Towards a Configurational Multimodal Urban Network Model: A Data-Driven Approach to Public Transport Modelling.

(article starts on next page)

#60.

Towards a Configurational Multimodal Urban Network Model:

A Data-Driven Approach to Public Transport Modelling

[1]Flavia Lopes; Spatial Morphology Group (SMoG) – Department of Architecture and Civil Engineering (ACE) – Chalmers University of Technology, Sweden; flavia.lopes@chalmers.se.
[2]Jorge Gil; Spatial Morphology Group (SMoG) – Department of Architecture and Civil Engineering (ACE) – Chalmers University of Technology, Sweden; jorge.gil @chalmers.se.
[3]Ioanna Stavroulaki; Spatial Morphology Group (SMoG) – Department of Architecture and Civil Engineering (ACE) – Chalmers University of Technology, Sweden; gianna.stavroulaki@chalmers.se.

_____

ABSTRACT

The development of multimodal urban network models, integrating public transport and the street network, is crucial for achieving more integrated and precise spatial analysis, supporting research and practice towards understanding and creating sustainable environments. Despite their importance, building these models still faces challenges. While a reproducible, data-driven approach is widely embraced for calculating travel times and accessibility analyses with high metric and temporal precision, a gap remains in generating simplified models to advance towards a configurational approach. To address this challenge, this paper introduces a data-driven approach for developing simplified, flexible, and interoperable multimodal public transport network models capable of supporting both time-based and configurational analysis. These models are constructed by aggregating data from the General Transit Feed Specification (GTFS) at different levels of simplification to cater for different types of analysis. The integration of multimodal public transport network models with specific street network models opens the potential configurational analysis that includes all modes of transportation. By capturing crucial information about the city across various scales, these models serve as a robust toolkit to support the endeavours of researchers and practitioners working towards more sustainable cities.

KEYWORDS

Multimodal urban network models, GTFS, public transport network models.

1   INTRODUCTION

The seamless integration of urban planning and transport planning is essential for promoting environmentally friendly mobility and achieving sustainable urban environments. In recent decades, substantial efforts have been dedicated to formulating an integrated approach that emphasizes the inherent multimodality of urban mobility. Described as multilayer networks, multimodal networks intricately connect various transport modes, proving to be a powerful tool in understanding and decision-making regarding the movement of goods and people across different spatial scales (Orozco et al., 2023).

Nevertheless, while a data-driven methodology has become commonplace for calculating travel times and accessibility using transport analysis packages such as *OpenTripPlanner* (Morgan et al., 2019), R5R (Pereira et al., 2021), and R5py (Fink et al., 2022), which generate multimodal networks from vast amounts of data supporting high metric and temporal precision, challenges persist in achieving the same level of robustness, effectiveness, and reproducibility for simplified models capable of capturing the configurational aspects of multimodal networks (Lopes et al., 2023).

Commonly used in traditional analyses of transport planning, these transport analysis packages demand high computational costs. Furthermore, due to the level of detail and complexity of the data they require to generate the analysis, they become inflexible to essential changes in urban design and planning processes, particularly when designing, modifying, and planning new connections within cities.

In this context, insights from the configurational approach used by the space syntax field can be very useful for creating multimodal network models suitable for capturing both detailed and configurational measures of the network systems. Recognizing this potential, significant efforts have been invested in building multimodal models using space syntax methods and tools (Chen and Karimi, 2022; Chiaradia et al., 2005; Gil, 2012, 2014, 2016; Law et al., 2012; Schwander, 2007). However, despite these efforts, achieving a consolidated modelling approach to describing the multimodality of the public transport network and its connections with street network remains challenging. Three main issues contribute to this:

- The lack of a systematic approach to building simplified multimodal public transport network models that is data-driven and reproducible.
- The lack of public transport models compatible with simplified street network models, such as the ones used in space syntax analyses.
- The lack of a measure that can adequately describe the centrality and, consequently, the potential movement of the public transport network in simplified models. The potential movement within the public transport network possesses different characteristics that require distinct measurement methods from pedestrian and private

vehicular movement, where angular and topological measures have proven effective (Berghauser Pont et al., 2019; Hillier and Iida, 2005; Penn et al., 1998).

To address some of these challenges, this paper introduces a data-driven method for creating simplified, flexible, interoperable, and reproducible multimodal public transport network models. These models are designed to be compatible with the simplified street network models used by space syntax and allow for both time-based and configurational analysis. The models are constructed by aggregating data from the General Transit Feed Specification (GTFS) at different levels. GTFS is a widely used public transport data standard known for its global coverage and availability worldwide. The dataset, often available through open access, consists of a collection of text files containing information about stop locations, routes, trips, and operating hours of various public transport modes, which are published by public transport operators.

This study introduces eight distinct public transport network models developed using the GTFS data, sourced from the Västra Götaland Region in Sweden. The models, created by incorporating all modes of public transport included within the GTFS dataset, exhibit varying levels of aggregation, ranging from detailed to simplified descriptions of the public transport network, to cater for different types of analysis.

