Computer Sciences Research Project Report

Analysis of city composition based on basic amenities including public transport

Author: Yubo Cai

Institute: Laboratoire d'informatique de l'École polytechnique (LIX) - Team AMIBio

Date: June 2nd, 2023

Supervisor: Sarah J. Berkemer

Contents

1 Introduction 2
2 Dataset: OSM and Filosofi 2
3 Methodology 3
4 Results 5
5 Discussion and Conclusion 7
6 Further work 8

Bibliography **9**

Appendix A Datasets, Links, and Tables **10**
1 Links to the dataset and package 10
2 Github file information 10
3 Table of the tags 10

Appendix B Graphs and Plots **12**

Acknowledgements

I would like to thank the kind professor Sarah J. Berkemer for hosting me and supervision during this project in the AMIBIO team at the Laboratoire d'informatique de l'École Polytechnique (LIX) the fall of 2023. This project is the extension work of the ENSAE BATO-MOUCHE project by Simon GENET, Leopold MAURICE, and Marie-Olive THAURY. I would like to thank Marie-Olive THAURY for her advice and feedback on the project and implementation of the method.

This project, building upon the foundation of Paper [11], introduced public transportation as a key consideration and developed a comprehensive set of calculation and statistical methods based on this addition. Therefore, in this project, we referenced certain methodologies from the paper, as well as data calculation and visualization code. I am grateful for their work, which provided a solid foundation and impetus for my project.

We initially considered three approaches: (i) treating public transport as a basic amenity category in itself; (ii) extending the radius due to the usage of public transportation; and (iii) potential weighting schemes for the accessibility to public transport. The first approach overlooks the intrinsic impact characteristics of transportation, thus leading to one-sided results. The second approach is not applicable to rail networks, as rail and street networks are distinct. Consequently, we opted for the third approach and incorporated the TOD model into our analysis.

Important links and code repository

1. **Github repository:** https://github.com/yubocai-poly/BATO-MOUTCHE-Extension.
2. **Link to the result data:** Here. Since the size of the result file is too large it can not be uploaded to Github.
3. **Github repository of the previous work:** https://github.com/LeoMaurice/BATO-MOUCHE-Stat-App.

More information and links please check

1 Introduction

Cities are crucial in societal transformations, and urban development significantly affects living standards and economic growth, particularly as many industrial sector workers reside in urban areas [1]. In this project, we study the accessibility of important amenities (e.g. essential shops and services, schools, etc.) with the influence of public transport through the concept of the X-minute city [8].

Building upon the foundation laid by Paper [11], this study incorporates public transportation networks into the accessibility measurement of amenities. Additionally, it utilizes accessibility as a key factor in analyzing the Transit-Oriented Development (TOD) degree of Paris. We initially conducted a descriptive statistical analysis of public transportation in Paris to gain an understanding of the basic situation. Subsequently, we employed the Two-Step Floating Catchment Area method (2SFCA) to measure accessibility with the influence of public transportation. Additionally, we calculated an Aggregated Amenity Score by applying importance weighting to each type of facility. Following this, we used the node-place model to calculate the TOD degree of the city and conducted regression analysis. Lastly, we integrated the TOD degree with our accessibility analysis for a comprehensive examination.

	Railway network		Share and Number of public transport trips in Paris (millions/percentage). 2019			
	(Km/mln inh.) 2019	(n. station/mln inh.) 2019	Metro	RER (RATP and SNCF)	Tramway (includes T4 and T11E)	Buses
Paris	104.9	142.5	1498 (42.4%)	1407 (39.8%)	340 (9.6%)	291 (8.2%)

Table 1: Structural transport system variables in Paris, source: INSEE: Freight, passengers and other transportation indicators - Vehicle registrations, Statistia: Number of public transport journeys in Paris and the Île-de-France region in 2019 and 2020

In our study, we mainly focused on the impact of the railway network (Metro, RER, Tramway) on public transportation. On one hand, railways are the most predominant mode (91.8% of the total number of public transport trips in Paris according to Table 1) of transportation for residents in Paris. On the other hand, unlike **bicycles, buses, and walking**, which all share the **street network** and primarily **differ in speed**, railways possess their unique transportation network. Therefore, they require a separate construction for analysis using distinct statistical methods.

2 Dataset: OSM and Filosofi

In our project, we used data from **OpenStreetMap** (OSM) [3] and **Filosofi** system. OSM is a free dataset that includes spatial information such as buildings, transportation networks, and points of interest. We primarily use the points of interest and transportation network data from the OSM dataset. Points of interest are categorized into various classes (public, health, leisure, catering, accommodation, shopping, money, tourism etc), and these categories are further divided into multiple tags (more information in Table A.1)

The **Filosofi** system, developed by INSEE, is a database focused on localized fiscal and social income. This database divides the territory into 200-meter square areas, breaking the boundaries of administrative divisions. Among other things, it provides variables such as the age pyramid of residents, residents' income, and the construction year of buildings. We merged the homogeneous grid dataset provided by Filosofi with OpenStreetMap to obtain the dataset we ultimately used.

We imported the OSM database using the **OSMnx** [2] Python package and constructed, visualized, and searched for optimal routes in transportation networks using **NetworkX**. Additionally, we calculated geographical spatial distances using **GeoPandas**. However, OSM also has some shortcomings. In fact, the database is filled by anonymous contributors, which can lead to errors and inconsistencies in information. Additionally, OSM lacks optimal route planning features similar to those found in software like Google Maps or CityMap, resulting in certain inaccuracies in our subsequent calculations of geographical distances.

