anpet
brazilian association for transportation research and education
TRANSPORTES
ISSN: 2237-1346
transportes.anpet.org.br

Can OpenStreetMap map Brazil's school transportation network? An analysis based on existing rural and urban routes

O OpenStreetMap pode mapear a rede de transporte escolar do Brasil? Uma análise baseada nas rotas rurais e urbanas existentes

Marcos Paulino Roriz Junior[1], **Willer Luciano Carvalho**[1], **Djailson Dantas Medeiros**[2]

[1]Federal University of Goiás, Aparecida de Goiânia, Goiás, Brasil
[2]National Fund for Educational Development, Brasília, Distrito Federal, Brasil

Contact: marcosroriz@ufg.br, (MPRJ); willer_carvalho@ufg.br, (WLC); djailson.medeiros@mdic.gov.br, (DDM)

Submitted:
19 April, 2025
Revised:
1 September, 2025
Accepted for publication:
28 October, 2025
Published:
4 January, 2026
Associate Editor:
Cira Souza Pitombo, University of São Paulo, Brazil

Keywords:
OpenStreetMap.
Brazil school transportation system.
Brazil school transportation.
Map completeness.
Rural routes.
OSM.

Palavras-chave:
OpenStreetMap.
Sistema de transporte escolar brasileiro.
Transporte escolar brasileiro.
Completude de mapa.
Rotas rurais.
OSM.

DOI: 10.58922/transportes.v34.e3118

ABSTRACT

A significant challenge in planning Brazil's school transportation system, especially in rural areas, is the lack of reliable data. Among the most critical datasets is the road network, which serves as input to numerous tasks, such as calculating walking distances and optimizing bus routes. In this context, OpenStreetMap (OSM), a free and collaborative mapping platform, offers a potential solution by providing open access to road data. However, parts of the network are missing or disconnected, particularly in rural regions. These gaps limit the effective use of OSM in transportation tools, as algorithms may fail to reach specific locations, especially in remote areas. To assess the network completeness, this study analyzed 7,159 real-world school bus routes from the Brazilian Electronic School Transport Management System. On average, 91.84% of the route's length is covered, but only 32% of them are fully mapped by OSM (100% of its length). The most frequent gaps appear on waterway routes, mixed routes, and last-mile sections, such as local roads leading to students' homes. Recognizing these limitations is crucial for enhancing the use of OSM in transportation tools and ensuring broader access to education for students who depend on public school transportation.

RESUMO

A falta de dados confiáveis é um dos principais desafios no planejamento do sistema de transporte escolar no Brasil, especialmente em áreas rurais. Entre os dados mais críticos está a malha viária, pois serve de base para calcular as distâncias percorridas pelos alunos e a otimização das rotas de ônibus. Nesse contexto, o OpenStreetMap (OSM), plataforma livre, surge como alternativa ao disponibilizar dados viários abertos. No entanto, partes da malha estão ausentes ou desconectadas, principalmente nas regiões rurais. Essas lacunas limitam o uso efetivo do OSM em ferramentas de transporte, já que algoritmos podem falhar ao tentar alcançar determinados locais, especialmente os mais remotos. Para avaliar a completude da rede, este estudo analisou 7.159 rotas de transporte escolar provenientes do Sistema Eletrônico de Gestão do Transporte Escolar brasileiro. Em média, 91,84% do comprimento das rotas está coberto, mas apenas 32% delas estão totalmente representadas no OSM (100% de sua extensão). As lacunas mais frequentes aparecem em rotas aquaviárias, mistas e nos trechos finais, como estradas locais que levam até as residências dos estudantes. Compreender essas limitações é essencial para aprimorar o uso do OSM em ferramentas de transporte e ampliar o acesso à educação para os alunos do transporte escolar.

1. INTRODUCTION

The Brazilian Federal Constitution guarantees the right to public school transportation to ensure universal access to basic education for all students (Brazil, 2022). This right must be provided regardless of the student's location, regardless of whether he or she resides in a rural, urban, or riverside area. This requirement increases the complexity of carrying out this task since rural and remote regions usually lack transportation infrastructure (Carvalho et al., 2010; FNDE and CEFTRU, 2010; Zhu et al., 2023).

According to the Brazilian Institute of Geography and Statistics (IBGE, 2023), 60.2% of Brazil's 5,570 municipalities are predominantly rural. Among this set, 54.6% are classified as adjacent rural areas, while 5.8% are remote rural areas. The rural nature of the municipalities and, consequently, the lack of infrastructure make it challenging to plan and operate the school transportation system, as city managers need to consider students with diverse realities. For example, routes need to be scheduled differently to accommodate students living in remote areas of the municipality, including farms and riverside regions, as well as those living near schools (Carvalho et al., 2010).

To assist municipalities in this task, the Brazilian government, in collaboration with FNDE and CECATE-UFG (2019), developed an open-source e-governance software called *Sistema Eletrônico de Gestão do Transporte Escolar* (SETE – Electronic School Transportation Management System). This system is designed to cater to municipalities with varying management needs, from those relying on local spreadsheets to those without digital control of their routes. To achieve this, SETE integrates various aspects of school transport management, including fleet management, student management, and routing, which are typically handled separately by other systems. It is important to note that SETE is not meant to replace existing systems, but rather complement them (FNDE and CECATE-UFG, 2019).

Among SETE's modules, route management stands out, as it provides ways to manage and understand how school bus routes work. It enables users to describe routes using various techniques, depending on the municipality's management maturity. In particular, at its lowest level, the manager will only provide a textual description of the route, including mileage and the number of students transported. It is also possible to digitize the route using georeferencing or drawing tools. In the latter case, SETE relies on the OpenStreetMap (OSM) collaborative platform as a means of assisting these tasks, since it is free and open-source (Foody et al., 2017; Haklay and Weber, 2008; Neis and Zielstra, 2014). The provided background map guides the managers in drawing the routes. Furthermore, unlike other systems, such as Google Maps and Waze, OSM maps can be freely integrated and processed by different applications, such as simulators (Lopez et al., 2018) and geographic information systems (Mooney and Minghini, 2017).

Compared to proprietary solutions, one advantage of the OSM platform is its ability to extract a mathematical graph representation of the road network from the system. This data is a crucial element for several transport planning activities, such as route optimization (Tong et al., 2021; Wang and Huang, 2017) and accessibility analysis (Benevenuto and Caulfield, 2020; Sdoukopoulos et al., 2024). However, since the platform is based on collaborative data, the extracted road network graph may present small gaps (Camboim et al., 2015; Mahabir et al., 2017), especially in rural areas, as in cases where there is no connection between a road and the entrance to a farm. Such failures can generate disconnections in the network, making it difficult to use the network in other systems. For example, in the cited case, a routing or accessibility algorithm may fail to reach the student's origin, as there is no link between them and the remaining nodes of the network.

Due to the high potential for reusing OSM data in other systems, other authors have investigated these weaknesses, i.e., the degree of completeness of the network graph. For instance, Camboim et al. (2015) observed a significant difference between the OSM rural roads with respect to the official mapping. The closer the municipalities were to the capital (Curitiba, Paraná, Brazil), the smaller the difference between the maps. Similarly, Elias et al. (2021) show a difference between OSM data and Salvador's (Bahia, Brazil) official cartography. However, it should be noted that the authors limited themselves to the main roads, such as those described in the IBGE database or official cartography, excluding rural and local roads.

