Analysis of city composition based on basic amenities including public transport

Authur: Yubo Cai
Supervisor: Sarah J. Berkemer

LIX AMIBIO Team, Ecole Polytechnique

2022-2023

Motivation

- Why analysis of city composition based on amenities is **needed**?
  - Urban development has a **high impact** on living quality as well as economics development.
  - Better urban planning and reducing the wealth gap.
- How does computer science help?
  - Efficient processing of **massive** amounts of statistical data and spatial information.
  - Conducting **statistics analysis** (regression, clustering) using programs.
  - Enhanced clarity in **visualization**.

This program focuses on urban science analysis through **statistical** and **data science** methods.

Contribution

- Build Two-step floating catchment area method (2SFCA) [1, Section 3] for accessibility analysis of amenities base on transport network.
- Proposal of Transit-Oriented Development (TOD) degree measurement methodology.
- Package development: https://github.com/yubocai-poly/BATO-MOUTCHE-Extension.
- Analysis and Conclusion of the result.

Datasets Information

We merged the homogeneous grid dataset provided by **Filosofi** with **Open Street Map** (OSM) to obtain the dataset we ultimately used.

Filosofi system: A database focused on localized fiscal and social data.

- Divides the territory into 200(or 1000)-meter square areas.
- Age pyramid of residents.
- Residents' income.
- Construction year of buildings

OSM: A free dataset that includes spatial information such as buildings, transportation networks, and points of interest.

- Points of interest are categorized into various classes (shopping, health, etc).
- Transportation information: Station, lines, categories.
- Cons:
  - Errors and Inconsistencies in information.
  - Lack of shortest path computation package.

Remark: In our project, we mainly consider the influence of the railway network, this will be explained later.

We divide the amenity into the following 6 types:

- Restaurant
- Culture and art
- Education
- Food shops
- Fashion and beauty
- Supply shops

More detailed information in Table A.1 in the report.

Accessibility scores for each type of amenity with railway network

– 2SFCA method description –

The 2SFCA approach can be described in two steps:

- **service catchment:** For each service (amenity), find all populations that fall within a threshold distance (dmax) and calculate the population-to-provider ratio.
- **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.

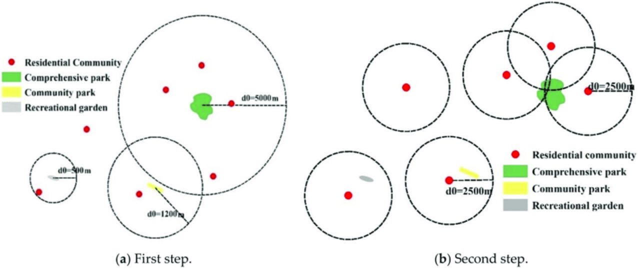

Figure: 2SFCA rationale illustration, originating from paper [2]

Accessibility scores for each type of amenity with railway network

– 2SFCA method description –

However, considering the influence of the *railway network*, we have several problems:

- How to set the threshold distance?
- How to the permeability coefficient of demand from k to square j?

Instead of using the threshold distance, we use the **traveling time** as our threshold.

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

There are two types of traveling time: by walking or by railway:

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

Remark: Railway includes Metro, RER, Tram. In the project, we set $\beta = 1.2$, the walking speed of 80m/min and the railway speed of 700m/min.

Accessibility scores for each type of amenity with railway network

– 2SFCA method description –

Weight between zone i and j:

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

We set our time threshold as 30 mins.

Aggregated accessibility score

How do we merge the score for each type of amenity together: Using importance weight.

$$f = \frac{N_p}{N} \quad \text{and} \quad w_p = 1 - \frac{N_p}{N} \quad \text{where } p \text{ is the category} \tag{3}$$

Why? Less number of the type $\Rightarrow$ More important in the region.

$$CS_i = \sum_{p=1}^{P} (1 - w_p) \times X_{i,p} \quad \text{where } p \text{ is the category}$$

TOD degree measurement

Transit-Oriented Development (TOD) degree: represents the spatial concentration of economic activities and population along the rail transit networks.

How to measure: Correlation of the node index and place index.