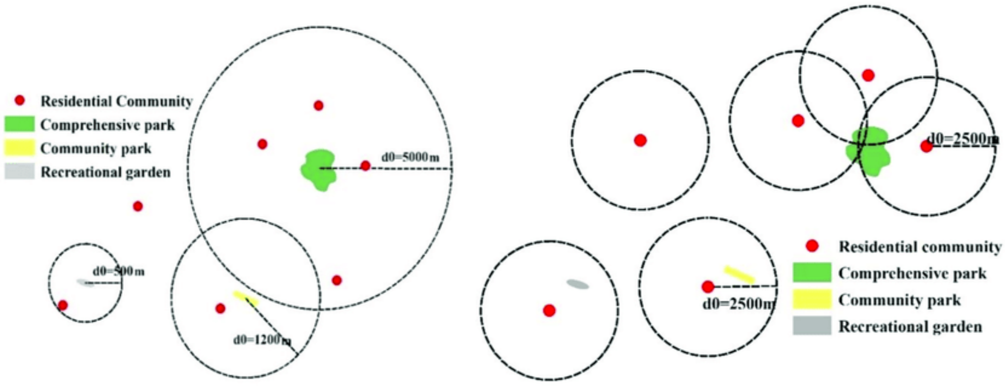
3 Methodology

3.1 Accessibility scores for each type of amenity with railway network

Accessibility scores are derived using the **two-step floating catchment area method** (2SFCA) [7, Section 3], originally utilized for assessing accessibility to health services. This method can be adapted for various extensive variables, such as the number of amenities. The 2SFCA approach involves evaluating the demand surrounding each service provider. Subsequently, for each individual or location seeking the service, the available supply is computed. The algorithm can be described in two steps:

1. **service catchment:** For each service, find all populations that fall within a threshold distance (dmax) and calculate the population-to-provider ratio
2. **population catchment:** For each population, find all services that fall within a threshold distance (dmax) and sum the population-to-provider ratios from step 1.

Here the provider is the number of each type of amenity. Figure 1 offers a more visual explanation of this process.



(a) First step. (b) Second step.

Figure 1: 2SFCA rationale illustration, originating from paper [4]

As we introduced in section 1, we keep using the X-minute city concept. In our research, we set the time threshold X as 30 mins, which is defined according to the approximate average time of journeys to work [9, Table 2] and other purposes.

Note *We compared the outcomes of transportation networks in 15-minute cities (which are previously used in [11]) and 30-minute cities. Ultimately, we chose 30 minutes as the threshold for the following reasons:*

1. *Under a 15-minute timeframe, rail transit does not significantly outperform walking in terms of covering longer distances. As a result, the impact of rail transit on the weighting process in the calculation of the accessibility score is limited (Check the comparison weights plot in the figure B.7).*
2. *Considering people's commuting behaviors, using rail transit is generally for work commutes or to reach farther areas. According to [9, Table 2], a 30-minute travel time threshold is defined based on the approximate average commute time in European cities. On the other hand, the choice of a 30-minute travel time does not significantly impact the average accessibility values for the study area [6].*

Delving deeper into the specifics, we designate the squares representing suppliers as j, while those consuming services are denoted as i. We further define k as the index for squares within a 30-minute accessible zone around j [1]. Let D_k represent the demand (measured in inhabitants) in square k, S_j the supply in square j (measured in the number of Points of Interest, POI), W_{kj} the permeability coefficient of demand from k to square j, and $\mathbb{P}_{kj}$ the probability that inhabitants in square k visit square j (for instance, $\mathbb{P}_{kj}\backslash$ proportional W_{kj}). We can then define the demand received by square j (counted in inhabitants) as follows (step 1):

$$\mathbb{D}_j = \sum_k \mathbb{P}_{kj} D_k$$

[1] Remark: Each square functions both as a supplier and a consumer.

From that aggregated demand taking into account the zone around square j, we can compute the supply per inhabitant with each amenity as step 2:

$$R_j = \frac{S_j}{\mathbb{D}_j} = \frac{S_j}{\sum_k \mathbb{P}_{kj} D_k}$$

Finally, we can compute the 2SFCA score for the square i counted in the number of POI accessible with the time threshold:

$$2SFCA_i = \sum_j \mathbb{P}_{ij} R_j$$

However, with the introduction of the railway network, calculating the permeability coefficient W_{kj} and determining whether there is 30-minute accessibility between two zones became a challenging aspect. Normally, the permeability coefficient is inversely proportional to the square of the distance. And with a constant speed, we have $d_{kj} \propto t_{kj}$. Therefore, we use the **traveling time** t_{kj} of zone k and j as the variable to measure the weights. We consider two types of traveling behaviors: walking or using rail transit. Correspondingly, there are two types of distances $d_{kj,\text{walk}}$ and $d_{kj,\text{railway}}$ to be considered. For the distance by railway, we refer to the method of calculating node index in [9] to design the following formula:

$$d_{kj,\text{railway}} = d\left[\text{AZ}_k, \text{AZ}_j\right] = \beta d\left[\text{AZ}_k, \text{station}_k\right] + d\left[\text{station}_k, \text{station}_j\right] + \beta d\left[\text{station}_j, \text{AZ}_j\right] \tag{1}$$

It consists of three terms: the access walking distances from AZ_k [2] to the nearest origin stations, the shortest path distance between station of origin and station of destination on the rail-based network, and the egress walking distances from the nearest destination station to AZ_j. Here β is the distance multiplier, which we use as 1.2 in our project. Then we compute the estimated travel time for walking and taking the railway (corresponding to roughly a walking speed of 80m/min and a railway speed of 700m/min).

$$\begin{cases} t_{\text{walk}} & = \frac{d_{kj,\text{walk}}}{v_{\text{walk}}} \\ t_{\text{railway}} & = \beta\frac{d[\text{AZ}_k,\text{station}_k]+d[\text{station}_j,\text{AZ}_j]}{v_{\text{walk}}} + \frac{d[\text{station}_k,\text{station}_j]}{v_{\text{railway}}} \end{cases} \tag{2}$$