The lack of knowledge of the completeness of rural roads regarding the OSM network also occurs in other countries. To exemplify, Minaei (2020) analyzed Iran's roads from 2008 to 2016. However, it does not delve into the length or the number of missing roads. Likewise, Zhang and Malczewski (2018) analyze Canada's roads, focusing on the number of missing tags instead of the road's track.

To address this issue, this work aims to investigate the completeness of the OSM road network compared to Brazilian school transport routes, particularly in the context of the rural road network. By comparing the OSM network with school transportation routes, the study aims to reveal whether other tools can rightly use the OSM data. It is essential to emphasize that this research is not limited to the scope of SETE, as identifying the quality of the network helps understand the possibility of using it for other activities and systems, such as route optimization and accessibility analysis. For this purpose, this work developed a computational method to analyze the completeness of the road network.

In addition to presenting the results concerning the overall completeness of the road network, the principal contribution of this study lies in the development of a methodology for systematically comparing and analyzing the completeness of school bus routes relative to OSM network data. The proposed framework is adaptable and can be applied to the evaluation of other types of routes, including those used by delivery services and freight transport. Moreover, the methodology is extensible to incorporate alternative road network datasets, including proprietary sources such as Google Maps and Azure/Bing Maps.

To detail the proposed method, this article is divided into five sections. Section 2 presents the theoretical framework that supports this work. Section 3 illustrates the steps of the proposed method, while Section 4 presents the experiment used and the analysis of the results obtained. Finally, Section 5 presents the conclusion and recommendations for future work to address the limitations of the proposed method.

2. FUNDAMENTAL CONCEPTS

To evaluate the completeness of the OSM network, it is necessary to understand both the school transportation phenomenon in its urban and rural context and the spatial and computational technologies involved in such tasks. In this sense, this section provides a concise review of these concepts.

2.1. School transportation system

The Brazilian School Transportation System is composed of urban and rural subsystems. In its rural context, the goal of the system is the movement of students from their residency to the public education network, which consists of schools that can be located in urban or rural areas

(Carvalho et al., 2010). The objective is to move students along the home-school-home route to facilitate their access and stay in schools, improving the conditions for public education provision (Nascimento and Andrade, 2022).

The right to education, as well as transportation, teaching materials, and food, is guaranteed by the Brazilian constitution through the availability of state supplementary programs (Medrano and Carvalho, 2021; Medrano et al., 2021). As a part of the Brazilian population still resides in rural areas, the state created several school transportation programs to guarantee this right. In the rural context, students usually live far away from schools, on farms, or in riverside areas (Carvalho et al., 2016). Nevertheless, Brazilian municipalities are obliged to comply with the constitution and offer rural school transport with the support of the federal government.

School transportation can be offered by the municipality's fleet or through a private-sector service concession. For this, complex planning is necessary. This can be a problem, as most municipalities do not have trained technicians, and services are not always optimized (Caldas et al., 2022). To mitigate these issues, FNDE (*Fundo Nacional de Desenvolvimento da Educação* – Brazil National Education Development Fund) seeks to support municipalities through public policies by providing training, financing, and management tools.

Computer systems can be used to support these municipalities, as they can provide functions to organize activities and guide people working in different areas. However, it was observed that most existing software systems focused on isolated management steps, such as routing or registering service order registration (FNDE and CECATE-UFG, 2019). Consequently, such systems are limited to municipalities with higher levels of administrative maturity, as they require several types of data as input. For example, the data and road network must be georeferenced to use a routing-based system. Obtaining this data, whether directly or indirectly, can be a significant barrier for a large portion of Brazilian municipalities that still lack the necessary technological capacity (Pinheiro Jr., 2019; Przeybilovicz et al., 2018).

The restrictions of existing software aimed at school transport management, such as the need for multiple input data, including the geolocation of students and schools, aligned with the focus on a specific facet of management, served as motivation for the construction of *Sistema Eletrônico de Gestão do Transporte Escolar* (SETE - Electronic School Transportation Management System). To support municipalities in route management, the system enables georeferencing via smartphone GPS. For example, the mobile application OsmAnd uses the smartphone's GPS sensor to obtain the route course (Novikova, 2022). If the user does not want to track the route, SETE provides means for him/her to draw the school bus route digitally on a map. In this case, the system uses the OpenStreetMap (OSM) mapping platform to guide the user in this task because OSM is free and open-source (Haklay and Weber, 2008; Senaratne et al., 2017).

Georeferenced routes enable managers to better understand their characteristics, such as mileage and time, and facilitate analysis for potential optimization and accessibility levels. However, to perform these tasks, such as optimizing or calculating accessibility, it is necessary to access the underlying road network. In this sense, the OSM platform has an advantage over other proprietary map systems. Because it is free and open-source, it is possible to extract a representation of the road network (Boeing, 2017; Giraud, 2022; Pereira et al., 2021).

Specifically, the road network can be seen as a graph $G=(V,E)$ with weights, where the edges (E) represent the paths, while the vertices (V) represent the connection nodes between them. Each edge has an attribute, which commonly represents the distance between nodes. Additionally, users can add other metadata, such as maximum speed, type of pavement, and number of lanes. Figure 1 illustrates an example of a city's OSM network represented as a graph. Note that each edge represents the connection between nodes and the distance between them. For example, the edge $(1,2,200m)$ means that there is a connection from vertex 1 to vertex 2 whose distance is 200 meters. Note that there is a lack of connection between nodes 6 and 7. This means there is a disconnection from vertex 7 to the rest of the network. Consequently, routing or accessibility analysis algorithms processing this graph will be unable to compute paths that reach vertex 7.

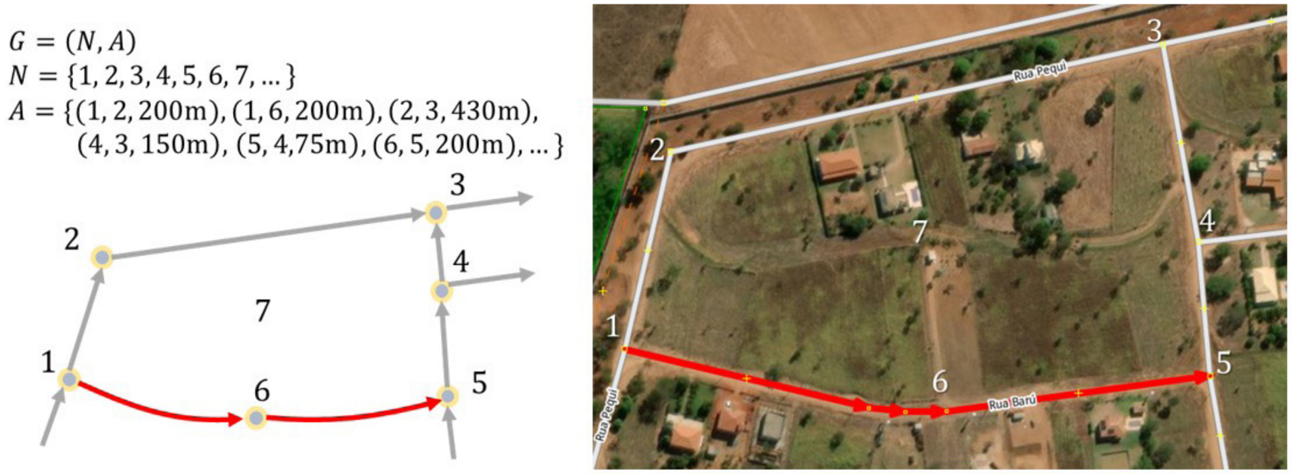

Figure 1. Example of a road network graph extracted from OSM.

