

# A Global Analysis of Bus Route Structure

Ed Manley<sup>\*1,2</sup>, Steven Gray<sup>3</sup> and Nick Malleson<sup>1,2</sup>

<sup>1</sup>School of Geography, University of Leeds

<sup>2</sup>Leeds Institute for Data Analytics, University of Leeds

<sup>3</sup>Centre for Advanced Spatial Analysis (CASA), University College London

March 7, 2020

## Summary

Bus networks serve a vital purpose in urban public transport, and ensuring accessibility to the bus service is an important transport planning duty for any city. However, we currently lack systematic approach to bus network analysis, that accounts for varying street network structures and city size. This paper introduces a methodology for a global analysis of bus network design. It uses General Transit Feed Specification (GTFS) data with freely available Open Street Map road network data to compare the features of the bus routes in different cities worldwide. Through clustering we are able to extract common trends in bus route design. Initial results demonstrate commonalities and differences between cities, and suggest that route network data holds promise as a means of estimating the quality of bus service provision.

**KEYWORDS:** Public Transportation; Bus Networks; Accessibility; Network Entropy; Clustering.

## 1 Introduction

The adequate provision of public transport, and local bus services in particular, has long been recognised as vital for a healthy society. For example, in the UK there is evidence that in many places people without cars face difficulties accessing essential services such as doctors' surgeries (Lovett et al., 2002) and shops (Guy et al., 2004; Lucas, 2006). These difficulties have been exacerbated by trends towards a greater reliance on car use at the expense of publicly-funded transport infrastructure, particularly in the 1970s and 1980s (Higgs and White, 1997), causing some people, particularly with low incomes in rural communities (Lowe et al., 1986), to face social exclusion (Pooley, 2016). In many places it has become "virtually impossible" to perform normal daily activities without access to a car (Lucas et al., 2001) and this situation will only get worse as the age structures in industrialised countries moves towards a larger proportion of elderly people who (for a number of reasons) will not use cars (Lucas, 2006).

---

<sup>\*</sup>e.j.manley@leeds.ac.uk

Previous attempts to measure accessibility and the impact of public transport provision include the use of the Public Transport Accessibility Level (PTAL) approach (Wu and Hine, 2003) and similar indices (Currie, 2010; Delbosc and Currie, 2011) as well as equity measures such as the Gini coefficient and Lorenz curve (Delbosc and Currie, 2011) and nuanced calculations of the connectivity of each transit stop to the wider network (Welch and Mishra, 2013). However, due to the difficulties in collecting, cleaning and analysing comparative data from large and diverse administrative areas, most previous studies focus solely on a sub-national regions. Although such regional analyses undoubtedly provide useful insights for their target areas, they prohibit analysis of wider (i.e. national) government policy or comparisons of different local authority regimes.

This preliminary study, instead, begins to use data from a number of different cities across the globe. Although the data are less rich than those that are often available at the regional level, the ability to compare public transport provision across different cities, let alone different countries, offers exciting opportunities for comparing how well different cities and countries serve their residents. This research collects data on bus routes from a total of 95 cities in 19 countries and calculates a number of different measures that characterise the structures of the bus bus routes. It then clusters the individual routes according to their similarity and explores the differences across cities. The results reveal high global variation in provision of different service types, primarily based on the proportions of ‘Express’ routes designed into the network. Although the metrics currently preclude an overall assessment of *accessibility*, the results highlight that the data and methods hold considerable potential for further global assessments of public service provision. Furthermore, by using publicly-available data and open source code, the work is entirely reproducible.

This paper has been structured as follows. Section 2 outlines the data sources used, Section 3 presents the details of the analysis and clustering results, and Section 4 draws conclusions.

## 2 Data Sources

The data was collected using a GTFS data service provided by OpenMobilityData ([transitfeeds.com](https://transitfeeds.com)) which provides an updated API of various public transit data from transit authorities worldwide. The service allows authenticated users, by way of an authentication key, to fetch data programmatically via the API and directly download the GTFS data that is collated by the service. In many cities there are multiple providers of transit data and as data is updated by these transit authorities on a regular basis a reproducible workflow has been created to fetch data from OpenMobility-

Country	Number of routes
USA	4601
Australia	1980
Canada	1789
Italy	853
New Zealand	777
Russia	709
Finland	575
Hungary	365
Poland	356
Austria	354
Czechia	328
Spain	282
Lithuania	182
France	175
Germany	172
Mexico	119

Table 1: The number of routes per country included in the analysis

Data to collect the data into one large collection for analysis.