Additionally, this paper outlines the potentials for connecting the multimodal public transport network with street network models, in order to create integrated multimodal urban network models to have a comprehensive overview of the movement of people and goods.

To present these models, the subsequent section explores developments within the space syntax field related to multimodal network models. The third section outlines the datasets and methods used to create the models, followed by the presentation of results in the fourth section. A discussion and conclusions are provided in the final section.

2   RELATED WORKS

Over the years, significant efforts have been made to model and analyse the configuration of multimodal models using space syntax methods and tools (Chen and Karimi, 2022; Chiaradia et al., 2005; Gil, 2012, 2014, 2016; Law et al., 2012; Schwander, 2007). Most of these studies consider the connection of the street network with only one or two modes of public transport (Chen and Karimi, 2022; Chiaradia et al., 2005; Gil, 2012; Law et al., 2012; Schwander, 2007), mainly with railways and metros. These modes of transport are known for their ability to capture movement on larger scales within both the city and the region due to their extensive coverage. However, by focusing solely on these two types of networks, the refinement for local-scale movement provided by buses, communal taxis, ferries, trams, among others, can be overlooked, resulting in biased analyses. The findings of Gil (2014; 2016) show that the

combination of all modes of public transport with the street network for motorized and non-motorized movement gives more accurate results regarding centralities and the potential of movement. Another limitation of these studies is that the public transport networks are constructed using local databases, except for Chen and Karimi (2022) who utilized GTFS data. These local databases pose challenges in replicating the same models with the same degree of refinement for other cities and regions.

Beyond these issues, discussions on how to integrate the public transport network with the street network, as well as the significant metrics to consider when combining them, are frequently raised. A common approach involves connecting the public transport network with the street network and conducting centrality analyses on a unified graph (Chen and Karimi, 2022; Gil, 2012). While this method can identify changes in certain structural aspects of the city where there's a concentration of public transport stops and stations, it fails to capture the true impact of public transport on street network movement. The distinction lies in the fundamental differences between public transport movement and the dynamics of movement in the street networks. Research indicates that movement potential within street networks is notably influenced by topological and angular aspects (Berghauser Pont et al., 2019; Hillier and Iida, 2005; Penn et al., 1998). Conversely, movement within public transport networks is influenced by additional factors such as route coverage, travel duration, and the frequency of transfers required during a journey.

Configurational analyses of integrated multimodal networks using different costs on the public transport network (i.e. topological, for stops and transfers) and the street network (i.e. angular) have been implemented at the regional scale (Gil, 2014; Gil and Read, 2014). Tests comparing the different network modes and travel cost combinations demonstrate the power of the more complex combinations over singular modes and costs (Gil, 2014). However, the combination of travel time on the public transport network and angular distance on the street network has not been tested, and still lacks a theoretical principle for combining these very different units. Other approaches, such as those by Chiaradia et al. (2005) and Law et al. (2012), aim to establish this connection by assessing the impact of public transport stops and stations on nearby streets or segments. This involves integrating data on the number of passengers that use the stops and stations over a specific time frame with the street network, to identify factors of attraction. However, it's important to note that this approach introduces an additional variable for analysis: the number of passengers. Obtaining accurate passenger data can be challenging and may not be readily available for many cities and regions. Nonetheless, considering the attraction factor is crucial in the discussion of how to combine public transport and street networks models to have an integrated multimodal urban network model.

Despite the valuable lessons learned from previous studies, achieving a consolidated and unified approach to construct a fully multimodal public transport network model, considering all modes of transport, as well as its integration with street network models, and identifying relevant measures that can describe all types of movement, remains a challenge.

3   DATASETS AND METHODS

This paper introduces eight multimodal public transport network models with varying levels of detail, offering the possibility of analyzing different aspects of the network spanning from time-based to configurational analysis. These models were constructed by aggregating GTFS data, processed using a flexible and interoperable automated approach. To clarify the process of constructing these models, this section provides details of the GTFS dataset and the data aggregation process.

3.1   GTFS Dataset

The GTFS[1] is a standard format essential for organizing public transportation schedules and associated geographic information. This structured data repository encompasses files detailing routes, stops, schedules, and geographical data. The dataset contains the following key components:

- **stops.txt**: Enlists all the stops situated along transit routes, featuring their unique identifiers, names, and precise geographic coordinates (latitude and longitude).
- **routes.txt**: Provides descriptions of the routes managed by transit agencies, including their unique identifiers, names, and other pertinent details.
- **trips.txt:** Offers insights into individual trips traversing transit routes, including trip IDs, route IDs, and service IDs.
- **stop_times.txt**: Precisely pinpoints the times of arrival and departure of vehicles at stops during trips. It includes detailed information such as trip IDs, stop IDs, arrival times, departure times, and stop sequences.
- **calendar.txt:** Outlines the dates for which transit services are operational, along with their respective schedules. This encompasses information such as service IDs, start dates, end dates, and operational days of the week.
- **calendar_dates.txt:** Enumerates exceptions to the regular service schedules outlined in the calendar.txt file. This encompasses dates when services are either added or removed from the schedule.

