City Transport Analyzer: A Powerful Qgis Plugin For Public Transport Accessibility And Intermodality Analysis

Gianmarco Naro1, Carlo Andrea Biraghi2, Emilia Lenzi3

1 Master of Science programme in Computer Science and Engineering, Politecnico di Milano - gianmarco.naro@mail.polimi.it
2 Department of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, Via Giuseppe Ponzio 31, 20133 Milano, Italy - (carloandrea.biraghi, emilia.lenzi)@polimi.it
3 Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via Giuseppe Ponzio 34/5, 20133 Milano, Italy

Keywords: GTFS, QGIS, Accessibility, Intermodality, Public Transport, Plugin

Abstract

This study presents the City Transport Analyzer (CTA) QGIS plug-in designed for analyzing urban public transportation (PT) networks, exploiting GTFS data to investigate their multimodal nature. GTFS, a standardized format for PT schedules and geographic information, enables accurate data sharing, crucial for service analysis. The plug-in extracts GTFS data and converts it into GIS layers for analysis. It constructs a graph data structure using NetworkX, representing PT routes, with edges indicating travel time and transport type. Two key analyses are conducted: Service Area for accessibility and Nearby Stops Path for intermodality. The accessibility analysis determines reachable areas within specified time frames via PT and walking from one or more source points. It provides comprehensive insights into service coverage and the potential reach of PT networks within urban areas. The intermodality analysis assesses street paths for modal changes beyond PT nodes, offering valuable information on connectivity and potential PT usage patterns. Tested on case studies in Milan and Rio de Janeiro, the plug-in demonstrates utility in understanding metropolitan PT networks, but also room for improving the overall accuracy, usability and performances of the tool. The CTA plugin offers a flexible and user-friendly tool for studying service coverage, accessibility, and connectivity, aiding urban infrastructure analysis and planning. Future work aims to extend analyses to the whole city, including other city contexts. Ultimately, the CTA plugin contributes to a deeper understanding of urban PT networks and areas, facilitating informed decision-making in urban planning and infrastructure management.

1. Introduction

1.1 Mobility, policy, and sustainability

The significance of transportation in sustainable development was initially acknowledged during the 1992 United Nations Earth Summit and further emphasized in its ensuing document, Agenda 21 (United Nations, 2021). Presently, transportation stands as the largest consumer of energy in developed nations and the fastest-growing sector in many developing regions. Notably, nearly a quarter of global greenhouse gas emissions from energy sources stem from transportation, a figure expected to rise significantly in the foreseeable future, as highlighted by the United Nations Framework Convention on Climate Change (UNFCCC).

The integration of sustainable transport within the 2030 Agenda for Sustainable Development is evident across various Sustainable Development Goals (SDGs) and associated targets, particularly those pertaining to food security, health, energy, economic growth, infrastructure, and urban settlements. Sustainable transportation holds promise in bolstering economic progress, fostering accessibility, and safeguarding the environment, thereby enhancing social equity, health, and urban resilience. Of particular note is SDG Target 11.2, which underscores the imperative to “provide access to safe, affordable, accessible and sustainable transport systems for all, improving road safety, notably by expanding public transport networks . . .” (https://sdgs.un.org/topics/sustainable-transport).

The imperative to transition toward sustainable urban mobility is underscored by the need to curb transport-related carbon emissions, achieve the Sustainable Development Goals, and develop urban transport systems that are economically viable, environmentally sustainable, and socially inclusive (https://sdgs.un.org/partnerships/sustainable-urban-mobility).

Numerous countries are tackling this issue through the adoption of Sustainable Urban Mobility Plans (SUMPs), with the European Union offering guidelines for their implementation (Comission of the European Communities, 2009). These guidelines, structured into four phases, emphasize the importance of data-driven analysis, strategic planning, and effective monitoring and evaluation (Torrisi et al. 2020). However, challenges such as funding constraints, inadequate modelling techniques, and suboptimal evaluation methodologies hinder the successful implementation of SUMPs (May, 2015).