Understanding the extent to which such disconnections affect school transport routes is the central research element of this paper. In this sense, the following section delves deeper into the concepts relating to the road network.

2.2. OpenStreetMap

As stated, the OpenStreetMap (OSM) platform is an open-source Volunteered Geographic Information (VGI) system for geographical information (Haklay and Weber, 2008; Senaratne et al., 2017). Similar to Wikipedia, users can edit the platform's maps to update or provide new spatial information, such as routes, courses, and building information. Each update to the network data is recorded in a history log, making it possible to reverse them. The community actively checks and monitors drastic changes (Choe et al., 2023; Neis et al., 2012). In fact, vandalism actions are not common and are generally limited to areas of territorial dispute, such as the Crimea region in Ukraine/Russia or roads in Jerusalem (Bull Floyd, 2019).

Several studies indicate that OSM maps are highly similar to government data and private platforms (Barrington-Leigh and Millard-Ball, 2017; Elias et al., 2021). One way to achieve this is that, in addition to receiving data from volunteers, OSM provides mechanisms that facilitate the insertion of government databases. For example, there are several initiatives to import and annually update databases from IBGE and other Brazilian government agencies into the platform (OSM, 2019, 2023a, 2023b).

Another positive aspect of OSM compared to proprietary platforms is the ability to add various annotations (tags) to enrich the edges and vertices of the road network. Figure 2 illustrates some of the existing tags. Note that, in addition to classifying roads, such as highways and residential areas, there is also the possibility of indicating the presence of barriers at network nodes, such as cattle grids and gates. It is also possible to specify the type of road surface, such as asphalt (paved) or compacted soil (unpaved), which serves as valuable input in route optimization and accessibility analysis tools.

(a) Cattle Grid (b) Federal highway with a police station

Figure 2. Example of tags used by an OSM road network (vertex and edges).

Unfortunately, filling all these tags is not mandatory for contributing to OpenStreetMap, as a newly created edge or node only needs a single tag (Davidovic et al., 2016). In fact, according to Mooney and Corcoran (2012), only 50% of objects use more than six tags. Nevertheless, despite these limitations, results indicate that, globally, OSM's global road network is over 80% complete when compared to real-world data. In the context of Brazil, a report from Camboim et al. (2015) indicates a completeness of over 80% for roads near Curitiba, the capital of the Paraná state in Brazil. However, the remainder of the roads present lower thresholds of completeness. It is essential to note that the study limited the ground truth to roads defined by the government's topographic database. This data can be outdated and does not include footpaths or access roads to farms and riverside areas.

The network graph can be freely exported to other systems (Geofabrik, 2023; Juhász et al., 2016) into various spatial data formats, including GeoJSON, Shapefile, and KML (Keyhole Markup Language). Such data is described as a graph, composed of vertices and arcs, allowing it to be processed by other systems. Typically, the exported data is first stored in one of these formats to be pre-processed in a spatial database.

2.3. Spatial databases

Spatial databases are commonly used to accelerate the processing of the graph extracted from the OSM data (Obe and Hsu, 2021). Among such databases, PostgreSQL stands out, as it is both free and open-source software, and it is capable of storing and processing large volumes of data. Furthermore, it is capable of providing various spatial operations, such as obtaining the intersection and union of spatial objects (Felício et al., 2022). Specifically, this is done through the PostGIS module, which adds functions for storing, searching, and indexing spatial data (PostGIS, 2023). Table 1 provides a brief description of some of these functions.

Table 1: Some spatial functions provided by PostGIS

Function	Description
ST_ApproximateMedialAxis(g)	Smoothes the path of *g* using the median axis of the spatial geometry
ST_Buffer(g, x)	Creates a buffer of x meters around a geometry *g*
ST_Difference(g1, g2)	Returns the geometry resulting from the difference between *g1* and *g2*
ST_Intersection(g1, g2)	Returns the geometry resulting from the intersection of *g1* and *g2*
ST_GeomFromGeoJSON(t)	Constructs a spatial geometry from data *t* in GeoJSON format
ST_GeomFromKML(t)	Constructs a spatial geometry from data *t* in KML format
ST_Length(g)	Calculates the distance in meters of a given geometry (*g*)
ST_Union(g1, g2)	Create a geometry from the union of *g1* and *g2*

The PostGIS extension employs spatial geometry to represent the combination of graph vertices and edges, allowing multiple paths to be merged into a single geometry. Vertices serve as connection points between edges and are indexed within the database table, allowing for efficient queries. For example, the *ST_Intersection* function can identify where edges cross, while *ST_Difference* retrieves portions of a geometry that do not overlap with another, which is crucial for detecting missing road segments.

3. METHODOLOGY

Based on the concepts previously described, this section presents a quantitative, applied, and descriptive method that uses real-world school bus routes to compute gaps in the OSM road network graph. The proposed method is divided into three stages: data loading, route processing, and road comparison, as described throughout this section.

3.1. Data loading

The first part of the method consists of loading the OSM ground-truth network data into a spatial database. Figure 3 illustrates the steps taken. We used the GeoFabrik service to download the data, which compiles and makes the OSM data available (GeoFabrik, 2023). The compressed dataset, totaling 1.8 GB, can be downloaded for free through this service.

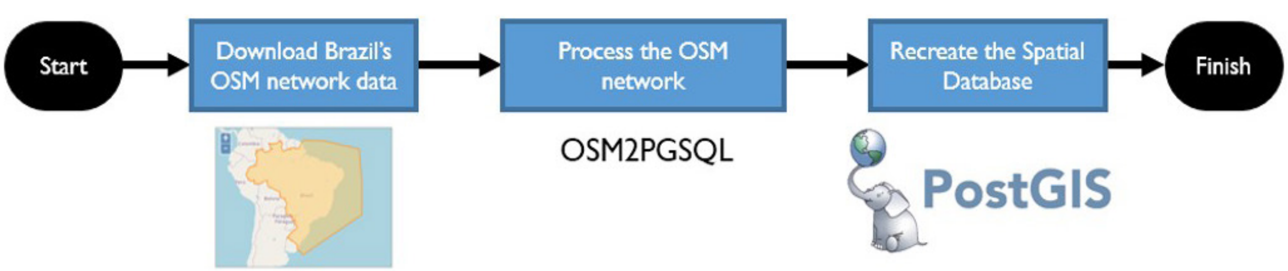

Figure 3. Sequence of steps for loading the OSM network into a spatial database.

The downloaded data is stored in a compressed XML (Extensible Markup Language) file defined by OSM (*.osm.pbf*). Therefore, first, it must be transformed into a spatial database, such as PostGIS (which extends PostgreSQL with support for geographic data). This format allows the network graph to be accessed and processed using programming and query languages like Python and SQL. To accomplish this, we utilized the OSM2PGSQL tool (OSM2PGSQL, 2025), a free and open-source software for handling OSM data. This software reads the raw XML data and recreates the spatial database.

As an alternative to recreating a spatial database, it is possible to load the OSM file directly into a Geographic Information System (GIS), such as QGIS or ArcGIS Desktop (ArcMap). However, considering that the network contains every single road in the country, as is the case in our scenario, this task can be computationally expensive. The advantage and recommendation for using a database arises from its ability to create indexes that allow the user to quickly retrieve and process only the parts he/she need of the network using instructions specified by the programming languages.