---

[1] More information in: www.gtfs.org.

These files collectively offer extensive insight into public transit systems, facilitating route planning, schedule management, and geographical visualization. For this paper, we utilized a GTFS dataset sourced from the Västra Götaland Region in Sweden[2], covering all trips of various modes of transport during the period from July to December 2022.

3.2 Buliding the Multimodal Network Models

To generate the multimodal public transport network models the GTFS was aggregated at different levels, resulting in eight distinct models. Essentially, text files comprising the GTFS were merged and filtered to generate new tables[3] encompassing information with varying levels of aggregation for the different models. The initial step involved establishing a geographic representation of the public transport network, containing information on all trips and time features of various modes of transport. This process laid the foundation for subsequent aggregations. To generate the tables, Python packages such as Pandas (version 1.5.3), Geopandas (version 0.12.2), NumPy (version 1.24.2), Shapely (version 2.0.1), and OSMnx (version 1.3.0) were utilized. Additional details regarding the models are elaborated on in the subsequent subtopics.

3.2.1 Model 0: The geographic representation of the GTFS

The first step was to create a 'baseline' model based for a geographical representation of the GTFS. In this process, two main geographic features were established: stops and links between stops. At this level, stops represent public transport stops and stations, while a link represents each trip of the different modes of transport covered by the GTFS.

To establish the stops data structure, the stops.txt file from the GTFS was merged with the geographic information from the OpenStreetMap (OSM) database via OSMnx. Subsequently, the links were created by merging the stops data previously constructed for this model with the routes.txt, tips.txt, and stop_times.txt files from GTFS (Figure 1 top).

3.2.2 Model 1: aggregating trips and introducing transfers and frequencies

The second step involved aggregating the 'baseline' model into the first model with analytical potential, where all the trips of one route were aggregated into a single link representing this route. In addition to tables for stops and links, tables for transfers between routes at the same stop, and transfers between routes at different but nearby stops were created to establish a fully connected network. In GTFS the stop_times.txt files, representing the time that each vehicle of each trip stops at each stop or station, provides the waiting times and the potential transfers between the vehicles. This informs the transfers at a stop and the frequency in the

---

links table, calculated from the number of trips within a three-hour interval. In this work, the transfers tables contain estimated waiting times based on the frequency.

Beyond the transfers at the same stop, to fully model the public transport network's potential, transfers between different stops were necessary. These transfers were created by identifying stops within a 50-meter buffer zone of each other. Time features for these transfers were calculated based on straight-line walkable distances (see Figure 1 bottom).
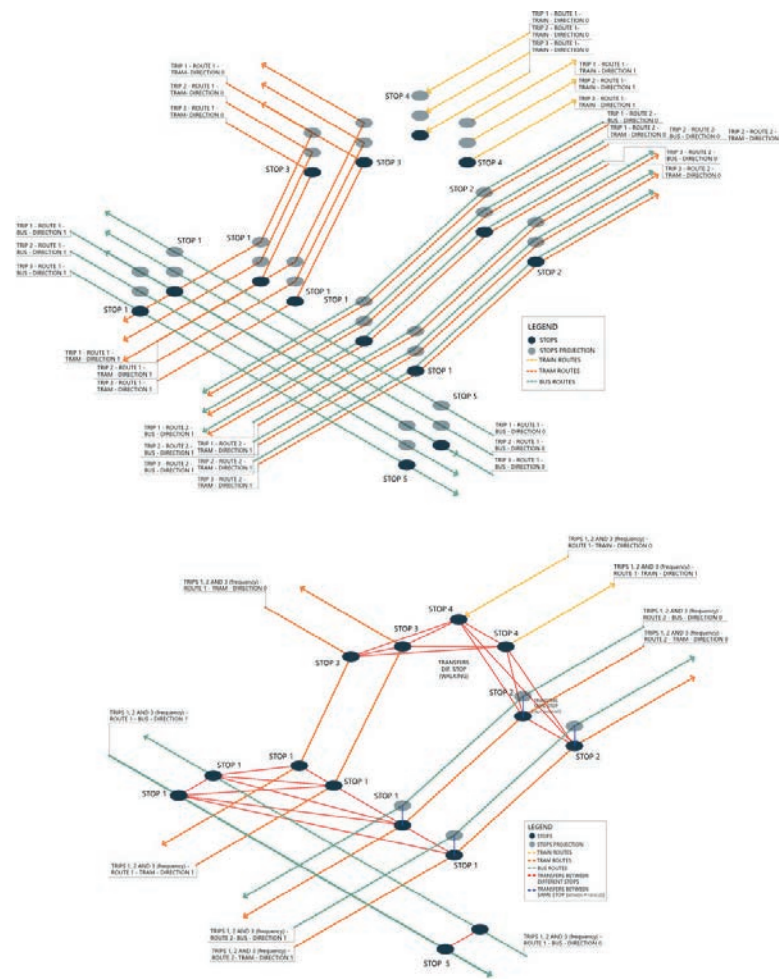

Figure 1: schema of the model 0 (top) and the model 1 (bottom).