In response to the evolving complexity of urban transportation systems, there is a growing emphasis on the analysis and optimization of public transportation (PT) services. Advanced technologies play a pivotal role in this endeavour, facilitating the study and enhancement of PT networks amid dynamic urban growth (Berg and Ihlström, 2019).

Addressing this multifaceted challenge, we present a novel QGIS plugin designed to provide a comprehensive tool for the detailed analysis of urban transportation networks, with a focus on Accessibility and Intermodality aspects.

2. Data and tools

2.1 Data standards

In the dynamic realm of public transportation, adherence to established standards is crucial for ensuring interoperability across various platforms and facilitating rigorous standardization of essential information.
Two prominent geospatial standards, the General Transit Feed Specification (GTFS, https://gtfs.org) and Transmodel (https://www.transmodel-cen.eu/), play crucial roles in structuring transportation-related data (CEN, 2019). GTFS, an open data format originally developed by Google, enables the distribution of detailed information about public transportation systems. Its simple structure, based on comma-separated values (CSV) text files, allows for easy creation and management of transit information, making it widely adopted by public transport agencies worldwide.

In contrast, Transmodel is a European reference data model for public transport information. While not a data exchange standard itself, Transmodel provides a conceptual framework for creating specific data exchange standards. Derived standards like NeTEx, TransXChange, and SIRI facilitate accurate and interoperable exchange of public transport information in various contexts.

GTFS focuses on standardized representation of public transport data in a simple format, while Transmodel provides a conceptual framework for creating specific data exchange standards, promoting interoperability. Both standards play crucial roles in facilitating access to and analysis of public transport data.

For this study, GTFS data is preferred due to its widespread availability, ease of access, compatibility with various data types, and regular updates, making it suitable for conducting thorough analyses and implementations within the research framework.

### 2.2 Tools for mobility assessment

Some of the tools for studying urban mobility based on GTFS data and related systems and software are now presented.

Fortin, Morency, & Trépanier (2016) study addresses the need for comprehensive tools to analyze transit networks. It introduces indicators structured using a graph-oriented technique, leveraging GTFS data to provide insights into network linearity and line overlapping. Our plug-in emphasizes GTFS data structure and transformation, inspired by this work, to enable more efficient analysis of transit networks.

In contrast, the detailed analysis of GTFS data from the Cityform Project's codebase offers insights into innovative data utilization methods. This method creates a “Road Map” using dictionaries, facilitating gravity index calculation and shortest path determination. Our plug-in incorporates aspects from this analysis to enhance efficiency in identifying frequent intersections on segments.

Brahmanage, Kandappu, & Zheng (2023) introduce MetroWatch, a predictable system utilizing AFC data for estimating passengers’ travel attributes within metro systems. Building on TripDecoder's methodology (Tian et al., 2021) MetroWatch provides precise route information and passenger density estimates, influencing our plug-in's predictive capabilities and practical insights for policymakers and urban planners.

The City Form Lab's GIS toolbox for ArcGIS Pro computes spatial accessibility via public transport using GTFS data. It measures access from origin to destination points within constraints chosen by the user, offering Reach and Gravity indices for urban planning (Sévouk et al., 2023). These studies illustrate the evolution of tools for urban mobility analysis, emphasizing data-driven methodologies in urban planning and transportation analysis. The advancements made by our plug-in integrate GTFS data seamlessly within FOSS GIS environments like QGIS, providing comprehensive insights into public transportation networks' accessibility and interoperability, contributing to ongoing efforts to define effective urban transportation analysis standards.

### 2.3 Content and repositories

This article presents the development of the City Transport Analyzer (CTA) plugin and its testing on two sample case studies: Milan and Rio de Janeiro. Section 2 presents the methods, from the pseudocode implementation of the analysis to the technicalities of the coding phase. Section 3 focuses on the plugin test on the case studies, presenting comparative maps and statistics, and discusses the current plugin performance including its limitations and the possible future research lines to address them. The data used in this research as well as the code of the plugin are publicly available at https://github.com/jianmarcoromani/city-transport-analyzer.

