2024-03-21

Simple Open Data Measures of Public Transit Service Availability

Planning and Temporal Variability

Emily C. Wilke

35xxxxxxxxx

emilycwilke@gmail.com

Ruprecht-Karls-Universität Heidelberg

xxxxxxx

Affidavits

Abstracts

English

test oh my oh my

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 [1], it has become part of a widespread political discourse around the so called *Verkehrswende* [2]. With increased awareness and concrete experiences of climate change this discourse has reached states of heated debate. Taunted benefits of a modal shift away from individual motorised transit towards public transit include a more efficient transport system both in terms of traffic throughput as well as in terms of energy usage, decreased local pollution [3], and a lessened dependency on cars favouring more healthy modes of transit and more livable environments for humans [4], [5], [2].

However appealing these pleas for an improved mobility regime may be, there is a distinct lack of easily accessible tools to measure how well a public transit system is serving its users, and the people who for some reason or another have decided that they will not use it. Public Transit Data beyond planned schedules is often hard to come by, and the often quite complex thoughts and models that go into service planning remain behind the closed doors of local transit agencies in a lot of cases [6], [7]. One option of addressing this gap is the use of reach data over larger areas to assess the coverage public transit provides for reaching different points of interest hopefully relevant to daily life of potential users e.g. [8]. These studies are based in schedule data and as in this example several assumptions about public transit usage. One of these assumptions is usually a time of day for travel time calculations, usually something like morning rush hour.

As general traffic volumes change over the course of the day the, so do public transit schedules. This temporal variability in transit accessibility is so far not generally addressed well in the literature and in available data sets [8]. So I shall make an attempt at addressing this gap.

1.1 Transit Accesibility Equity and Equality

To address this gap as identified above it is paramount to first define, what is even meant by terminology like public transit accessibility or reach. Then it is necessary to expand on the short motivation I have given above to open up the questions i seek to answer in this thesis. Furthermore, I will give a almost certainly incomplete overview of the landscape of literature and research surrounding this topic.

1.1.1 Terminology

- There are various different, and branching understandings of *access* and *accessibility*.

-
- Especially accesibility often used in the sense of the disabled community [9]
- general measures of access play an important role in empirical studies of transportation systems. There *access*, as opposed to *mobility*, measures the ease of reaching valued destinations [9].

Specifically in North American literature, spurred by rules of the Federal Transport Administration on transit accessibility studies, there seem to be two general modes of thinking about transit access *Access to Transit* and *Access with Transit* [10]. Access to transit means that transit itself becomes the target destination of such an access analysis. And while access to transit measures can be as simple as the average proximity to transit stops in an area. These measures however fall short in explaining the mobility decisions in day to day life as they fail to consider desirable destinations [7].

- This paper takes a geographical view into *access* and uses it interchangeably with *accessibility*.
- This paper also tries to consider access with transit above access to transit.
- In the end such analysis aims to show gaps and inequalities in usable access to the transportation system.
- Looking specifically at transit time as an impedance or cost factor and it's variability over time.

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 [6].
  - statistical routing data based on conveyal engine [11]

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 [6]
- Traveltime Datasets such as [12] and [8]

1.3 Methodological Approach

- explorative data analysis based on openly available and openly licensed data, with open source software tools

1.3.1 Data Acquisition

1.3.2 origins

- hexgrids from h3pandas [13] based on uber's implementation of them

1.3.2.1 Transport Data

- osm files from geofabrik [14] downloadad with [15]
- gtfs files from various transit companies [16]–[18] but ultimately settled on cropping the german weekly transit dataset from DELFI [19] using gtfs-general.
- Content descriptions for all of these and their usage

1.3.2.2 Population Data

1.3.3 Destinations

- no point of interests due to complexity
- h3 pandas [13]
- h3 cell to h3 cells with populations excluded or not

1.3.4 Case Studies

- Selected based on data availability, personal familiarity: Heidelberg

2 Transit Access

2.1 Measures of Access

There is plenty of ways to operationalise access with transit then.

2.1.1 Isochrones as a Measure of Reach

An easy tool to asses a basic measure of access are isochrone maps. They are comprised of an area that can be reached within a certain time, where the actual isochrone delimits this area. The isochrone is the line that denotes equal (greek: *iso*) time (greek: *chronos*) and would transform into a circle in a coordinate system transformed by travel time. This can be generalised to a access measure: Destination opportunities outside this circle are considered reachable and counted, opportunities without are left out. This measure then encapsulates "all possible destinations within [a] travel cost threshold" [9]. That then is what [8] would call an cummulative accesibility measure. A modern tool that provides such isochrone maps is for example ors [20]. Using Isochrones as an easy to understand and interpretable measure of access in geoinformation systems has a much longer history however [21], [22], [9].