3.2.3   Model 2: aggregating links by directions

The goal of the third step was to aggregate the two directions of each route into a single link. To achieve this, the public transport stops data was aggregated. Instead of having two public transport stops representing both directions for each route, they were combined to have only one stop located between the previous two.

After aggregating the stops, the links of both directions were also aggregated into a single link. In taking this step, transfers at the same stop also needed to be aggregated following the same logic. The frequency for transfers at the same stop was calculated using the average of the

frequencies for both directions of each route. Transfers between different stops were aggregated as well, and the straight-line walkable time was recalculated (Figure 2 top).

3.2.4   Model 3: aggregating stops by names

The subsequent stage entails constructing a model where the stops were aggregated by names, while maintaining the separation of links by routes.

As depicted in Figure 2, all stops labelled 'stop 01' were joined into a single stop, whereas those with distinct names remained disaggregated. The links between stops and transfers were established using the new stop locations. The frequency followed the same rule as the previous model. Transfers between different stops were calculated once again based on straight-line walkable time (Figure 2 bottom).
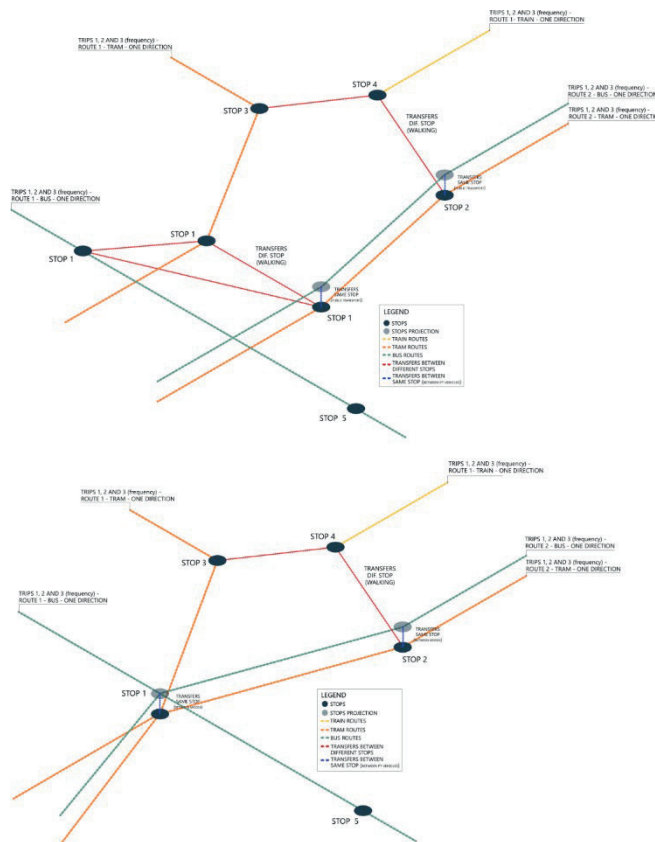

Figure 2: schema of the model 2 (top) and the model 3 (bottom).

3.2.5   Model 4: aggregating links by modes

At this stage, the links are aggregated based on public transport modes. Instead of having various links representing routes between stops, this model only differentiates links between modes, such as bus routes, tram routes, train routes, etc. In this model, transfers are only possible between two vehicles of different modes that share the same stop.

The frequency of public transport was determined by identifying the highest frequency among vehicles of a specific mode. This approach ensures the route with the shortest waiting time was considered.

Transfers between different stops were subsequently generated, with calculations based once again on straight-line walkable time (Figure 3 top).

3.2.6   Model 5: aggregating stops by stop areas

This stage involved aggregating at the same location stops that do not share the same name but are sufficiently close to each other and have different modes passing through them. This occurs in specific areas of a city, like central stations or other types of transfer terminals. To define which stops is part of a stop area, a buffer zone of 50m was created. If the buffers touched each other and the stops were connected to different modes, they were aggregated under a new name that characterizes the stop area. If the buffers did not touch each other directly but were connected with each other in a chain via the buffers of other stops, they were also considered part of a stop area, as illustrated in Figure 3 (bottom).

