

# Seasonal catchment areas using an attribute based fuzzy lattice data structure

*Michelle de Klerk and Inger Fabris-Rotelli*

Department of Statistics, University of Pretoria, Pretoria, South Africa

Seasonality impacts various industries and sectors, influencing agricultural cycles, economic planning, and healthcare resource allocation. We propose a novel approach using an attribute based fuzzy lattice data structure to create overlapping catchment areas using the fundamentals of label propagation and graph clustering. This approach considers both the link structure and attribute similarities between nodes in a network, where the nodes are points of interest in a road network. Nodes may be close or far apart based on connectivity and shared attributes, such as common interests or in a geographical application considering topography features. In this study, we incorporate static and seasonal attributes for geographical nodes, allowing us to explore seasonal catchment areas and provide a more realistic view of accessibility throughout the year. This integrated approach offers a comprehensive framework for assessing spatial accessibility and understanding seasonal variations in regions to enhance planning for essential services.

*Keywords:* Attribute augmented graph, Fuzzy lattice data, Seasonal catchment areas.

## 1. Introduction

Accessibility to essential facilities and services is a fundamental component of societal well-being and equitable development as highlighted in Wang (2014). Ensuring equal access to resources like healthcare, education, and infrastructure is critical for promoting social equity, particularly for disadvantaged and under-served populations. Unfortunately, accessibility is often hindered by a range of geographic and non-geographic barriers, resulting in significant disparities in service provision as discussed in Green et al. (2016). This is particularly evident for disadvantaged groups, where factors such as income, minority status, and geographical remoteness can exacerbate accessibility issues (Rader et al., 2022). Globally, there is a stark contrast between high-income and low- and middle-income countries when it comes to accessing basic services, with the latter often facing considerable barriers and limited infrastructure (Peters et al., 2008). In low-resource settings, capturing accurate catchment areas is essential for estimating population needs and optimising resource distribution. Effective resource planning and demand estimation hinge on understanding both supply and demand dynamics.

Spatially accessible areas are regions where services or facilities can be reached by a population within a defined geographic boundary, based on factors like distance, travel time or ease of access.

---

*Corresponding author:* Michelle de Klerk ([mich.botes@gmail.com](mailto:mich.botes@gmail.com))

*MSC2020 subject classifications:* 62H11, 05C90, 05C81

These areas are characterised by their proximity to a point of interest (POI), making them available for potential or actual use by the community (Wang, 2014). Potential spatially accessible areas are regions where services or facilities theoretically offer access to the population, implying service availability but not guaranteeing actual utilisation. Realised spatially accessible areas represent regions where potential access translates into actual usage, capturing the utilisation of services by the population. This depends on various individual, social, and economic factors (Khan, 1992).

Spatial catchment areas are geographical regions within which services or facilities can be accessed by a population. A catchment area represents the geographical zone from which a POI attracts communities to utilise its services or products (Tao et al., 2018; Luo and Wang, 2003; Shao and Luo, 2022). It is typically defined by the maximum distance or travel time users are willing to cover to reach the POI and can be determined naturally (e.g., by geographic boundaries) or as a predefined establishment (Green et al., 2016; Luan et al., 2020).

There is a minimal set of conditions from both the supply and demand side which a catchment area should meet as discussed in Macharia et al. (2021). On the supply side, this involves considering the various services provided by POIs, determining whether they are specialised, accounting for competing service providers, assessing the capacity of POIs and identifying any natural or man-made barriers. On the demand side, it is important to identify areas that experience significant fluctuations throughout the year and account for time-sensitive factors such as weather changes and population flows.

Seasonality impacts a variety of industries and sectors. In agriculture, it can influence harvesting times, soil quality, and disease outbreaks (Anyamba et al., 2014). In economics, seasonal trends affect resource planning for tourism and retail sales, requiring careful resource allocation (Corluka, 2019). In healthcare, optimising resources is crucial for addressing seasonal infectious diseases, such as colds and flus and gastrointestinal diseases (Musengimana et al., 2016).