Then we have the weight (permeability coefficient) between the accessible zone i and j such that

$$W_{kj} = \frac{30^2}{\min(t_{\text{walk}}, t_{\text{railway}})^2} \mathbb{1}\left(\min(t_{\text{walk}}, t_{\text{railway}}) \leq 30 \text{ mins}\right)$$

3.2 TOD degree Measurement

Transit-Oriented Development (TOD) degree in urban structure refers to the spatial concentration of economic activities and population along the rail transit networks. We employed the node-place model method outlined in the paper to measure the TOD degree. In detail, for every Analysis Zone (AZ), we assessed both a "node index" and a "place index". Following this, we examined the distribution of these indices using a bivariate scatterplot on an xy graph, depicting the node and place index values for each AZ within our study area. The TOD degree in the urban structure was then ascertained by examining the correlation strength between these two indices.

The **node index** of an AZ (Equation 3) corresponds to the closeness centrality index

$$\text{node index }_{AZi} = \frac{N-1}{\sum_{j-1}^{N} \min(t_{ij,\text{walk}}, t_{ij,\text{railway}})} \tag{3}$$

where N is the number of AZ in the map and the method to compute $\min(t_{ij,\text{walk}}, t_{ij,\text{railway}})$ is the same as equation 2.

The **place index** of an AZ is the amenity score per unit area. For each type of amenity, we assigned a frequency and a weight. The frequency is determined by the ratio of the number of items in that category (N_p) to the total number of amenities (N) and the weight is the opposite of the frequency:

$$f = \frac{N_p}{N} \quad \text{and} \quad w_p = 1 - \frac{N_p}{N} \quad \text{where } p \text{ is the category} \tag{4}$$

[2] Remark: AZ_k means Analysis Zone k

The underlying concept here is that if a particular type of amenity is less abundant in a certain area, its relative importance in that area becomes higher. After performing the calculations, we obtained the weights for each type of facility:

IdINSPIRE	**weight restaurant**	**weight culture and art**	**weight education**	**weight food shops**	**weight fashion beauty**	**weight supply shops**
CRS3035RES200mN 2893400E3763200	0.641828	0.946926	0.965579	0.845934	0.795573	0.80416

Table 2: The weights for each type of amenity

For each AZ, we can compute the amenity score (AS) and place index with the following formula:

$$AS_i = \sum_p N_p w_p \quad \text{and} \quad \text{place index}_i = \frac{AS_i}{\mathrm{A}_{\mathrm{AZ}_i}} \tag{5}$$

where p is the category, i is the index of AS, and $\mathrm{A}_{\mathrm{AZ}_i}$ is the area of AZ_i. To facilitate comparison across various case studies, we applied min-max normalization to calculate standardized node and place indices for each study.

3.3 Measurement of aggregated accessibility score

In section 3.1 we introduced the method to compute the accessibility for each type of amenity. It naturally led us to consider how to calculate the Aggregated 2SFCA. Thanks to the method for calculating the importance weight of each amenity that we proposed in equation 4, we were also able to calculate the aggregated accessibility score in a similar manner:

$$CS_i = \sum_{p=1}^{P} (1 - w_p) \times X_{i,p}$$

where $X_{i,p} = \frac{2SFCA_{i,p} - \min_j 2SFCA_{j,p}}{\max_j 2SFCA_{j,p} - \min_j 2SFCA_{j,p}}$ is the min-max normalization of the accessibility score for zone i and amenity $p \in P$.

4 Results

We first provide an overview of public transportation in Paris. We primarily consider three types of public transport: rail (encompassing metro, RER, and tramways), buses, and shared bicycles. Paris has an abundance of transportation resources, far surpassing other major European cities (compare with Table 1 and Table 2 in [9]), with a total of 308 rail transit stations. The rail network within Paris spans approximately 226.7 kilometers.

	Railway (RER, Metro, Tailway)	Bus	Bicycle Rental
Numbers of Station	308*	2801+	992+

Table 3: Number of stations for the three types of public transportation in Paris. * source: Open Data RATP, + sourse: OpenStreetMap.

We also compared the public transportation resources between different districts. Generally, the availability of these resources aligns with the number of amenities in each area (Figure 2 and Figure 3.1 in [11]). More central and affluent neighborhoods tend to have more abundant public transportation options.

4.1 Aggregated 2SFCA Result

We calculated the aggregated accessibility score both with (Figure B.5) and without (Figure B.6) amenity-type weights w_p. In reality, the difference between the two is not substantial, but we were still able to discern subtle variations. For instance, in some zones of the 16th arrondissement, the scores increased after considering weights. This is because the 16th arrondissement is home to a large number of schools (like the Paris Dauphine University, EJM) and art galleries (such as the Musée d'Art Moderne de Paris), which are facilities with higher weights. Therefore, we believe that the amenity score adjusted for weights is more reasonable.