Subsequently, the links and transfers at the same stop were regenerated to accommodate the new stop areas. Transfers between different stops no longer exist at this stage, since nearby stops are aggregated at the same point (Figure 3 bottom).
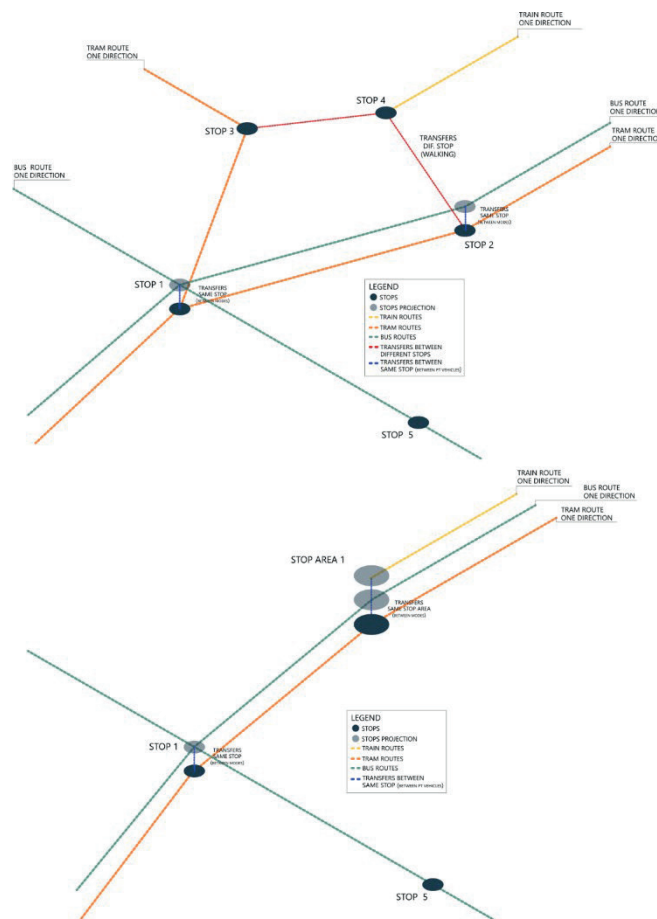

Figure 3: schema of the model 4 (top) and the model 5 (bottom).

3.2.7   Model 6: aggregating links representing modes into a single link

In this model, a complete simplification of the links is introduced. The stops in this model were the same as those in the previous 'stop areas' model (Model 5). The links corresponding to different modes were aggregated into a single link.

Transfers at the same stop no longer exist; instead, transfers became a single attribute of the stops table, indicating whether transfers are possible, the quantity of transfers feasible at the stop, and an average waiting time (that can be used as a weight) (Figure 4 top).

3.2.8   Model 7: excluding "trivial nodes"

To simplify Model 6 and creating a model consisting only of the most significant stops, involves retaining only those stops with multiple routes and modes connected to them, while excluding the "trivial nodes" in the network. A trivial node is defined as having degree/connectivity equal to 2. An iterative process was implemented to identify and remove all "trivial nodes" until no node with a degree/connectivity equal to 2 remained. Subsequently, the most simplified links were generated again, considering the updated stops. Information regarding transfers at the same stop followed the same rule as in the previous model (Figure 4 bottom).
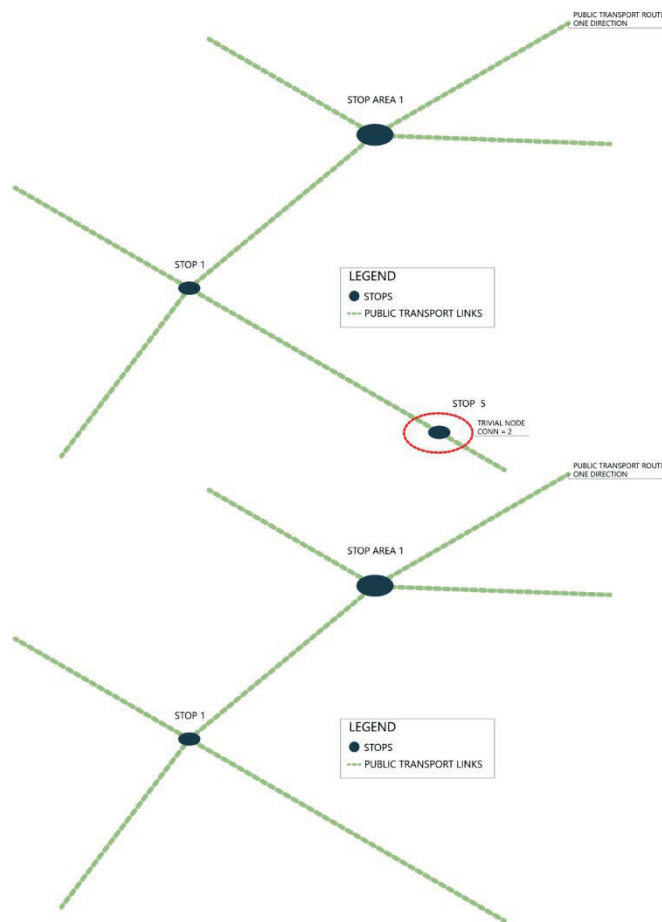

Figure 4: schema of the model 6 (top) and the model 7 (bottom).