Catchment areas can be classified as either non-overlapping or overlapping (Challen et al., 2022). Non-overlapping catchment areas provide a simplified model of spatial access where each geographical region is uniquely associated with a single service facility. While this approach simplifies modelling, it may misallocate demand, particularly when multiple facilities could serve the same region, potentially neglecting user choice and specific needs (Challen et al., 2022). Overlapping catchment areas offer a more realistic approach, allowing multiple facilities to serve the same region, acknowledging user choice, and providing a better representation of service accessibility. Overlapping catchments demonstrate how demand can be shared across facilities, potentially reducing disparities in service provision (Wang, 2014). However, ensuring equitable distribution and avoiding demand concentration at a few nodes can remain challenging.

Various methods exist for identifying catchment areas in spatial accessibility studies. These range from simpler approaches like circular buffers (Andersen and Landex, 2008) and Euclidean buffer-based methods (Lin et al., 2020) to more advanced techniques such as the floating catchment method, the two-step floating catchment area (2SFCA) method (Radke and Mu, 2000; Luo and Wang, 2003), gravity-based models (Wang, 2014), distance decay functions, and variable catchment areas, as summarised in Tao et al. (2018).

A probabilistic approach to defining overlapping catchment areas has been explored through the use of a fuzzy lattice data structure, as discussed by de Klerk and Fabris-Rotelli (2024). This method represents spatially accessible regions while promoting a more balanced distribution of demand and

supply. Fuzzy lattice catchment areas are generated using label propagation principles and follow a probabilistic framework that enables proportional demand allocation. In this context, "fuzzy" refers to assigning membership degrees as probabilities between 0 and 1, ensuring a more realistic representation of spatial accessibility. Drive-time thresholds (Green et al., 2016) are applied to ensure the probability structure accounts for both the geographical boundaries of POIs and the surrounding population. By contrasting this method with traditional approaches, it underscores the impact of using a weighted distribution of resources and demand compared to uniform weighting.

The approach outlined in de Klerk and Fabris-Rotelli (2024) however primarily focuses on the structural connections within the network and does not incorporate attributes such as environmental or topographic factors that, while not explicitly represented in structural links, still connect nodes through shared properties. To address this limitation, we expand on the method by developing an attribute based fuzzy lattice data structure that integrates both the network's link structure and node attributes (Zhou et al., 2009) to iteratively propagate labels. Using a semi-supervised learning approach to propagate labels to neighbouring nodes, the entire geographical region can be represented as an attribute based fuzzy lattice structure, with each node having a set of probabilities associated with various labels. This allows unlabelled nodes to retain multiple labels received from their neighbours, reflecting overlapping catchment areas. The degree of this overlap is captured probabilistically, allowing to create overlapping catchment areas. Using a real-world application in healthcare resource allocation across seasons, we compare seasonal catchment areas to static catchment areas created with a fuzzy lattice data structure. This comparison highlights how accessibility and supply-demand ratios fluctuate seasonally, offering insights into the dynamic nature of catchment areas in regions with changing environmental conditions.

The remainder of this paper is structured in the following order. Section 2 covers the methodology, where we define an attribute augmented graph and the transition probability matrix which is then used to create an attribute based fuzzy lattice data structure and seasonal catchment areas. In Section 3, we apply this methodology in a practical scenario, demonstrating how the integration of attributes can influence demand patterns and the supply-demand ratio. This is followed by a discussion and conclusion, offering final remarks and insights on the analysis.

## 2. Methodology

This section covers the fundamentals of a network graph, including nodes, edges, and weights, and introduces a set of attributes associated with the nodes in a graph  $G$ . In Section 2.1, we define an attribute-augmented graph  $G_a$  that incorporates these attributes. Section 2.2 proposes an attribute-based fuzzy lattice data structure that leverages both the link structure and node attributes to iteratively propagate labels. Finally, Section 2.3 applies this attribute based fuzzy lattice to a simple toy example, illustrating how the supply-demand ratio and accessibility are impacted.

