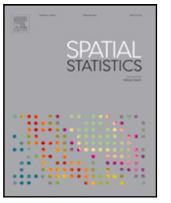
Spatial Statistics 68 (2025) 100911

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta



Attribute based spatial segmentation for optimising POI placement

M. de Klerk *, I. Fabris-Rotelli

University of Pretoria, Department of Statistics, Faculty of Natural and Agricultural Sciences, Pretoria, Gauteng, South Africa

ARTICLE INFO

Keywords:
Attribute augmented graph
Spatially disjoint
Catchment areas
Spatial segmentation

ABSTRACT

Effective spatial planning and resource optimisation require precise demarcation of potential spatial accessible areas and optimal placement of points of interest (POIs). Our approach introduces a novel attribute based spatial segmentation methodology that utilises an iterative clustering approach to create unique macro-regions, each associated with key structural and attribute specific properties. By integrating a probabilistic attribute based structure with k-means clustering, we adaptively segment spatial regions to balance area based attributes and topological characteristics. The full geographical network is segmented into attribute based macro-regions for all spatially accessible and spatially disjoint regions. Attribute based spatial segmentation offers insights into why certain areas may be spatially disjoint and if it is identified as potential spatially accessible areas to determine which POIs can be placed to maximise accessibility. This approach transforms city planning and resource allocation by aligning POI placement with regional needs and characteristics.

1. Introduction

When considering spatial planning and resource optimisation, determining accurate service catchment areas and POI placement is crucial for city planning. As discussed in Macharia et al. (2021), ideal catchment areas should balance both supply and demand, account for geography and physical barriers to access, and differentiate by service type and capacity. Additionally, they should be adaptable to seasonal changes, ensuring service delivery aligns with local needs and infrastructure limitations, particularly in low-resource settings.

In de Klerk and Fabris-Rotelli (2024), a fuzzy lattice data structure was introduced to identify catchment areas based on spatial proximity and probabilistic membership. This methodology was expanded to consider not only spatial proximity but also key attributes, such as topography and environmental factors, to create seasonal catchment areas (de Klerk and Fabris-Rotelli, 2025). Both approaches primarily focused on spatially accessible regions by assessing the accessibility of surrounding areas from existing POIs. In contrast, this work reverses the perspective by utilising the transition probabilities from an attribute augmented graph structure (de Klerk and Fabris-Rotelli, 2025) to identify where new POIs should ideally be placed by creating attribute based spatial segments which will assist in addressing gaps in service coverage and planning for future needs.

Attribute based spatial segmentation creates macro-regions, each characterised by a unique structural layout and attribute properties, including social factors, topographical features or environmental characteristics. For regions that are spatially disjoint, where communities lack access to a specific POI among a set of POIs providing a particular product or service (Wang, 2014), these attribute segments offer a framework for identifying potential spatial areas. Effective clustering methodologies are key to defining these macro-regions, which are associated with distinct attribute segments. Traditional clustering techniques often rely on either shared edges or attribute based properties but rarely integrate both in a unified geographical framework.

* Corresponding author.
E-mail address: dklerkm@gmail.com (M. de Klerk).

https://doi.org/10.1016/j.spasta.2025.100911
Received 6 December 2024; Received in revised form 15 April 2025; Accepted 6 June 2025
Available online 21 June 2025

In this work, clustering plays a vital role in grouping regions based on shared characteristics or proximity, enabling planners to delineate coherent regions for efficient service distribution. We leverage an iterative clustering process applied to a transition probability matrix derived from an attribute augmented graph to create attribute based spatial segments. These segments reflect both structural connectivity and attribute similarity, ensuring the formation of spatially coherent macro-regions. When incorporated into a spatial autoregressive (SAR) model, attribute based spatial segmentation can model supply, identify regions with the greatest need or potential benefit from new POIs and transform potential spatial regions into realised spatial areas. This comprehensive methodology bridges the gap between spatial accessibility needs and facility planning, aligning potential POI locations with regional attributes and improving resource distribution.

The remainder of this paper is structured as follows. First, we review the historical developments in graph clustering and its application to spatial planning. This is followed by a detailed discussion of the transition probability matrix derived from an attribute augmented graph, which forms the foundation for initial clustering. We then describe the iterative clustering process used to optimise and produce spatially coherent macro-regions. Following this, the attribute based spatial segmentation methodology is applied to a real-world case study, predicting average POI sales and identifying ideal POI placements. Finally, we examine regions that remain spatially disjoint and discuss the implications of this approach for improving resource allocation and accessibility.

2. Literature review

Clustering on networks has been extensively explored as a means to reveal underlying structures, identify communities, and optimise connectivity within complex systems. Approaches such as modularity maximisation (De Meo et al., 2011), spectral clustering (Ng et al., 2001), and graph partitioning algorithms (Pothen, 1997) have been instrumental in detecting cohesive subgroups within various networks.

Spectral clustering (Ng et al., 2001) has gained significant attention for its ability to partition graphs by leveraging eigenvectors of the graph Laplacian. It identifies clusters based on the similarity between data points in a transformed spectral space. However, spectral clustering often relies on the assumption of linear separability in the spectral space, which can limit its effectiveness in complex or highly non-linear network structures. Additionally, its reliance on global eigenvector calculations introduces computational inefficiencies for large-scale networks (Von Luxburg, 2007), making it less suitable for applications with extensive geographic or contextual variability.

SimRank (Jeh and Widom, 2002) introduced a similarity measure for graph nodes based on their structural context, where two objects are considered similar if they are related to similar objects. This approach relies on iterative computation of similarity scores using a graph theoretic model. Graph mincuts (Blum and Chawla, 2001), a graph based semi-supervised learning technique, uses min-cut optimisation to classify data by exploiting both labelled and unlabelled data, partitioning the graph into regions that minimise cut costs while respecting label constraints. Another approach, the k-means clustering enhancement through a feature weight self-adjustment (FWSA) mechanism (Tsai and Chiu, 2008), adaptively assigns weights to features to improve clustering quality by minimising intra-cluster separations and maximising inter-cluster separations. These methods optimise how nodes are grouped by only considering either structural, attribute or feature-based similarities to achieve improved clustering or segmentation outcomes.

A method that balances both structural and attribute based similarities is the SA-Cluster (structural and attribute) algorithm (Zhou et al., 2009). This approach uses a unified distance measure and incorporates a neighbourhood random walk model to generate cohesive clusters with homogeneous attributes and dense connectivity. However, the random walk parameters must be carefully specified, as they directly influence the extent and localisation of connectivity exploration within the network. In geographic contexts, determining appropriate parameter values can be particularly challenging due to the inherent complexity and variability of spatial regions, necessitating careful tuning to ensure meaningful clustering results.

While the methods discussed above primarily focus on network graph structures, they lack direct applicability to geographic contexts. To address this gap, we propose a novel attribute based spatial segmentation methodology that integrates structural and attribute based similarities to cluster geographic nodes into spatially coherent regions. This is done by applying an iterative clustering process directly to the transition probability structure derived from an attribute augmented graph, as detailed in de Klerk and Fabris-Rotelli (2025). By doing so, it eliminates the reliance on computationally intensive spectral transformations, offering a more efficient and adaptable solution. This approach balances structural connectivity and attribute based similarities, enabling the formation of spatially coherent areas that reflect diverse regional characteristics. Furthermore, the methodology is highly scalable, requiring only a single parameter, the number of clusters, allowing for practical implementation without the complexity of extensive parameter tuning.

Our approach is particularly impactful in grouping the full geographical region into attribute based spatial segments, encompassing both spatially accessible and spatially disjoint regions. By identifying areas lacking POIs within an accessible drive-time threshold, our method highlights geographic barriers or limited accessibility, categorising such regions as spatially disjoint. This segmentation not only identifies regions that could benefit from new POIs but also explains why certain areas remain inaccessible. The proposed methodology transforms otherwise inaccessible regions into potential service catchment areas, providing a powerful tool for improving resource allocation, accessibility planning and service delivery across urban and rural contexts.

3. Methodology

In this section, the methodology of attribute based spatial segments will be discussed, beginning with the creation of the transition probability matrix, $\mathbf{P}_a$, derived from an attribute augmented graph structure, as illustrated with a small toy example in Section 3.1. Following this, the initial clustering on $\mathbf{P}_a$ will be explored to obtain preliminary cluster memberships and macro-regions, as described in Section 3.2. Based on the results from Section 3.2, it is evident that the macro-regions in the toy example can be further refined. This leads to the exploration of spatially optimal cluster memberships and attribute based segmentation in Section 3.3.

3.1. Transition probability matrix of an attribute augmented graph

Let $G = G(V, E)$ be a network graph, where V represents the set of nodes and E the edges connecting them Bhagat et al. (2011). In this context, nodes typically correspond to entities such as individuals within a social network, while edges depict relationships like friendships or interactions. In this study, we focus specifically on spatial graphs, where nodes correspond to geographical locations.

The graph's structural connectivity, referred to as its link structure, describes how nodes are interconnected via edges, which can be quantified using transition probabilities. Beyond this structural framework, graphs can be enriched by incorporating attributes, denoted as Λ, that capture additional properties of the nodes V. In an augmented graph G_a, these attributes are represented by attribute nodes, distinct from the structural nodes that embody the graph's spatial layout and link structure. Attribute nodes create connections between structural nodes based on shared characteristics, forming an integrated representation (Zhou et al., 2009).

To incorporate attribute nodes into the node connectivity, we define an attribute augmented graph as $G_a = (V \cup V_a, E \cup E_a, \mathbf{W}_a^*)$, following the definition in Zhou et al. (2009). Here, V represents the structural nodes, V_a denotes the attribute nodes, E refers to the structural edges, E_a indicates the attribute edges and $\mathbf{W}_a^*$ the structural and attribute weights with $\mathbf{W}_a^* = \begin{pmatrix} \mathbf{W}_s & 0 \\ 0 & \mathbf{W}_a \end{pmatrix}$ where $\mathbf{W}_s$ is an $N \times N$ matrix of structural edge weights and $\mathbf{W}_a$ is an $T \times T$ diagonal matrix for the attribute edge weights.

Consider the structural nodes $V = \{v_1, \dots, v_N\}$, which represents the full geographical region that is subdivided into N grid cells, with the centres of these cells serving as the node representative locations. A structured edge $(v_i, v_j) \in E$ exists if and only if node $v_i \in V$ shares a boundary point with node $v_j \in V$ for $i \neq j$ in the geographical area, as defined by queen's contiguity (Wang, 2014). Each node in V is associated with a set of attributes $\Lambda = \{\lambda_1, \dots, \lambda_A\}$. If each attribute $\lambda_i \in \Lambda$ has z_i levels, where $i \in \{1, \dots, A\}$, then let $A = \{a_1, \dots, a_T\}$ represent the T-dimensional set of indicator attributes associated with each level of $\lambda_i \in \Lambda$, where $T = \sum_{i=1}^{A} z_i$. The indicator attribute information for each structural node v_i is represented as a T-dimensional binary vector $\mathbf{b}_i$ such that attribute matrix $\mathbf{B}$ is structured as

$$\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_N \end{pmatrix} = \begin{pmatrix} a_1(v_1) & a_2(v_1) & \dots & a_T(v_1) \\ a_1(v_2) & a_2(v_2) & \dots & a_T(v_2) \\ \vdots & \vdots & \ddots & \vdots \\ a_1(v_N) & a_2(v_N) & \dots & a_T(v_N) \end{pmatrix} \tag{1}$$

where $a_j(v_i) = 1$ signifies the presence of an attribute edge in E_a if node v_i is associated with indicator attribute a_j otherwise, $a_j(v_i) = 0$ (Lin et al., 2021). Let $V_a = \{v_{a_1}, \dots, v_{a_T}\}$ be the set of indicator attribute nodes resulting in the combined node set $V \cup V_a$ to be of size $N + T$.

In an attribute augmented graph structure, transitions between nodes can occur either through structural edges (connecting structural nodes) or attribute edges (linking nodes to their associated attributes). The transition probability matrix $\mathbf{P}_a$ for G_a is an $(N + T) \times (N + T)$ matrix

$$\mathbf{P}_a = \begin{pmatrix} \mathbf{P}_s & \mathbf{P}_{sa} \\ \mathbf{P}_{as} & \mathbf{0} \end{pmatrix} \tag{2}$$

where $\mathbf{P}_s$ is an $N \times N$ submatrix representing transition probabilities between structural nodes, $\mathbf{P}_{sa}$ an $N \times T$ submatrix representing the transition probabilities from structural to attribute nodes, $\mathbf{P}_{as}$ is a $T \times N$ submatrix for transitions from attribute to structural nodes and the transition probability matrix between two attributes nodes is captured in the $T \times T$, $\mathbf{0}$ submatrix since there is no shared edges between two attribute nodes (de Klerk and Fabris-Rotelli, 2025).