The workflow consists of various interconnected scripts which fetch and process the individual zip files for each transit authority for all the cities in the OpenMobilityData collection and creates an index for these downloaded files. Each file has an associated checksum created within the workflow so that in future iterations of the workflow the data can be refreshed if there is an update to the source GTFS data. At this stage each zip file is unpacked, checked for validity and cleaned if any errors are found within the source file. The cleaned data, line by line is, inserted into a master file which after processing all of the collection is, in turn, inserted into a database for later analysis. This workflow results with a final collection of 3,337 agencies, 12.3 million routes consisting of 1.6 million stops.

After data cleaning and removing circular routes, the analysis covers 95 cities in 19 countries. As might be expected, there is an over-representation of countries in the Global North and as illustrated in Table 1.

### 3 Bus Route Clustering

After the GTFS data were downloaded, a number of metrics were calculated to describe the different characteristics of each bus route. Table 2 outlines these features. All parameters measures are normalised by characteristics of the city or route structure. This ensures that we control for varying city size and street network arrangement, which naturally impact bus route design. These normalised measures should more closely reflect the relative route design preferences imposed by transport planners in each city.

Table 2: Features of the bus routes used in the classification. All features are normalised by city or route features.

Feature	Description
Route Length	The length of the bus route, normalised by the maximum route length for all routes in the city that the route is part of
Stop Distance	Number of stops on route, normalised by route length
Sinuosity	Measure of actual the route length divided by the Euclidean distance between origin and destination
Curvature	Sum angular deviation of route normalised by relative entropy of city street network (Boeing, 2019)
Turns	The number of major turns (greater than 60 degrees), normalised by route length

K-Means clustering is a commonly used approach and is appropriate here as a means of clustering the bus routes according to the similarities of the features outlined in Table 2. Before running the analysis, however, an appropriate number of clusters used must be determined. If there are too many clusters then similar routes will not be grouped together, whereas using too few clusters risks creating overlapping clusters. The Silhouette Coefficient,  $S$ , can be used to determine the most appropriate number of clusters.  $S$  is a measure of the mean distances that each bus route is from its

constituent cluster centroid. Values of 1 indicate that all routes are close to their assigned cluster and values of 0 suggest that there are overlapping clusters.

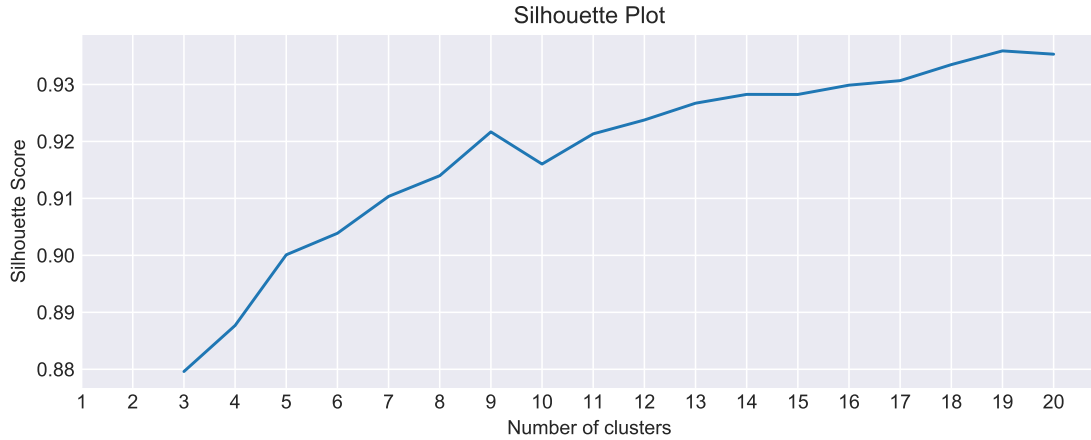


Figure 1: The Silhouette score with varying number of clusters.