Besides cummulative there's also individual accessibility measures that are relevant for locations where not the access to a variety of locations is important, but the access to a specific location is of interest. This can be true for services that are common but not generally interchangeable either due to restrictions on their use or scarcity. For these measures the minimum travel cost is of relevance [8].

2.1.2 Mean Travel Time

Despite these measures not being mathematically complex this research approaches the question from a different point of view and tries to adapt a measure originally adopted in network centrality analysis for the measure of reach.

- active measure of reach, that is I measure how easy or hard it may be to move from one cell to another, as oppossed how easy it is for a cell to be reached [9].
- travel time as a common cost measure for transit accessibility

2.2 Temporal Variability

Transit access then however depends on temporal aspects as well, both because different destinations offer various time constraints as well as the transport network changing over the course of the day [9].

2.3 Processing

2.3.1 Travel Matrices

- enough for basic reach analyses, isochrone itself not important
- calculated with r5py [23] as used in [12], based on the conveyal engine [11], [24]

2.3.2 Temporal Variability

- conveyal approach [24]
- also used in [8] for metrics spanning the UK, but identified gap in temporal variability of transport choices
- automatic clustering using u-map, pca and k-means

2.3.3 Clustering

- manually by neighbourhoods
- manually by time chunks

Optional bits:

- Dimensionality reduction PCA or UMAP [25] based on the maths from [26]
- UMAP clustering prone to confabulations [27], [28].
- Clustering K-Means or HDBSCAN [29] based on an algorithm proposed by [30]

2.3.4 Detailed Views

2.4 Results

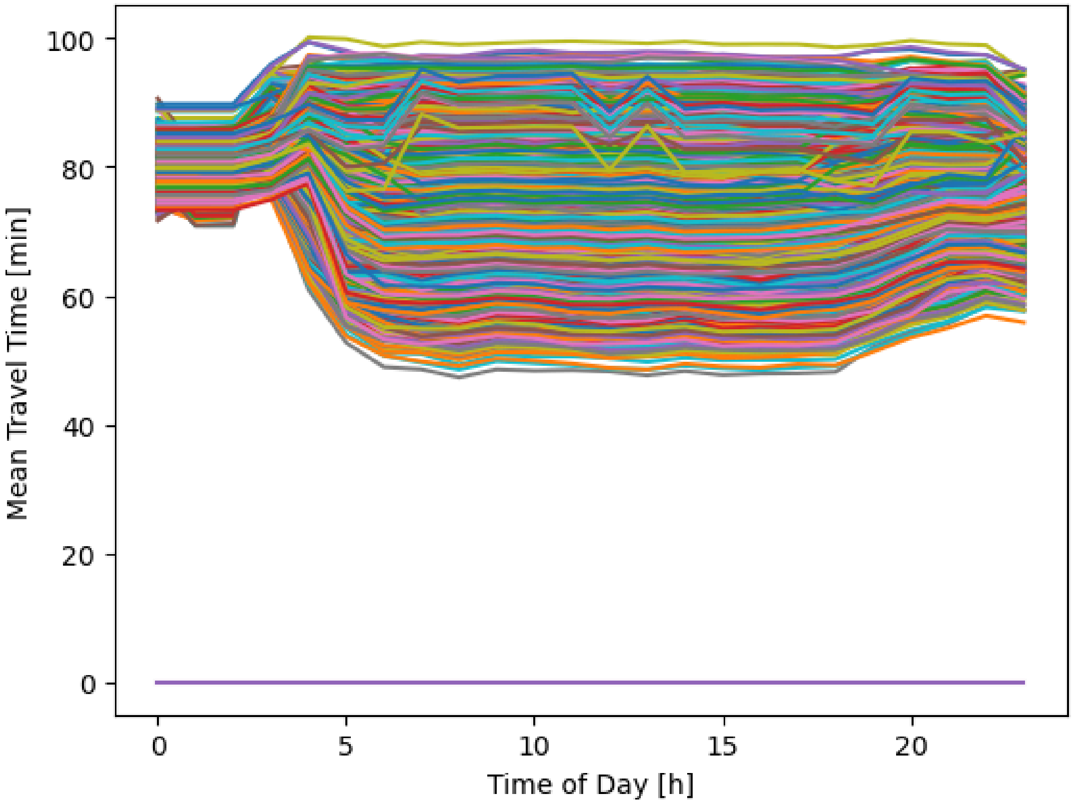

Figure 1: Plot of average travel times in Heidelberg from cell to cell with no population mask, over the course of a weekday.