Consider Fig. 1, which illustrates a simple example with structural nodes $V = \{v_1, \dots, v_{20}\}$. Each structural node corresponds to a geographical grid, with attributes captured in Λ. In this case, the structural nodes are associated with two attributes, λ_1 and λ_2, comprising two levels each. Attribute λ_1 consists of levels "Group1" and "Group2" and attribute λ_2 of levels "Group3" and "Group4" such that $A = 2$, $z_1 = 2$, $z_2 = 2$, $N = 20$ and $T = 4$. Let $a_1(v_i) = 1$ signify that structural node v_i is associated to "Group1" and $a_1(v_i) = 0$ otherwise, similarly for "Group2" to a_2, "Group3" to a_3 and "Group4" to a_4. For attribute λ_1, nodes $v_4, v_5, v_6, v_{10}, v_{12}, v_{16}$, and v_{17} are associated with level "Group2", while the remaining nodes are linked to level "Group1". When considering attribute λ_2, nodes $v_6, v_7, v_{10}, v_{12}, v_{13}$, and v_{15} are associated with level "Group4", while the remaining nodes are linked to level "Group3". From Eq. (1) the attribute matrix $\mathbf{B}$ for the simple example illustrated in Fig. 1 is provided by

$$\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \\ \mathbf{b}_4 \\ \vdots \\ \mathbf{b}_{20} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

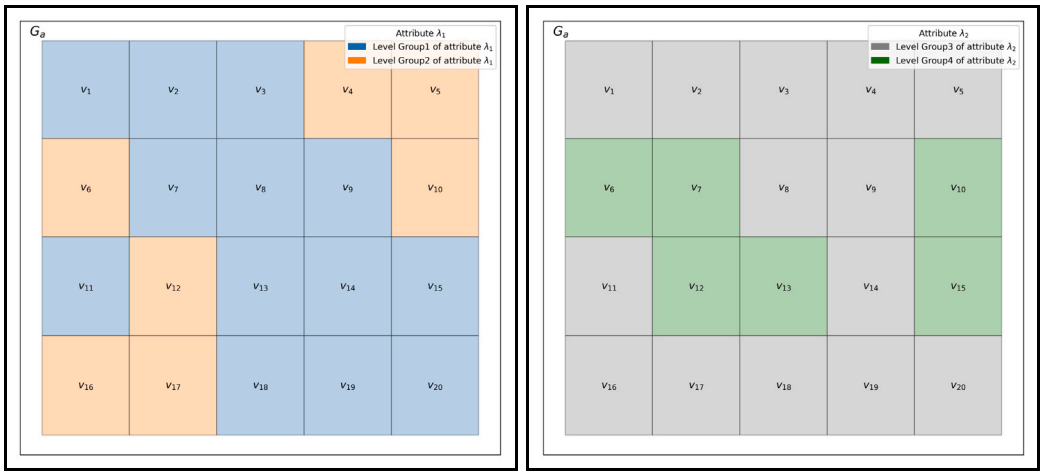


Fig. 1. Graph G_a with structural nodes $v_1, \ldots, v_{20}$, associated with attributes λ_1 and λ_2.

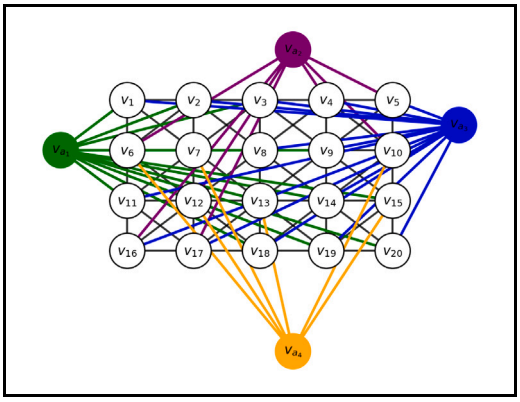


Fig. 2. The network structure of G_a: The structural nodes $(v_1, \ldots, v_{20})$ are depicted as black-outlined circles, with structural edges represented by black lines. Indicator attribute nodes $(v_{a_1}, v_{a_2}, v_{a_3}, v_{a_4})$ are shown as circles in green, purple, blue, and yellow, respectively, with their corresponding attribute edges displayed in the same colours as the nodes.

The set V_a contains indicator nodes $\{v_{a_1}, v_{a_2}, v_{a_3}, v_{a_4}\}$ reflecting associations with "Group1", "Group2", "Group3" and "Group4" respectively. Assume all structural and attribute weights in $\mathbf{W}_a^*$ are set to 1. Consequently, $\mathbf{W}_s$ is simplified to the adjacency matrix that defines structural neighbouring nodes based on shared boundary points between grids. The network structure of G_a is illustrated in Fig. 2.

The transition probability matrix $\mathbf{P}_a$ for G_a is a 24×24 matrix, as illustrated in Eq. (3)

$$\mathbf{P}_a = \begin{pmatrix} \mathbf{p}_{a,1} \\ \mathbf{p}_{a,2} \\ \vdots \\ \mathbf{p}_{a,N+T} \end{pmatrix} = \left(\begin{array}{cccc|cccc} 0 & \frac{1}{5} & \cdots & 0 & \frac{1}{5} & 0 & \frac{1}{5} & 0 \\ \frac{1}{7} & 0 & \cdots & 0 & \frac{1}{7} & 0 & \frac{1}{7} & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \frac{1}{5} & 0 & \frac{1}{5} & 0 \\ \hline \frac{1}{13} & \frac{1}{13} & \cdots & \frac{1}{13} & 0 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{14} & \frac{1}{14} & \cdots & \frac{1}{14} & 0 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \end{array}\right) \tag{3}$$

where the embedded representation of every node i is captured in the $N+T$-dimensional feature vector $\mathbf{p}_{a,i}, i = 1, \ldots, N+T$.

3.2. Cluster memberships and macro-regions of $\mathbf{P}_a$

Clustering is a fundamental technique in data analysis aimed at grouping objects into homogeneous categories based on their similarity. Clustering uncovers the inherent structures of high-dimensional data, such as the transition probabilities captured in $\mathbf{P}_a$, without relying on predefined hypotheses or models (Wierzchoń and Kłopotek, 2017).

Cluster analysis will be applied on the feature vectors $\mathbf{p}_{a,i}, i = 1, \dots, N+T$ in $\mathbf{P}_a$ to identify natural patterns in the nodes based on shared attribute and structural connections. The feature vectors, $\mathbf{p}_{a,i}, i = 1, \dots, N+T$, will be split into $K < N+T$ clusters $C = (C_1, \dots, C_K)$, where every cluster needs to contain at least one feature vector and each feature vector will belong to only one cluster.

Initially, clusters are formed using the k-means (Steinley, 2006) algorithm by partitioning $\mathbf{p}_{a,i}, i = 1, \dots, N+T$ into K clusters. The optimal number of clusters, K, is determined using the elbow method (Jain, 2010), where the point at which the rate of decrease in inertia slows significantly is identified, indicating diminishing returns from additional clusters. Initial cluster centres, $\boldsymbol{\mu}_j, j = 1, \dots, K$ are randomly selected from $\mathbf{P}_a = (\mathbf{p}_{a,1}, \dots, \mathbf{p}_{a,N+T})^T$. The remaining feature vectors are assigned to the respective clusters based on which has the closest cluster centre by minimising the distance between each feature vector $\mathbf{p}_{a,i}$ and its assigned cluster centroid $\boldsymbol{\mu}_j$ i.e. $\sum_{i=1}^{N+T}\sum_{j=1}^{K} u_{ij}(\mathbf{p}_{a,i} - \boldsymbol{\mu}_j)(\mathbf{p}_{a,i} - \boldsymbol{\mu}_j)^T$. $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_{N+T})^T = [u_{ij}]$ is the $(N+T) \times K$ matrix indicating the assignment of the ith feature vector, $\mathbf{p}_{a,i}$, to the jth class i.e. $u_{ij} = 1$ if feature vector i is assigned to cluster j for $i = 1, \dots, N+T; j = 1, \dots, K$ and $u_{ij} = 0$ otherwise. After the remaining feature vectors are assigned to one of the K clusters $C = (C_1, \dots, C_K)$, new cluster centres are calculated by $\boldsymbol{\mu}_j = \frac{1}{n_j}\sum_{i=1}^{N+T} u_{ij}\mathbf{p}_{a,i}$ where n_j is the number of feature vectors assigned to cluster $j, j = 1, \dots, K$. This process continues until there is no more changes in the assignment of feature vectors to clusters. This process provides the initial cluster assignments, $\mathbf{U}(0)$ with $u_{ij} \in \{0, 1\}$ a crisp partition (Wierzchoń and Kłopotek, 2017) as outlined in algorithm 1.

Algorithm 1 Initial k-means clustering on P_a

Require: $\mathbf{P}_a = (\mathbf{p}_{a,1}, ..., \mathbf{p}_{a,N+T})^T$ - dataset, K - number of clusters
Ensure: $C = \{C_1, C_2, \dots, C_K\}$ - set of clusters, initial cluster assignment $\mathbf{U}(0)$
Select K initial centroids $\boldsymbol{\mu}_j, j = 1, ..., K$ randomly from $\mathbf{P}_a$
repeat
  Create K empty clusters $C_1, C_2, \dots, C_k$
  for $i = 1$ to $N+T$ **do**
    Assign $\mathbf{p}_{a,i}$ to the cluster C_j with the closest centroid $\boldsymbol{\mu}_j$
    for $j = 1$ to K **do**
      Update centroid $\boldsymbol{\mu}_j$ as the mean of all points in cluster C_j
    end for
  end for
until no more changes in the assignment of points to clusters, $\mathbf{U}$

Applying this initial clustering process on the simple example illustrated in Fig. 1 the optimal number of clusters obtained using the elbow method was $K = 5$ clusters. The final cluster assignments $C_1 = \{\mathbf{p}_{a,4}, \mathbf{p}_{a,5}\}$, $C_2 = \{\mathbf{p}_{a,9}, \mathbf{p}_{a,14}, \mathbf{p}_{a,15}, \mathbf{p}_{a,18}, \mathbf{p}_{a,19}, \mathbf{p}_{a,20}\}$, $C_3 = \{\mathbf{p}_{a,16}, \mathbf{p}_{a,17}\}$, $C_4 = \{\mathbf{p}_{a,6}, \mathbf{p}_{a,7}, \mathbf{p}_{a,8}, \mathbf{p}_{a,10}, \mathbf{p}_{a,12}, \mathbf{p}_{a,13}, \mathbf{p}_{a,21}, \mathbf{p}_{a,22}, \mathbf{p}_{a,23}, \mathbf{p}_{a,24}\}$ and $C_5 = \{\mathbf{p}_{a,1}, \mathbf{p}_{a,2}, \mathbf{p}_{a,3}, \mathbf{p}_{a,11}\}$ is indicated in $\mathbf{U}(0)$ in Eq. (4) and a graphical representation is provided in the first image of Fig. 3.

$$\mathbf{U}(0) = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{4}$$

From the cluster assignments illustrated in Fig. 3, it can be observed that there are structural nodes $v_i \in V$ which are adjacent to each other and belong to the same cluster. Since this clustering is applied to spatial data, the goal is to create spatially coherent regions rather than clusters of aspatial data. To achieve this, a new spatial entity, termed a macro-region, is constructed.

Macro-regions are formed by merging structural nodes $v_i \in V$ that share a structural edge $(v_i, v_j) \in E$ and whose feature vectors $\mathbf{p}_{a,i}$ and $\mathbf{p}_{a,j}$ belong to the same cluster C_k. Specifically, two nodes v_i and v_j are grouped into the same macro-region if they are connected by a structural edge and their feature vectors satisfy the condition $u_{ik} = u_{jk} = 1$, where $u_{ik} = 1$ indicates that the feature vector $\mathbf{p}_{a,i}$ is assigned to cluster C_k. The attribute nodes from V_a are not included when creating the macro-regions, as they have no geographical properties.