Figure 1 illustrates how the *silhouette score* varies as the number of clusters is incremented up to 20. Although there is no completely clear choice, approximately 5 clusters is appropriate because the Silhouette score begins to plateau. With 5 clusters the score of  $S \approx 0.9$  is close to 1 which indicates high differentiation. Figure 2 illustrates how the values of each of the factors in Table 2 varies by each cluster.

The next stage in the research will be to begin to label and further analyse the clusters. In particular, it will be interesting to begin to compare different city structures in different countries to see whether the local bus routes share similar characteristics. To this end, Figure 3 shows, for a subset of all the cities, what proportion of the routes in the city correspond to each cluster. These patterns are potentially interesting. For example, some European cities such as Belgium, Austria and Italy contain routes that predominantly belong to Cluster 0. This cluster is notable for having relatively short routes with typical sinuosity. At the other end of the scale, cities such as Auckland, Washington DC, Toulouse and Budapest have very few Cluster 0 routes but many more Cluster 2 routes. Cluster 2 routes have very large numbers of stops and large numbers of turns compared to the other clusters. Like Cluster 0, Cluster 3 routes are common throughout our set of cities, being the longer and straighter routes with few stops.

While this initial analysis is not yet sufficiently advanced to make any firm statements about the approaches taken by different national or regional jurisdictions to their bus transport planning, it provides a useful basis for international comparison and measuring the effect of bus route design on outcomes.