4   CASE STUDY: ANALYSING THE MULTIMODAL PUBLIC TRANSPORT NETWORK MODELS

To demonstrate the capabilities of the public transport network models, we conducted time-based accessibility and centrality analyses using the multimodal public transport models developed for the Västra Götaland Region in Sweden. In the time-based accessibility analyses, we assessed the number of public transports stops reachable within a 15-minute trip while navigating solely within the public transport network on a typical Monday morning. For the centrality analysis, we calculated closeness centrality weighted by travel times. Both analyses were performed using Igraph (version 0.10.4), a versatile network analysis package in Python that offers various graph manipulation tools and analysis algorithms.
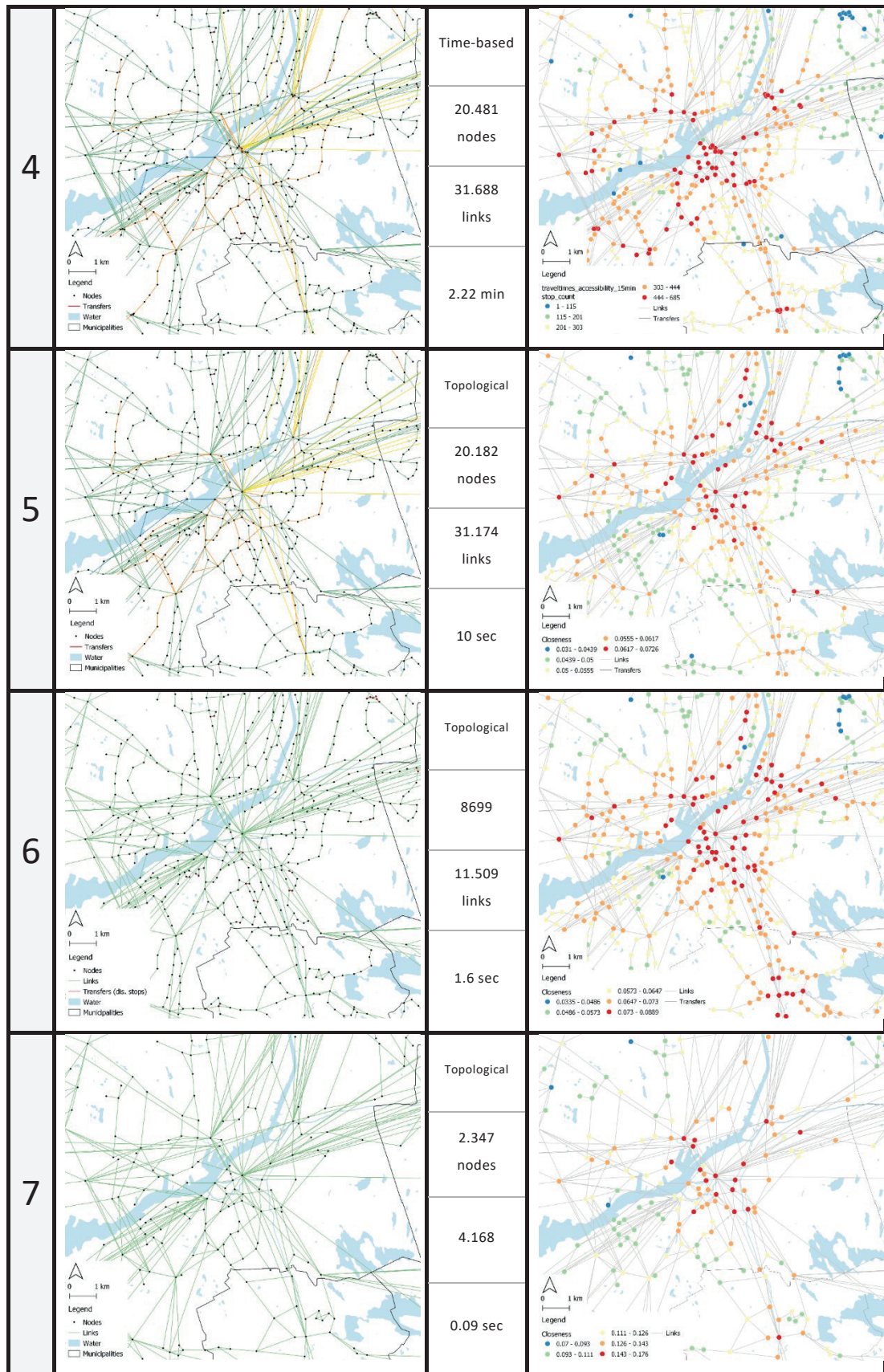
4.1   Results

The results comprise eight distinct public transport network models, with seven of these models generating graphs of varying sizes, offering different analytical possibilities (Table 1). Models 1 to 4 involve topological aggregation, consolidating stops and links while retaining original geographical characteristics. Conversely, from models 5 to 7, the aggregations move towards a higher level of geographical abstraction, prioritizing the relationships between components over geographic specifics. As the models become more aggregated, their capacity for detailed analysis diminishes, particularly in scenarios requiring geographical and temporal refinement. However, increased aggregation introduces new possibilities, such as the ability to alter the network by design and measure descriptive metrics with lower processing time costs.

