2024-01-31

Simple Open Data Measures of Public Transit Service Availability

Temporal Variability

Emily C. Wilke
35xxxxxxxx
emilycwilke@gmail.com
Ruprecht-Karls-Universität Heidelberg
xxxxxxx

Affidavits

Abstracts

English

German

Contents

1 Introduction

In recent years, but for decades by now, the demand for a paradigm shift in transportation infrastructure and service has become louder and louder. While calls for a shift away from car centric mobility are nothing new and were a well established part of German Academic discourse in the 1990s already (Holzapfel, 1993), it has become part of a widespread political discourse around the so called *Verkehrswende* (Holzapfel, 2020). With increased awareness and concrete experiences of climate change this discourse has reached states of heated debate. Benefits of

1.1 Transit Accesibility Equity and Equality

1.1.1 Terminology

1.1.2 Motivation

- Traditional transport planning centering on men?
  - German Transport Planning post world war 2?
- Transit planning and identifying demand in public transit networks is a complicated process, that takes into account a plethora of data that's hard to access or acquire (Pieper, 2021).
  - statistical routing data based on conveyal engine (Conway et al., 2017)

1.1.3 Research Question

- How temporal variability in transit accessibility maps on to spatial usage patterns?
- are cities really better connected? edge times?
- are rural areas always worse off in connections or are there times when it's actual actions?

1.2 Related Work

- Network Centrality Measures
  - road networks
  - public transit networks
  - bipartite networks
- Transit Equity Studies
  - US
  - Network Planning (Pieper, 2021)
- Traveltime Datasets such as (Tenkanen & Toivonen, 2020) and (Verduzco Torres & McArthur, 2024)

1.3 Methodological Approach

1.3.1 Data Acquisition

- explorative data analysis

1.3.2 origins

- hexgrids from h3pandas (Dahn, 2023) based on uber's implementation of them

1.3.2.1 Transport Data

- osm files from geofabrik (Geofabrik GmbH, 2018)
- gtfs files from various transit companies (DELFI, 2023; Rhein-Neckar-Verkehr GmbH, 2023; VRS, 2023; VVS, 2023).

1.3.2.2 Population Data

1.3.3 Destinations

- Usage of openly available data, preferably from osm .. extracted with pyrosm (Tenkanen, 2023)
- specific data if necessary, eg secondary school data not mapped in osm (Ministerium für Schule und Bildung NRW, 2016)

1.3.4 Case Studies

- Selected based on data availability, personal familiarity.

1.3.4.1 secondary schools

- see (Verduzco Torres & McArthur, 2024)
- data from (Ministerium für Schule und Bildung NRW, 2016)

1.3.4.2 sports clubs

- osm data

1.3.4.3 hexgrid cells

- h3 pandas (Dahn, 2023)

2 Transit Reach

2.1 Measures of Reach

2.1.1 Isochrones as a Measure of Reach

- ors (HeiGIT, 2023)
- cummulative or individual accesibility measures from (Verduzco Torres & McArthur, 2024)

2.1.2 Mean Travel Time

2.2 Temporal Variability

- conveyal approach (Conway et al., 2018)
  - also used in (Verduzco Torres & McArthur, 2024) for metrics spanning the UK, but identified gap in temporal variability of transport choices
- automatic clustering using u-map, pca and k-means

2.3 Processing

2.3.1 Travel Matrices

- enough for basic reach analyses, isochrone itself not important
- calculated with r5py (Fink et al., 2022) as used in (Tenkanen & Toivonen, 2020), based on the conveyal engine (Conway et al., 2017; 2018)

2.3.2 clustering

- Dimensionality reduction PCA or UMAP (McInnes, 2018) based on the maths from (McInnes et al., 2020)
- Clustering K-Means or HDBSCAN (McInnes et al., 2016) based on an algorithm proposed by (Campello et al., 2013)

2.4 Results

3 Transit Access and Planning

3.1 Conveyal Percentiles

- see (Verduzco Torres & McArthur, 2024)

3.2 Processing

3.3 Results

4 Results

5 Discussion

5.1 General Limitations

- Lack of real world measures as Comparisons
- Focuses solely on door to door travel times and neglects
  - reliability and delay Data
  - public transit fare structures (Conway & Stewart, 2019)
- lacks data including
  - comparisons to cars
  - ride hailing services see (Barajas & Brown, 2021)
  - related on demand services (trial at rohrbach)
- *inequality* being silly at times (Graeber & Wengrow, 2022).

5.2 Methodological short commings

- UMAP clustering prone to confabulations (2018; Schubert, 2017).

6 Final Remarks

6.1 Conclusion

6.2 Outlook

Bibliography

(2018, June). https://stats.stackexchange.com/q/263539

Barajas, J. M., & Brown, A. (2021). Not minding the gap: Does ride-hailing serve transit deserts?. *Journal of Transport Geography*, *90*, 102918. https://doi.org/10.1016/j.jtrangeo.2020.102918

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 160–172). Springer. https://doi.org/10.1007/978-3-642-37456-2_14

Conway, M. W., & Stewart, A. F. (2019). Getting Charlie off the MTA: a multiobjective optimization method to account for cost constraints in public transit accessibility metrics. *International Journal of Geographical Information Science*, *33*(9), 1759–1787. https://doi.org/10.1080/13658816.2019.1605075