Macro-regions, denoted as $\mathbf{M} = \{\mathcal{M}_1, \dots, \mathcal{M}_Q\}$, are thus formed by the union of all structural nodes that meet this criteria. Formally, a macro-region $\mathcal{M}_q$, where $q = 1, \dots, Q$, is defined as the set of all connected structural nodes that belong to the same

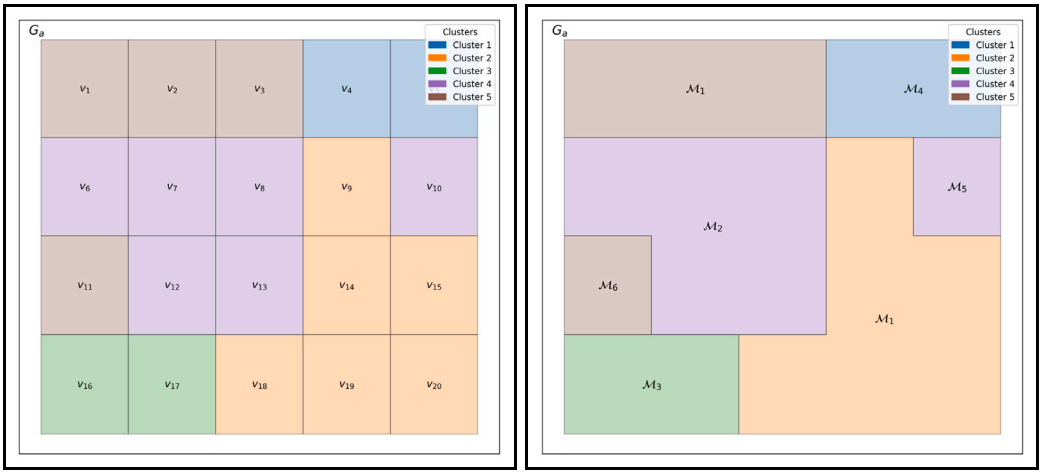


Fig. 3. The left image illustrates the initial cluster assignments $\mathbf{U}(0)$ on $\mathbf{P}_a$ with $K = 5$, while the right image shows macro-regions based on $\mathbf{P}_a$ with $Q = 7$.

cluster, expressed as

$$\mathcal{M}_q = \bigcup_{\substack{(v_i,v_j)\in E \\ k\in\{1,\dots,K\}, u_{ik}=u_{jk}=1}} \{v_i, v_j\}. \tag{5}$$

Each macro-region is uniquely associated with one of the clusters $C_k \in C$, but a single cluster can produce multiple macro-regions if its constituent nodes form disconnected subgraphs within the network G_a. Consequently, the total number of macro-regions Q satisfies $K \leq Q$, where K is the number of clusters. This framework ensures that macro-regions are defined both by spatial connectivity and clustering assignments, providing a cohesive representation of spatially coherent regions.

Using Eq. (5) to construct macro-regions for the simple example shown in Fig. 1, $Q = 7$ macro-regions where identified as shown in the second image of Fig. 3.

Upon close inspection of the macro-regions created in Fig. 3, it becomes evident that there are many smaller-sized macro-regions, such as $\mathcal{M}_5$ and $\mathcal{M}_6$. This is a result of the k-means clustering of $\mathbf{P}_a$, which primarily focuses on minimising the distance between each feature vector $\mathbf{p}_{a,i}$ and its assigned cluster centroid $\boldsymbol{\mu}_j$. However, this approach does not fully account for spatial dependencies and neighbouring nodes.

Although the feature vectors for the attribute nodes $\mathbf{p}_{a,21}, \mathbf{p}_{a,22}, \mathbf{p}_{a,23}, \mathbf{p}_{a,24}$ are not directly part of the macro-regions, they play a crucial role in clustering the nodes by incorporating both structural and attribute links. Notably, $\mathbf{p}_{a,21}, \mathbf{p}_{a,22}, \mathbf{p}_{a,23}, \mathbf{p}_{a,24}$ were all assigned to cluster 4. By examining the attributes in Fig. 1 and comparing them to the macro-regions in Fig. 3, it can be observed that regions $\mathcal{M}_2$ and $\mathcal{M}_5$ are not associated with all levels of attributes λ_1 and λ_2. This discrepancy suggests that while $\mathbf{p}_{a,21}, \mathbf{p}_{a,22}, \mathbf{p}_{a,23}, \mathbf{p}_{a,24}$ and macro-regions $\mathcal{M}_2$ and $\mathcal{M}_5$ are all assigned to cluster C_4, the initial clustering does not fully account for all attribute and structural links to achieve spatially optimal solutions. In the following section, an iterative process will be explored to spatially smooth the initial clusters obtained in $\mathbf{U}(0)$ and generate more spatially coherent macro-regions.

3.3. Optimised cluster memberships and macro-regions

The macro-regions generated through the clustering of $\mathbf{P}_a$ provide an insightful yet preliminary segmentation of the graph. While the k-means approach effectively partitions the data based on feature similarity, its focus on minimising intra-cluster distances, neglects the spatial relationships between neighbouring nodes. As a result, the initial macro-regions may lack spatial coherence, as seen with smaller macro-regions, which emerge due to localised feature similarities rather than broader spatial patterns. To address this limitation and improve spatial contiguity, further refinement of the cluster memberships is required. This refinement leverages the inherent network structure and spatial dependencies embedded in $\mathbf{P}_a$. The next section explores a method to iteratively enhance the spatial cohesion of these macro-regions while maintaining their distinctiveness.

To identify more spatially coherent macro-regions, we need to re-cluster the memberships obtained in $\mathbf{U}(0)$ by incorporating the network structure represented in $\mathbf{P}_a$. Let $\mathbf{S}(t) = \mathbf{P}_a \cdot \mathbf{U}(t-1)$ be an $(N+T) \times K$ matrix of the probability distribution of the cluster memberships based on the connectivity structure in $\mathbf{P}_a$. Each K-dimensional feature row vector in $\mathbf{S}(t)$, $\mathbf{S}(t) = (\mathbf{s}_1, \dots, \mathbf{s}_{N+T})^T$, provides weighted averages of the nodes' affiliations to their respective clusters. $\mathbf{S}(t) = [s_{ij}(t)]$ is the $(N+T) \times K$ matrix indicating the assignment of the ith node to the jth cluster membership at iteration t with $s_{ij}(t) \in [0,1]$ which indicates a fuzzy or probabilistic membership as opposed to $u_{ij}(t) \in \{0,1\}$ which indicates a crisp membership of 0 or 1. If the nodes in $\mathbf{P}_a$ are more tightly grouped around one cluster centroid, higher probabilities will be observed for that cluster in the corresponding rows of $\mathbf{S}(t)$.

For $t = 1$, the cluster membership from the initial cluster assignment $\mathbf{U}(0)$ is $\mathbf{S}(1) = \mathbf{P}_a \cdot \mathbf{U}(0)$. When applied to the example illustrated in Fig. 1, the cluster memberships of $\mathbf{U}(0)$ are

$$\mathbf{S}(1) = \begin{pmatrix} 0 & 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 0.714 & 0.286 \\ 0.143 & 0.143 & 0 & 0.57 & 0.143 \\ 0.143 & 0.143 & 0 & 0.57 & 0.143 \\ 0.2 & 0.2 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 0.571 & 0.429 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.143 & 0.357 & 0.143 & 0.071 & 0.286 \\ 0 & 0.167 & 0 & 0.833 & 0 \end{pmatrix}. \tag{6}$$

From the results of $\mathbf{S}(1)$, it can be observed that nodes strongly associated with a particular cluster (rows dominated by a single value close to 1) are more tightly grouped around one cluster centroid. For example, nodes v_1 to v_6 exhibit a high affinity to cluster 4, as reflected in the fourth column of $\mathbf{S}(1)$. This is because most nodes are initially assigned to cluster 4, which dominates the probability structure in $\mathbf{S}(1)$.

If we re-cluster $\mathbf{S}(1)$ using the k-means algorithm described in Algorithm 1, but with $\mathbf{S}(t) = (\mathbf{s}_1, \dots, \mathbf{s}_{N+T})^T$ as the input data and initial centroids $\boldsymbol{\mu}_j, j = 1, \dots, K$ chosen randomly from $\mathbf{S}(t)$, the cluster memberships change to

$$\mathbf{U}(1) = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{7}$$

and the resulting cluster membership probability distribution matrix $\mathbf{S}(2) = \mathbf{P}_a \cdot \mathbf{U}(1)$ becomes

$$\mathbf{S}(2) = \begin{pmatrix} 0.4 & 0 & 0 & 0.6 & 0 \\ 0.286 & 0 & 0.286 & 0.429 & 0 \\ 0.286 & 0 & 0.429 & 0.286 & 0 \\ 0.143 & 0.143 & 0.571 & 0 & 0.143 \\ 0.2 & 0.2 & 0.4 & 0 & 0.2 \\ 0 & 0 & 0.143 & 0.429 & 0.429 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0.214 & 0.5 & 0.143 & 0.143 \\ 0 & 0.5 & 0 & 0.333 & 0.167 \end{pmatrix}. \tag{8}$$

As the iteration progresses, the memberships in $\mathbf{S}(t)$ begin to balance between clusters, reflecting both structural and attribute based dependencies as can be seen in $\mathbf{S}(2)$ where the cluster memberships for $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_6$ are now different from $\mathbf{s}_4, \mathbf{s}_5$ as when compared to $\mathbf{S}(1)$.

Continuing the process, the cluster memberships between $\mathbf{U}(t)$ and $\mathbf{U}(t+1)$ will gradually stabilise. This stabilisation reflects the influence of the connectivity structure in $\mathbf{P}_a$. The final clusters account for transition probabilities between nodes and smooth cluster memberships based on the underlying graph structure.

The iterative process continues until convergence is achieved in terms of mutual information (Tishby and Slonim, 2000) between $\mathbf{U}(t)$ and $\mathbf{U}(t+1)$ given by

$$I(\mathbf{U}(t), \mathbf{U}(t+1)) = \sum_{j=1}^{K} \pi_{s_j}(t) \sum_{i=1}^{N+T} u_{ij}(t+1) \log \frac{u_{ij}(t+1)}{\pi_{u_i}(t+1)}. \tag{9}$$

where $\pi_{s_j}(t) = \frac{1}{N+T} \sum_{i=1}^{N+T} s_{ij}(t), j = 1, 2, \dots, K$ is the prior probabilities of the cluster memberships at iteration t and $\pi_{u_i}(t+1) = \frac{1}{N+T} \sum_{j=1}^{N+T} \delta\left(\mathbf{u}_i(t+1), \mathbf{u}_j(t+1)\right)$ is the average cluster membership for each node i at iteration $t+1$ with $\delta\left(\mathbf{u}_i(t+1), \mathbf{u}_j(t+1)\right) = 1$ if $\mathbf{u}_i(t+1) = \mathbf{u}_j(t+1)$ and 0 otherwise.

When the iterative process outlined in Eq. (9) converges, the optimal crisp cluster memberships will be captured in $\hat{\mathbf{U}}$ with corresponding fuzzy membership in $\hat{\mathbf{S}}$.

By computing mutual information during each iteration, we ensure that the current clustering structure captures as much information as possible about the distribution of the input data, gradually approaching an optimal clustering solution.

Since the probabilities are split as fractions for structural and attribute nodes in $\mathbf{P}_a$, the likelihood of observing the same pattern of probabilities per node is high. This can result in two or more optimal solutions, which may cause the iterative process in Eq. (9) to oscillate. If this happens, priority will be given to the solution where more nodes shares a structure edge as the primary

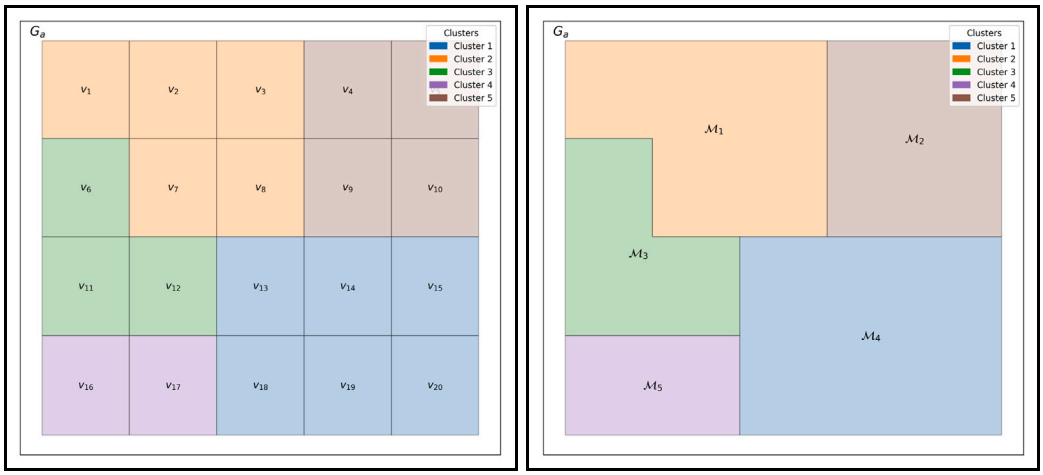


Fig. 4. The left image illustrates the optimised cluster assignments $\hat{\mathbf{U}}$ on $\mathbf{P}_a$, while the right image shows macro-regions based on structural node cluster assignments on $\hat{\mathbf{U}}$.

objective of the clustering is to form mutual spatial regions. To formalise this, let $\sum_{i=1}^{N+T}\sum_{j=1}^{N+T}\delta\left(\mathbf{u}_i,\mathbf{u}_j\right)\mathbf{A}_{ij}$ be a spatial coherence score, where $\mathbf{A}$ is the adjacency matrix of all nodes in the set $V \cup V_a$ and $\delta\left(\mathbf{u}_i,\mathbf{u}_j\right)$ indicates whether the cluster assignment for node i and node j are the same with $\delta\left(\mathbf{u}_i,\mathbf{u}_j\right) = 1$ if $\mathbf{u}_i = \mathbf{u}_j$ and 0 otherwise. This score is calculated for all solutions obtained through Eq. (9), if more than one optimal solution was identified, and the cluster membership with higher structural spatial coherence will be assigned as $\hat{\mathbf{U}}$.