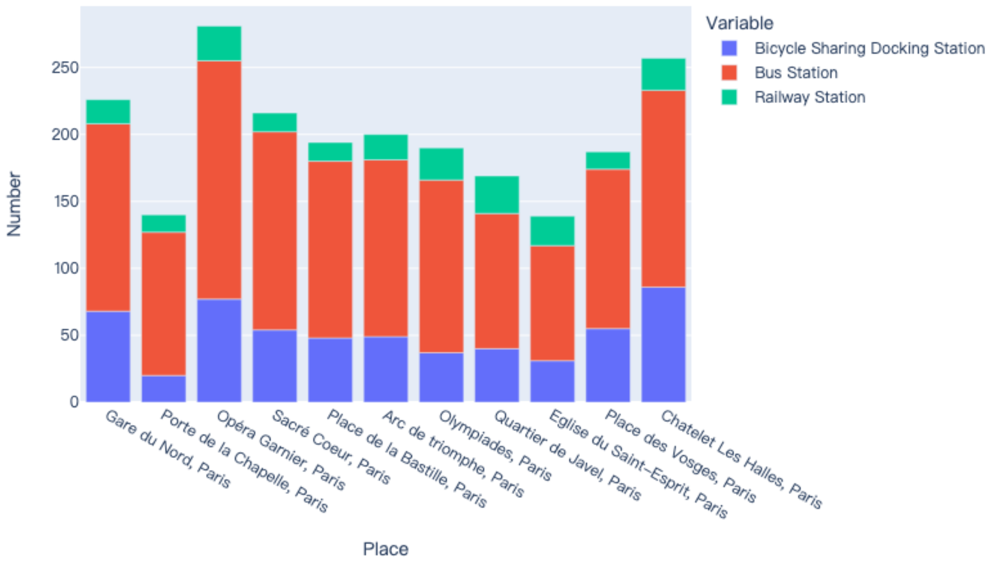


Figure 2: Comparison of the composition of transportation in 11 districts (with 1km radius) of Paris

We also observed that Paris has three central areas with the highest accessibility scores. These are from Opéra to the Louvre, Châtelet-Les Halles, and the Latin Quarter, each representing the centers of art, education, and commerce, respectively. Moreover, overall, the scores generally decrease in concentric circles moving outward from these central areas, which aligns with our expectations.

4.2 TOD degree analysis result

The Node index and Place index respectively reflect the degree of transportation convenience and the abundance of facility resources within an Analysis Zone (AZ) under the influence of rail transit. The results are visually represented by Figures 3 and 4 respectively. The Node index image significantly shows a high index in the central urban area, which diminishes in

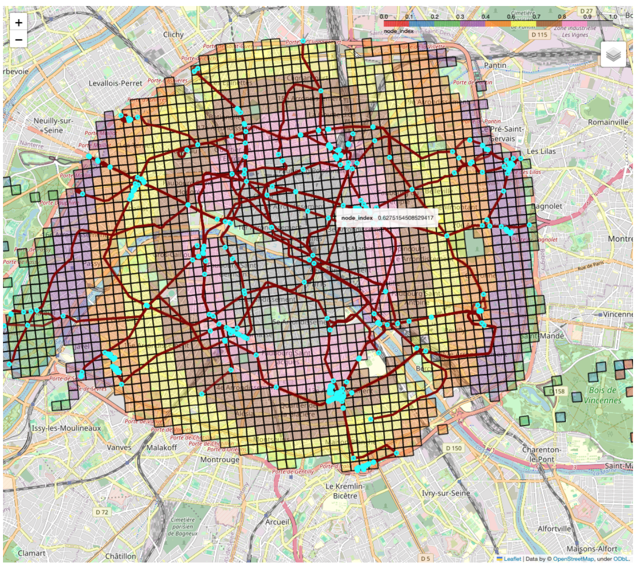


Figure 3: Node Indices of Paris

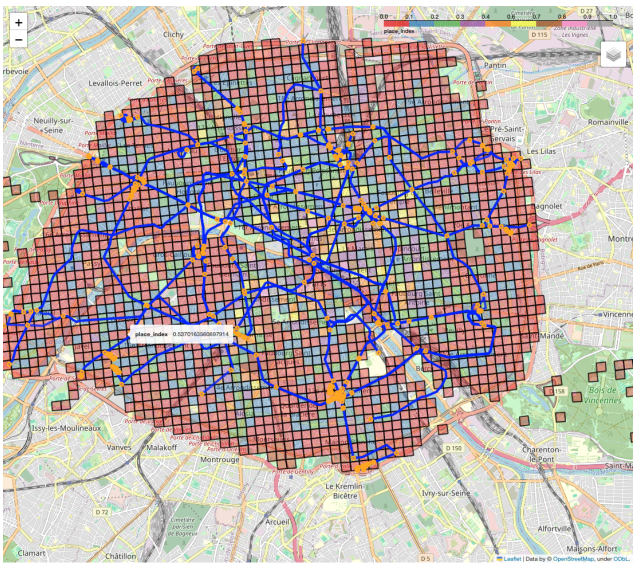


Figure 4: Places Indices of Paris

concentric circles moving outward. This pattern is primarily due to two factors. On one hand, the Parisian rail transit has a certain circular structure. On the other hand, because our Analysis Zones (AZ) are 200x200 meters and the Paris region as a whole is relatively small, walking constitutes a significant portion of people's transportation behavior. This contributes to the pronounced circular structure observed in the Node index image. In contrast, the Place index appears more scattered and lacks a distinct structural characteristic. Therefore, we aim to conduct regression tests to calculate the TOD degree.

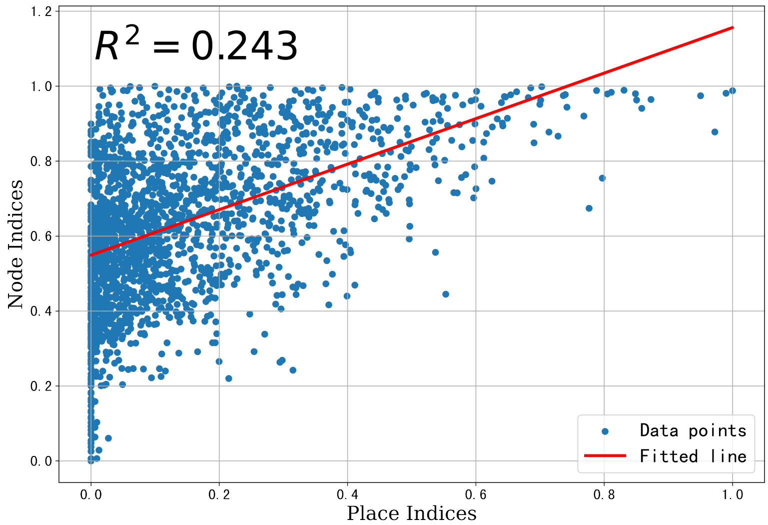


Figure 5: Scatter Plot of Place Index vs Node Index with Correlation Analysis

	R	R^2	45-degree slope difference (degrees)	Line slope	Mean value of node index norm	Mean value of place index norm
Paris	0.493	0.243	-23.219	0.399	0.639	0.149

