Analyzing Accessibility to Citizenship Services
A Network Analysis and Machine Learning Approach

Angela Stefania Bergantino[1], Mario Intini[1], Gianluca Monturano[1,2]

[1]Department of Economics, Management and Business Law - University of Bari
[2]Department of Economics - University of Modena and Reggio Emilia

European week of Cities and Regions Close to you
Napoli, November 26, 2024

- **Data-Driven Decision Making:** Machine learning transforms how we understand complex economic systems by enabling data-driven insights and decision-making.
- **Predictive Analytics:** ML techniques are crucial for forecasting economic trends, consumer behavior, and market dynamics with high accuracy.
- **Automation and Efficiency:** Automation of routine tasks in economic analysis reduces errors and increases efficiency, allowing economists and businesses to focus on strategy and innovation.
- **Personalization:** ML supports the customization of financial services and products to individual needs, enhancing customer satisfaction and engagement.
- **Risk Management:** Advanced algorithms help in predicting and mitigating risks by analyzing large volumes of data, crucial for financial institutions.

Supervised Learning

- **Definition:** Learns from labeled data to predict outcomes for new data.
- **Usage:** Used in applications where historical data predicts future events, such as credit scoring and sales forecasting.
- **Examples:** Regression, Decision Trees, SVM.
- **Requirement:** Requires a dataset with input-output pairs.

Unsupervised Learning

- **Definition:** Identifies patterns and relationships in data without labels.
- **Usage:** Useful for segmenting data into clusters and discovering the underlying structure of datasets.
- **Examples:** Clustering, Principal Component Analysis (PCA).
- **Requirement:** Does not require labeled data, works directly with input data.

- Income levels (Bloise, Brunori, Raitano, 2021)
- Health (Carrieri, Lagravinese, Resce, 2024)
- Public Spending and Local Sustainability (Resce, 2022)
- Education (Delogu, Lagravinese, Paolini Resce, 2023)
- Delay of Infrastructure Coeshion Policies (Coco, Monturano, Resce, 2024)
- Resilience to shocks (Bonacini, Gallo & Patriarca, 2021)

- This research aims to evaluate the accessibility of essential services across Italian municipalities (a cross-sectional dataset for 2023)

- By employing advanced **Network Analysis** and **Machine Learning** techniques, we seek to:
  - Assess the **current state** of accessibility to key services like **health**, **education**, and **mobility**
  - Identify underserved areas and predict **future accessibility** challenges
  - Propose data-driven **policy interventions** to enhance the equity of service distribution.

- The study integrates large-scale data, geospatial analysis, and predictive modeling to provide systematic insights into reducing territorial inequalities.

Spatial heterogeneity:

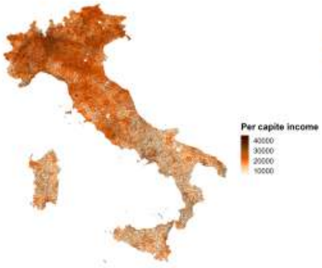
- **one of major determinants of country differences in income inequality** (Bourguignon and Morrisson, 1998),
- **negative association with growth**(Alesina and Rodrik, 1994; Clarke, 1995; Persson and Tabellini, 1994)
- **driver of populist backslash taking place around the world** (Albanese et al., 2022; Rodrıguez-Pose, 2018)

To compensate for such imbalances many national and international institutions have been developing a consistent policy framework with the stated aim of reducing the economic issues

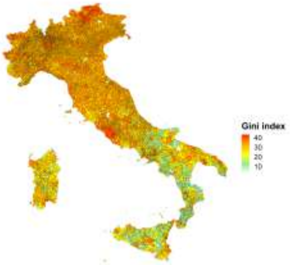
- Accessibility to citizenship services, such as health, education, and mobility, is crucial for regional development **(Crescenzi & Rodriguez-Pose, 2011; Rodriguez-Pose, 2018; Bergantino et al. 2024)**
- Disparities in access lead to inequalities in socio-economic conditions and quality of life **(Becker et al., 2018; Iammarino et al., 2019)**
- This research aims to analyze accessibility using Network Analysis and Machine Learning **(Acemoglu et al., 2012; Resce et al., 2023)**
- Restricted access to services can deepen regional inequalities and hinder human capital development. **(Putnam, 2000; Florida, 2002)**
- Addressing spatial disparities in service accessibility is essential for fostering inclusive growth and social cohesion. **(Sen, 1999; Storper, 2020)**