### 2.1 Attribute Augmented Graph

Traditional graph-based models primarily rely on structural connectivity, where in a geographical application nodes represent spatial locations (grid cells, administrative units, etc.), and edges define adjacency relationships. However, this representation fails to capture meaningful similarities between regions that share common attributes but are not directly connected. This limitation makes it difficult

to model complex spatial patterns where both proximity and attribute similarity play a role. To address this, we expand on the attribute augmented graph as discussed in Zhou et al. (2009) by integrating it in a geographical application to consider both geographical locations as well as shared characteristics between regions.

Using node classification formulation as illustrated in Bhagat et al. (2011), consider a graph  $G = G(V, E, W)$ , with  $V$  the set of nodes,  $E$  the set of edges and  $W$  the edge weights. Nodes typically represent entities such as individuals in a social network and edges represent the relationships between these entities, such as friendships, or interactions. In this paper we focus on a spatial graph, that is, the nodes represent geographical areas.

A structural node refers to the position of a node within the graph, independent of any specific attributes, and represents geographic regions such as delineated grid cells, whose centres then represent the nodes. These nodes are connected by edges that denote spatial adjacency, meaning that a connection (edge) exists between two nodes if they share a border or are within a predefined distance threshold. The structure of the graph itself provides information about the network, such as how densely connected certain nodes are, adjacent nodes and how central a node is in comparison to the other nodes. The link structure refers to how the nodes are connected through the edges and can be quantified in a statistical manner by the transition probability from one node to the other. Herein, the link structure uses label propagation such as in Raghavan et al. (2007) to propagate labels across the network to unlabelled nodes, so that structurally close nodes are more likely to share similar labels.

In addition to the link structure in  $G$ , consider a set of attributes  $\Lambda$  which is associated with the nodes  $V$  in  $G$ . Attribute nodes are added to represent specific characteristics or properties of the nodes in a network. For instance, in a co-authorship network, an attribute node might represent a research topic. If multiple researchers (nodes) publish in the same domain, they will be connected to this common attribute node, creating additional links beyond the standard co-authorship connections as illustrated in Zhou et al. (2009). In a geographical application, structural nodes such as regions can be connected not only by shared boundaries but also by common attributes. These attributes may include population size, primary road access, average house price, and other socio-economic factors. Such connections extend beyond geographic adjacency, capturing relationships between regions based on shared characteristics. These nodes are different from structural nodes, which reflect the network's topology. In an augmented graph  $G_a$ , attribute nodes are added to express similarity based on shared properties. Connections, or attribute edges, are drawn between structural nodes and attribute nodes when the structural nodes share a similar attribute.

Incorporating the attribute nodes into the connectivity between nodes, let  $G_a = (V \cup V_a, E \cup E_a, \mathbf{W}_a^*)$  be an attribute augmented graph as defined in Zhou et al. (2009) with  $V$  the structural nodes,  $V_a$  the attribute nodes,  $E$  the structural edges,  $E_a$  the attribute edges and  $\mathbf{W}_a^*$  the attribute and structural weights. The attribute augmented graph  $G_a = (V \cup V_a, E \cup E_a, \mathbf{W}_a^*)$  is constructed using the union operator ( $\cup$ ) to integrate attribute nodes and attribute edges, where  $V \cup V_a$  is the combined node set of both attribute and structural nodes, and  $E \cup E_a$  is the combined set of edges which expands the set of structural edges by introducing attribute edges that connect attribute nodes to structural nodes.  $\mathbf{W}_a^*$  is the weights matrix which includes both structural weights and attribute weights.

Let the structural nodes  $V = \{v_1, \dots, v_N\}$  represent the full geographical region which is subdivided

into  $N$  smaller grid cells, whose centres then represent the node locations. A structured edge  $(v_i, v_j) \in E$  iff structure node  $v_i$  shares a line segment, as defined by Rook's contiguity (Wang, 2014), with structure node  $v_j$  in the geographical area. Rook's contiguity is similar to the movement of the rook in chess, which can move horizontally or vertically across the board but not diagonally. In a grid layout, this means each cell is adjacent to its immediate neighbours and can be connected to a maximum of four neighbours: up, down, left, and right, forming what is known as a Rook's neighbourhood (Ver Hoef et al., 2018). Let the nodes in  $V$  be associated with the attributes  $\Lambda = \{\lambda_1, \dots, \lambda_A\}$ . If each attribute  $\lambda_i \in \Lambda$  has  $k_i$  levels,  $i \in \{1, \dots, A\}$ , then let  $A = \{a_1, \dots, a_T\}$  be the  $T$ -dimensional set of indicator attributes that is associated with each level for attribute  $\lambda_i \in \Lambda$ , where  $T = \sum_{i=1}^A k_i$ . Let  $V_a = \{v_{a_1}, \dots, v_{a_T}\}$  be the set of indicator attribute nodes. The size of node set  $V \cup V_a$  is now  $N + T$ .