Table 4: TOD degree of the urban structure in Paris and other indices describing the scatterplots

Figure 5 illustrates the correlation between these two indices through a scatter plot, providing a visual representation of how the node and place indices relate to each other across the different AZs of Paris. The R-value of 0.493 and the R^2 value of 0.243 indicate a moderate positive correlation, meaning that areas with higher connectivity tend to have a higher density and diversity of amenities, which is consistent with the principles of TOD. However, compared to other European cities, Paris's R-value remains relatively low. We believe this is due to the relatively dispersed nature of amenities in Paris, and the overall high density and even distribution of rail transit, which does not show a distinct difference between suburban and urban areas.

5 Discussion and Conclusion

Although the analysis conducted in this project yields relevant and satisfactory results and is more akin to an analytical report than other research papers, it is important to acknowledge the presence of certain shortcuts and flaws in the study, which pose twofold main challenges:

1. To the best of our knowledge, there is currently no literature that provides a systematic analysis and calculation method for accessibility analysis using the 2SFCA under the influence of public transport, particularly rail transit. We combined the node-place model method for calculations and proposed a systematic methodology and code implementation, which is one of the main contributions of this project. However, there is still considerable room for improvement and refinement in the methodology, such as the formula for calculating weights, as well as how to select more reasonable coefficients, including using 3SFCA [5, 10] (respectively, fixed point SFCA method) to improve upon the 2SFCA.
2. On the other hand, the challenge lies mainly in the implementation of the code. We first needed to integrate the data from OpenStreetMap with the grid data from Filosofi, which involved extensive data processing work. The actual computation of the Node Index required distance measurements, for which we used OSMnx and NetworkX to calculate the shortest paths. Considering that Paris is divided into more than 1700 zones, each measuring 200x200

meters, calculating the optimal rail distance between each zone presents a significant computational challenge. However, the optimal path algorithms for maps are extremely complex, and the granularity of rail transit information in OpenStreetMap is insufficient. We did not take into account spatial information and the impact of different routes, merely treating railways as a network for computation, thereby introducing errors in this area.

OpenStreetMap is a widely used database in urban science research. However, its inaccuracies can affect our research. For instance, in our analysis of bus stops in the Paris area, both "highway=bus_stop" and "public_transport=platform" tags indicate bus stops in the database. This overlap can result in data duplication or omission, posing challenges for accurately assessing public transportation infrastructure.

In summary, under the influence of public transportation, there are differences in overall convenience in Paris, but these differences are not as pronounced as one might expect. This is primarily due to the highly dense public transportation network in Paris, which mitigates the inequality in accessibility. Additionally, the relatively compact size of the city amplifies the role of the transportation network (for example, allowing people to quickly travel from the northern to the southern end of Paris via metro or RER).

During the TOD degree analysis, Paris did not exhibit particularly favorable results, and the correlation between the node index and place index was relatively low. This is largely attributed to the unique urban characteristics of Paris. Paris is a city with clearly defined boundaries, generally referring to arrondissement 1 to 20. It has a high concentration of resources, whether in education, art, culture, etc. Additionally, public transportation, such as the metro, is confined primarily to the Parisian city area. Few cities in the world can boast such a dense concentration of resources within such a small area. When compared with other cities, the disparity in urban scale is a significant factor contributing to the considerable differences in TOD degree. However, if we extend our analysis to include the Petite Couronne, we might observe an imbalance in the distribution of amenities and transportation resources, given the significant difference in urban structure between Paris and its surrounding suburbs. In the Petite Couronne, public transportation primarily consists of the RER, L line, J line, and buses, with a lower density compared to central Paris. Viewed holistically, this scenario results in a high concentration of transportation resources in central Paris, with the railway network, primarily the RER, radiating outward and public transport facilities becoming less dense. Ultimately, this pattern contributes to more favorable TOD degree analysis outcomes.

6 Further work

Future work and improvements to the project should focus on the following areas:

1. Enhancing and refining the statistical methods used, such as incorporating the fixed point SFCA method, and taking into account the social and demographic composition. For example, assigning a higher weight to schools for younger populations and giving more importance to medical resources for older age groups.
2. Expanding the scope of the study from the Paris region to include the Petite Couronne, and possibly even the entire Île-de-France region.
3. Employing methods like Principal Component Analysis (PCA) and clustering to explore the impact of different types of transportation.
4. Developing more precise algorithms for optimal route calculation, possibly by integrating commercial map databases such as the Google Maps API into the project to obtain more accurate path length calculations. Also, considering the spatial information of public transportation, especially vertical aspects (like underground railways, elevated bridges, tunnels, etc.).

Bibliography

[1] Mark Abrahamson. *Globalizing Cities: A Brief Introduction*. Routledge, 2019.

[2] Geoff Boeing. "OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks". In: *Computers, Environment and Urban Systems* 65 (2017), pp. 126–139.

[3] Mordechai Haklay and Patrick Weber. "OpenStreetMap: User-Generated Street Maps". In: *IEEE Pervasive Computing* 7.4 (2008), pp. 12–18. doi: 10.1109/MPRV.2008.80.