3.2. Data processing

After loading the OSM data into a spatial database, the next step is to compare the school bus routes with this network. Figure 4 illustrates the stages of this process. First, we need to obtain the data for each bus route we intend to analyze. As such, we extracted all georeferenced routes available in SETE.

Commonly, these routes are stored in spatial formats, such as GeoJSON and Shapefile. SETE, the Brazilian Electronic School Transportation Management System, uses GeoJSON to store the bus routes. Hence, we chose this format as it can be easily processed and imported into the database using the Python programming language. However, we highlight that other spatial formats can be used. The only caveat is that it needs to be able to transform them into a database geometry, which is valid for several spatial formats, such as GeoJSON, KML, and Shapefile. To transform the data into a geometry, we used the PostGIS function *ST_GeomFromGeoJSON*. As a result, we obtain the school bus geometry.

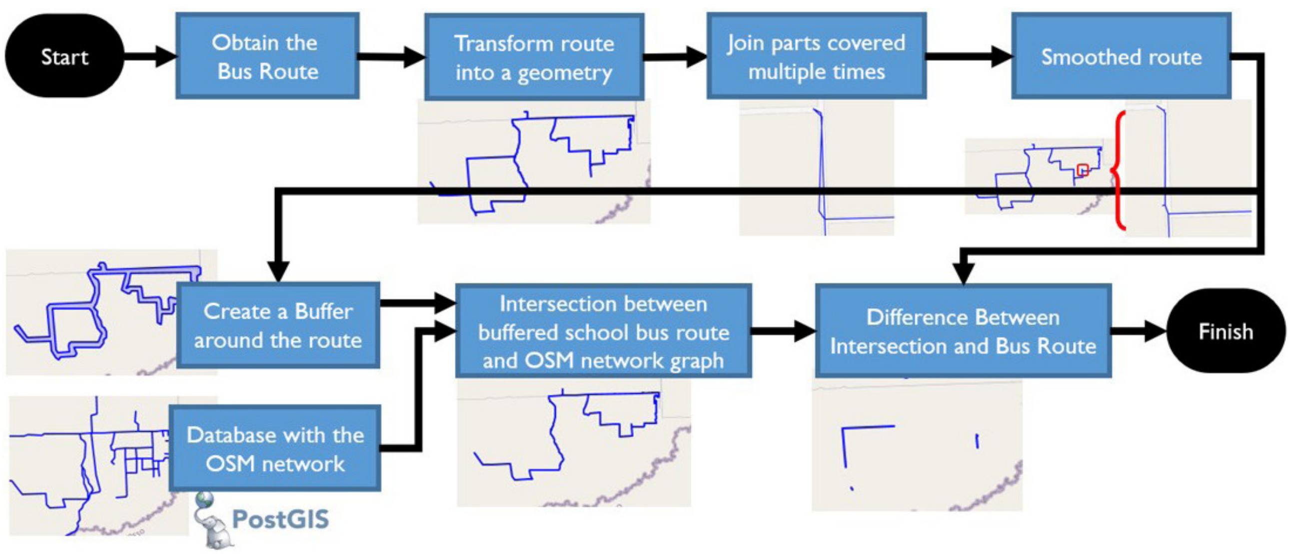

Figure 4. Sequence of steps for computing gaps in the OSM network using school bus routes.

Before proceeding, we observed minor variations in the collected bus routes. As many municipalities use GPS devices to obtain their routes, overlapping lines can appear when traveling part of the route multiple times due to GPS errors, as a given route may pass through the same location numerous times. It is essential to remove this inaccuracy; otherwise, the traveled road will be counted multiple times if it does not exist in OSM. For example, if the local road that gives access to the farm is not present in the OSM network, it will be computed twice: once when the vehicle enters to pick up the student and once when it travels back to continue the trip.

To address this issue, we simplified the route path by joining overlapped road segments in the school bus route. To exemplify our solution, consider Figure 5, where a school bus picks up a student and then returns to the main avenue to continue the route (once to the student's home and once back to the main road). These variations can cause the resulting geometry to present unaligned and overlapping tracks in the same trip.

To fix this issue, we apply some spatial operations to merge and smooth the bus route. First, we load the data into PostGIS by using the *ST_GeomFromGeoJSON* function to obtain a spatial geometry from GeoJSON data. Then, we create a buffer around the road segments with the *ST_Buffer* operation. Finally, using the resulting buffer geometry, we create a smooth track by applying the *ST_ApproximateMedialAxis* function, which returns the median line of the geometry. In our algorithm, we use a threshold of 20 meters to join them, as the track is typically generated by the same GPS device, and this threshold is sufficient to accommodate errors in Brazilian rural areas (Oliveira et al., 2019). Therefore, we create a buffer of 20 meters, join them, and afterward compute the smooth road using the *ST_ApproximateMedialAxis* function.

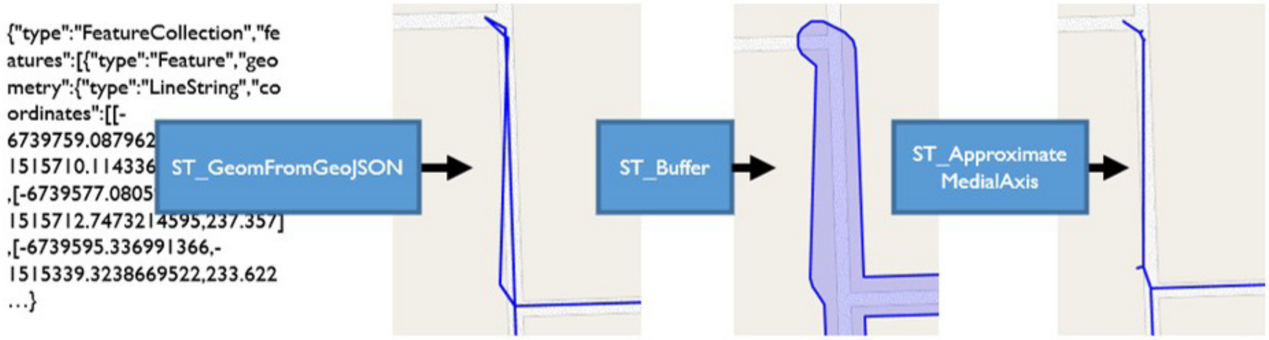

Figure 5. Example of PostGIS operations to smooth the route data.

After obtaining the smoothed route, the next step is to find how close the bus route is to the underlying OSM network. Precisely, we want the intersection between these two graphs. To do this, we first create a buffer for the school bus route under analysis. In our case, a 50-meter buffer was created around each lane of the bus route. We used a more significant threshold, 50 meters instead of 20 meters, as there can be a higher discrepancy between the vehicle's GPS device and the one provided by the OSM collaborator (Oliveira et al., 2019).

3.3. Road comparison

To calculate the completeness of the OSM road network, we intersected the buffered GPS-based school bus routes with the OSM features. Completeness was evaluated according to the ISO (2013) standard, specifically through a direct external evaluation of the completeness omission element. The resulting metrics include both the total length and the percentage of route segments not represented in the mapped network, indicating missing features in the dataset. We perform a complete inspection of the dataset, that is, we do not sample the 7,159 school bus routes but analyze the dataset in its entirety.