- Night time travel in Figure 1 shows a consistent travel time spread. Around 4 am in the morning this fans out however, and while a majority of cells have faster connections a small part actually have longer average median travel times.
- Outlier one cell on the eastern station throat of RH with no connections to any other cells and therefore an average travel time of 0 minutes.

3 Transit Access and Planning

3.1 Motivation

- see [8]

3.2 Processing

- essentially the same processing as for mean travel times Section 2.3 but taking the difference between the 90th and 10th percentile of r5py travel times.

3.3 Results

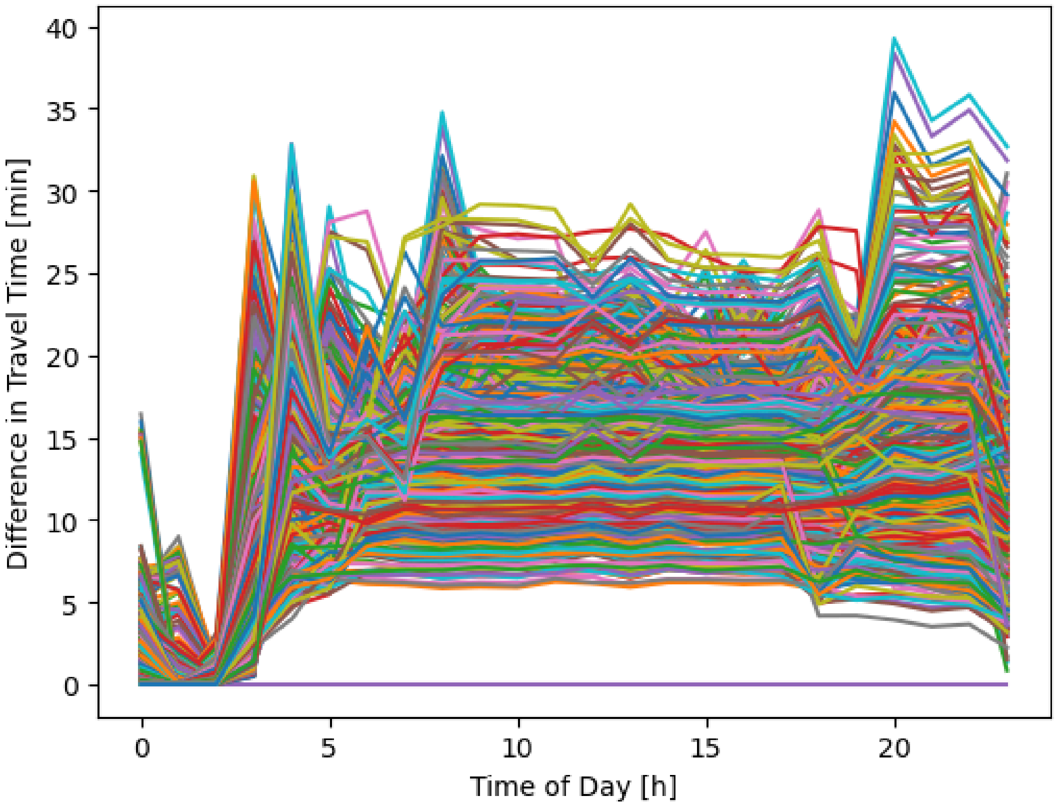

Figure 2: Plot of differences in travel times between 90th and 10th percentile in Heidelberg from cell to cell with no population mask, over the course of a weekday.

- As Figure 2 shows the impact of planning over the course of the day is much more variable than than the average travel times based on a median travel time over the course of an hour (compare also Figure 1).

4 Results

4.1 Travel Times

4.1.1 Temporal Variation

4.2 planning data

5 Discussion

5.1 Temporal Variations of Mean Travel Times

- as expected more central places within the city boundaries also feature lower average travel times throughout the day.

This accessibility measure only considers other populated cells, and assumes a scenario in which destinations such as friends and other people, that lie generally within populated areas but don't have a specific location in general, are of interest. It does not distinguish between the impact of the spatial population distribution and the impact of the specific public transit network however (compare [9]). Travel time here is assumed to be an objective cost measure, [9] however suggests that percieved travel time might substantially vary from the objectively needed travel time. And even hypothetical travelers with objective time perception have different ideas about the cost a certain travel time entails based on the mode of transit. That is they may have preferences on where and how they like to spend their travel time. This may go so far as taking objectively worse routes to reach their destinations to avoid wait times at interchanges.