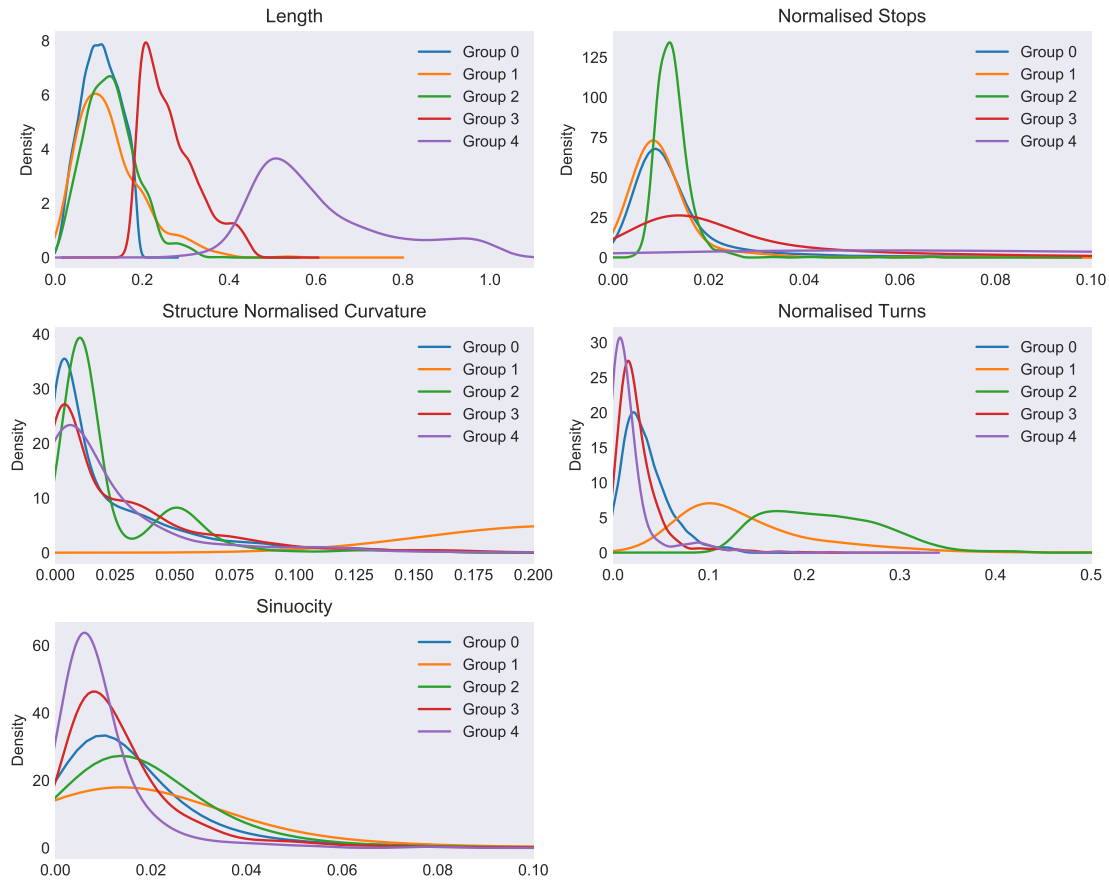


Figure 2: Cluster profiles.

## 4 Conclusions

This paper has demonstrated how General Transit Feed Specification (GTFS) data, that are published globally by a growing number of national and regional governments, can be combined with Open Street Map data to allow international comparisons of the structures of different bus routes. Clusters of routes were created and some preliminary analysis of these was conducted. Ultimately these data hold promise as a means of estimating the quality of bus service provision in global cities.

## Biographies

**Ed Manley** is Professor of Urban Analytics in the School of Geography, University of Leeds, and Turing Fellow at the Alan Turing Institute for Data Science and Artificial Intelligence. Ed's research aims to deepen our quantitative understanding of human behaviour in cities, and how these behaviours shape urban dynamics. He is author of the book 'Agent-based Modelling and Geo-