This is a time-consuming process, as the intersection happens between the buffered polygon and the entire road network. As a result, a subroute network is obtained, representing all edges in the OSM network that correspond to the school bus route under analysis. The subroute generated by the intersection of the bus route with OSM serves as input to calculate the gap in the network. This is achieved by comparing the intersected part with OSM (subroute) to its original shape (smoothed bus route).

To do so, we use the *ST_Difference* function, which returns all geometries in the school bus route that do not exist in the OSM network. To compute the distance of this gap, the *ST_Length* function is used, which returns the size of each gap in meters. Again, note the importance of smoothing the route before beginning this process. If this is not done, a greater distance may eventually be obtained due to the existence of back-and-forth lines in the network relating to the school bus route's geometry.

4. RESULTS

The proposed method was executed with data from school bus routes extracted by SETE in October 2023. The data includes routes traced by GPS devices and also those drawn by Brazilian managers in the system. Although the database contains 27,121 routes, only 7,159 (33.89%) had been georeferenced by users at that time. The 7,159 georeferenced routes were identified across a total of 482 distinct municipalities.

4.1. Experiment requirements

Before presenting the results, we detail the computational infrastructure used to execute the proposed method. We implemented the algorithms using the Python programming language and the PostgreSQL database with the PostGIS extension enabled for data storage. The experiments were run on a computer with the following specifications: (a) Operating System: Microsoft Windows 11; (b) Processor: Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz, 1992 Mhz, 4 Cores; (c) Memory: 24 GB DDR4 @ 2133 MHz; (d) Storage: 500 GB SSD M.2

The complete OSM network data for the entire country of Brazil required approximately 12 GB of storage, while the SETE roads occupied around 174 MB. On average, it took 14.51 seconds to analyze each school bus route.

4.2. Data results

In total, the results show that the road network covered by the 7,159 school bus routes amounts to 189,689,671.407 kilometers. Of this total, approximately 9.78% of the road segments, 18,547,390.25 kilometers, are not represented in OSM.

To better understand how the map network covers the school bus routes, we built a Pareto diagram, illustrated by Figure 6. The graph shows the proportion of routes that are fully covered by OSM as the coverage threshold is progressively relaxed, offering insights into the spatial completeness of the map data relative to the transportation network.

The results indicate that the OSM network can cover 100% of the 32.92% of the bus routes, that is, 2,357 routes out of 7,159. When we consider a lower coverage, for instance 90% $\left[90\%,95\%\right)$, we can see that the network contains at least 90% of 76.78% of the routes. A plateau is formed when we consider a network coverage of 70%, this percentage includes 90% of the school bus routes.

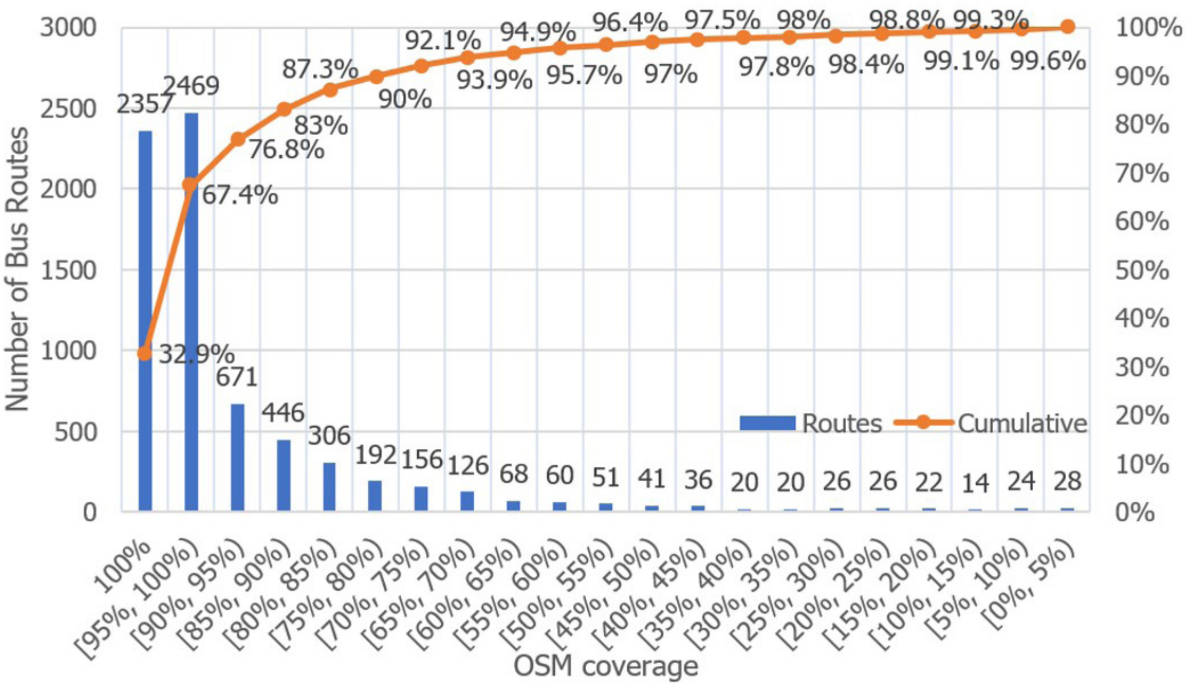

Figure 6. Pareto diagram that illustrates how the percentage of the total analyzed route is present in OSM increases as we relax the coverage threshold.

To better understand the possible gaps in the network, Table 2 summarizes the routes analyzed, considering their location. The routes' average mileage and percentage in the OSM are also included. It is essential to highlight that these values consider the route path after the smoothing operation, that is, after combining the parts where the vehicle traveled several times.

The data shows that the OSM network covers 90% of the bus routes in most Brazilian states. Specifically, the results suggest that, on average, 91.84% of the routes are present in the OSM network. When grouping the data by region, it is observed that the north and northern part of the country has the lowest completeness index, with Amazonas at 32.27% and Maranhão at $71,30\%$. In contrast, the Federal District and the state of Espírito Santo bus routes have 99.78% and 97.74% coverage, respectively.

When the numbers for the state of Amazonas are broken down, we observe that many of the gaps occur due to the routes being waterways or mixed. Figure 7 (a) illustrates an example of this type of route. Note that the only part present in OSM refers to the path located on land. The occurrence of floods and droughts can modify the network, making it difficult to map these roads in map systems.

A concentration of gaps was also observed on the fringes (branches) of the analyzed routes, that is, in the last miles. Figures 7 (b), (c), and (d) illustrate such scenarios. These fringes often refer to local roads leading up to a farm entrance or riverbank regions.