Italy is a unique case to study accessibility because it has significant economic inequality among its regions **(Alesina, 1996; Stiglitz, 2012; Piketty, 2014; Atkinson, 2015; Tridico, 2018; Iammarino, 2019; Liberati and Resce, 2022)**

- Significant disparities in access between Northern and Southern Italy. **(Svimez, 2019; Barca, 2009)**
- Unequal distribution of health, education, and transport services contributes to gaps in quality of life. **(Costa et al., 2018; Iacus et al., 2020)**
- Southern Italy faces challenges such as higher unemployment and lower quality of infrastructure. **(Trigilia, 2019)**
- Population decline is linked to insufficient services in rural areas. **(Barca et al., 2014; Dijkstra et al., 2020)**


Per capite income
40000
30000
20000
10000


Share of Poverty
70
60
50
40
30
20


Gini index
40
30
20
10

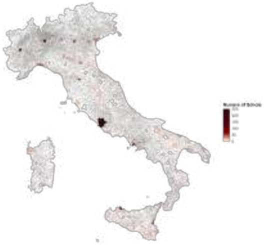


Figure: Map of Schools

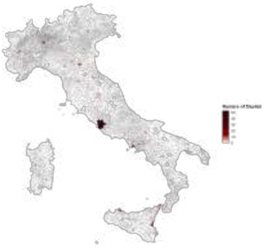


Figure: Map of Hospitals

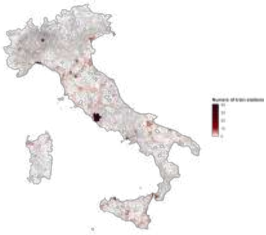


Figure: Map of Stations

- **SNAI (2014) is a place-based policy** primarily aimed at improving accessibility to citizenship services such as health, education, and mobility, but only in inner and remote areas (**Barca et al., 2014**)
- it is organized through a polycentric mapping aimed at overcoming the historical dichotomies in the economic and statistical territorial study, for example, South-North, Coastal and Hinterland
- Promotes equitable socio-economic development by enhancing infrastructure and ensuring equal access to **Citizenship Services**
- Implementation challenges include gaps in infrastructure and inefficiencies
- Previous studies show the ineffectiveness of the depopulation policy (e.g., **Vendemmia, 2021; Monturano et al., 2023**)

Spatial distribution of municipalities according to SNAI

Limitations of the National Strategy for Inner Areas (SNAI): The SNAI policy, while aiming to address accessibility issues in inner areas, lacks a detailed analysis of distances and travel times between essential services, both within individual municipalities and across neighboring ones. This study fills that gap by examining these spatial connections in detail.

To evaluate the accessibility of citizen services within and across municipalities, we used a **Network Analysis** **(Newman, 2010; Borgatti et al., 2009; Giraud & Lambert, 2018)**.

- **Nodes**: Represent individual services (e.g., hospitals, schools, train stations)
- **Edges**: Represent connections between nodes (distances, duration)

Network analysis helps to identify gaps, redundancy, and structural weaknesses in service delivery. This approach provides actionable insights to optimize public infrastructure investments, making it a powerful tool for improving access to essential services.

1. **Health services**: number of hospitals (public/private), IRCCS, and health research centers,
2. **Care services** nursing homes
3. **Educational services**: schools categorized by level (nursery, elementary school, high school), i.e. when a municipality has all the school offerings and type.
4. **Mobility services**: location of railway stations

Total Count of Each Service Type:

Service Type	Total Count
Schools	10995
Stations	2533
Nursing Homes	558
Care Centers	500

Total Number of Municipalities: 7901, of these:

- **5261** municipalities have **no services**
- **2640** municipalities have at least one service, distributed as follows:

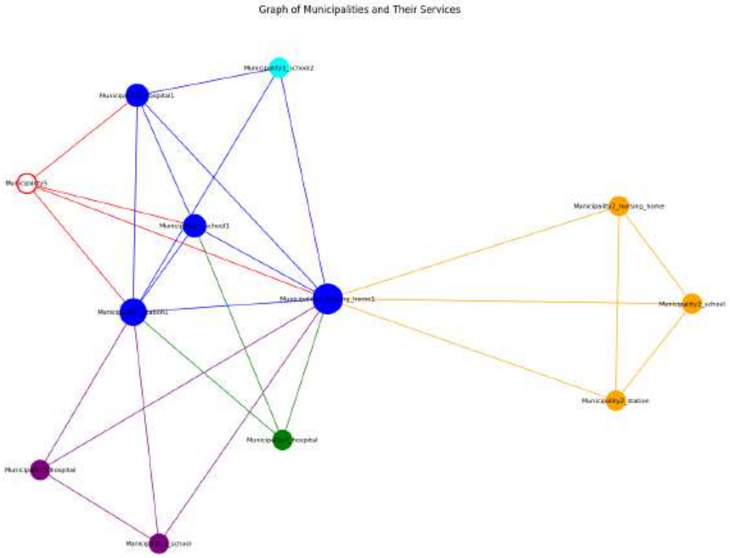
Number of Services	Number of Municipalities
1 Service	1702
2 Services	600
3 Services	235
4 Services	103

Network Analysis

- If a municipality has at least one service, it closes its Network Graph internally

- If a municipality lacks at least one service, it connects all existing services to the absent service(s) in the nearest municipality(s) within 20km of road, if it exists, following the SNAI theory

- If a municipality has no services, it connects its centroid to the nearest existing service(s) within 20km, if it exists, following the SNAI framework


Graph of Municipalities and Their Services
Municipality2_school

- **Euclidean Distance**: A measure between two points in a plane.

$$d_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$$

- **Road Distance and Duration (using OSRM)**: Calculated with the Open Source Routing Machine (OSRM) package within the R programming environment, providing both the drivable distance and estimated travel time between two points (**Giraud & Lambert, 2018; Ferster et al., 2022**).

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

where: $d_{\text{street}}(u, v)$ represents the shortest path in the graph between two services (nodes) u and v; $l(e_i)$ is the length of each arc e_i along the path.

Figure: Average Distance Map (km)

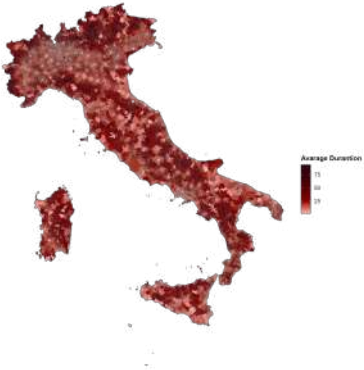


Figure: Average Duration Map (minutes)

From the Network we obtain the Feature Engineering:

- **Node Degree (k)**: $k_i = \sum_{j \in \mathcal{N}(i)} a_{ij}$
  Where a_{ij} is 1 if there is a connection between nodes i and j.
- **Betweenness Centrality**: $BC(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$ $\sigma_{st}(v)$ is the number of shortest paths from s to t passing through v.
- **Closeness Centrality**: $C(v) = \frac{1}{\sum_{t \neq v} d(v,t)}$
  Measures node proximity to others in the network.

Services Dataset: Distances and Connections

From	To	Distance (km)	Duration (min)	Health Services	Education Services	Transport Services	Connection
Municipality1_hospital1	Municipality1_hospital2	0.3	1	2	0	0	1
Municipality1_hospital1	Municipality1_school1	0.5	1	2	2	0	1
Municipality1_hospital1	Municipality1_school2	0.7	1	2	2	0	1
Municipality2_nursing_home	Municipality2_school	0.5	1	1	1	0	1
Municipality2_nursing_home	Municipality2_station	0.8	2	1	0	1	1
Municipality2_school	Municipality2_station	0.6	1.5	0	1	1	1
Municipality3_school	Municipality3_hospital	1.2	2	1	0	0	1
Municipality3_school	Municipality2_nursing_home	1.8	3	0	1	0	1
Municipality4_hospital	Municipality1_hospital1	4.5	10	1	0	0	1
Municipality5_centroid	Municipality1_hospital1	6.0	12	1	0	1	1
Municipality5_centroid	Municipality2_nursing_home	7.0	14	1	0	0	1
Municipality5_centroid	Municipality3_school	5.0	10	0	1	0	1
Municipality5_centroid	Municipality4_hospital	8.0	15	1	0	0	1