### 3. Method

#### 3.1 Purpose of the analysis

For distances under 2km, walking and cycling can be competitive, but for longer distances, public transportation (PT) is the primary alternative. While direct PT connections can sometimes be faster than driving, switching modes inevitably impacts commuting time. Therefore, strategies to enhance multi-modal transport attractiveness are essential. This importance is underscored by the establishment of a High-Level Advisory Group on Sustainable Transport (HLAG-ST) by the UN Secretary-General in 2014, representing various transport modes including road, rail, aviation, marine, ferry, and urban public transport providers.

Intermodality research has seen significant growth in recent years, focusing on interchange hubs as nodes, promoting the integration of different functions following Transit-Oriented Development (TOD) principles.

Our plugin leverages the multimodal nature of PT networks modeled as GTFS, providing insights into urban systems' dynamics related to sustainable mobility through Service Area Analysis (accessibility) and Nearby Stops Analysis (intermodality).

#### 3.1.1 Service Area (accessibility):

The Service Area Analysis estimates the accessibility between an area and the rest of the urban system via PT. GTFS data is used as PT data in simple vector format do not constitute a graph, thus cannot be continuously navigated. This analysis complements walkability assessment through isochrones on the pedestrian network.

Users input a starting point layer and a time interval. The plugin then finds the closest point on the PT route network to the user-provided input and draws a service area line layer on the route graph, considering the maximum allowed displacement by line or combination of lines in each direction. Default speed values for each mode can be modified in the plugin source code. A convex hull of the service area is generated with the aim of evaluating the reachability of diverse services or amenities.
3.1.2 Nearby Stops Paths (intermodality): The Nearby Stop Paths Analysis introduces novelty in assessing the intermodality dimension by considering the potential impact of intermodal displacements on the street network. This perspective identifies street links with a higher potential probability of being crossed for modal switch opening to complementary investigations of their walkability level according to different parameters.

This study proposes a systematic method to detect "potential multimodal routes" (PMR) for each PT stop, identifying stops of different lines within a reasonable walking distance defined by the user. Stops with at least one line in common with the source are removed, assuming PT connections are faster than walking. The shortest path on the pedestrian street network is then computed between each source and its targets, ranking street links according to their probability of being crossed by multi-modal displacements.

Users provide a starting stops layer and a search radius in meters. GTFS is used to derive information about lines passing through each node, while the shortest paths are drawn on the pedestrian graph.

3.2 Technicalities and plug-in implementation

The plugin’s development relied heavily on Python, chosen for its extensive library support and capacity to manage intricate data and graphical representations. Specifically, the OSMnx (Boeing, 2017) library, designed for extracting, modeling, and analyzing urban geographical data from OpenStreetMap, was employed. OSMnx’s robust tools for processing and visualizing urban networks, along with its compatibility with NetworkX graphs, facilitated efficient extraction of street networks from supplied datasets.

The algorithms employed included Breadth-First Search (BFS) (Kozen, 1992) and Dijkstra’s algorithm (Dijkstra, 1959), tailored for analyzing network structures and determining optimal pathways within the plugin. The modified BFS algorithm operated within predefined time constraints for weighted graph scenarios, prioritizing traversal of edges based on time constraints. Dijkstra’s algorithm played a crucial role in determining the shortest paths between nodes, providing accurate distance computations essential for identifying optimal routes.

The initial step of the plugin involved generating a graph from GTFS data to dynamically create spatially aware urban networks reflecting the complex nature of public transportation systems. This graph, created using NetworkX, was tailored to ensure that each PT route exclusively utilized its corresponding edge, with distinct overlapping edges created for multiple public lines traversing the same edge. Each edge of the graph denoted the transportation mode it served, along with average travel time and length.