Once the optimal solution $\hat{\mathbf{U}}$ along with $\hat{\mathbf{S}}$ is obtained, the macro-regions can be determined as formulated in Eq. (5). Revisiting the example illustrated in Fig. 1, the newly defined macro-regions based on $\hat{\mathbf{U}}$ are shown in Fig. 4.

Continuing with this example, the stabilised value $\hat{\mathbf{S}}$ in Eq. (10) provides cluster assignments coherent with the connectivity structure in $\mathbf{P}_a$,

$$\hat{\mathbf{S}} = \begin{pmatrix} 0.2 & 0.4 & 0.4 & 0 & 0 \\ 0.143 & 0.571 & 0.286 & 0 & 0 \\ 0.143 & 0.429 & 0.143 & 0 & 0.286 \\ 0 & 0.286 & 0.143 & 0.143 & 0.429 \\ 0 & 0 & 0.2 & 0.2 & 0.6 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.286 & 0.286 & 0.071 & 0.143 & 0.214 \\ 0.333 & 0.167 & 0.333 & 0 & 0.167 \end{pmatrix}. \tag{10}$$

When comparing the macro-regions in Figs. 3 and 4, it can be noted that the number of macro-regions has decreased from $Q = 7$ to $Q = 5$ while maintaining the same number of clusters $K = 5$. The regions are also larger in size. Although the number of macro-regions has been reduced, there is still a combination of key attributes that are repeated across different clusters.

To further understand and describe the spatial configuration, attribute segmentation can be introduced as a higher-level categorisation of the macro-regions based on shared prominent attribute combinations. Unlike macro-regions, which are defined spatially, attribute segmentations focus on grouping macro-regions with similar modal attribute profiles into categorical values. For example, if $\mathcal{M}_1$ and $\mathcal{M}_4$ share the same dominant attribute combination "Group1" and "Group3", they will belong to the same attribute segmentation. Similarly, $\mathcal{M}_2$ and $\mathcal{M}_5$ may share a different attribute combination "Group2" and "Group3", forming a separate segmentation. This results in fewer attribute segmentations than macro-regions, as multiple macro-regions can map to a single attribute segmentation.

Formally, let Q denote the number of macro-regions, R the number of attribute segmentations, and K the number of clusters. Then, $R \leq Q$, where each macro-region is associated with exactly one attribute segmentation, but multiple macro-regions can map to the same attribute segmentation.

Since the optimal clusters are based on structural links between nodes and shared attributes, the final clusters can share similar attributes but can be located in different geographical regions. Hence, attribute segmentation provides a meaningful, higher-level description for each macro-region.

Clusters with identical modal attributes across all attribute levels are grouped into the same attribute segmentation to facilitate visualisation and to serve as covariates in a model. Attribute segmentation applied to the example in Fig. 4 provides the results shown in Fig. 5.

The macro-regions that share the most prominent attributes (or levels of an attribute) are grouped together in the same attribute based spatial segmentation. For example, macro-regions $\mathcal{M}_1$ and $\mathcal{M}_4$ both share the attribute combination "Group1" and "Group3", $\mathcal{M}_2$ and $\mathcal{M}_5$ both share "Group2" and "Group3" and $\mathcal{M}_3$ is characterised by "Group2" and "Group4".

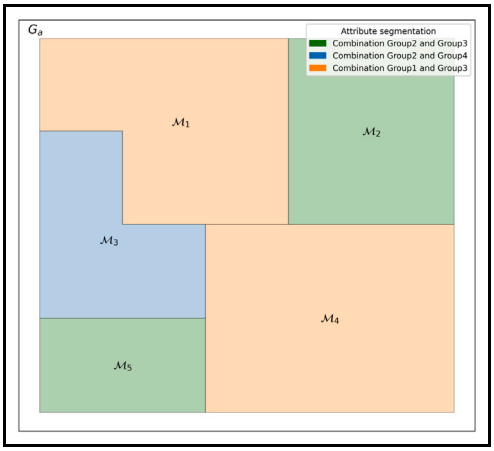


Fig. 5. Attribute segmentation of clusters.

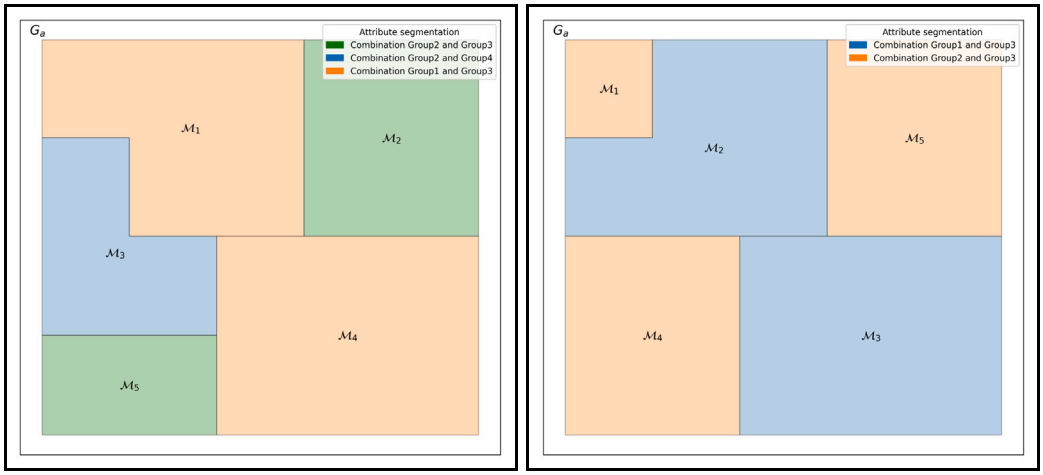


Fig. 6. The left image illustrates attribute segments with two levels in λ_1, namely "Group1" and "Group2" and two levels in λ_2, namely "Group3" and "Group4". The image on the right illustrates attribute segments with two levels in λ_1, namely "Group1" and "Group2" and only one level in λ_2, namely "Group3".

3.4. Adaptability based on attributes

The adaptive capability of the proposed method is embedded in the iterative refinement process applied after the initial k-means clustering on the attribute augmented transition probability matrix, $\mathbf{P}_a$. Initially, macro-regions are formed based on spatial and attribute similarities, but these can be fragmented or lack spatial coherence. To improve this, the method uses a dynamic re-clustering mechanism that updates cluster memberships at each iteration based on the connectivity structure encoded in $\mathbf{P}_a$, thereby refining regions by incorporating both attribute level and topological dependencies.

As the process progresses, grid cells may shift between clusters, merging or separating from neighbouring cells based on their structural connections and shared attributes. This iterative adjustment continues until convergence is achieved in terms of mutual information, resulting in final cluster memberships $\hat{\mathbf{U}}$ and more spatially coherent macro-regions.

This adaptiveness is verified experimentally in the toy example, where the initial macro-regions (based on $\mathbf{U}(0)$) include several small, fragmented regions. After applying the iterative refinement, the resulting macro-regions are larger and more spatially contiguous, as shown by the reduction in the number of macro-regions (from $Q = 7$ to $Q = 5$) and improved spatial coherence.

To illustrate the adaptability of this model when there are changes in attributes, let the toy example be revised such that attribute λ_2 only has one level, specifically "Group3". In this case, the only input that would need to be adjusted is the optimal number of clusters K, which is determined as $K = 4$ using the elbow method. Re-clustering the transition probabilities with this adjustment will result in new macro-regions, as illustrated in Fig. 6.

In the revised scenario, the iterative process will adaptively regroup the nodes based on shared attribute patterns, even though variability in λ_2 has been reduced to a single level. The new attribute segmentations capture only the distinct combinations across the remaining attributes, specifically "Group1 and Group3" and "Group2 and Group3". In Fig. 6, the image on the left shows the

same attribute segments as those in Fig. 5, where both λ_1 and λ_2 have multiple levels. In contrast, the image on the right closely resembles the pattern in the left panel of Fig. 1, where only λ_1 varies and λ_2 is fixed at "Group3". In this case, λ_2 has no influence on the attribute segments. This comparison demonstrates the model's ability to flexibly adjust to changes in attribute information, generating meaningful macro-regions and attribute segments that reflect the underlying spatial-attribute structure.

3.5. Methodology conclusion

In this work, we developed an enhanced clustering framework that integrates structural and attribute based information to identify spatially coherent macro-regions. Starting from the initial clustering of $\mathbf{P}_a$ using k-means, it was observed that traditional clustering methods primarily focus on minimising intra-cluster distances in the feature space, often neglecting spatial dependencies between neighbouring nodes. This limitation resulted in smaller, disjointed macro-regions and inconsistent spatial coherence. To address this, we introduced an iterative refinement process that incorporates the structural relationships embedded in $\mathbf{P}_a$, ensuring that the final clusters are both spatially contiguous and meaningful in terms of their feature vectors.

The refinement process was guided by a mutual information framework, which evaluates the similarity between cluster membership matrices across consecutive iterations, ensuring convergence to stable and interpretable solutions. By integrating structural links and attribute based information, the framework allows for a unified clustering approach that is capable of capturing complex spatial and attribute level dependencies. This novel methodology bridges the gap between traditional clustering approaches and practical applications, offering a scalable solution for spatial planning and geographic analysis. The integration of structural coherence and attribute alignment lays the foundation for future applications in optimising resource allocation, defining service areas, and segmenting spatial systems with multimodal data.

4. Application

To evaluate the proposed methodology in a practical application we consider region G_a to be the district municipality City of Cape Town[1] Region G_a is subdivided into 2042 square grids of size 1 km $\times$ 1 km whose centres will be the structural nodes.

The POIs in this study are pharmaceutical stores, part of a subset from one of South Africa's largest pharmaceutical chains. The supply data for these POIs consists of the average summer sales (December, January, and February) of an affordable over-the-counter diarrhoea medication, recorded from 2015 to 2019. In South Africa, a low- to middle-income country, access to affordable medication is vital for public health. The over-the-counter medication used in this application is widely available and commonly used to manage diarrhoea, a condition with significant health implications, particularly in regions facing challenges related to water quality and sanitation (Musengimana et al., 2016).

This section will explore the attributes associated with the nodes in G_a, followed by the identification of clusters, macro-regions and attribute segments. The application will conclude by evaluating the performance of a SAR model to predict the average sales per POI when incorporating attribute segmentations into a covariate-dependent spatial weights matrix.

4.1. Attributes

Attributes play a central role in shaping the outcomes of the proposed segmentation framework, as they inform the spatial variation in demand and accessibility to POIs. By incorporating attributes that reflect real-world environmental and socio-economic conditions, the model moves beyond structural connectivity alone, enabling a more nuanced understanding of local service needs. In the context of this study, the attributes must capture both the factors that drive demand and those that influence access to pharmaceutical services.

The following subsections detail the rationale behind attribute selection and the methods used to preprocess and categorise each attribute for use in the segmentation model.

4.1.1. Attribute selection

The selection of attributes in this study was guided by their data availability and practical relevance with pharmaceutical sales of diarrhoea medication in the City of Cape Town. The overall modelling framework aims to capture the spatial variation in demand and accessibility to POIs, and as such, the attributes chosen reflect key environmental and socio-economic drivers influencing medication use. It is important to note that the choice of attributes must be application-specific and context-sensitive. For instance, models supporting the placement of schools may prioritise the distribution of young populations, while those for healthcare infrastructure would consider disease prevalence or travel time to the nearest facility.