TOD degree measurement

Node index: the closeness centrality index.

$$NI_i = \frac{N-1}{\sum_{j-1}^{N} \min(t_{ij,\text{walk}}, t_{ij,\text{railway}})}$$

Place index: amenity score per unit area. The amenity score (AS) and place index with the following formula:

$$AS_i = \sum_p N_p w_p$$

$$\text{place index}_i = \frac{AS_i}{\mathrm{A}_{\mathrm{AZ}_i}}$$

Remark: We applied **min-max normalization** to calculate standardized node and place indices for each study.

Basic Statistics

	Railway	Bus	Bicycle Rental
Numbers of Station	308*	2801+	992+

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

	Share and Number of public transport trips in Paris (millions/percentage). 2019			
	Metro	RER (RATP and SNCF)	Tramway (includes T4 and T11E)	Buses
Paris	1498 (42.4%)	1407 (39.8%)	340 (9.6%)	291 (8.2%)

Table: Structural transport system variables in Paris, Railway takes 91.8% of the total number of public transport trips in Paris

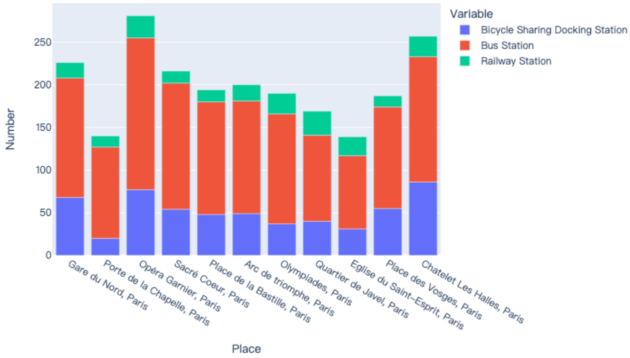

Figure: Comparison of the composition of transportation in 11 districts (with 1km radius) of Paris

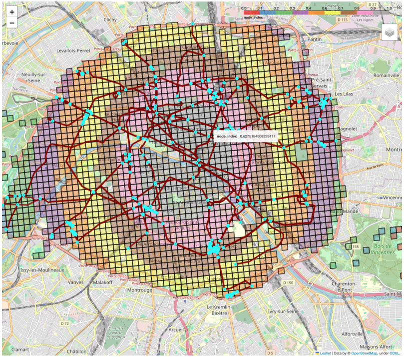

Figure: Node Indices of Paris

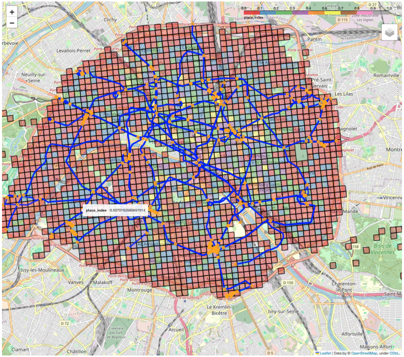

Figure: Places Indices of Paris

TOD degree analysis result

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

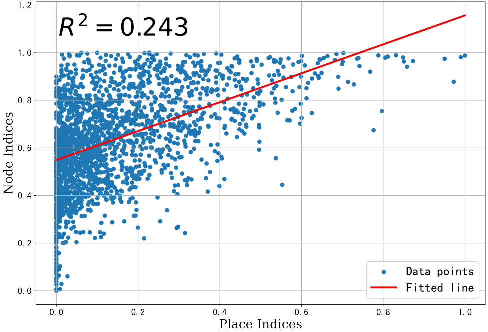

Figure: Scatter Plot of Place Index vs Node Index with Correlation Analysis

Conclusion

- There are differences in overall convenience in Paris due to its dense public transportation network, but these differences are less pronounced than expected.
- During the TOD degree analysis, Paris did not show particularly favorable results, with a low correlation between the node index and the place index.
- The high concentration of transportation resources in central Paris, with the railway network, primarily the RER, radiating outwards, and less dense public transport facilities.
- Highly centralized resources in Paris, lack of resources outside Paris.

Future work

- **Enhancing and refining the statistical methods used**, such as incorporating the fixed point SFCA method, and taking into account the social and demographic composition.
- Expanding the scope of the study from the Paris region to include the Petite Couronne.
- Employing methods like Principal Component Analysis (PCA) and clustering to explore the impact of different types of transportation.
- Developing more precise algorithms for optimal route calculation.

Reference

[1] 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.

[2] 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.