Table 2: Description of the routes analyzed by each Brazilian state

State	# Cities	# Routes	Average km of the Routes (after smoothing)	Average % of OSM network coverage
Acre	1	1	18.34 km	83.29%
Alagoas	9	155	14.93 km	96.50%
Amapá	1	3	7.15 km	96.66%
Amazonas	3	39	11.55 km	32.27%
Bahia	38	748	17.97 km	89.58%
Ceará	7	136	14.32 km	93.07%
Distrito Federal	1	1	25.25 km	99.78%
Espírito Santo	7	232	12.59 km	97.74%
Goiás	15	272	52.62 km	87.89%
Maranhão	5	19	16.07 km	71.30%
Mato Grosso	20	175	49.51 km	82.55%
Mato Grosso do Sul	11	131	50.36 km	84.58%
Minas Gerais	93	1150	30.33 km	92.55%
Pará	18	171	21.65 km	84.93%
Paraíba	8	43	18.04 km	91.16%
Paraná	32	751	24.58 km	97.35%
Pernambuco	21	817	16.12 km	92.70%
Piauí	11	177	19.19 km	84.92%
Rio de Janeiro	7	36	11.00 km	96.62%
Rio Grande do Norte	10	128	14.97 km	96.03%
Rio Grande do Sul	59	456	31.31 km	91.62%
Rondônia	26	424	31.83 km	93.79%
Santa Catarina	35	222	23.09 km	97.16%
São Paulo	26	369	28.47 km	96.13%
Sergipe	3	133	14.89 km	96.63%
Tocantins	15	370	36.04 km	85.76%
TOTAL	482	7159	—	—

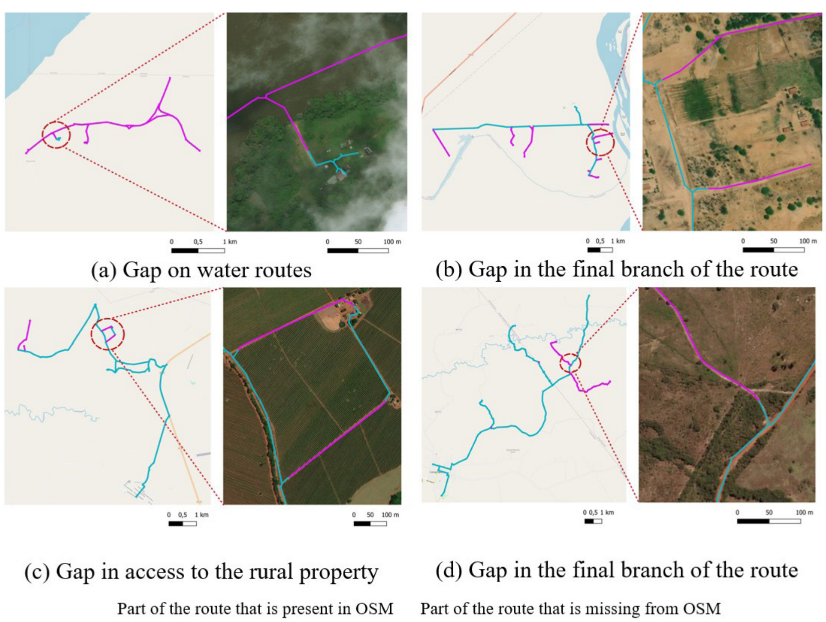

Figure 7. Example of gaps found in the OSM network.

In principle, gaps in the last mile do not entirely harm the OSM network since the generated graph is still continuous, and there is no disconnection between substantial parts of the network. However, considering these locations are commonly used for picking up students, it is difficult to readily use the network for routing algorithms or accessibility methods, as their origins are lost. Alternatively, users can project the student's boarding or disembarking location to the nearest road.

5. CONCLUSION AND FUTURE WORK

This research analyzed a relevant sample to measure the OpenStreetMap (OSM) completeness, considering more than 7 thousand real-world school bus routes from different parts of Brazil. The results show that, despite still needing some updates, the OSM network can cover a significant part of the roads used by the Brazilian school transportation system.

The results suggest that the existing OSM network fully covered more than 32% of the routes analyzed, and more than 87% of the routes had at least 80% coverage. The main gaps were identified in sections related to waterways and access to the students' homes, mainly those located in rural areas, such as local roads to farms and riverbank regions, which are branches of local roads.

Hence, methods that utilize waterways or require access to students' home locations, such as route optimization and accessibility analysis, need to update the network or preprocess the data to address these issues. In some cases, such as for generating origin-destination matrices, these updates can be critical for obtaining a correct network; otherwise, a bus route may fail to reach the last mile and, as a result, the student's home. To do so, one approach is for users to utilize the Java OpenStreetMap Editor (JOSM) tool, a graphical software that facilitates the insertion, editing, or removal of data from OSM.

Regarding preprocessing approaches, another solution is to artificially relocate the user location to the closest road in OSM. This approach can be useful in scenarios where only a few parts of the road are missing, but it will fail to provide a useful network if the missing part is sufficiently large, which will impact route generation and accessibility analysis tools.

On the other hand, a promising solution is the use of machine learning to automatically map missing road segments. Recently, Facebook's (META) AI lab released the "RapiD" editor (*https://rapideditor.org/*), which completes these roads by extracting paths from high-resolution satellite images (Schröder-Bergen et al., 2021). Figure 8 illustrates a screenshot of the tool, where road segments detected from satellite images are highlighted in purple. By selecting these segments (highlighted in yellow and purple), users can choose to add them manually to the OSM database. Notice that not all road segments are detected automatically (area pointed by the green text). That is why the tool does not automatically insert the new road segment into the OSM database. Instead, it is intended to support and enhance users in quickly identifying and adding missing roads with a single click. In a recent study, Fila, Štampach, and Herfort (2025) reviewed the current limitations of such AI-assisted approaches and discussed how companies and researchers are working to improve machine learning models to enhance this process.

It is important to highlight that the presented method has some hard assumptions. In particular, we considered that the paths covered by the school bus route under analysis actually exist. Since the GPS routes were drawn or tracked by the city managers, they may travel through places that are not existing roads, such as shortcuts or inside a farm. For example, when a route passes through the middle of a farm, the method considers that this part of the route should be a road. As it is not present in the municipality's road infrastructure, the route will be considered in the comparison with the OSM road network, causing a possible gap.