In this application, which examines summer sales of affordable diarrhoea medication, the selected attributes are ease of access, demand and temperature. These were chosen based on both their influence on diarrhoea medication demand and the availability of high-resolution data across the entire study area. Demand is a fundamental consideration in determining the optimal placement of POIs, as it reflects the potential need for services within a given area. Including demand ensures that regions with little to no population are deprioritised, while areas with concentrated populations are more likely to be identified as suitable candidates

[1] District municipalities consists of multiple local municipalities. They are administrative divisions which is accountable for providing basic services within the area. Source: Education and Training Unit (ETU),www.etu.org.za (accessed February 15, 2024) located in the Western Cape province in South Africa.

for service provision. This is particularly important in resource-constrained settings, where it is crucial to align service delivery with actual community needs (Macharia et al., 2021). Temperature was included due to its strong association with the incidence of diarrhoea, especially during the warmer summer months. High temperatures promote bacterial and viral growth in water and food sources, increasing the risk of diarrhoeal outbreaks (Musengimana et al., 2016). Ease of access captures the structural and infrastructural characteristics of the built environment. It is particularly relevant in low- to middle-income settings where travel costs, public transport availability, and road quality can significantly affect access to POIs (Macharia et al., 2021).

While other factors such as socio-economic status, healthcare supply chain efficiency and access to clean running water are likely to play a role, these were excluded due to a lack of comprehensive spatial coverage. Future studies could integrate additional datasets as they become available, potentially improving the granularity and predictive power of the segmentation. Nonetheless, the three selected attributes provide a strong foundation for evaluating spatial demand patterns and informing region-specific planning for access to essential medication.

In Section 4.1.2, ease of access, demand, and temperature will be unpacked in more detail and categorised into distinct levels to enable their integration into the attribute based segmentation model.

4.1.2. Attribute prepossessing and categorisation

Key attributes influencing sales, which are allocated to all structural nodes in G_a include ease of access, demand and temperature. These attributes were selected not only for their relevance to diarrhoea medication demand, but also for their ability to capture spatial variability in factors known to impact public health and access to services. Each attribute contributes a distinct layer of contextual information to the segmentation process, allowing the model to differentiate between nodes.

The first attribute, ease of access, categorises nodes based on the primary type of road access, which has three levels: principal, suburb, and other.[2] Nodes with principal road access are either connected by major roads, such as highways, or main roads that serve as key routes within towns or cities. These roads offer the highest level of accessibility and play a crucial role in connecting POIs to densely populated or high-traffic areas. Nodes classified as suburb roads are connected by secondary roads or local streets, typically found in residential or suburban areas. The other category includes unpaved paths, trails, or less accessible roads, indicating lower accessibility for POIs in these locations.

The second attribute, demand, is represented by the average population size from 2015 to 2019 associated with each node, derived from overlaying census and deeds office data provided by Lightstone (Pty) Ltd.[3] To segment demand, we used quartiles based on the distribution of population data in City of Cape Town, defining four categories: no demand for population totals below Q1, low demand for totals between Q1 and Q2, mid demand for totals between Q2 and Q3, and high demand for totals equal to Q3 or more. This approach ensures that regions with similar population-driven demand characteristics are grouped, simplifying the integration of this attribute and removing the complexity of continuous population data. By focusing on quartiles, the attribute reflects both high-density urban areas and sparsely populated regions, capturing the heterogeneity of demand across the City of Cape Town.

The final attribute, temperature, includes two levels: hot days and normal days. High temperatures are particularly relevant as they can have a direct impact on public health and infrastructure. Hot days, defined as days with maximum temperatures exceeding a threshold indicative of heat stress conditions, promote the growth of bacteria, viruses, and parasites in food and water, thereby increasing the risk of waterborne diseases such as diarrhoea (Musengimana et al., 2016). Normal days, on the other hand, represent typical temperature conditions that are less likely to exert such extreme impacts. The levels for all attributes on all structural nodes in G_a is illustrated in Fig. 7.

Temperature data was sourced from the South African Weather Services (SAWS), which provided daily minimum and maximum temperatures recorded at various weather stations in the Western Cape. Since only 10 SAWS stations were located in the City of Cape Town, OpenWeather[4] was used as an additional data source to ensure adequate variability across the city. The combined temperature inputs from the two sources were used to represent the spatial variability across the region, but the coverage was insufficient to represent the entire geographical network G_a. Weather stations were mapped to $4\,\text{km} \times 4\,\text{km}$ grids because weather patterns do not vary significantly over smaller spatial scales, making this resolution appropriate for temperature interpolation. Out of the total 177, $4\,\text{km} \times 4\,\text{km}$ grids covering the City of Cape Town, observed minimum and maximum temperatures were available for only 101 grids. This meant that predictions were required for the remaining 76 grids to fill gaps in the temperature dataset.

To address this, Ordinary Kriging (Cressie, 1990) was applied on a grid size of $4\,\text{km} \times 4\,\text{km}$, and the results were subsequently mapped down to the corresponding $1\,\text{km} \times 1\,\text{km}$ grids to estimate minimum and maximum temperatures. The midpoints of the $4\,\text{km} \times 4\,\text{km}$ grids that overlapped with a SAWS or OpenWeather station were used as the x, y coordinates of the observed value $z(x_i)$, while the remaining grids required interpolation using Ordinary Kriging, defined by $\hat{z}(x_0) = \sum_{i=1}^{N} \lambda_i z(x_i)$ where $z(x_i)$ are the observed values at neighbouring locations, and λ_i are the Kriging weights. These weights are determined by solving a system of linear equations that incorporate the distances between observed points, the distances between observed and unobserved locations and the spatial correlation structure defined by the variogram. For maximum temperatures, a Gaussian variogram structure was used, with a partial sill of 0.004, a range of 0.5 and a nugget of 0.0005. For minimum temperatures an Exponential variogram structure was a better fit, with a smaller range of 0.25, as shown in Fig. 8. These parameters ensure that λ_i effectively captures spatial dependence, assigning higher weights to closer observed points and lower weights to farther ones.

[2] Department of Health South Africa, *TRH 26: South African Road Classification and Access Management Manual*, https://wayleave.tshwane.gov.za/ (accessed August 15, 2024).

[3] Lightstone (Pty) Ltd procures its data directly from the Deeds office and is comprised of a snapshot of all South Africa property ownership as at 1993, with a full history of all transactions to augment with census data. Ethics clearance number: NAS003/2023.

[4] Open Weather, https://openweathermap.org/ (accessed October 17, 2024).

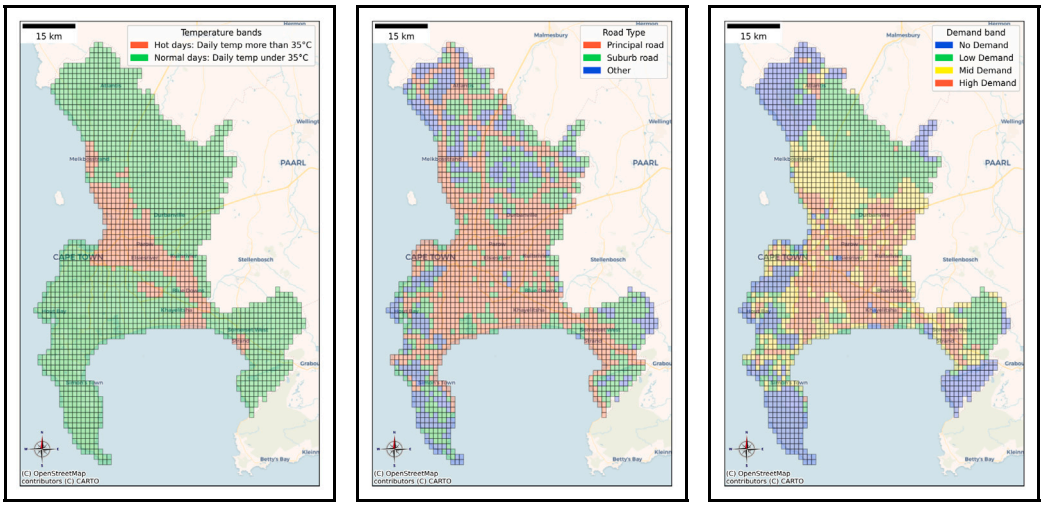


Fig. 7. Visualisation of the temperature (left), ease of access (middle), and demand (right) attributes at node level. The map highlights the spatial distribution of these attributes across individual nodes. Map data © OpenStreetMap contributors, under the Open Database License (ODbL).

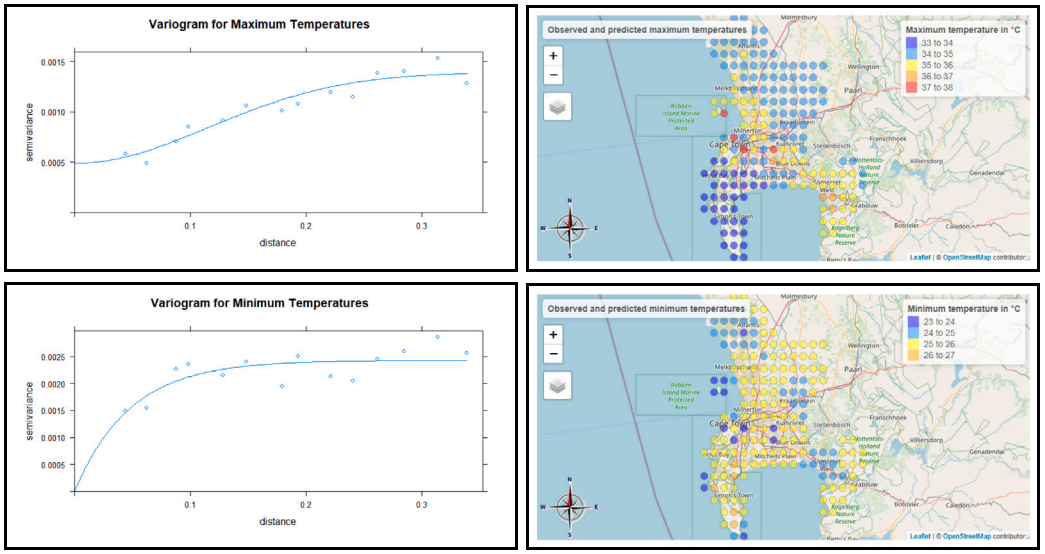


Fig. 8. The top left panel shows the variogram for maximum temperatures. The top right panel illustrates the observed and predicted maximum temperature values across the City of Cape Town. The bottom left panel displays the variogram for minimum temperatures. The bottom right panel shows the observed and predicted minimum temperature values. Map data © OpenStreetMap contributors, under the Open Database License (ODbL).

A combination of both maximum and minimum temperatures are used to create the final temperature attribute. Nodes are classified as "hot days" if they experience maximum summer temperatures above 35 °C and minimum summer temperatures above 25 °C, aligning with criteria identified by Heat Watch Cape Town[5] for high-temperature zones. All other nodes are classified into the "normal days" category.

4.2. Clusters and macro-regions

To illustrate the varying outcomes of clusters and macro-regions based on different attributes, multiple scenarios are investigated. The first scenario focuses on forming clusters and macro-regions by considering one attribute at a time, utilising $\hat{\mathbf{U}}$ to assess the

[5] StoryMap by CAPA Strategies, *Heat Watch Cape Town: Community Heat Mapping in Cape Town, South Africa*, https://storymaps.arcgis.com/stories/35fd2bf7e70c448ab0a58245d2f2cd0b (accessed August 20, 2024).

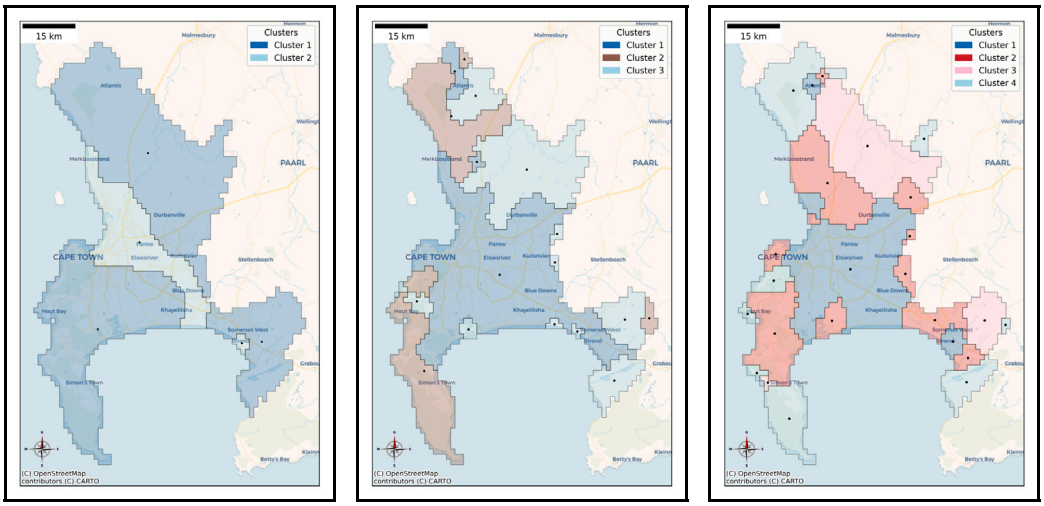


Fig. 9. Representation of the temperature attribute (left), ease of access attributes (middle) and demand attributes (right) at both cluster and macro-region levels. The maps illustrate how these attributes are aggregated into clusters and macro-regions, based on structural edges and spatial connections. Clusters are shown in different shaded colours, with the midpoints of the macro-regions indicated by the black dots. Map data © OpenStreetMap contributors, under the Open Database License(ODbL).

impact of individual attributes. The second scenario examines the combined influence of all attributes, comparing the clusters generated under $\mathbf{P}_a$ with those derived from $\hat{\mathbf{U}}$.