[4] Shunwei Ji et al. "How to find vacant green space in the process of urban park planning: case study in Ningbo (China)". In: *International Journal of Environmental Research and Public Health* 17.21 (2020), p. 8282.

[5] Huilin Liang et al. "Using an improved 3SFCA method to assess inequities associated with multimodal accessibility to green spaces based on mismatches between supply and demand in the metropolitan of Shanghai, China". In: *Sustainable Cities and Society* 91 (2023), p. 104456.

[6] Wei Luo and Fahui Wang. "Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region". In: *Environment and planning B: planning and design* 30.6 (2003), pp. 865–884.

[7] Matthew R McGrail and John S Humphreys. "Measuring spatial accessibility to primary care in rural areas: Improving the effectiveness of the two-step floating catchment area method". In: *Applied Geography* 29.4 (2009), pp. 533–541.

[8] Carlos Moreno et al. "Introducing the "15-Minute City": Sustainability, resilience and place identity in future post-pandemic cities". In: *Smart Cities* 4.1 (2021), pp. 93–111.

[9] Enrica Papa and Luca Bertolini. "Accessibility and transit-oriented development in European metropolitan areas". In: *Journal of Transport Geography* 47 (2015), pp. 70–83.

[10] R Shanmathi Rekha et al. "Accessibility analysis of health care facility using geospatial techniques". In: *Transportation Research Procedia* 27 (2017), pp. 1163–1170.

[11] GENET Simon, MAURICE Leopold, and THAURY Marie-Olive. "Analysis of the composition of big cities around Points-Of-Interest". In: *ENSAE Stat'App Report* (2023).

Appendix A Datasets, Links, and Tables

1 Links to the dataset and package

Providing links and resources that can assist in accessing the fundamental databases required to run this project.

1. **Filosofi dataset:** Données carroyées – Carreau de 200m. Here we mainly use the dataset with the name of Carreau 200m – Geopackage – Métropole - Martinique - La Réunion. Or you can download it more straightforward through this link. Then also provide the dataset with 1km $\times$ 1km grid in Données carroyées – Carreau de 1km. You can find all the explanation for the variables in this webpage.
2. **OSMnx documentation:** https://osmnx.readthedocs.io/en/stable/ - The document to introduce **OSMnx** package with Python.
3. **Course on OSMnx by G. Boeing (created OSMnx):** https://github.com/gboeing/osmnx-examples - The examples contain most basic usage of the package.
4. **OpenStreetMap Wiki:** https://wiki.openstreetmap.org/wiki/Main_Page - In this website, you can find the tags of each node, facility, amenity, transportation in the map, which you need those tags to absorb the data you need for the OSM.
5. **Geopandas:** https://geopandas.org/en/stable/
6. **NetworkX:** https://networkx.org/
7. **OSMnx:** OSMnx utilizes the Overpass API to access OpenStreetMap (OSM) data. It encapsulates this access, simplifying and rendering OSM network graphs more realistically. Additionally, OSMnx incorporates algorithms for network analysis. As of now, OSMnx stands as arguably the most efficient tool for accessing OSM data.

2 Github file information

You can find the code of the project in https://github.com/yubocai-poly/BATO-MOUTCHE-Extension.

1. basicstatistics.py: Statistics tools and visualization functions.
2. transportationtool.py: Code implementation for section 3 methodology.
3. TOD_analysis.ipynb: This is the result file of section 4.2.
4. accessibility_analysis.ipynb: This is the result file of section 4.1.
5. basic_stats.ipynb All the basic statistics result of the public transportation in Paris.

3 Table of the tags

In order to analyse the city composition around POIs, we define aggregated categories of POIs, based on *OpenStreetMap* tags :

[1] **Restaurants:** all types of restaurants including cafes, bars, fast-foods, pubs, ice-cream shops.
[2] **Culture and art:** shops and amenities related to literature, music, cinema, plastic arts, performances, video games, games.
[3] **Education:** primary schools, middle schools, high schools, colleges, universities.
[4] **Food shops:** including supermarkets as well as specialist food shops (e.g. bakeries, butchers, dairy shops, seafood shops, wine shops...).
[5] **Fashion and beauty:** all shops related to clothes, fashion accessories (e.g. jewellery, watches), beauty care (e.g. cosmetics, hairdresser, massage, hair removal, perfumery).
[6] Supply shops: everyday life shops apart from food shops (e.g. insurance, sport shops, furniture shops, household appliance shops ...).

Details of the **OpenStreetMap** tags in each category are listed in the Table A.1

Category	OSM key	OSM tags
Restaurant	amenties	bar, biergarten, cafe, fast food, food court, ice cream, pub, restaurant
Culture and art	amenties	arts centre, cinema, conference centre, events venue, library, music school, planetarium, public bookcase, studio, theatre, toy library
	shops	anime, antiques, art, books, camera, collector, craft, frame, games, model, musical instrument, music, photo, ticket, trophy, video, video games
Education	amenties	college, kindergarten, school, university
Food shops	shops	alcohol, bakery, beverages, brewing supplies, butcher, cheese, chocolate, coffee, confectionery, convenience, dairy, deli, farm, frozen food, greengrocer, ice cream, pasta, pastry, seafood, spices, tea, wine, water, supermarket
Fashion and beauty	shops	bag, boutique, beauty, clothes, cosmetics, erotic, fabric, fashion accessories, hairdresser, hairdresser supply, jewelry, leather, massage,perfumery, sewing, shoes, tailor, tattoo, watches, wool
Supply shops	shops	agrarian, appliance, atv, baby goods, bathroom furnishing, bed, bicycle, boat, bookmaker, candles, cannabis, car, caravan, car parts, carpet, car repair, charity, chemist, computer, copy-shop, curtain, department store, do-it-yourself, doors, dry cleaning, e-cigarette, electrical, electronics, energy, fireplace, fuel, fishing, flooring, florist, fuel, funeral directors, furniture, garden centre, garden furniture, gas, general, gift, glaziery, golf, groundskeeping, hardware, health food, hearing aids, herbalist, hifi, household linen, housewar, hunting, insurance, interior decoration, jetski, kiosk, kitchen, laundry, lighting, locksmith, lottery, mall, medical supply, military surplus, mobile phone, money lender, motorcycle, newsagent, nutrition supplements, optician, outdoor, outpost, paint, party, pawnbroker, pest control, pet, pet grooming, pyrotechnics, radiotechnics, religion, scuba diving, security, ski, snowmobile, sports, stationery, storage rental, swimming pool, telecommunication, tiles, tobacco, toys, trade, trailer, travel agency, tyres, vaccum cleaner, weapons, window blind