5.2 Differences to Planning

5.3 temporal variation in planning data

- so what/what does this indicator describe that we don't get from the pure schedule already?

5.4 General Limitations

- Lack of real world measures as Comparisons
- special point of interests like school data
- Focuses solely on door to door travel times and neglects other aspects of such journeys.

There's no reliability and delay data included in this analysis as such data in a disaggregated form is rarely openly available and often requires setting up data scraping [31]. We also excluded public transit fare structures as implemented in [32]. While these are a common cost factor used next to travel time [9], they are complicated to implement for a simple analysis like this.

- lacks data including
  - comparisons to cars
  - ride hailing services see [33]
  - related on demand services (trial at rohrbach and schlierbach)

- *inequality* being silly at times [34].

6 Final Remarks

6.1 Conclusion

- Well this turned out to be a bit boring …

6.2 Outlook

- implementing a analysis accounting for delay and cancellation data would help to get a better picture of lived realities
- implementing an analysis for scenarios based on more specific scenarios of day to day life
- planning data can be helpful for planning new services, including needed night services, or stops
- introducing points of interest to the analysis for clearer scenarios

Bibliography

[1] H. Holzapfel, “Hat das Auto in der Stadt noch etwas zu suchen?,” *Strategien gegen den Verkehrsinfarkt.* in Deutsche-Bank-Research. Schäffer-Poeschel, Stuttgart, pp. 63–80, 1993.

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

[3] D. Teufel, “Der Autoverkehr als Umweltfaktor,” *Strategien gegen den Verkehrsinfarkt.* in Deutsche-Bank-Research. Schäffer-Poeschel, Stuttgart, pp. 41–61, 1993.

[4] C. Rissel, N. Curac, M. Greenaway, and A. Bauman, “Physical Activity Associated with Public Transport Use—A Review and Modelling of Potential Benefits,” *International Journal of Environmental Research and Public Health*, vol. 9, no. 7, pp. 2454–2478, Jul. 2012, doi: 10.3390/ijerph9072454.

[5] M. Stevenson *et al.*, “Land use, transport, and population health: estimating the health benefits of compact cities,” *The Lancet*, vol. 388, no. 10062, pp. 2925–2935, Dec. 2016, doi: 10.1016/S0140-6736(16)30067-8.

[6] F. Pieper, “Der Kreislauf der Aufgaben – Leistungsplanung und Leistungserstellung im Betrieb,” *Grundwissen Personenverkehr und Mobilität.* GRT Global Rail Academy and Media GmbH, Leverkusen, pp. 234–283, 2021.

[7] A. Karner, “Assessing public transit service equity using route-level accessibility measures and public data,” *Journal of Transport Geography*, vol. 67, pp. 24–32, Feb. 2018, doi: 10.1016/j.jtrangeo.2018.01.005.

[8] J. R. Verduzco Torres and D. P. McArthur, “Public transport accessibility indicators to urban and regional services in Great Britain,” *Scientific Data*, vol. 11, no. 1, p. 53–54, Jan. 2024, doi: 10.1038/s41597-023-02890-w.

[9] D. M. Levinson and H. Wu, “Towards a general theory of access,” *Journal of Transport and Land Use*, vol. 13, no. 1, pp. 129–158, Jun. 2020, doi: 10.5198/jtlu.2020.1660.

[10] P. R. Carleton and J. D. Porter, “A comparative analysis of the challenges in measuring transit equity: definitions, interpretations, and limitations,” *Journal of Transport Geography*, vol. 72, pp. 64–75, Oct. 2018, doi: 10.1016/j.jtrangeo.2018.08.012.

[11] M. W. Conway, A. Byrd, and M. van der Linden, “Evidence-Based Transit and Land Use Sketch Planning Using Interactive Accessibility Methods on Combined Schedule

and Headway-Based Networks," *Transportation Research Record*, vol. 2653, no. 1, pp. 45–53, 2017, doi: 10.3141/2653-06.

[12] H. Tenkanen and T. Toivonen, "Longitudinal spatial dataset on travel times and distances by different travel modes in Helsinki Region," *Scientific Data*, vol. 7, no. 1, p. 77–78, Mar. 2020, doi: 10.1038/s41597-020-0413-y.

