BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2026
http://excel.fit.vutbr.cz

System for Evaluating the Quality of Public Transport Networks Using GTFS Datasets

Bc. Jan Kvapil*

Abstract

Publishing of schedule data by public transport agencies is becoming increasingly common, with the General Transit Feed Specification (GTFS) being the preferred standard for doing so. Agencies often publish these datasets as part of their open data programs, enabling anyone to browse them, run their local trip planners, or, in the context of this work, analyze the networks they describe.

This project develops a reproducible pipeline for evaluating and comparing the quality of public transport networks using GTFS feeds, the JRC-ESTAT Census Population Grid 2021, the Lissy API, and region borders published as open data to trim the networks' outreach into other regions. The workflow automates dataset retrieval, extraction, and parsing of core GTFS tables and applies region-specific processing to ensure consistent representation of spatial and public transport features. The resulting metrics, intermediate analysis products, and visualizations are designed to support exploratory analysis and reporting through Python Notebooks and a lightweight web-based visualizer, with the analysis focusing on population, spatial, and overall network statistics derived from static data. The workflow is demonstrated on four region-based datasets: IDS JMK, Brno (subset of IDS JMK), PID/IDSK, and Prague (subset of PID/IDSK).

*xkvapi19@stud.fit.vut.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Public transport GTFS feeds provide standardized data that is valuable for evaluating network quality, despite regional implementation differences and subjective definitions of what constitutes a good service level. To yield meaningful performance indicators, raw feeds must be properly aligned and combined with supplementary datasets, such as administrative boundaries or population census data.

This work presents a repeatable, data-driven workflow for assessing and comparing public transport networks using static open data (Figure 1). By automating the integration of GTFS feeds with population grids and boundary data, the proposed method transforms raw tables into consistent, region-bounded metrics that characterize network structure, relate service provision to where people live, and are easily extensible to new areas as evaluation needs evolve.

The four key areas of this analysis are:

1. Supply-demand analysis, utilizing the JRC-ESTAT Census Population Grid 2021 [1]
2. Stop density analysis, utilizing the Lissy API[1] by Ing. Juraj Lazúr [2] for the IDS JMK network or the `shapes.txt` for GTFS feeds that include it
3. Travel time analysis, utilizing the OpenTripPlanner 2[2] routing engine and OpenStreetMap[3] data
4. Static metrics analysis based on GTFS feeds and data obtained during the aforementioned steps

2. Analysis Approach

2.1 Supply-Demand Analysis

The core and most expansive analysis approach of this work is supply-demand analysis. It aims to answer the question: "Is the network adequately serving the population?" The analysis uses the GTFS dataset of the respective area, together with the JRC-ESTAT Census Population Grid 2021 [1] with 100m resolution. Working with the established catchment area radius of 400m for bus stops, and 800m for rail-based stations (Figure 2), the analysis first identifies stops within walking dis-

[1] https://github.com/Jorgen98/Lissy
[2] https://docs.opentripplanner.org/en/latest/
[3] https://www.openstreetmap.org/

tance of each 100x100m area (a "cell"). Next, for each stop, the number of departures is gathered for a full working day, as well as for each hour of a working day, Saturday, and Sunday. Using inverse-distance weighting, the number of departures is then normalized. This ensures that cells with multiple stops in their walking distance do not have an inflated or double-counted service frequency. Finally, the "transit deficit" is calculated for each cell, following the formula $td = pop/dep$, where pop is the population of a given cell, and dep is the number of departures calculated for each cell.

Visualization of these cells in a map can then show either areas that are relatively underserved (using a median-based scale, Figure 3), served less than the public agency sees as necessary (using a settable threshold, Figure 4), or not served at all (displayed universally in blue).

2.2 Stop Density Analysis

Using the Lissy API[4] by Ing. Juraj Lazúr [2] (approach used for IDS JMK) or by processing the `shapes.txt` file, should the GTFS dataset possess it, the routing of public transport lines is obtained in the form of polylines. Using these polylines, distances between stops can be calculated. From this data, the average distance between stops can be calculated based on actual routing data, making it far more accurate than the geodesic distance between two stops.

These polylines can again be visualized in a map, showing inter-stop segments that may be too short (meaning that a connection stops perhaps too frequently), or too long (usually belonging to an express or long-distance connection, which stops infrequently). As with the supply-demand analysis, relatively short inter-stop segments (using a median-based scale, Figure 5) can be displayed, and a user-defined threshold can be set (Figure 9), resulting in color-coded inter-stop segments being displayed (green for segments with a length below the threshold, and red for all other segments, Figure 6).

2.3 Travel-Time Analysis

While the number of possible Origin-Destination pairs is $N \times (N-1)$ for N stops on the network, a large percentage of people travel from their nearest stop to a nearby hub. These hubs are usually expansive in size; the number of platforms and departures is significantly larger than at an ordinary stop. Using OpenTripPlanner 2[5], the travel time to N geographically closest hubs is calculated, with the shortest time being recorded for visualization and analysis.

[4] https://github.com/Jorgen98/Lissy

[5] https://docs.opentripplanner.org/en/latest/

The obtained data is visualized by color-coding each stop based on the calculated travel time. Optionally, the polylines for these journeys can be displayed, which can show core arteries of the transit system (Figures 7 and 8).

2.4 Static Metrics Analysis

Static numerical data can be obtained from the GTFS dataset or using the aforementioned analysis approaches. Based on the census dataset, the total number of citizens in the analyzed area can be calculated, as well as the number and percentage of citizens without service within walking distance. Based on the stop density analysis, an average distance between stops can be calculated to determine whether stops are optimally spaced. Based on the travel time analysis, the average travel time to the nearest hub can be calculated, which can be used to determine whether public transport vehicles and line routing are appropriate for efficient commuting (Table 1). Additionally, the aforementioned datasets can be used to calculate the number of departures provided by the system, the percentage of barrier-free services, the percentage of barrier-free stops, the average number of people served by a single line, and much more.

3. Conclusion

This project presents a tool that can be used to analyze public transport networks in the same way across Europe, with only minor preprocessing of the GTFS dataset being needed to run the workflow and obtain statistics. Due to its universality, the tool can be used to compare public transport networks across cities or regions, as well as to pinpoint weaknesses and obtain statistics about individual networks.

Acknowledgements

I would like to thank my supervisors, Ing. Jiří Hynek, Ph.D., and Ing. Juraj Lazúr, for their oversight, guidance, feedback, and support during the making of this work.

References

[1] Cristian Pigaiani, Sergio Freire, and Filipe Batista. JRC-ESTAT Census Population Grid 2021. Dataset, 2026. PID: http://data.europa.eu/89h/98336641-fd1c-4992-8c7b-c470dd5eb81e.

[2] Juraj Lazúr. Analýza a vizualizace dat hromadné dopravy města Brna. Diplomová práce, Vysoké učení technické v Brně, Brno, 2023.