Conway, M. W., Byrd, A., & Eggermond, M. van. (2018). Accounting for uncertainty and variation in accessibility metrics for public transport sketch planning. *Journal of Transport and Land Use*, *11*(1). https://doi.org/10.5198/jtlu.2018.1074

Conway, M. W., Byrd, A., & Linden, M. van der. (2017). Evidence-Based Transit and Land Use Sketch Planning Using Interactive Accessibility Methods on Combined Schedule and Headway-Based Networks. *Transportation Research Record*, *2653*(1), 45–53. https://doi.org/10.3141/2653-06

Dahn, J. (2023, November). *h3pandas: Integration of H3 and GeoPandas*. https://github.com/DahnJ/H3-Pandas

DELFI. (2023, December). *Deutschlandweite Sollfahrplandaten (GTFS)*. https://www.opendata-oepnv.de/ht/de/organisation/delfi/startseite?tx_vrrkit_view%5Baction%5D=details&tx_vrrkit_view%5Bcontroller%5D=View&tx_vrrkit_view%5Bdataset_formats%5D%5B0%5D=ZIP&tx_vrrkit_view%5Bdataset_name%5D=deutschlandweite-sollfahrplandaten-gtfs&cHash=01414d5793fcd0abb0f3a2e35176752c

Fink, C., Klumpenhouwer, W., Saraiva, M., Pereira, R., & Tenkanen, H. (2022, September). *r5py: Rapid Realistic Routing with R5 in Python*. Zenodo. https://doi.org/10.5281/zenodo.7060438

Geofabrik GmbH. (2018). *Geofabrik Download Server*. http://download.geofabrik.de/

Graeber, D., & Wengrow, D. (2022). *The Dawn of Everything. A New History of Humanity*. Penguin Books.

HeiGIT. (2023). *Openrouteservice API*. Heidelberg Institute of Geoinformation Technology. https://openrouteservice.org/

Holzapfel, H. (1993). Hat das Auto in der Stadt noch etwas zu suchen?. In H.-J. Frank & N. Walter (Eds.), *Strategien gegen den Verkehrsinfarkt* (pp. 63–80). Schäffer-Poeschel.

Holzapfel, H. (2020). *Urbanismus und Verkehr: Beitrag zu einem Paradigmenwechsel in der Mobilitätsorganisation*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-29587-5

McInnes, L. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. https://umap-learn.readthedocs.io/en/latest/index.html

McInnes, L., Healy, J., & Astels, S. (2016). *The hdbscan Clustering Library — hdbscan 0.8.1 documentation*. https://hdbscan.readthedocs.io/en/latest/index.html

McInnes, L., Healy, J., & Melville, J. (2020, September). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. arXiv. https://doi.org/10.48550/arXiv.1802.03426

Ministerium für Schule und Bildung NRW. (2016, January). *Grunddaten der Schulen und Schulaufsicht in NRW*. https://www.schulministerium.nrw.de/BiPo/OpenData/Schuldaten/schuldaten.csv

Pieper, F. (2021). Der Kreislauf der Aufgaben – Leistungsplanung und Leistungserstellung im Betrieb. In VDV-Akademie (Ed.), *Grundwissen Personenverkehr und Mobilität* (pp. 234–283). GRT Global Rail Academy and Media GmbH.

Rhein-Neckar-Verkehr GmbH. (2023). *Aktueller GTFS*. https://www.opendata-oepnv.de/ht/de/organisation/verkehrsunternehmen/rnv/openrnv/datensaetze?id=1405&tx_vrrkit_view[dataset_name]=soll-fahrplandaten-rnv&tx_vrrkit_view[action]=details&tx_vrrkit_view[controller]=View

Schubert, E. (2017, February). *Answer to "Clustering on the output of t-SNE"*. https://stats.stackexchange.com/a/264647

Tenkanen, H. (2023, October). *pyrosm*. https://pyrosm.readthedocs.io/en/latest/index.html

Tenkanen, H., & Toivonen, T. (2020). Longitudinal spatial dataset on travel times and distances by different travel modes in Helsinki Region. *Scientific Data*, *7*(1), 77. https://doi.org/10.1038/s41597-020-0413-y

Verduzco Torres, J. R., & McArthur, D. P. (2024). Public transport accessibility indicators to urban and regional services in Great Britain. *Scientific Data*, *11*(1), 53. https://doi.org/10.1038/s41597-023-02890-w

VRS. (2023, January). *Soll-Fahrplandaten VRS*. opendata-oepnv.de. https://www.opendata-oepnv.de/ht/de/organisation/verkehrsverbuende/vrs/startseite?tx_vrrkit_view%5Baction%5D=details&tx_vrrkit_view%5Bcontroller%5D=View&tx_vrrkit_view%5Bdataset_formats%5D%5B0%5D=ZIP&tx_vrrkit_view%5Bdataset_name%5D=soll-fahrplandaten-vrs&cHash=5db6d1227f9456ffed0b252688eadec0

VVS. (2023, December). *Soll-Fahrplandaten VVS 2024 Jahresfahrplan*. https://www.opendata-oepnv.de/ht/de/organisation/verkehrsverbuende/vvs/startseite?tx_vrrkit_view%5Baction%5D=details&tx_vrrkit_view%5Bcontroller%5D=View&tx_vrrkit_view%5Bdataset_formats%5D%5B0%5D=ZIP&tx_vrrkit_view%5Bdataset_name%5D=soll-fahrplandaten-vvs&cHash=77fbc8e1cfc3643518ca99625acb8ff1