[13] J. Dahn, "h3pandas: Integration of H3 and GeoPandas." Accessed: Jan. 22, 2024. [Online]. Available: https://github.com/DahnJ/H3-Pandas

[14] Geofabrik GmbH, "Geofabrik Download Server." Accessed: Dec. 17, 2023. [Online]. Available: http://download.geofabrik.de/

[15] H. Tenkanen, "pyrosm." Accessed: Jan. 18, 2024. [Online]. Available: https://pyrosm.readthedocs.io/en/latest/index.html

[16] VRS, "Soll-Fahrplandaten VRS." Accessed: Jan. 10, 2024. [Online]. Available: 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

[17] VVS, "Soll-Fahrplandaten VVS 2024 Jahresfahrplan." Accessed: Dec. 17, 2023. [Online]. Available: 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

[18] Rhein-Neckar-Verkehr GmbH, "Aktueller GTFS." Accessed: Nov. 27, 2023. [Online]. Available: 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

[19] DELFI, "Deutschlandweite Sollfahrplandaten (GTFS)." Accessed: Dec. 20, 2023. [Online]. Available: 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%5

Bdataset_name%5D=deutschlandweite-sollfahrplandaten-gtfs&cHash=01414d5793fcd0abb0f3a2e35176752c

[20] HeiGIT, "Openrouteservice API." Accessed: Jul. 01, 2023. [Online]. Available: https://openrouteservice.org/

[21] D. O'Sullivan, A. Morrison, and J. Shearer, "Using desktop GIS for the investigation of accessibility by public transport: an isochrone approach," *International Journal of Geographical Information Science*, vol. 14, no. 1, pp. 85–104, Jan. 2000, doi: 10.1080/136588100240976.

[22] V. Bauer, J. Gamper, R. Loperfido, S. Profanter, S. Putzer, and I. Timko, "Computing isochrones in multi-modal, schedule-based transport networks," in *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, Irvine California: ACM, Nov. 2008, pp. 1–2. doi: 10.1145/1463434.1463524.

[23] C. Fink, W. Klumpenhouwer, M. Saraiva, R. Pereira, and H. Tenkanen, "r5py: Rapid Realistic Routing with R5 in Python." Accessed: Jan. 18, 2024. [Online]. Available: https://zenodo.org/records/7060438

[24] M. W. Conway, A. Byrd, and M. van Eggermond, "Accounting for uncertainty and variation in accessibility metrics for public transport sketch planning," *Journal of Transport and Land Use*, vol. 11, no. 1, Jul. 2018, doi: 10.5198/jtlu.2018.1074.

[25] L. McInnes, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." Accessed: Jan. 29, 2024. [Online]. Available: https://umap-learn.readthedocs.io/en/latest/index.html

[26] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." Accessed: Jan. 29, 2024. [Online]. Available: http://arxiv.org/abs/1802.03426

[27] "Clustering on the output of t-SNE." Accessed: Jan. 29, 2024. [Online]. Available: https://stats.stackexchange.com/q/263539

[28] E. Schubert, "Answer to "Clustering on the output of t-SNE"." Accessed: Jan. 29, 2024. [Online]. Available: https://stats.stackexchange.com/a/264647

[29] L. McInnes, J. Healy, and S. Astels, "The hdbscan Clustering Library — hdbscan 0.8.1 documentation." Accessed: Jan. 29, 2024. [Online]. Available: https://hdbscan.readthedocs.io/en/latest/index.html

[30] R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-Based Clustering Based on Hierarchical Density Estimates," in *Advances in Knowledge Discovery and Data*

Mining, J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2013, pp. 160–172. doi: 10.1007/978-3-642-37456-2_14.

[31] D. Kriesel, "BahnMining - Pünktlichkeit ist eine Zier." Accessed: Dec. 19, 2023. [Online]. Available: https://media.ccc.de/v/36c3-10652-bahnmining_-_punktlichkeit_ist_eine_zier

[32] M. W. Conway and A. F. Stewart, "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*, vol. 33, no. 9, pp. 1759–1787, 2019, doi: 10.1080/13658816.2019.1605075.

[33] J. M. Barajas and A. Brown, "Not minding the gap: Does ride-hailing serve transit deserts?," *Journal of Transport Geography*, vol. 90, p. 102918–102919, Jan. 2021, doi: 10.1016/j.jtrangeo.2020.102918.

[34] D. Graeber and D. Wengrow, *The Dawn of Everything. A New History of Humanity*. Dublin: Penguin Books, 2022.