny, Gedalia Kott, Bruce, Javed Nissar, et al. 2022. *A/B Street*. Zenodo. <https://doi.org/10.5281/ZENODO.6331922>.
- Cooper, Crispin H. V. 2018. "Predictive Spatial Network Analysis for High-Resolution Transport Modeling, Applied to Cyclist Flows, Mode Choice, and Targeting Investment." *International Journal of Sustainable Transportation* 12 (10): 1–11. <https://doi.org/10.1080/15568318.2018.1432730>.
- 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>.
- Gao, Hong, Zhenjun Yan, Xu Hu, Zhaoyuan Yu, Wen Luo, Linwang Yuan, and Jiyi Zhang. 2021. "A Method for Exploring and Analyzing Spatiotemporal Patterns of Traffic Congestion in Expressway Networks Based on Origin–Destination Data." *ISPRS International Journal of Geo-Information* 10 (5): 288. <https://doi.org/10.3390/ijgi10050288>.
- Guo, Diansheng, and Xi Zhu. 2014. "Origin-Destination Flow Data Smoothing and Mapping." *IEEE Transactions on Visualization and Computer Graphics* 20 (12): 2043–52. <https://doi.org/10.1109/TVCG.>
- He, Biao, Yan Zhang, Yu Chen, and Zhihui Gu. 2018. "A Simple Line Clustering Method for Spatial Analysis with Origin-Destination Data and Its Application to Bike-Sharing Movement Data." *ISPRS International Journal of Geo-Information* 7 (6): 203. <https://doi.org/10.3390/ijgi7060203>.
- Horni, Andreas, Kai Nagel, and Kay W. Axhausen, eds. 2016. *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press. <https://doi.org/10.5334/baw>.
- 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 (1): 19–26. <https://doi.org/10.3141/2498-03>.
- Katranji, Mehdi, Etienne Thuillier, Sami Kraiem, Laurent Moalic, and Fouad Hadj Sem. 2016. "Mobility Data Disaggregation: A Transfer Learning Approach." *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, November, 1672–77. <https://doi.org/10.1109/itsc.2016.7795783>.

- Leurent, Fabien, Vincent Benezech, and Mahdi Samadzad. 2011. "A Stochastic Model of Trip End Disaggregation in Traffic Assignment to a Transportation Network." *Procedia - Social and Behavioral Sciences* 20: 485–94. <https://doi.org/10.1016/j.sbspro.2011.08.055>.
- Li, Haojie, Yingheng Zhang, Hongliang Ding, and Gang Ren. 2019. "Effects of Dockless Bike-Sharing Systems on the Usage of the London Cycle Hire." *Transportation Research Part A: Policy and Practice* 130 (December): 398–411. <https://doi.org/10.1016/j.tra.2019.09.050>.
- Lindsey, Greg, Steve Hankey, Xize Wang, and Junzhou Chen. 2013. "The Minnesota Bicycle and Pedestrian Counting Initiative: Methodologies for Non-Motorized Traffic Monitoring." *Minnesota Department of Transportation*. <https://www.lrrb.org/media/reports/201324.pdf>.
- Liu, Qiliang, Jie Yang, Min Deng, Ci Song, and Wenkai Liu. 2021. "SNN_flow: A Shared Nearest-Neighbor-Based Clustering Method for Inhomogeneous Origin-Destination Flows." *International Journal of Geographical Information Science*, 1–27.
- 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>.
- Martin, David, Christopher Gale, Samantha Cockings, and Andrew Harfoot. 2018. "Origin-Destination Geodemographics for Analysis of Travel to Work Flows." *Computers, Environment and Urban Systems* 67 (January): 68–79. <https://doi.org/10.1016/j.compenvurbsys.2017.09.002>.
- 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* 48 (6): 1684–96. <https://doi.org/10.1177/2399808320942779>.
- Openshaw, S. 1977. "Optimal Zoning Systems for Spatial Interaction Models." *Environment and Planning A: Economy and Space* 9 (2): 169–84. <https://doi.org/10.1068/a090169>.
- Opie, Keir, Jakub Rowinski, and Lazar N. Spasovic. 2009. "Commodity-Specific Disaggregation of 2002 Freight Analysis Framework Data to County Level in New Jersey." *Transportation Research Record* 2121 (1): 128–34. <https://doi.org/10.3141/2121-14>.
- Shi, Xiaoying, Fanshun Lv, Dewen Seng, Baixi Xing, and Jing Chen. 2019. "Exploring the Evolutionary Patterns of Urban Activity Areas Based on Origin-Destination Data." *IEEE Access* 7: 20416–31. <https://doi.org/10.1109/access.2019.2897070>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. New York, NY: Springer.
- Zheng, Xinhua, Wei Chen, Pu Wang, Dayong Shen, Songhang Chen, Xiao Wang, Qingpeng Zhang, and Liuqing Yang. 2016. "Big Data for Social Transportation." *IEEE Transactions on Intelligent Transportation Systems* 17 (3): 620–30. <https://doi.org/10.1109/tits.2015.2480157>.

SUPPLEMENTARY MATERIALS

ZIP of Supplementary Information

Download: <https://findingspress.org/article/33873-jittering-a-computationally-efficient-method-for-generating-realistic-route-networks-from-origin-destination-data/attachment/86187.zip>