Let $\mathbf{P}_a$ be constructed for three different scenarios: the first includes only temperature as an attribute, the second considers only ease of access as an attribute, and the third includes only demand as an attribute in G_a. When examining each attribute separately, the optimal number of clusters, determined by partitioning $\mathbf{S}(t) = (\mathbf{s}_1, \dots, \mathbf{s}_{N+T})^T$ until the optimal clusters are obtained, corresponds to the number of levels in that attribute. For temperature, $K = z_1 = 2$ clusters are identified; for ease of access, $K = z_2 = 3$ clusters are identified; and for demand, $K = z_3 = 4$ clusters are identified. This is because there is no overlap or interference between attributes and transition probabilities between nodes. As a result, the alignment between the optimal number of clusters and the levels in the attribute is very distinct.

The number of macro-regions (Q) formed when considering only ease of access as an attribute is $Q = 19$, for demand $Q = 24$, and for temperature $Q = 5$. In all cases, the number of clusters (K) is less than the number of macro-regions ($K < Q$) because macro-regions are formed using spatial connections and clusters. When a cluster contains nodes that belong to the same cluster but are not spatially adjacent, these nodes are assigned to separate macro-regions.

With a single attribute, Eq. (9) converges rapidly, and the final solution for $\hat{\mathbf{U}}$ closely resembles the initial solution $\mathbf{U}(0)$, obtained when clustering $\mathbf{P}_a$. The macro-regions, formulated using structural edges and cluster membership as indicated in Eq. (5), are illustrated in Fig. 9. When comparing Figs. 7 to 9, the macro-regions align naturally with the attributes on node level, as only one attribute is included in each case.

However, when considering all attributes together, with $T = z_1 + z_2 + z_3 = 9$ different attribute levels across the three, defining macro-regions becomes more challenging. The optimal number of clusters, as identified from the elbow method is $K = 15$, from which $\mathbf{U}(0)$ is initially derived. Using the cluster memberships from $\mathbf{U}(0)$ together with structural edges $(v_i, v_j) \in E$, $Q = 358$ macro-regions are obtained, as shown in the first image in Fig. 10. Although the k-means method minimises the distance between the initial cluster centroids and the transition probabilities of the nodes, the resulting solution is not spatially optimal. It leads to very small macro-regions that reflect minor differences in the data, rather than smoothing the spatial solution to create more realistic and applicable regions.

When applying the iterative process outlined in Eq. (9), the process converges at $t = 33$, and the macro-regions obtained using structural edges $(v_i, v_j) \in E$ and $\hat{\mathbf{U}}$ are shown in the second image of Fig. 10. The resulting macro-regions are more spatially smoothed out, which leads to a reduction of $Q = 358$ to $Q = 55$ macro-regions. The macro-regions in the second image of Fig. 10 form spatially cohesive areas that reflect underlying structural links and combinations of temperature, ease of access, and demand patterns.

4.3. Attribute segmentation

An attribute segment is a description of a cluster based on the modal attributes observed within that cluster (and consequently the macro-regions associated with that cluster) when considering all levels of the attributes. For this application, attribute segmentation provides a unique description of the combination of all levels observed for temperature, ease of access and demand.

Even though $K = 15$ clusters are identified in the final cluster membership $\hat{\mathbf{U}}$, as shown in Fig. 10, multiple clusters share the same prominent attributes when considering all levels across temperature, ease of access and demand. In total, there are $R = 9$ distinct attribute segments, which span over the $K = 15$ clusters and are classified as shown in Table 1.

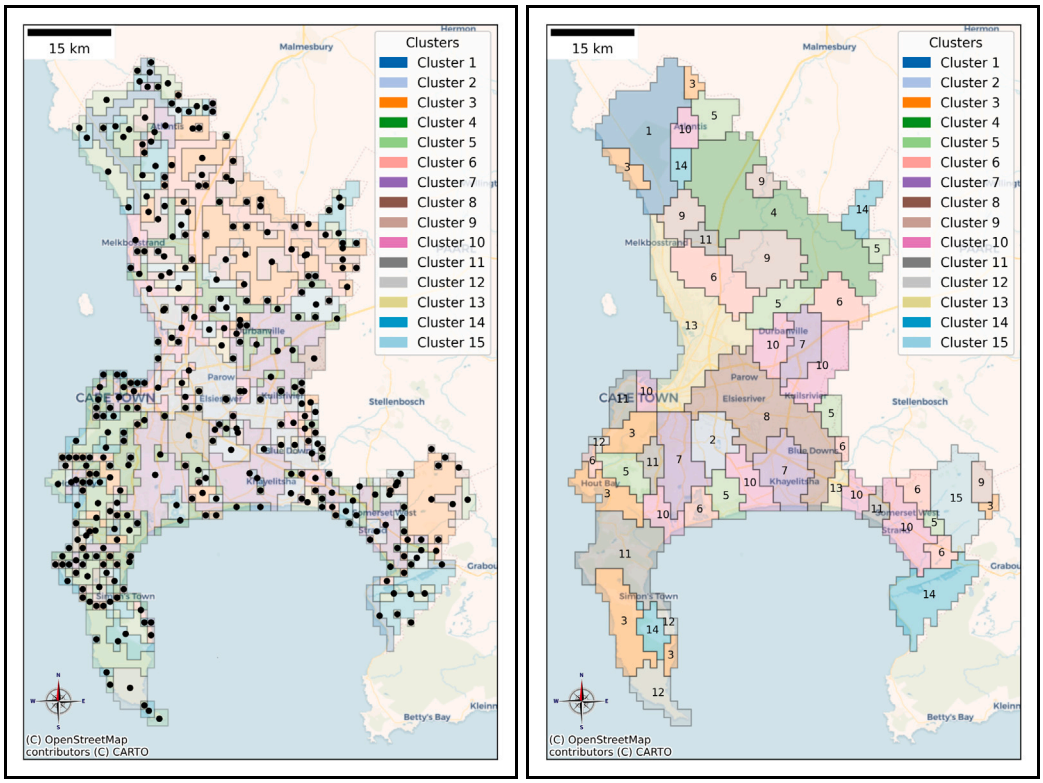


Fig. 10. Macro-regions constructed using the initial cluster membership ($\mathbf{U}(0)$, left) and the final cluster membership ($\hat{\mathbf{U}}$, right). Clusters are shown in different shaded colours. On the left, the midpoints of the initial $Q = 358$ macro-regions are marked with black dots, while on the right, the midpoints of the reduced $Q = 55$ macro-regions are labelled with their corresponding cluster numbers. Map data © OpenStreetMap contributors, under the Open Database License (ODbL).

Table 1
Distinct clusters with mode value of ease of access, temperature, and demand.

Cluster	Road type	Temp	Demand	Attribute segmentation
C_1	Other	Normal	None	ONN
C_2	Principal	Normal	High	PNH
C_3	Other	Normal	None	ONN
C_4	Suburb	Normal	Low	SNL
C_5	Suburb	Normal	Low	SNL
C_6	Suburb	Normal	Mid	SNM
C_7	Principal	Normal	High	PNH
C_8	Principal	Hot days	High	PHH
C_9	Principal	Normal	Low	PNL
C_{10}	Principal	Normal	High	PNH
C_{11}	Principal	Normal	Mid	PNM
C_{12}	Other	Normal	None	ONN
C_{13}	Principal	Hot days	Mid	PHM
C_{14}	Suburb	Normal	None	SNN
C_{15}	Suburb	Normal	Low	SNL

The assignment of the attribute segments can be described as indicated in Table 2.

As part of an exploratory descriptive analysis focused on the clusters, macro-regions and attribute segments, the potential correlations between these and the average sales for POIs are explored. When examining the macro-regions with which the POIs overlap, it is observed that only the macro-regions linked to the set of clusters $\{C_2, C_7, C_8, C_{10}, C_{11}, C_{13}\}$ overlap with POIs. These clusters are associated with only four attribute segments: PNH, PHH, PNM, and PHM. This suggests that the macro-regions from a location perspective and attribute perspective exhibit unique qualities that align closely with POI placement. The attribute segmentation, average sales, and total number of stores for each cluster that overlaps with at least one POI are summarised in Table 3.

It is noted that all clusters listed in Table 3 have principal roads as their primary road type, indicating high accessibility. Notably, both clusters C_8 and C_{13}, with average sales above 7.5, are associated with hot days in the temperature attribute, suggesting a

Table 2
Description of each attribute segment based on road type, temperature and demand level.

Attribute segment	Description
ONN	Other roads, normal temperature and no demand.
PNH	Principal roads, normal temperature and high demand.
SNL	Suburb roads, normal temperature and low demand.
SNM	Suburb roads, normal temperature and mid-level demand.
SNN	Suburb roads, normal temperature and no demand.
PHH	Principal roads, high temperature (hot days) and high demand.
PNM	Principal roads, normal temperature and mid-level demand.
PHM	Principal roads, high temperature (hot days) and mid-level demand.
PNL	Principal roads, normal temperature and low demand.

Table 3
POI average sales per cluster and attribute segmentation.

Cluster	Attribute segmentation	Average sales	Total stores
C_2	PNH	8.98	2
C_7	PNH	6.65	11
C_8	PHH	8.18	3
C_{10}	PNH	5.98	19
C_{11}	PNM	5.29	4
C_{13}	PHM	7.5	9

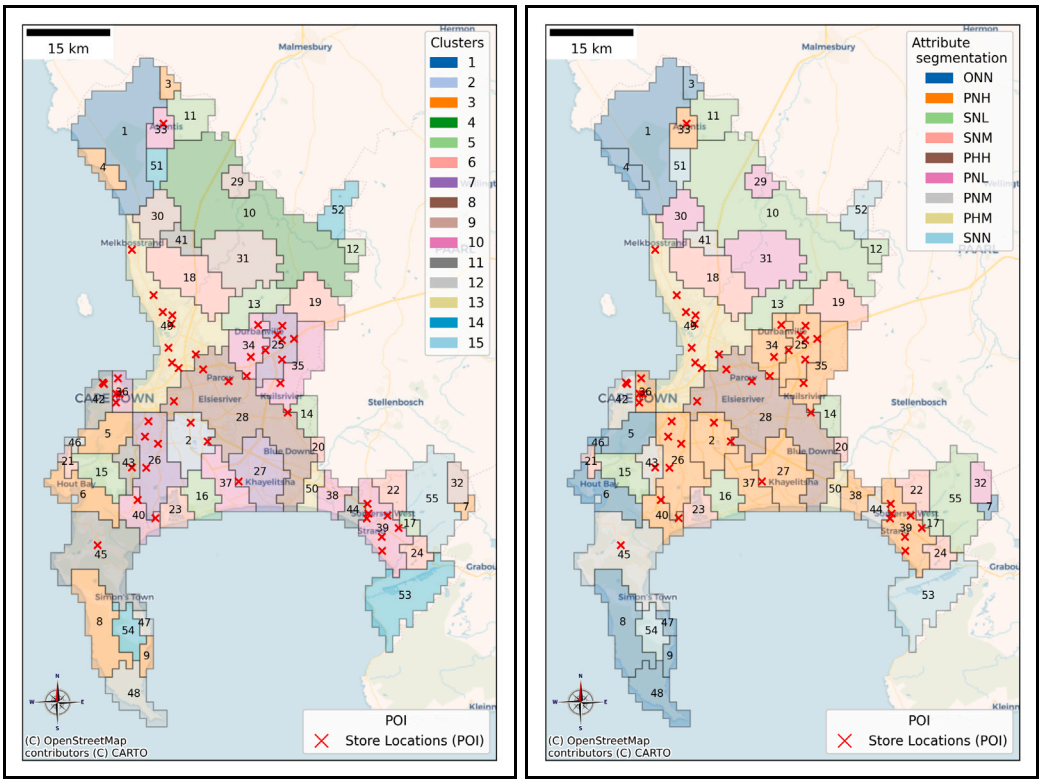


Fig. 11. POI locations are indicated with an x, shown over macro-regions in relation to clusters (left) and attribute segmentation (right). The left image displays POI locations with clusters shaded in different colours and the midpoints of the macro-regions labelled with their corresponding macro-region numbers. The right image shows POI locations over attribute segmentations, shaded in different colours, with the midpoints of the macro-regions labelled with their corresponding macro-region numbers. Map data © OpenStreetMap contributors, under the Open Database License (ODbL).

potential link between elevated temperatures and increased demand. The demand levels across these clusters are generally mid or high. The placement of POIs overlaid with clusters and attribute segments shown at the macro-regions level is illustrated in Fig. 11.