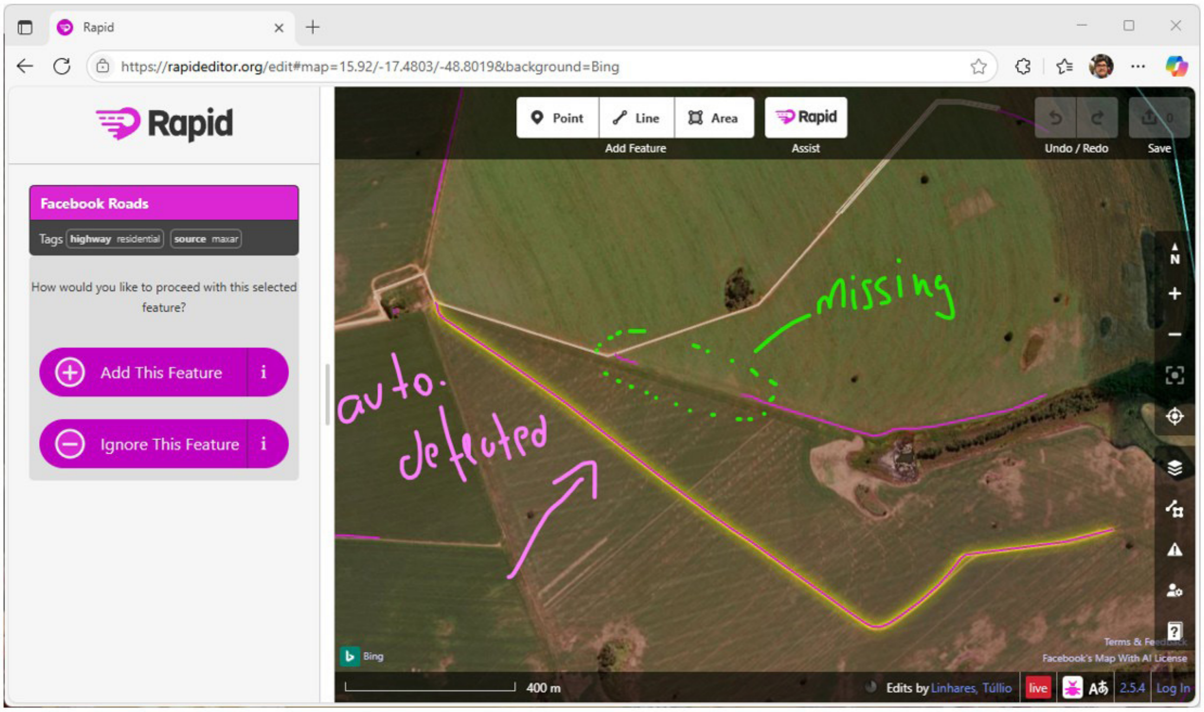

Figure 8. Screenshot of Facebook's Rapid editor, the highlighted road in purple/yellow is extracted from a high-quality satellite image.

To address this issue, in future work, we intend to briefly delve into existing bus routes analyzed to ensure they are correct using satellite images. In addition, we intend to investigate the completeness of the road tags since the presence of attributes such as surface (type of pavement) and maximum speed (maxspeed) can assist in the analysis of network data by other systems. Another improvement we plan to investigate is the use of map-matching algorithms during the route processing phase, particularly when identifying the portions of the buffered school bus routes that align with the OSM road network. Our goal is to assess whether the current buffering-based approach might include additional road segments beyond those actually traversed.

AUTHORS' CONTRIBUTIONS

MPRJ: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing; WLC: Conceptualization, Methodology, Validation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition; DDM: Conceptualization, Validation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

CONFLICTS OF INTEREST STATEMENT

The authors declare no conflict of interest.

USE OF ARTIFICIAL INTELLIGENCE-ASSISTED TECHNOLOGY

The authors declare that no artificial intelligence tools were used in the research reported here or in the preparation of this article.

DATA AVAILABILITY STATEMENT

The data supporting this study were provided by the National Fund for Educational Development (FNDE), and the authors do not have the right to redistribute them. Requests for access to the data should be directed to the data provider.

REFERENCES

Barrington-Leigh, C. and A. Millard-Ball (2017) The world's user-generated road map is more than 80% complete. *PLoS One*, v. *12*, n. 8, p. e0180698. DOI: 10.1371/journal.pone.0180698. PMid:28797037.

Benevenuto, R. and B. Caulfield (2020) Measuring access to urban centres in rural Northeast Brazil: A spatial accessibility poverty index. *Journal of Transport Geography*, v. *82*, p. 102553. DOI: 10.1016/j.jtrangeo.2019.102553.

Boeing, G. (2017) OSMnx: A Python package to work with graph-theoretic OpenStreetMap street networks. *Journal of Open Source Software*, v. *2*, n. 12, p. 215. DOI: 10.21105/joss.00215.

Brazil, Federal Supreme Court (2022) Constitution of the Federative Republic of Brazil. Brasília, DF: STF.

Bull Floyd, S.Q. (2019) Understanding threats to crowdsourced geographic data quality through a study of OpenStreetMap contributor bans. In Valcik, N.A. (ed.) *Geospatial Information System Use in Public Organizations*. Routledge, p. 80-96.

Caldas, L.; R. Martinelli and B. Rosa (2022) Solving a school bus routing problem in rural areas: an application in Brazil. *In* Jesica, A.; H. Ramalhinho and V. Stefan (eds.) *Computational Logistics: 13th International Conference*. Cham: Springer International Publishing, p. 162–176. DOI: 10.1007/978-3-031-16579-5_12.

Camboim, S.; J. Bravo and C. Sluter (2015) An Investigation into the Completeness of, and the Updates to, OpenStreetMap Data in a Heterogeneous Area in Brazil. *ISPRS International Journal of Geo-Information*, v. *4*, n. 3, p. 1366-88. DOI: 10.3390/ijgi4031366.

Carvalho, W.L.; R.O. Cruz; M.T. Câmara et al. (2010) Rural school transportation in emerging countries: The Brazilian case. *Research in Transportation Economics*, v. *29*, n. 1, p. 401-9. DOI: 10.1016/j.retrec.2010.07.051.

Carvalho, W.L.; Y. Yamashita and J.J.Q. Aragao (2016) Rural School Transportation in Brazil as an Essential Factor For The Education Policy. *International Journal of New Technology and Research*, v. *2*, n. 1, p. 6-12.

Choe, Y.; M. Tomko and M. Kalantari (2023) Assessing Mapper Conflict in OpenStreetMap Using the Delphi Survey Method. *In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. New York: Association for Computing Machinery, p. 1–17. DOI: 10.1145/3544548.3580758.

Davidovic, N.; P. Mooney; L. Stoimenov et al. (2016) Tagging in volunteered geographic information: An analysis of tagging practices for cities and urban regions in OpenStreetMap. *ISPRS International Journal of Geo-Information*, v. *5*, n. 12, p. 232. DOI: 10.3390/ijgi5120232.

Elias, E.N.N.; F.R. Amorim; M.A.R. Schmidt et al. (2021) Discovering spatiotemporal patterns on data quality assessment in collaborative mapping: a preliminary study in an area of Brazil. *Abstracts of the ICA*, v. 3, p. 73. DOI: 10.5194/ica-abs-3-73-2021.

Felício, S.; J. Hora; M.C. Ferreira et al. (2022) Handling OpenStreetMap georeferenced data for route planning. *Transportation Research Procedia*, v. *62*, p. 189-96. DOI: 10.1016/j.trpro.2022.02.024.

Fila, M.; R. Štampach and B. Herfort (2025) AI-generated buildings in OpenStreetMap: frequency of use and differences from non-AI-generated buildings. *International Journal of Digital Earth*, v. *18*, n. 1, p. 2473637. DOI: 10.1080/17538947.2025.2473637.

FNDE and CEFTRU (2010) *Manual de Planejamento do Transporte Escolar Rural*. Brasília: FNDE.

FNDE and CECATE-UFG (2019) *Relatório de Engenharia do Sistema Eletrônico de Gestão do Transporte Escolar (SETE)*. Goiânia: Universidade Federal de Goiás.

Foody, G.; L. See; S. Fritz et al. (2017) *Mapping and the Citizen Sensor.* London: Ubiquity Press. DOI: 10.5334/bbf

Geofabrik (2023) *OpenStreetMap Data Extracts: Brazil*. Available at: <https://download.geofabrik.de/south-america/brazil.html> (accessed 10/28/2025).

Giraud, T. (2022) osrm: interface between R and the OpenStreetMap-based routing service OSRM. *Journal of Open Source Software*, v. *7*, n. 78, p. 4574. DOI: 10.21105/joss.04574.

Haklay, M.M. and P. Weber (2008) OpenStreetMap: user-generated street maps. *IEEE Pervasive Computing*, v. *7*, n. 4, p. 12-8. DOI: 10.1109/MPRV.2008.80.

IBGE (2023) P*roposta Metodológica para Classificação dos Espaços do Rural, do Urbano e da Natureza no Brasil*. Rio de Janeiro: IBGE, 174 p. ISBN. 978-85-240-4575-2.