Municipal Dataset

Municipal Network Features:

- Min Geodetic Distance (km)
- Max Geodetic Distance (km)
- Average Geodetic Distance (km)
- Min Distance (km)
- Max Distance (km)
- Average Distance (km)
- Min Duration (min)
- Max Duration (min)
- Average Duration (min)
- Min Distance Reverse (km)
- Max Distance Reverse (km)
- Average Distance Reverse (km)
- Min Duration Reverse
- Max Duration Reverse
- Average Duration Reverse
- Total Services per Municipality
- Schools per Municipality
- Stations per Municipality
- Nursing Homes per Municipality
- Hospitals per Municipality
- School Connection
- Station Connection
- Nursing Home Connection

Municipal Socio-economic Features:

- Number of Residents
- Per Capita Income
- Share of Poverty
- Territorial Surface
- Altitude of Center (meters)
- Coastal Municipality
- Isolated Municipality
- Per Capita Land Consumption
- Population Density
- Share Resident Foreign Population
- Population Over 18
- Local Units
- Employees
- Female Mayor
- Mayor with Degree
- Majority Councilors with Degree
- Majority Assessors with Degree
- Mayor's Age
- Average Age of Councilors
- Average Age of Assessors

To make the features of the network and socio-economic indicators homogeneous in terms of levels and units of measurement, we use a double normalization technique, for each municipality

Step 1: Computing values as densities with respect to the municipal surface

$$Y_i^* = \frac{Y_i}{Area_i} \quad \forall i = 1, \ldots, N = 7975$$

where

- Y_i is the socio-economic and neatworks indicator
- $Area_i$ is the area of municipality i,

Step 2: Robust standardization using quantiles values

$$Y_i^{**} = \frac{Y_i^* - Median(Y_i^*)}{{Y_i^*}_{0.75} - {Y_i^*}_{0.25}} = \frac{Y_i^* - Median(Y_i^*)}{IQR(Y_i^*)} \qquad \forall i = 1, \ldots, N = 7975$$

Machine Learning
Unsupervised Learning

- **Machine Learning:** A branch of artificial intelligence (AI) focused on developing systems that can learn from and make decisions based on data (Cerulli, 2023; Goodfellow et al., 2016; Murphy, 2012). Machine learning algorithms are designed to identify patterns and make predictions or decisions without being explicitly programmed for each task (Maranzano et al., 2024; Bishop, 2006).

- **Unsupervised Learning:** A type of machine learning where the algorithm learns patterns and structure from unlabeled data (Cerqueti et al., 2024; MacKay, 2003). Unlike supervised learning, unsupervised learning does not use labeled input/output pairs but rather identifies natural groupings or patterns within the data (Maranzano et al., 2024; Hastie et al., 2009).

- **Applications:** Common applications of unsupervised learning include clustering, anomaly detection, and dimensionality reduction. These techniques are often used to segment data, discover relationships, and gain insights into the data without prior labels (Cerulli, 2023; Cerqueti et al., 2024; Hastie et al., 2009).

- **K-Means Clustering**
  - Partitions data into k clusters by minimizing the variance within each cluster

$$S = \sum_{i=1}^{k} \sum_{x \in C_i} (x - \mu_i)^2$$

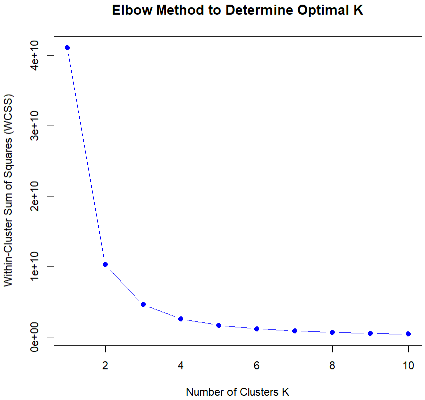
  - Iteratively assigns data points to the nearest cluster centroid and updates centroids until convergence.
- **Hierarchical Clustering**
  - Builds a hierarchy of clusters using either agglomerative (bottom-up) or divisive (top-down) approaches.
  - Forms a dendrogram to visualize nested clusters at different levels.
  - Uses various linkage methods, such as single, complete, or average linkage (for distance d between clusters A and B):

$$d(A, B) = \min_{a \in A, b \in B} d(a, b) \quad \text{(Single linkage)}$$

$$d(A, B) = \max_{a \in A, b \in B} d(a, b) \quad \text{(Complete linkage)}$$

$$d(A, B) = \frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a, b) \quad \text{(Average linkage)}$$

Elbow Method for Optimal Number of Clusters (K)



The **Elbow Method** is used to determine the optimal number of clusters K by identifying the "elbow point" where the Within-Cluster Sum of Squares (WCSS) significantly decreases and starts to level off.