The findings indicate consistency in the results when comparing models 1 through 4 and models 5 through 7. This suggests that employing models 3 or 4 to measure time-based travel can yield similar results to those achieved with model 1, albeit with a smaller graph size and reduced processing times (see Table 1). For instance, Model 01 comprises 56,302 nodes and 91,595 links, with a processing time of 2 hours and 5 minutes. In comparison, models 3 and 4 feature 31,646 nodes, 68,579 links, and a processing time of 16.46 minutes; and 20,481 nodes, 31,688 links, and 2.22 minutes of processing time, respectively.

Similarly, the same logic applies when comparing model 7 to model 5. However, given that model 5 already exhibits low processing times, it is prudent to consider the intended purpose of the analysis.

Table 1: summary of the multimodal models and analysis.

| M O D E L S	R E S U L T S		
	Map	Suitable Analysis Number of nodes (graph) Number of links (graph) Processing Time	Processing Time (min) (15 min. time-based accessibility) Results
1		Time-based 56.297 nodes 91.595 links 2h5min	
2		Time-based 31.646 nodes 68.579 links 17.44 min	
3		Time-based 31.646 nodes 68.579 links 16.46 min	

5 TOWARDS AN INTEGRATED MULTIMODAL URBAN NETWORK MODEL

By understanding the analytical possibilities offered by multimodal public transport network models, it is conceivable that integrating them into the street network could enhance the potential to comprehensively capture the movement of goods and people in cities and regions. However, before proceeding with the integration, several key factors must be considered. Firstly, it is essential to align the spatial resolutions of public transport and street network models to ensure model compatibility. This alignment prevents biased or unrealistic results. For Models 1 through 4, which offer high detail, a correspondingly detailed street network model like Road Centre Lines (RCL) is suitable. Models 5 and 6, with some topological aggregation, can be paired with segment maps. Model 7, with the highest level of simplification, benefits from integration with both segment and axial maps. This strategic alignment ensures the accuracy and reliability of analyses.

In addition to model compatibility, it is necessary to consider how the public transport network will be connected to the street network. Our study suggests that the most effective way to combine both networks is through a two-step process. Essentially, the public transport network is processed first to generate a weight, which is then transferred to the street network. Subsequently, the street network is analysed considering these weights (see Figure 5). This weight is conceived as a representative measure for the transport network, such as the number of routes passing through each stop, accessibility within a certain time frame, the quantity of transfers, or a combination of these variables.
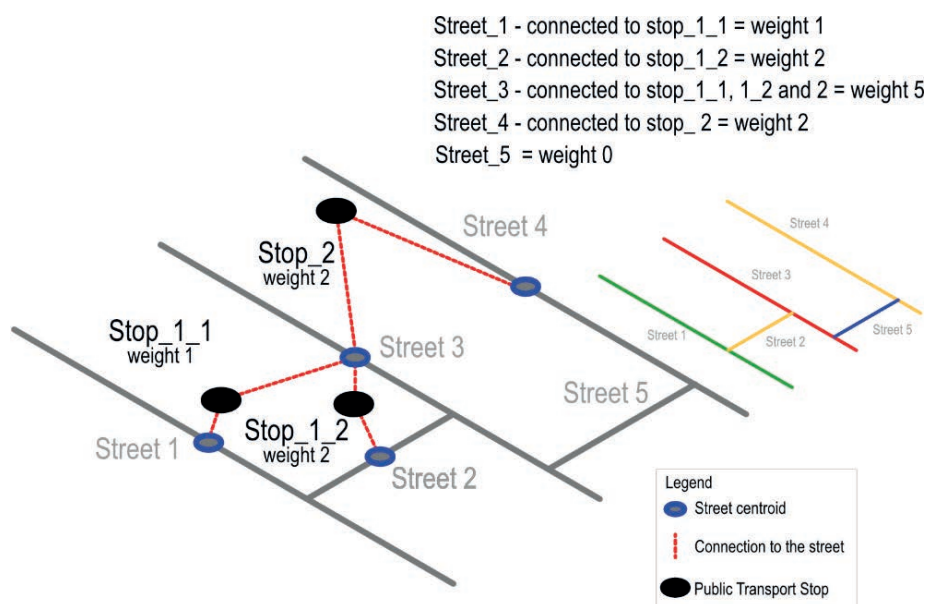
Figure 5: schema of the connection of the public transport networks model with a street network model.

6   DISCUSSION AND CONCLUSIONS

In this paper, we introduce a data-driven approach to create flexible, interoperable, and reproducible multimodal public transport network models. To demonstrate the capabilities of these models, we conducted time-based and configurational analyses. The results illustrate a range of possibilities, from models offering high spatial-temporal resolution to those with varying levels of topological aggregation. Each model presents advantages and disadvantages that must be considered when utilizing them. For example, models with high spatial and temporal accuracy entail greater computational costs, whereas more aggregated models consume less computational resources.

