A spatial Evaluation of Essential Citizenship Services at municipal level. The contribution of Network Analysis and Machine Learning Methodologies

Angela Stefania Bergantino[1], Mario Intini[1], Gianluca Monturano[1]

Department of Economics, Management and Business Law,University of Bari Aldo Moro, Italy, Largo Abbazia S. Scolastica 53, Bari, BA, Italy

1. INTRODUCTION

This research focuses on evaluating the accessibility of essential services in Italian municipalities. Using advanced machine learning and network analysis techniques, we assess the current state of accessibility and predict future developments, aiming to provide valuable insights for improving the distribution of services across the country.

1.1. Literature Review and Research Context

Inequalities in access to services reflect broader socioeconomic gaps, affecting factors such as education, health outcomes, and economic opportunities. Key studies [1] highlight the crucial role of geographic accessibility in economic productivity and social cohesion. Rodriguez-Pose [2] and Barca [3] emphasize the importance of targeted regional strategies, while Florida [4] and Acemoglu and Autor [5] explore the dynamics between regional inequalities and technological changes. So far, most assessments of accessibility to services have relied predominantly on causality analyses using counterfactual econometric estimators. However, predictive approaches, less used so far, are becoming increasingly essential to anticipate future trends and dynamics. There is a growing need for data-driven methodologies that integrate complex systems to better predict changes in accessibility and guide policy making.

This project aims to contribute to the existing literature through a polycentric study of the Italian territory, overcoming traditional territorial dichotomies such as North-South and coast-hinterland, providing a more granular analysis of service accessibility [6]. Using unsupervised machine learning [7] and network analysis, we aim to identify underserved areas and develop innovative strategies to improve accessibility and territorial resilience.

Our approach promotes sustainable development and equitable access to services, thus contributing to more effective public policies and urban planning.

2. DATA, METHODS, AND EMPIRICAL STRATEGY

2.1. Data Collection and Spatial Mapping

The methodological process begins with a detailed mapping of Italian municipalities, aimed at identifying and locating essential public services such as schools, hospitals, and railway stations. We use geospatial data to determine the exact coordinates of these services across Italian municipalities, thus establishing the foundation for subsequent analyses.

2.2. Network Analysis with Spatial Calculation of Road Distances

We calculate the road distances between services using and iterating functions from the Open Source Routing Machine (OSRM) package developed within the R programming environment (Giraud & Lambert, 2018; Ferster et al., 2022).

The generic formula for calculating the distance in a graph, using and implementing OSRM, can be expressed with the following basic equation:

$$d_{street}(u,v) = min(\sum_{i=1}^{n} l(e_i)$$

- where $d_{stret}(u, v)$ represents a path in the graph between two services (nodes)u and v; $l(e_i)$ the length of the arc; e_i the shortest path u, v.

Practically, through distance-network analysis, we calculate the road distances between the selected services for each Italian municipality (see Table 1 for an example based on simulated random data). Consequently:

- For municipalities with all services, the graph connects the distances between services within the same municipality;
- For municipalities with at least one missing service, the remote network algorithm connects the incomplete municipality to the nearest municipality with the missing service, within 20 minutes, optimizing the shortest road route.

Table 1. Network Database

From	To	Distance	Duration)	Health	Education	Transport	Connection
Municipality1_hospital1	Municipality1_school1	0.5	2	1	1	0	Direct
Municipality1_hospital1	Municipality1_station1	0.6	3	1	0	1	Direct
Municipality1_hospital1	Municipality1_school2	0.7	2	1	1	0	Direct
Municipality1_school1	Municipality1_station1	0.3	1	0	1	1	Direct
Municipality1_school1	Municipality1_school2	0.4	1	0	2	0	Direct
Municipality1_station1	Municipality1_school2	0.8	2	0	1	1	Direct
Municipality2_school	Municipality2_station	0.6	1.5	0	1	1	Direct
Municipality4_service3	Municipality1_hospital1	2.5	5	1	0	0	Indirect
Municipality5	Municipality1_hospital1	3	6	1	0	1	Indirect
Municipality5	Municipality1_school1	3.2	6.5	0	1	1	Indirect
Municipality5	Municipality1_station1	3.4	7	0	0	1	Indirect
Municipality5	Municipality4_service3	4	8	1	0	0	Indirect

2.3. Machine Learning and Clustering for Service Accessibility

From the distance-network analysis, we derive municipal-level average engineering features, which we combine with granular socio-economic and institutional variables (e.g., per capita income, resident population, number of local businesses, mayor's educational background). We use these datasets to cluster Italian municipalities based on service accessibility. To achieve this, we employ unsupervised machine learning algorithms, which by design do not require labeled data. Specifically, we implement algorithms such as:

- **K-Means:** minimizes the variance within clusters. $S = \sum_{i=1}^{k} \sum_{x \epsilon C_i} (x - \mu_i)^2$. Whereμ_i is the centroid of the cluster C_i.
- **DBSCAN**: identifies clusters based on point density. $DBSCAN(D, \epsilon, MinPts)$. Where D is a dataset, ϵ it is the proximity radius, $MinPts$ is the minimum number of points in a neighborhood to form a dense region.
- **Hierarchical Clustering**: creates a hierarchy of clusters that can be visualized as a dendrogram. There is no single formula, as the method proceeds through levels of aggregation based on distance or similarity.
- **Spectral Clustering**: uses the spectral properties (eigenvalues) of the dataset's similarity matrix to reduce the dimensionality before clustering. $L = D^{\frac{1}{2}} - (D - W)D^{\frac{1}{2}}$. Where L è the normalized Laplacian matrix, D is the degree matrix and W is the adjacency matrix.

- **Mean Shift**: proceeds towards the data density peaks, based on the kernel density estimate $m(x) = \frac{\sum_{xi \in N(x)} K(xi-x)xi}{\sum_{xi \in N(x)} K(xi-x)}$. Where $m(x)$ is the mean shift vector, N it's the neighborhood of x, and K is a kernel.

3. PRELIMINARY RESULTS

The map in Figure 1, which applies the K-Means clustering algorithm with k=3 selected through the Elbow Method, divides Italian municipalities into three distinct clusters based on accessibility to citizenship services, such as schools, hospitals and transport infrastructure. From the analysis of the map, the clusters corresponding to the largest urban areas, such as Rome and Milan, are highlighted, which show a high density of services, indicating better accessibility. These urban clusters, densely populated and with a wide range of services, contrast markedly with those of more rural or less densely populated areas, where services are more scattered and less accessible. This geographical distinction highlights the importance of targeted policies that address the specific needs of different areas. For example, it may be necessary to improve the transport infrastructure in rural areas or increase the number of services in municipalities with less accessibility. Furthermore, it may also be appropriate to strengthen existing services in large cities to maintain and improve the effectiveness of access to services in the face of growing demand. In summary, the classification of municipalities into clusters based on the accessibility of services provides a useful guide for the development of public policies and urban planning, aiming to ensure fair and effective access to essential services for all citizens, regardless of their geographical location.

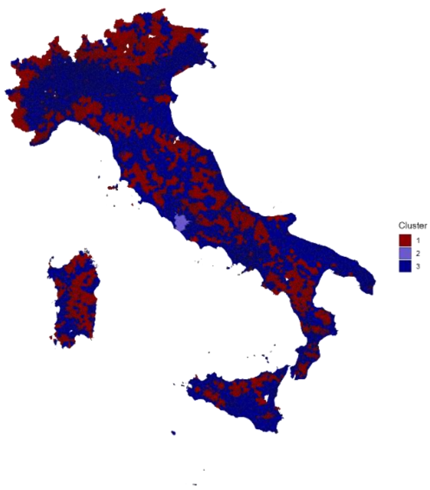

Figure 1. Cluster Analysis of Citizenship Service Accessibility

REFERENCES

[1] Krugman, P. (1991). *Geography and Trade*. MIT Press.

[2] Rodriguez-Pose, A. (2018). *The revenge of the places that don't matter (and what to do about it)*. *Cambridge Journal of Regions, Economy and Society*, 11(1), 189–209.

[3] Barca, F. (2009). *An agenda for a reformed Cohesion Policy: A place-based approach to meeting European Union challenges and expectations*. Bruxelles.

[4] Florida, R. (2002). *The Rise of the Creative Class*. Basic Books.

[5] Acemoglu, D., & Autor, D. (2011). *Skills, tasks and technologies: Implications for employment and earnings*. In *Handbook of Labor Economics*, vol. 4, pp. 1043–1171.

[6] Bergantino, A. S., Buonarota, M., Buongiorno, A., & Intini, M. (2023). *Regional multimodal accessibility: Policies and strategies for sustainable tourism destinations in coastal areas*. *Research in Transportation Business & Management*, 48, 100872.

[7] Cerulli, G. (2023). *Optimal Policy Learning with Observational Data in Multi-Action Scenarios: Estimation, Risk Preference, and Potential Failures*. arXiv preprint arXiv:2403.20250.