ISO (2013) *ISO 19157:2013 Geographic information — Data quality*. Geneva: ISO. Available at: <https://www.iso.org/standard/32575.html> (accessed 10/28/2025).

Juhász, L.; A. Rousell and J. Jokar Arsanjani (2016) Technical guidelines to extract and analyze VGI from different platforms. *Data*, v. *1*, n. 3, p. 15. DOI: 10.3390/data1030015.

Lopez, P.A.; E. Wiessner; M. Behrisch et al. (2018) Microscopic Traffic Simulation using SUMO. *In 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. Maui: Institute of Electrical and Electronics Engineers, p. 2575–2582. DOI: 10.1109/ITSC.2018.8569938.

Mahabir, R.; A. Stefanidis; A. Croitoru et al. (2017) Authoritative and volunteered geographical information in a developing country: a comparative case study of road datasets in Nairobi, Kenya. *ISPRS International Journal of Geo-Information*, v. *6*, n. 1, p. 24. DOI: 10.3390/ijgi6010024.

Medrano, R. and W. Carvalho (2021) Factors that influence the evaluation of the "Caminho da Escola" program: an analysis based on the perception of managers, school principals and members of the social monitoring and control council. *Case Studies on Transport Policy*, v. *9*, n. 4, p. 1677-87. DOI: 10.1016/j.cstp.2021.09.002.

Medrano, R.M.A.; W.L. Carvalho; Y. Yamashita et al. (2021) A resource allocation model of national school transportation program for municipalities. *Case Studies on Transport Policy*, v. *9*, n. 4, p. 1660-71. DOI: 10.1016/j.cstp.2021.08.011.

Minaei, M. (2020) Evolution, density and completeness of OpenStreetMap road networks in developing countries: the case of Iran. *Applied Geography (Sevenoaks, England)*, v. *119*, p. 102246. DOI: 10.1016/j.apgeog.2020.102246.

Mooney, P. and P. Corcoran (2012) The annotation process in OpenStreetMap. *Transactions in GIS*, v. *16*, n. 4, p. 561-79. DOI: 10.1111/j.1467-9671.2012.01306.x.

Mooney, P. and M. Minghini (2017) A review of OpenStreetMap data. In Foody, G.; L. See; S. Fritz et al. (eds.) *Mapping and the Citizen Sensor*. London: Ubiquity Press, p. 37–59. DOI: 10.5334/bbf.c.

Nascimento, M.V.L.A. and M.O. Andrade (2022) School transportation program as means to improve public education in a minor rural town in Northeastern Brazil. *Ensaio*, v. *30*, n. 114, p. 182-206. DOI: 10.1590/s0104-40362021002903093.

Neis, P.; M. Goetz and A. Zipf (2012) Towards automatic vandalism detection in OpenStreetMap. *ISPRS International Journal of Geo-Information*, v. *1*, n. 3, p. 315-32. DOI: 10.3390/ijgi1030315.

Neis, P. and D. Zielstra (2014) Recent developments and future trends in volunteered geographic information research: the case of OpenStreetMap. *Future Internet*, v. *6*, n. 1, p. 76-106. DOI: 10.3390/fi6010076.

Novikova, A. (2022) *9 Best GPX Tracker Apps for Android & iOS*. Free Apps For Me. Available at: <https://freeappsforme.com/gpx-tracker-apps/> (accessed 10/28/2025).

Obe, R.O. and L.S. Hsu (2021) *PostGIS in action* (3rd ed.). Shelter Island: Manning Publications.

Oliveira, G.A.; J. Almir Cirilo; P. Lustosa Brito et al. (2019) Qualidade do posicionamento em aplicativos VGI obtido por sensores de localização em smartphones. *Revista Brasileira de Cartografia*, v. *71*, n. 3, p. 806-31. DOI: 10.14393/rbcv71n3-49478.

OSM (2019) *Importação das Redes Geodésicas do IBGE*. OpenStreetMap Wiki. Available at: <https://wiki.openstreetmap.org/wiki/Importa%C3%A7%C3%A3o_das_Redes_Geod%C3%A9sicas_do_IBGE> (accessed 10/28/2025).

OSM (2023a) *Map features*. OpenStreetMap Wiki. Available at: <https://wiki.openstreetmap.org/wiki/Map_features> (accessed 10/28/2025).

OSM (2023b) *Import/Guidelines*. OpenStreetMap Wiki. Available at: <https://wiki.openstreetmap.org/wiki/Import/Guidelines> (accessed 10/28/2025).

Pereira, R.H.M.; M. Saraiva; D. Herszenhut et al. (2021) r5r: rapid realistic routing on multimodal transport networks with R5 in R. *Findings*. DOI: 10.32866/001c.21262.

Pinheiro Jr., L. (2019) The Brazilian smart cities: a national literature review and cases examples. In Anthopoulos, L. (ed.) *Smart City Emergence*. Amsterdam: Elsevier, p. 351–365. DOI: 10.1016/B978-0-12-816169-2.00017-1.

PostGIS (2023) *PostGIS, spatial and geographic objects for postgreSQL*. Available at: <https://postgis.net/> (accessed 10/28/2025)

Przeybilovicz, E.; M.A. Cunha and F. DE S. Meirelles (2018) O uso da tecnologia da informação e comunicação para caracterizar os municípios: quem são e o que precisam para desenvolver ações de governo eletrônico e smart city. *Revista de Administração Pública*, v. *52*, n. 4, p. 630-49. DOI: 10.1590/0034-7612170582.

Schröder-Bergen, S.; G. Glasze; B. Michel et al. (2021) De/colonizing OpenStreetMap? Local mappers, humanitarian and commercial actors and the changing modes of collaborative mapping. *GeoJournal*, v. *87*, n. 6, p. 5051-66. DOI: 10.1007/s10708-021-10547-7.

Sdoukopoulos, A.; E. Papadopoulos; E. Verani et al. (2024) Putting theory into practice: a novel methodological framework for assessing cities' compliance with the 15-min city concept. *Journal of Transport Geography*, v. *114*, p. 103771. DOI: 10.1016/j.jtrangeo.2023.103771.

Senaratne, H.; A. Mobasheri; A.L. Ali et al. (2017) A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, v. *31*, n. 1, p. 139-67. DOI: 10.1080/13658816.2016.1189556.

Tong, P.; W. Du; M. Li et al. (2021) Last-mile school shuttle planning with crowdsensed student trajectories. *IEEE Transactions on Intelligent Transportation Systems*, v. *22*, n. 1, p. 293-306. DOI: 10.1109/TITS.2019.2956786.

Wang, J. and X. Huang (2017) Routing school bus for better student learning. *In 2017 25th International Conference on Geoinformatics*. Buffalo: Institute of Electrical and Electronics Engineers, p. 1–7. DOI: 10.1109/GEOINFORMATICS.2017.8090947.

Zhang, H. and J. Malczewski (2018) Accuracy evaluation of the Canadian OpenStreetMap road networks. *International Journal of Geospatial and Environmental Research*, v. *5*, n. 2, p. 1-22. DOI: 10.13140/RG.2.2.11349.73440.

Zhu, M.; Z. Wang; H. Cui et al. (2023) Situation and hotspot analysis on rural transport infrastructure from the perspective of bibliometric analysis. *Journal of Traffic and Transportation Engineering*, v. *10*, n. 6, p. 1074-98. DOI: 10.1016/j.jtte.2023.04.009.