The issues of spatial and temporal resolution are aspects that require careful consideration, particularly when connecting multimodal public transport networks to a street network. Once connected, as shown in the results, the analysis process occurs in two stages, where the public transport network is initially processed and then generates a weight that influences the street network analysis. Therefore, computational cost and model flexibility regarding changes need to be considered.

In the detailed and operational phases of urban and transport planning, where precise spatiotemporal data is critical, models with finer levels of granularity, such as 1 and 2, are preferred despite their higher computational demands. This preference arises from the necessity of accurately capturing spatial and temporal dynamics during this phase. Conversely, in strategic planning phases, where the focus shifts towards designing overarching connections between locations rather than fine operational details, simpler models like 6 and 7, which facilitate cost-effective adjustments, become more relevant.

The support these models provide to decision-making processes is crucial. However, it's important to acknowledge that this work is still in progress with some limitations, such as the absence of a representative measure for determining the weight the public transport network should assign to street network models, as well as the lack of analysis of the integrated models. Both aspects are under discussion and testing but are not yet ready and are being considered as subsequent phases of this research.

Despite these limitations, this paper introduces a new perspective for generating public transport network models, which integrate with street networks allowing configurational analysis, and open the possibility of capture the movement of people and goods within cities and regions. This represents an important tool to support decision-making at various levels of urban and transport planning, ultimately contributing to the creation of liveable, healthier, and sustainable cities.

REFERENCES

Berghauser Pont M, Stavroulaki G, Bobkova E, et al. (2019) The spatial distribution and frequency of street, plot and building types across five European cities. *Environment and Planning B: Urban Analytics and City Science* 46(7): 1226–1242.

Chen PN and Karimi K (2022) Analysis and modelling of the multilevel transport network : the metro and railway system in greater London. Glasgow: University of Strathclyde Publishing, pp. 354–362. Available at: https://strathprints.strath.ac.uk/80473/ (accessed 19 June 2023).

Chiaradia A, Moreau E and Raford N (2005) Configurational Exploration of Public Transport Movement Networks: A Case Study, The London Underground. *Proceedings of 5th International Space Syntax Symposium*.

Fink, C., Klumpenhouwer, W., Saraiva, M., Pereira, R. and Tenkanen, H.: r5py: Rapid Realistic Routing with R5 in Python, doi:10.5281/zenodo.7060437, 2022.

Gil J (2012) *Integrating Public Transport Networks in the Axial Model*. *Proceedings of 8th International Space Syntax Symposium*. Santiago, Chile, 2012, p. 8103:1- 8103:21. Santiago de Chile: PUC.

Gil J (2014) Analyzing the Configuration of Multimodal Urban Networks. *Geographical Analysis* 46: 368–391.

Gil J and Read S (2014) Patterns of sustainable mobility and the structure of modality in the Randstad city-region. A|Z ITU Journal of the Faculty of Architecture 11(2): 231–254.

Gil J (2016) *Urban Modality: Modelling and Evaluating the Sustainable Mobility of Urban Areas in the City-Region*.

Hillier B and Iida S (2005) Network and Psychological Effects in Urban Movement. In: *Spatial Information Theory* (eds AG Cohn and DM Mark), Berlin, Heidelberg, 2005, pp. 475–490. Lecture Notes in Computer Science. Springer.

Law S, Chiaradia A and Schwander C (2012) Towards a multi-modal space syntax analysis. A case study of the London Street and underground network. *Proceedings of 8th International Space Syntax Symposium*. Santiago, Chile, 2012, p. 8021:1- 8021:20. Santiago de Chile: PUC.

Lopes F, Gil J and Stavroulaki I (2023) Simplified geodata models for integrated urban and public transport planning. *AGILE: GIScience Series* 4. Copernicus GmbH: 1–9.

Morgan, M., Young, M., Lovelace, R. and Hama, L.: OpenTripPlanner for R. J, Open Source Softw, 4, 1926. https://doi.org/10.21105/joss.01926, 2019.

Orozco LGN, Alessandretti L, Saberi M, et al. (2023) Multimodal urban mobility and multilayer transport networks. *Environment and Planning B: Urban Analytics and City Science* 50(8): 2038–2070.

Penn A, Hillier B, Banister D, et al. (1998) Configurational Modelling of Urban Movement Networks. *Environment and Planning B: Planning and Design* 25(1). SAGE Publications Ltd STM: 59–84.

Pereira, R., Saraiva, M., Herszenhut, D., Braga, C.K., Braga, V., Conway, M. and Wigginton, M.: r5r: Rapid Realistic Routing on Multimodal Transport Networks with R5, R, Transp. Find, https://doi.org/10.32866/001c.21262, 2021.

Schwander C (2007) Network Analysis Applied: the railway network in South East England. In: *Proceedings of the 6th International Space Syntax Symposium*, Istanbul, Turkey, June 2007, p. 076:1-076:17. Istanbul Technical University.