Table A.1: OSM tags selected in each category

Types	**function**	**OSM Key**	**Tags**
Bus	stop	public_transport	platform
	stop	highway	bus_stop
Railway	stop position	public_transport=stop_position + train=yes + railway=stop	
	platform	public_transport=platform + railway=platform	
Trams	stop position	public_transport=stop_position + tram=yes + railway=tram_stop	
	platforms	public_transport=platform + railway=platform	
Cable cars, chair lifts, gondolas, etc	stations	aerialway	station
Bicycle Rental	amenity	bicycle_rental	docking_station

Table A.2: OSM tags of the public transport

Appendix B Graphs and Plots

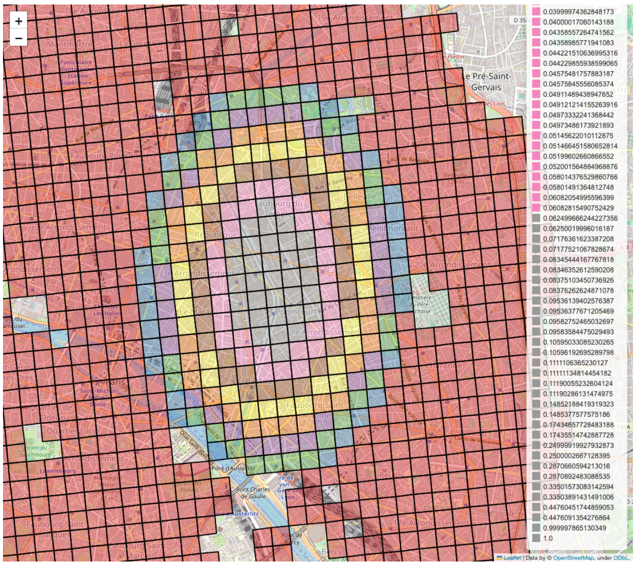


Figure B.1: X=30

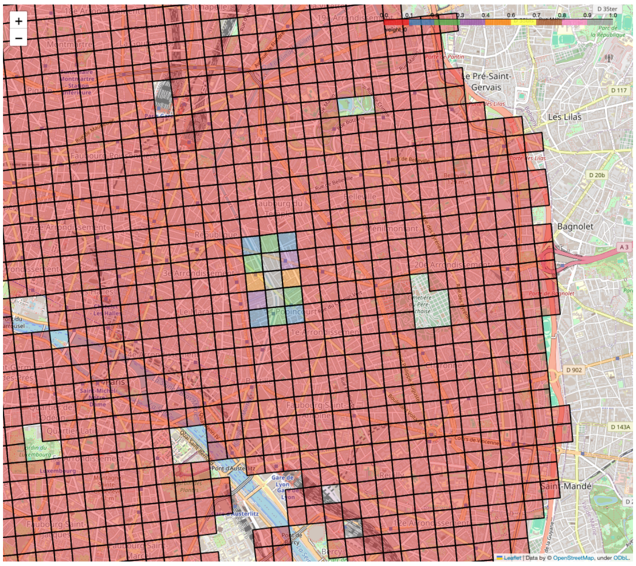


Figure B.2: X=15

Figure B.3: Weights of each square for the grey square in the center within X-mins threshold, comparison images

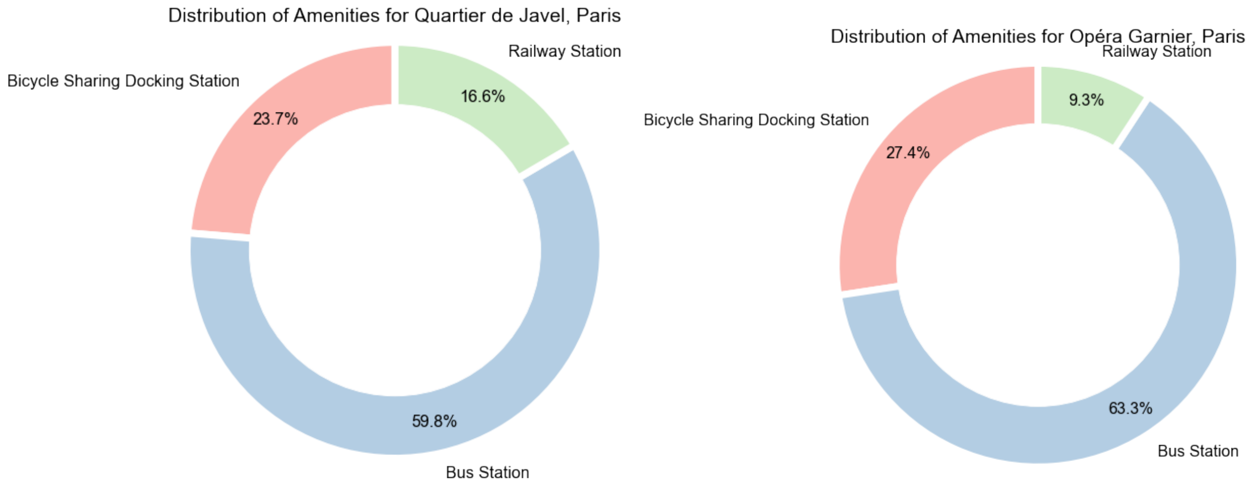


Figure B.4: The composition of two Parisian districts, Quartier de Javel and Opéra Garnier, presents contrasting characteristics. Quartier de Javel is primarily considered a residential area, while Opéra Garnier is a major commercial hub. However, determining their nature based solely on the proportion of public transportation infrastructure from the provided diagrams proves challenging.