Cluster C_2 exhibits a high average sales value despite not being associated with an attribute segment typically linked to elevated sales. The increased sales observed in C_2 could be attributed to other influencing factors, such as the size of the store, its operating hours, the availability of stock or the store's location.

The effect of incorporating attribute segmentation to predict supply at a POI level will be tested using a SAR model. This will involve incorporating the location of a macro-region, or more specifically, the location of a store within a macro-region, along with attribute segments. Attribute segments will be included as aspatial covariate inputs, with their influence further integrated into the covariates spatial weights matrix. This approach will be applied in the next section to predict the average number of sales per POI.

4.4. Spatial autoregressive model with attribute segmentation

A spatial autoregressive (SAR) model (Cressie, 1990) is utilised to account for spatial dependence in the data by incorporating a spatial weights matrix ($\mathbf{W}$) that captures the relationships or proximities between different regions. In this study, we apply a SAR model to assess whether attribute based spatial segments impact the prediction of the summer average sales of a diarrhoea medication across 48 pharmaceutical POIs, $P = \{P_1, \dots, P_M\}, M = 48$, located in the City of Cape Town.

Attribute based spatial segmentation is expected to contribute to the predictive power of the model by grouping regions based on shared characteristics. However, since attribute segmentation alone does not capture store-level differences, the quantitative variable operating hours (OpHours), representing the total number of hours a store is open during the week, is included as an additional covariate. Incorporating OpHours accounts for variations in operational time across stores, which likely influences sales performance. This combination of regional-level and store-level covariates provides a comprehensive basis for the spatial regression model.

Before fitting the SAR model, a general linear model was first applied to predict log-transformed sales, per POI, using the covariates OpHours and the attribute segmentations (PHH, PNM, PHM, with PNH as the reference category). Moran's I statistic (Cressie, 1990) was calculated based on a 5-nearest neighbour structure for the residuals of this model, yielding $I = 0.143$, with a variance of 0.006 and a p-value of 0.019. These results indicate statistically significant spatial autocorrelation in the residuals, justifying the use of a spatial regression approach. The SAR model builds upon the spatial regression framework (Ver Hoef et al., 2018) defined as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} + \boldsymbol{\epsilon} \tag{11}$$

where $\mathbf{y}$ is the $M \times 1$ vector of log-transformed average sales for each POI, $\mathbf{X}$ is the $M \times p$ matrix of explanatory variables including OpHours and attribute segmentations (PHH, PNM, PHM, with PNH as the reference category), $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression coefficients and p the number of parameters, $\mathbf{u} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ is the latent spatial random error and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_\epsilon^2 \mathbf{I})$ is the independent error term.

For SAR models, the covariance structure is defined as

$$\boldsymbol{\Sigma} = \sigma_u^2 \left((\mathbf{I} - \mathbf{C})(\mathbf{I} - \mathbf{C})^\top\right)^{-1}$$

where $\mathbf{C} = \rho\mathbf{W}$, with $\mathbf{W}$ the spatial weights matrix and ρ the SAR smoothing parameter (Ver Hoef et al., 2018). The weights matrix $\mathbf{W}$ incorporates both spatial and attribute based proximities between regions, as defined by Ejigu and Wencheko (2020)

$$w_{ij} = \exp\left(-\left(\alpha u_e + (1-\alpha) d_{ij}\right)\right) \tag{12}$$

where d_{ij} represents the standardised Euclidean distance between locations i and j and u_e reflects the absolute difference in the environmental covariates between two locations. Traditionally, in Eq. (12), u_e is based on the absolute difference of a quantitative environmental covariate. However, in this study, the environmental covariates are non-ordinal qualitative values. Consequently, we construct u_e based on attribute segmentation similarity, such that $u_{e_{ij}} = 0$ if POI i and j belong to the same attribute segmentation (associated with the overlapping macro-region) and $u_{e_{ij}} = 1$ if they belong to different attribute segmentations.

The parameter α controls the balance between spatial distance and attribute similarity ($\alpha \in [0, 1]$). For $\alpha = 0$, the weights depend entirely on spatial distance, while for $\alpha = 1$, they depend solely on attribute similarity. The SAR model for this application is formulated as

$$y = \beta_0 + \beta_1 \text{OpHours} + \beta_2 \text{PHH} + \beta_3 \text{PNM} + \beta_4 \text{PHM} + u + \epsilon. \tag{13}$$

This model is tested across all levels of α, starting at 0 (indicating that only spatial distance is considered) and increasing in increments of 0.05 up to 1 (indicating that only the attribute segmentation environmental effect is considered). In addition to evaluating the model's performance, the significance of the variables is assessed at all levels of α. The operating hours (OpHours) were found to be significant at all levels of α. However, to test the effect of including attribute segmentation, we cannot rely solely on the p-values from the model output, as these values are dependent on the chosen reference category. Therefore, to quantify the significance of including attribute segmentation, we use a likelihood ratio test (Dunn and Smyth, 2018) to compare Model A, which has p_A parameters, with Model B, which has p_B parameters, as shown in the following equation

$$\begin{aligned} &\text{Model A: } y = \beta_0 + \beta_1 \text{OpHours} + u + \epsilon \\ &\text{Model B: } y = \beta_0 + \beta_1 \text{OpHours} + \beta_2 \text{PHH} + \beta_3 \text{PNM} + \beta_4 \text{PHM} + u + \epsilon. \end{aligned} \tag{14}$$

Under H_0, we are testing whether the categorical covariate, attribute segmentation, does not have a significant effect on the model, i.e., $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$. Under H_0, assuming that Model A is sufficient, denote the fitted values as $\hat{y}_A$, which produces the log-likelihood l_A. Similarly, for Model B with fitted values $\hat{y}_B$, the log-likelihood is l_B. The log-likelihood ratio test (LRT) statistic (Dunn and Smyth, 2018) for comparing Models A and B is

$$L = 2(l_B - l_A) \sim \chi^2_{p_B - p_A}. \tag{15}$$

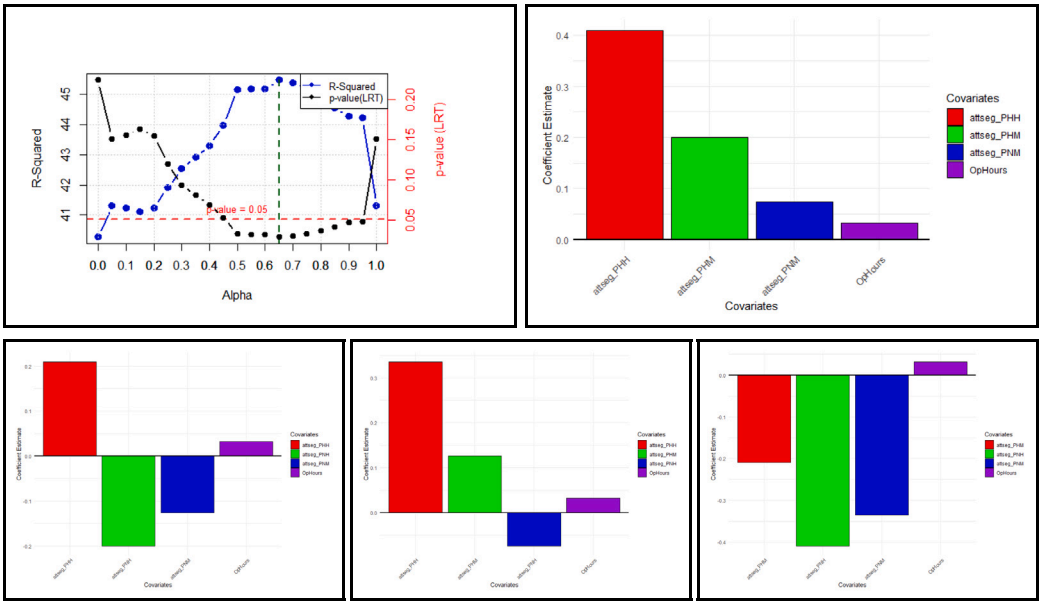


Fig. 12. Top-left: R-squared and p-value (LRT) trends for varying α in the SAR model. Top-right: Coefficient estimates with reference level PNH. Bottom-left: Coefficient estimates with reference level PHM. Bottom-middle: Coefficient estimates with reference level PNM. Bottom-right: Coefficient estimates with reference level PHH.

Testing both the model performance and significance of the covariates as indicated in Eqs. (14) and (15), the final results of the SAR model outlined in Eq. (13) are presented in Fig. 12 for all values of α ranging from 0 to 1 in increments of 0.05.

From the results in Fig. 12 it can be noted that at $\alpha = 0$, we observe the lowest $R^2 = 40.27$, and the attribute segmentation covariates are not significant in the model (p-value(LRT) = 0.52). As the value of α increases, the R^2 value also increases, reaching a maximum of 45.476, after which it begins to decrease again. The attribute segmentation covariates are significant for values of $0.5 \leq \alpha \leq 0.95$. From the results in Fig. 12, $\alpha = 0.65$ produces the best model, with the highest R^2 and all variables being significant.

When considering the effect of all covariates in Fig. 12 with the reference level for the attribute segmentation covariate being PNH (top right image), it can be observed that all other covariates have a positive effect on the log-transformed value of average sales. The PHH level has the strongest effect, followed by PHM, both of which are associated with the "hot days" attribute. When the reference level is changed to PHM, PNM, and PHH, as shown in Fig. 12, the operating hours consistently have a positive effect, while the impact of PHH and PHM remains substantial. In the last image, where PHH is the reference level, all other levels exhibit a negative effect relative to PHH, which is expected since PHH is associated with the highest average sales.

Upon investigation of the $Q = 55$ macro-regions, five have been identified as having the potential to exhibit positive average sales despite currently lacking allocated POIs. Among these, three macro-regions are classified as spatially inaccessible regarding access to the specific POI in question, suggesting high sales potential if accessibility issues are addressed as shown in the second row of images in Fig. 13.

Conversely, other spatially inaccessible regions, such as those segmented under the ONN attribute segment, will not benefit from adding a POI due to their characteristics, including limited road access, typical temperature ranges, and a lack of demand as indicated on Fig. 13. This segmentation provides a clear rationale for why certain areas remain spatially disjoint and do not warrant further resource allocation.

4.5. Effect of α, attributes, weights selection, label connectivity and sample size

While the α parameter plays an important role in balancing structural and attribute influences in the model, it is not the only factor that affects segmentation outcomes. Several other modelling choices also impact performance and interpretability. One such factor is the selection of attributes and their relevance to the specific spatial problem under investigation. In this study temperature, demand and ease of access were chosen as attributes based on data availability and their empirical correlation with observed sales as discussed in Section 4.1.1.

In addition to attribute selection, the treatment of edge weights plays a crucial role. Structural and attribute edge weights represent the strength of connections between nodes in the graph. In this research, both types of edge weights were uniformly set to 1 to indicate the presence of a connection. However, future studies could investigate the impact of quantifying structural and attribute edge weights between regions, using an iteratively adjusted self-learning mechanism. This dynamic adjustment process would modify the weights based on how effectively an attribute contributes to clustering. Attributes that strongly differentiate

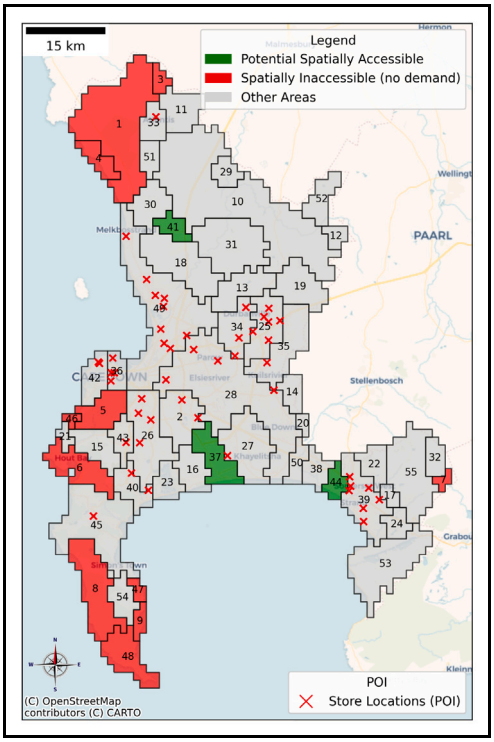


Fig. 13. POI locations are indicated with an x, shown over macro-regions (with midpoints labelled by macro-region numbers) in relation to attribute segmentation. Red-shaded macro-regions indicate spatially inaccessible areas that fall within an ONN or SNN attribute segmentation with no demand. Green-shaded regions represent macro-regions with high sales potential, classified under PNH, PHH, PNM, or PHM attribute segmentations, but with no allocated POI. The remaining grey-shaded macro-regions either already contain a POI or fall within one of the remaining attribute segments that did not indicate high sales potential. Map data © OpenStreetMap contributors, under the Open Database License (ODbL).

clusters would be assigned higher weights, while those with weaker clustering tendencies would receive lower weights, as discussed in Zhou et al. (2009).