To utilize the plug-in effectively, users must provide two key datasets: the GTFS file containing transit information and a polygon file delineating the geographic area of interest. GTFS data undergo processing and storage in an SQLite database for optimized querying efficiency, with only relevant tables included to streamline database operations. The polygonal data define the study area and are used for generating the pedestrian graph using the OSMnx library. Upon completion of the analysis, users gain access to various layers providing comprehensive information about the analyzed transportation networks. These layers include for the Service Area analysis are:

- Starting Points: The points from the graph that serve as the basis for the analysis are contained in a point layer. The coordinates of these sites are given in detail. They are the closest points on the route graph respect to the input points provided by the user.
- Service Area: A layer that shows the created service area with all edges that can be visited during the period of time specified. The nodes at each edge’s extremities, length, travel time, and related mode of transportation are all listed.
- Convex Hull: A layer representing the convex hull derived from the service area.

The output layers for the Nearby Stops Path analysis are:

- Starting Stops: The point layer containing the stops from which the analysis starts. It provides comprehensive information about each stop, including ID, name, and the type of transit that utilize it.
- Circular Buffer: This layer shows the circular buffer that will be used to define the region from which stops will be selected later. Its visual portrayal facilitates comprehension of the area under analysis.
- Selected Stops: The point layer that shows the chosen stops inside the circular buffer. Every stop has a number, the ID of the first stop it’s connected to, a name, the means of transportation that pass through it, and a boolean field that indicates if it has been chosen for more examination. Selected stops are displayed as yellow squares while non-selected one are red.
- Shortest Paths: This layer maps the shortest paths between one stop and another. It includes details like the path’s length, the ID and name of the starting stop, the ID and name of the destination stop.
- Intersections: This layer identifies all road segments intersected by the shortest paths. The number of times the shortest paths intersect each of these segments is counted and shown for each of them.

The plugin interface, designed from scratch to accommodate multiple subsequent actions and alternative pathways to achieve desired results, ensures a user-friendly interaction and an efficient analysis processes. Figure 1 presents a collage of the interface of each step, from data preparation to the setting of the preferences for the analyses.
4. Results and discussion

4.1 Milan & Rio de Janeiro case studies

The choice of using Milano and Rio de Janeiro as case studies moves from the involvement of one of the authors in the ongoing REMIRIO (https://www.immdesignlab.com/remirio) project (Integrated and Sustainable Urban Regeneration Milan Rio), a city-to-city cooperation initiative funded by the Europe Aid - Partnerships for sustainable cities in Latin America and the Caribbean. In addition, both cities have a freely downloadable GTFS file and an articulated PT network including diverse modes.

Milan and Rio de Janeiro differ notably in size, population, and density and compactness. Milan, situated in northern Italy, features a relatively small size with a dense population.

<table>
<thead>
<tr>
<th></th>
<th>Rio</th>
<th>Milan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area (km²)</td>
<td>1221</td>
<td>181</td>
</tr>
<tr>
<td>Population (M)</td>
<td>6,21</td>
<td>1,37</td>
</tr>
<tr>
<td>Density (ab/km²)</td>
<td>5526</td>
<td>7555</td>
</tr>
<tr>
<td>N° Stops</td>
<td>7'185*</td>
<td>4'955</td>
</tr>
<tr>
<td>N° Routes</td>
<td>133'369*</td>
<td>137'062</td>
</tr>
<tr>
<td>N° Modes</td>
<td>2*</td>
<td>4</td>
</tr>
<tr>
<td>N° Lines</td>
<td>401*</td>
<td>163</td>
</tr>
<tr>
<td>Routes Length (km)</td>
<td>21'843*</td>
<td>14'296</td>
</tr>
<tr>
<td>Street length (km)</td>
<td>12'247*</td>
<td>11'388</td>
</tr>
</tbody>
</table>

Table 1. General information and GTFS consistency for Rio and Milan. *real values are higher as the official GTFS only includes BRT and Onibus

Its urban core is concentrated, surrounded by rural and recreational void areas, except for the north side, which exhibits continuous conurbations with adjacent municipalities.

In contrast, Rio de Janeiro in Brazil is significantly larger, approximately seven times the size of Milan, and hosts a population 4.5 times larger. Its development encompasses vast natural areas, including mountains, resulting in a more fragmented and articulated shape. Looking at their street and PT networks, it is interesting to notice than while the pedestrian ones have similar extension, Rio’s PT routes are 1.5 times longer than the Milanese ones. Table 1 reports both general information and GTFS consistency for the two cities.