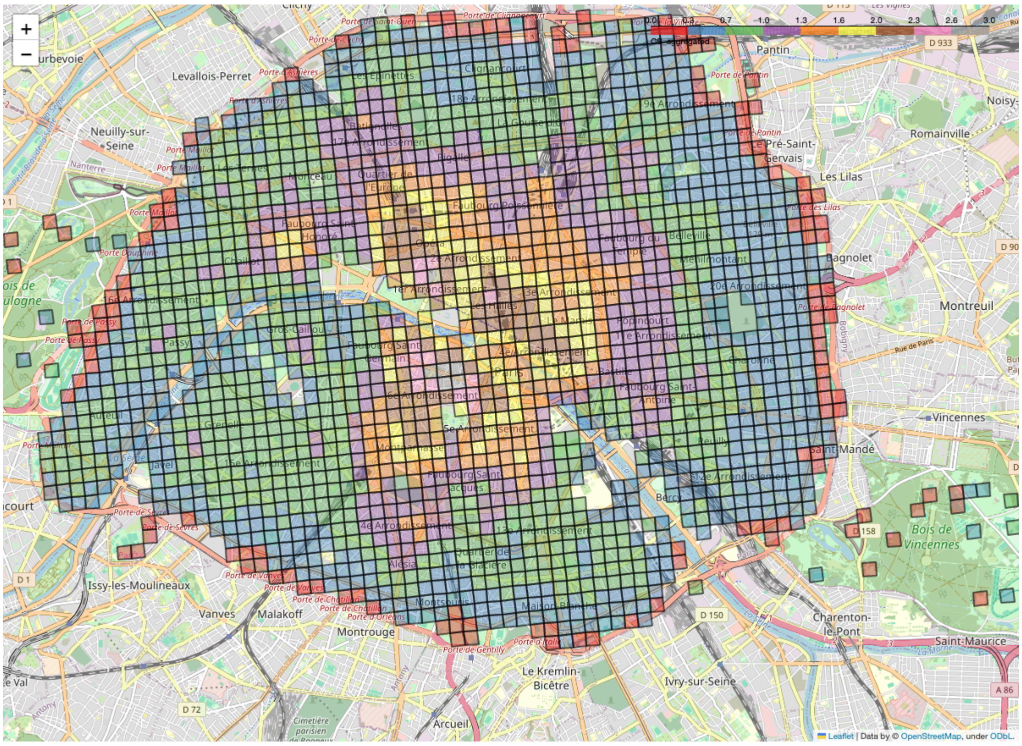


Figure B.5: With weight w_p

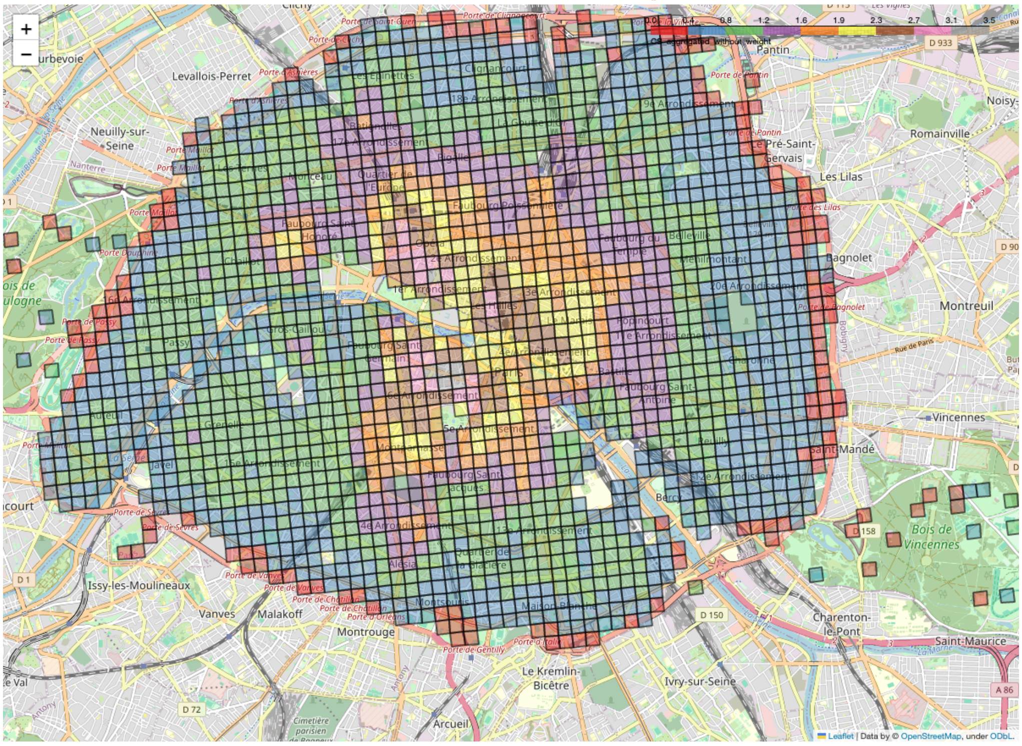


Figure B.6: Without weight w_p

Figure B.7: Aggregated 2SFCA under the influence of public transportation with and without amenity weights for Paris