Label connectivity is also a foundational assumption in the segmentation framework. All regions analysed in this study were spatially connected and permitted the computation of probabilities across all nodes. In areas with fragmented infrastructure or natural barriers, disconnected graphs may emerge. These cases require special treatment to preserve internal consistency within clusters and may demand modified segmentation strategies. Addressing these factors through adaptive weighting, robust attribute selection, and attention to connectivity will be essential for applying the framework to diverse domains and geographic contexts. Alternative methods for labelling disconnected graphs, as discussed in Gallian and Stewart (2015), can also be explored.

Even though the sample size used in this study was limited to 48 pharmaceutical POIs, the SAR model was implemented with several safeguards to reduce potential estimation bias. As highlighted by Smith (2009), strongly connected spatial weights matrices, where every location is linked to every other, can lead to downward-biased estimates of spatial dependence and reduce the effectiveness of diagnostics such as Moran's I, which may fail to detect spatial autocorrelation under excessive network density. In this study, the spatial weights matrix $\mathbf{W}$ used in the SAR model was fully connected (label connected) to allow for the integration of attribute based similarity, however, the presence of meaningful spatial dependence was independently confirmed using a Moran's I test based on a 5-nearest neighbour structure, prior to fitting the SAR model. This yielded a Moran's I value of 0.143 ($p = 0.019$), confirming statistically significant spatial autocorrelation that was not a result of over-connectedness. Furthermore, the SAR model incorporated store-level covariates, specifically OpHours, to account for operational heterogeneity across POIs, which helps to reduce potential estimation bias by controlling for known non-spatial effects. Future research could enhance model robustness by incorporating resampling methods, such as bootstrapping, or expanding the dataset to finer spatial or temporal resolutions, especially when evaluating the influence of attribute informed spatial weights in small-sample contexts (Itiveh and Aronu, 2025).

5. Discussion

The advantages of attribute based spatial segmentation lie in its ability to generate clusters that are tailored to the specific characteristics and needs of different regions. By incorporating spatial proximity along with environmental and other relevant attributes, this approach ensures that resource allocation aligns closely with local demands. The methodology supports realistic macro-regions, creating attribute segments that reflect area similarities and enable targeted planning and optimisation efforts.

The application of this methodology to predict the average sales of an over-the-counter diarrhoea medication in South Africa provides valuable insights into its practical utility. The use of attribute segmentation captures the interactions between environmental, spatial, and demand-related factors, revealing patterns that would not be evident through traditional clustering approaches.

Beyond the SAR models applied in this research, other spatial regression methods could also be explored as potential alternatives. For instance, geographically weighted regression (GWR) (Fotheringham et al., 2003) allows for localised relationships between predictors and outcomes, enabling the analysis of spatial heterogeneity across regions. Similarly, spatial error models (SEM) and spatial lag models (SLM) (Darmofal, 2015) address spatial dependence in error terms or outcomes. These models are particularly relevant when interactions between regions play a critical role in shaping outcomes, as is often the case in spatial planning applications. The application of this work, however, lies in the integration of attribute based spatial segmentation as covariates in the spatial prediction, enhancing the adaptability of resource planning models to diverse and geographically nuanced contexts.

This methodology can be applied to any geographical region, G_a, provided the areas exhibit similar environmental or attribute based characteristics to ensure consistent and meaningful clusters. For instance, in South Africa, Gauteng and the Western Cape demonstrate vastly different patterns in terms of population size and climate, necessitating separating the regions to capture their unique regional dynamics effectively.

The grid and area sizes can be adjusted to a finer scale for more precise POI optimisation within towns or cities. However, the current application was limited by insufficient records for smaller geographical regions, preventing thorough testing at finer scales. When working with smaller areas, it is crucial to ensure adequate variability in selected attributes, as environmental factors, such as temperature, may show minimal variation within small regions.

While this study focused on a single pharmaceutical chain in South Africa, multiple smaller pharmacies that sell the same over-the-counter medication could be incorporated to improve demand predictions. Including suburban pharmacies could capture broader sales patterns, including areas with lower demand, and enhance the overall prediction model.

Further enhancements could include the integration of temporal dynamics to capture the evolving nature of spatial regions. Regional characteristics, such as economic development, infrastructure expansion, population mobility, and environmental shifts, can significantly change over time, requiring adaptive regional divisions. The proposed methodology is well-suited to accommodate this dynamism. By incorporating temporal data streams and developing temporal catchment areas, the model could continuously update segmentation outputs to reflect real-time or seasonal variations in accessibility, demand, or environmental conditions. This approach would be particularly beneficial in contexts where infrastructure, such as road networks, changes frequently, affecting regional boundaries and accessibility calculations.

The integration of mixed data clustering techniques with spatial regularisation, such as the Fuzzy C-Medoids Clustering for Mixed Data model with spatial constraints (FCMd-MD-SP) (D'Urso et al., 2025), also presents an avenue for further research. This model handles datasets containing both quantitative and qualitative attributes by combining dissimilarity measures for each attribute type through a weighting system. The weights are optimised during the clustering process, reflecting the relevance of each attribute type, while a spatial penalty term ensures that membership degrees are spatially smoothed.

This approach also holds potential for extending to other domains, such as environmental and public health studies. For instance, attribute based spatial segmentation could be used to investigate how spatial variation in vegetation type, temperature, rainfall and elevation contributes to ecologically distinct zones, supporting conservation planning or climate adaptation strategies in more complex spatial patterns. Additionally, incorporating spatial attributes like access to clean water, sanitation infrastructure and population density could help identify areas at risk of disease outbreaks, informing early intervention strategies.

6. Conclusion

Our research presents a novel framework for spatial planning and resource optimisation, leveraging attribute based spatial segmentation. This study integrates probabilistic clustering structures with iterative refinement processes to produce tailored macro-regions that align with both regional attributes and geographic needs. The methodology stands out by ensuring the precise delineation of service catchment areas, optimising POI placements, and addressing spatially accessible as well as disjoint areas through adaptive segmentation.

Unlike traditional connectivity-based clustering approaches, our model incorporates spatial and attribute based proximities, dynamically adapting to local conditions and providing actionable insights into the unique characteristics of diverse geographic regions. By integrating spatial regression analysis, including a SAR model, the framework not only accounts for spatial dependence but also quantifies the role of attribute segmentation and store-level characteristics, such as operating hours, in predicting POI-level outcomes.

The results demonstrate the potential of this methodology to significantly improve resource allocation and service optimisation, particularly in heterogeneous environments. The adaptability of the proposed framework ensures its applicability across various industries and regions, making it a robust tool for informed decision-making in urban planning, retail, and healthcare resource management.

Acknowledgements

The work reported herein was made possible through funding by the South African Medical Research Council through its Division of Research Capacity Development under the Biostatistics Capacity Development partnership with the Belgian Development Agency (Enabel) under its framework of Building Academic Partnerships for Economic Development(BAPED). Additionally, this research received support from the National Research Foundation of South Africa (Grant Number 137785 and COE-MaSS Grant Number 2022-086-LIF-COVID-19) . The content and opinions expressed herein are the sole responsibility of the authors and do not necessarily represent the official views of the NRF, SAMRC or the funders.

References

Bhagat, S., Cormode, G., Muthukrishnan, S., 2011. Node classification in social networks. Soc. Netw. Data Anal. 115–148.
Blum, A., Chawla, S., 2001. Learning from labeled and unlabeled data using graph mincuts. In: Proceedings of the 18th International Conference on Machine Learning. pp. 19–26.
Cressie, N., 1990. The origins of Kriging. Math. Geol. 22, 239–252.
Darmofal, D., 2015. Spatial Analysis for the Social Sciences. Cambridge University Press, UK.
de Klerk, M., Fabris-Rotelli, I., 2024. Hospital Accessibility Catchment Areas as a fuzzy lattice data structure. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. X-4-2024, pp. 99–106.
de Klerk, M., Fabris-Rotelli, I., 2025. Seasonal Catchment Areas using an attribute based fuzzy lattice data structure. South African Statist. J. 59 (1), 1–16.
De Meo, P., Ferrara, E., Fiumara, G., Provetti, A., 2011. Generalized louvain method for community detection in large networks. In: 2011 11th International Conference on Intelligent Systems Design and Applications. IEEE, pp. 88–93.
Dunn, P.K., Smyth, G.K., 2018. Generalized Linear Models with Examples in R. Springer, New York.
D'Urso, P., De Giovanni, L., Federico, L., Vitale, V., 2025. Fuzzy clustering of mixed data with spatial regularization. Spat. Stat. 65, 100874.
Ejigu, B.A., Wencheko, E., 2020. Introducing covariate dependent weighting matrices in fitting autoregressive models and measuring spatio-environmental autocorrelation. Spat. Stat. 38, 100454.
Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, UK.
Gallian, J.A., Stewart, D., 2015. Properly even harmonious labelings of disconnected graphs. AKCE Int. J. Graphs Comb. 12 (2–3), 193–203.
Itiveh, F., Aronu, C., 2025. Enhancing spatial autoregressive models with bootstrap techniques: A methodological investigation into bias, precision, and sample size effects. Earthline J. Math. Sci. 15 (3), 381–399.
Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognit. Lett. 31 (8), 651–666.
Jeh, G., Widom, J., 2002. Simrank: A measure of structural-context similarity. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 538–543.
Lin, H., Zhan, Y., Zhao, Z., Chen, Y., Dong, C., 2021. Overlapping community detection based on attribute augmented graph. Entropy 23 (6), 680.
Macharia, P.M., Ray, N., Giorgi, E., Okiro, E.A., Snow, R.W., 2021. Defining service catchment areas in low-resource settings. BMJ Glob. Heal. 6 (7).
Musengimana, G., Mukinda, F.K., Machekano, R., Mahomed, H., 2016. Temperature variability and occurrence of diarrhoea in children under five-years-old in cape town metropolitan sub-districts. Int. J. Environ. Res. Public Heal. 13 (9), 859.
Ng, A., Jordan, M., Weiss, Y., 2001. On spectral clustering: Analysis and an algorithm. Adv. Neural Inf. Process. Syst. 14.
Pothen, A., 1997. Graph partitioning algorithms with applications to scientific computing. In: Parallel Numerical Algorithms. Springer, pp. 323–368.
Smith, T.E., 2009. Estimation bias in spatial models with strongly connected weight matrices. Geogr. Anal. 41 (3), 307–332.
Steinley, D., 2006. K-means clustering:A half-century synthesis. Br. J. Math. Stat. Psychol. 59 (1), 1–34.
Tishby, N., Slonim, N., 2000. Data clustering by Markovian relaxation and the information bottleneck method. Adv. Neural Inf. Process. Syst. 13.
Tsai, C.-Y., Chiu, C.-C., 2008. Developing a feature weight self-adjustment mechanism for a k-means clustering algorithm. Comput. Statist. Data Anal. 52 (10), 4658–4672.
Ver Hoef, J.M., Peterson, E.E., Hooten, M.B., Hanks, E.M., Fortin, M.-J., 2018. Spatial autoregressive models for statistical inference from ecological data. Ecol. Monograph. 88 (1), 36–59.
Von Luxburg, U., 2007. A tutorial on spectral clustering. Stat. Comput. 17, 395–416.
Wang, F., 2014. Quantitative Methods and Socio-Economic Applications in GIS. CRC Press, United States.
Wierzchoń, S., Kłopotek, M., 2017. Modern Algorithms of Cluster Analysis. Springer International Publishing, Germany.
Zhou, Y., Cheng, H., Yu, J.X., 2009. Graph clustering based on structural/attribute similarities. Proc. the VLDB Endow. 2 (1), 718–729.