The GTFS of Rio available on the data.rio geoportal (https://www.data.rio/datasets/gtfs-do-rio-de-janeiro/about) covers only two (BRT and Onibus) out of the existing 9 modes (Barcas, Bonde, Metro, STPL, Telefericos, Trem and VLT). In support of the REMIRIO activities, the existing GTFS was integrated by adding Metro, Telefericos and VLT while the remaining modes will be added in the next future. More details on the differences in terms of number of features between the two version of Rio GTFS are reported in Table 2, together with the Service area results.

For Milan the ATM (Milanese Transport Agency) GTFS, provided by AMAT (Mobility, Environment and Territory Agency), downloadable at https://dati.comune.milano.it/gtfs.zip, was used. It includes Bus, Filobus, Tram and Metro lines. Figure 2 shows the consistency of the GTFS files.

4.2 Experiments

Both Service area and Nearby stop analysis were tested on both cities. For the Nearby stop, a single starting point was considered, respectively in Duomo square (NIL 1) for Milan, and in the centroid of Bairro Cidade Nova for Rio. From these points a circular buffer of 1.1km radius was used to select the starting stops. A search radius of 500m was set for both cases. As no major changes were visible for the selected area of Rio between the two versions of the GTFS, only the result of the updated GTFS is presented.
The Service area analysis for Rio was run on the same starting point of the Nearby Stops for the two available GTFS. In Milan, the centroid of each Municipio (local administrative boundaries) was used for a total of 9 points. In both cases, two time frames of 15 and 30 minutes were tested.

The comparison between the two Nearby stop maps (Figure 3 and 5) show interesting results. The range of the shortest path count in the two context is similar despite the differences among the contexts. This results from a balance between the few stops (28) with many lines (up to 70) in Rio, and many stops (98) with few lines (up to 7) in Milan. The number of stops selected with the 500m buffer increased by similar values, 55% in Rio and 65% in Milan. The links in Milan draw a sort of ring with hotspots in P.za Duomo (west), with the axis of Via Orefici on the north and Corso Italia on the south, P.ta Romana (south) and P.ta Vittoria (east). These last two system of streets are particularly evident for Municipio 5 (south), whose centroid is in an almost rural area, and the existing PT infrastructures lays in separate linear direction separated by large voids. This highlights a limit of the algorithm used. For deepening this aspect, A* algorithm could be tested to see if the problem persists.

A visual comparison between the two time thresholds for both GTFS versions (baseline in red, edited in blue) is presented in Figure 4. The 15min map is clear and shows an increment of 38% in the covered area thanks to the introduction of the new new modes. In the 30min map some unexpected results occurred. Some portions resulting as accessible in the baseline are not considered in the richer GTFS version. The accessible area increases only by 10% because of this issue. Coherently, also the SA length and link N for 30min are lower than the correspondent values on the baseline GTFS. The reason for this could lay in digitizing errors during the GTFS editing or in the nature of the algorithm used. For deepening this aspect, A* algorithm could be tested to see if the problem persists.

Figure 6 and Table 3 present the results for Milan. It must be reminded that these results don’t draw a picture of the overall Municipio potential accessibility but only that of a single point for each of them. It follows that the huge differences between them are due on one side to the effective structure of the PT network, and on the other to the distance of the starting point from the closest PT route. This is particularly evident for Municipio 5 (south), whose centroid is in an almost rural area, and the existing PT infrastructures lays in separate linear direction separated by large voids. This highlights a limit of the actual code, as the west side PT lines, reachable within 30min but more distant than the east-side ones, result as not accessible. This happens because once the closest entry to the routes network is found, no other entries are searched.