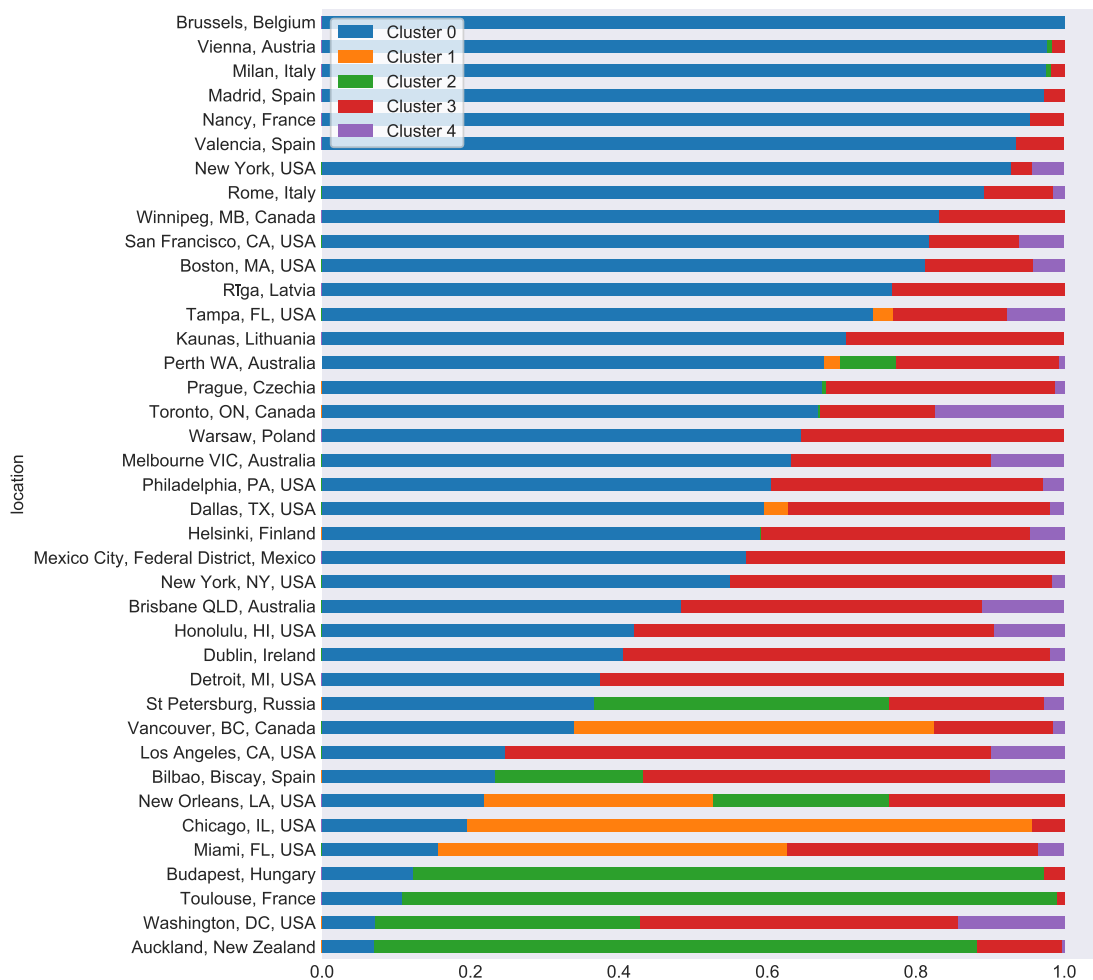


Figure 3: Proportions of different route types within a subset of cities

graphical Information Science’, recently published by Sage, Associate Editor of the Applied Spatial Analysis and Policy journal, and chairs the GIScience Research Group at the Royal Geographical Society.

**Steven Gray** is a Principal Teaching Fellow and a Spatial Software Engineer in the Centre for Advanced Spatial Analysis, University College London. As a Computing Scientist at heart, Steve’s research focuses on Large scale data collection, distributed systems (within the cloud) and data analysis, Human Computer Interaction, mobile development,, Web based Mapping and Ubiquitous Computing. With over 20 years’ professional software development under his belt, he has built multiple award-winning systems and his work has been featured in various worldwide media outlets.

**Nick Malleson** is a Professor of Spatial Science at the Centre for Spatial Analysis and Policy at

the School of Geography, University of Leeds, UK. He has a PhD in Geography and undergraduate degrees in Computer Science (BSc) and Multidisciplinary Informatics (MSc). Most of his research focuses on the development of computer models that help to understand and explain social phenomena. He has a particular interest in simulations of crime patterns, and in models that can be used to describe the flows of people around cities. More recently, he has become interested in how ‘big’ data, agent-based modelling, and smart cities initiatives can be used to better understand the daily dynamics of cities and reduce the impacts of phenomena such as pollution or crime.

## References

- Boeing, G. (2019). Urban spatial order: Street network orientation, configuration, and entropy. *Applied Network Science*, 4(1):67.
- Currie, G. (2010). Quantifying spatial gaps in public transport supply based on social needs. *Journal of Transport Geography*, 18(1):31–41.
- Delbosc, A. and Currie, G. (2011). Using Lorenz curves to assess public transport equity. *Journal of Transport Geography*, 19(6):1252–1259.
- Guy, C., Clarke, G., and Eyre, H. (2004). Food retail change and the growth of food deserts: A case study of Cardiff. *International Journal of Retail & Distribution Management*, 32(2):72–88.
- Higgs, G. and White, S. (1997). Changes in service provision in rural areas. Part 1: The use of GIS in analysing accessibility to services in rural deprivation research. *Journal of Rural Studies*, 13(4):441–450.
- Lovett, A., Haynes, R., Sünnerberg, G., and Gale, S. (2002). Car travel time and accessibility by bus to general practitioner services: A study using patient registers and GIS. *Social Science & Medicine*, 55(1):97–111.
- Lowe, P., Bradley, T., and Wright, S., editors (1986). *Deprivation and Welfare in Rural Areas*. Geo Books, Norwich, UK.
- Lucas, K. (2006). Providing transport for social inclusion within a framework for environmental justice in the UK. *Transportation Research Part A: Policy and Practice*, 40(10):801–809.
- Lucas, K., Grosvenor, T., and Simpson, R. (2001). *Transport, the Environment and Social Exclusion*. York Pub., Layerthorpe. OCLC: 48929200.
- Pooley, C. (2016). Mobility, Transport and Social Inclusion: Lessons from History. *Social Inclusion*, 4(3):100.
- Welch, T. F. and Mishra, S. (2013). A measure of equity for public transit connectivity. *Journal of Transport Geography*, 33:29–41.
- Wu, B. M. and Hine, J. P. (2003). A PTAL approach to measuring changes in bus service accessibility. *Transport Policy*, 10(4):307–320.