- **X-axis**: Number of Clusters K
- **Y-axis**: Within-Cluster Sum of Squares (WCSS)
- The "elbow" appears around $K = 3$ or $K = 4$, suggesting these as potential values for the optimal number of clusters.

Figure: K-Means Clustering $K = 3$

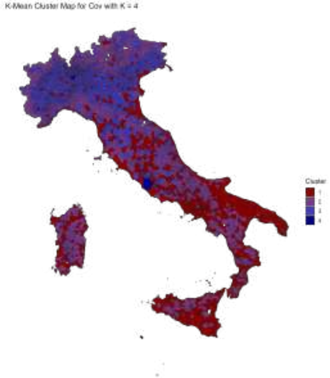


Figure: K-Means Clustering $K = 4$

Figure: K-Means Clustering $K = 3$

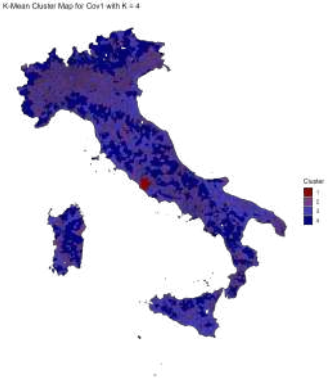


Figure: K-Means Clustering $K = 4$

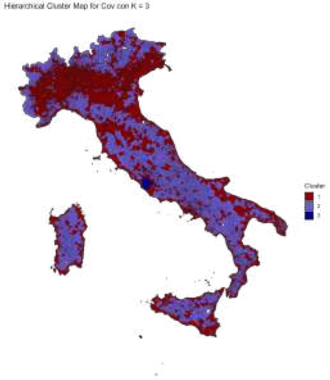


Figure: Hierarchical Clustering with Total Components

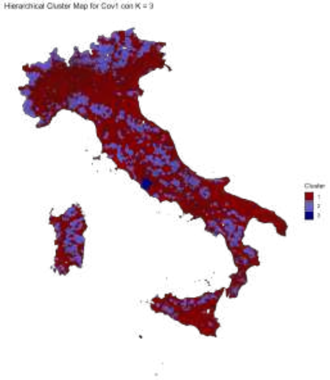


Figure: Hierarchical Clustering with Accessibility Component Only

The clustering analysis reveals distinct patterns based on the total feature set and accessibility component alone. Key findings include:

- **Regional Differences:** There are clear differences in clustering among Inner areas, Centers, and Cities. These areas tend to form separate clusters, indicating varying levels of access to services and socio-economic characteristics across regions.
- **Influence of Accessibility:** Accessibility plays a critical role in defining the clusters, especially when analyzed as a standalone component. The patterns suggest that municipalities with better accessibility tend to cluster together, while more remote or less accessible areas form distinct groups.
- **Impact of Spatial Dimension:** The spatial arrangement of municipalities significantly influences clustering outcomes. Spatial proximity and geographic factors contribute to the grouping, underscoring the importance of the spatial dimension in both total and accessibility-focused analyses.
- **Policy Implications:** The observed clustering of internal areas versus urban centers can guide targeted policy interventions. Improving accessibility and infrastructure in remote areas could help bridge the gap between urban and rural municipalities, reducing socio-economic disparities.

- **Integration of Unsupervised Machine Learning Algorithms**
  - Leverage new algorithms to identify hidden patterns and relationships in accessibility data.
- **Identification of Critical Areas in Italy**
  - Focus on specific regions facing significant accessibility challenges to target priority interventions.
- **Policy Adjustments for Sustainable Development**
  - Propose targeted policy reforms to ensure equitable service access and promote sustainable development.

Thank you for your attention