<table>
<thead>
<tr>
<th>Municipio</th>
<th>SA 15min (km²)</th>
<th>SA 30min (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Central</td>
<td>29.54</td>
<td>37.30</td>
</tr>
<tr>
<td>2 North-East</td>
<td>21.87</td>
<td>47.01</td>
</tr>
<tr>
<td>3 East</td>
<td>7.59</td>
<td>23.41</td>
</tr>
<tr>
<td>4 South-East</td>
<td>28.62</td>
<td>27.26</td>
</tr>
<tr>
<td>5 South</td>
<td>12.78</td>
<td>57.69</td>
</tr>
<tr>
<td>6 South-West</td>
<td>4.51</td>
<td>25.68</td>
</tr>
<tr>
<td>7 West</td>
<td>1.13</td>
<td>12.12</td>
</tr>
<tr>
<td>8 North-West</td>
<td>3.89</td>
<td>11.54</td>
</tr>
<tr>
<td>9 North</td>
<td>1.76</td>
<td>11.98</td>
</tr>
<tr>
<td>Mean</td>
<td>12.40</td>
<td>28.22</td>
</tr>
</tbody>
</table>

Table 3. Service Area (SA) comparison between the 9 Municipi in Milan for the 15min and 30min thresholds.
Figure 3. Nearby stop analysis for Rio, Cidade Nova Bairro.

Figure 4. Service area analysis for Rio. Above 15 minutes; below 30 minutes.

Figure 5. Nearby stop analysis for Milan, P.za Duomo, NIL 1.

Figure 6. Service area analysis for Milan Municipi. Above 15 minutes; below 30 minutes.
4.3 Conclusions

The City Transport Analyzer plugin enables novel (intermodality) and well-routed (service area) relevant analysis for urban public transportation networks based on the GTFS data standard. The plugin can deliver timely information regarding the urban transportation network. The information extracted by the described plugin provides both quantitative and qualitative insights, delineating the transportation network, including its pedestrian aspects. Moreover, the plugin serves as an initial example of how GTFS data can be utilized in creating a weighted and connected graph beneficial for this type of analysis. However, at the current stage of development, it has some limitations and great room for making it an even better tool for the FOSS community.

However, at its current developmental stage, the plugin exhibits certain limitations yet possesses substantial potential for further enhancement, thus positioning it as an increasingly valuable tool within the FOSS community. The main limitations concern scalability and adaptability to different operating systems.

Although the plugin demonstrates commendable efficiency in importing GTFS data and generating graphs, its performance diminishes notably during analyses involving a high volume of features (exceeding 100). This is reasonable given the type of analysis, but it reduces the plugin’s flexibility across different case studies. Moreover, performance degradation was observed during testing on certain computer configurations, with instances of QGIS becoming unresponsive when the plugin runs concurrently with other applications. To enhance its efficiency, particularly when analyzing specific segments of the urban network, a potential strategy involves precluding unnecessary network portions. This could be achieved by spatially clipping the network using a circular buffer whose radius is determined by the combination of user-input minutes (e.g., 15 minutes) and the speed of the fastest commuting modes available (e.g., metro). At present, this option is not included in the plugin to avoid burdening its interface, but it is a task that can be easily performed by users familiar with QGIS, the target audience of this plugin.

Future works can also focus on expanding the analysis flexibility. Indeed, the intermodality analysis could be generalized and extended to other urban elements like points of interest, civic numbers, or entrances. Facilitating the input of simple vector files and granting users greater control over selecting shared attributes for point feature inclusion (or exclusion) will vastly augment the tool’s flexibility for diverse applications in urban studies. This enhancement offers advantages for PT analysis, especially in scenarios where GTFS data is unavailable or where superior datasets exist.

Augmenting the attributes stored in output layers could involve incorporating established and experimental metrics to evaluate the PT network comprehensively, both at nodal (stops) and linear (routes) scales.

Looking ahead, functionalities for GTFS editing or creation starting from standard vector formats are desirable to cover the entire process comprehensively. Should this prove unfeasible while maintaining the plugin’s simplicity and user-friendliness, the option of implementing multiple menu items is recommended. The continued development of the CTA plugin will address these issues in the forthcoming year.

References