The indicator attribute information for structural node  $v_i$  is regarded as a  $T$ -dimensional binary vector,  $\mathbf{x}_i$ , and can be represented in an attribute matrix  $\mathbf{X}$  as

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{pmatrix} = \begin{pmatrix} a_1(v_1) & a_2(v_1) & \dots & a_T(v_1) \\ a_1(v_2) & a_2(v_2) & \dots & a_T(v_2) \\ \vdots & \vdots & \ddots & \vdots \\ a_1(v_n) & a_2(v_n) & \dots & a_T(v_n) \end{pmatrix},$$

where  $a_j(v_i) = 1$  indicates an attribute edge in  $E_a$  if node  $v_i$  is associated with indicator attribute  $a_j$ , else  $a_j(v_i) = 0$  (Lin et al., 2021). The structural and attribute weights of  $G_a$  is given by

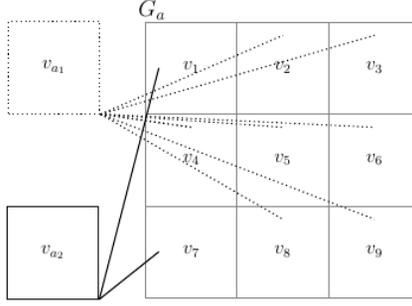
$$\mathbf{W}_a^* = \begin{pmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_a \end{pmatrix},$$

with  $\mathbf{W}$  an  $N \times N$  matrix indicating the structure edge weights  $w_{ij}$  between structure nodes  $v_i$  and  $v_j$  and  $\mathbf{W}_a$  is a  $T \times T$  diagonal matrix consisting of the attribute weights  $w_{a_i}$ ,  $i = 1, \dots, T$ . Weights represent the relative importance of connections between nodes in the graph, where heavily weighted edges encourage strong connections. The structure edge weights reflect the importance of shared boundaries between structural nodes, whereas the attribute weights quantify the importance or influence of a shared attribute between structural nodes. If attribute connections are more important (similar economic profiles, income levels, etc.), attribute edges will receive larger weights than structural edges, whereas if shared boundaries are more significant (natural boundaries, access to a service is explicitly limited to a designated neighbourhood, etc.), structural connections will have larger weights than attribute connections. If both structural and attribute connections carry equal importance, they will be assigned equal weights to ensure a balanced influence. An edge weight can either be relative to a quantity, for example in a social network the number of interactions (messages, common friends, etc.) among users, or it can simply be set to 1 to indicate that a link is present, as discussed in Bhagat et al. (2011).

When considering an attribute augmented graph structure, not only can a node transition via a structural edge, but also via an attribute edge. Consider the  $(N + T) \times (N + T)$  transition probability matrix  $\mathbf{P}_a$  of  $G_a$  constructed as

$$\mathbf{P}_a = \begin{pmatrix} \mathbf{P}_s & \mathbf{P}_{sa} \\ \mathbf{P}_{as} & \mathbf{0} \end{pmatrix}, \quad (1)$$

where  $\mathbf{P}_s$  is an  $N \times N$  submatrix containing the transition probabilities for structural node  $v_i$  to reach



**Figure 1.** Graph  $G_a$  with structural nodes  $v_1, \dots, v_9$  and attribute nodes  $v_{a_1}$  and  $v_{a_2}$

structural node  $v_j$ ,

$$p_{v_i, v_j} = \begin{cases} \frac{w_{ij}}{w_i + \sum_{i=1}^T w_{a_i}}, & \text{if } (v_i, v_j) \in E, \\ 0, & \text{otherwise.} \end{cases}$$

$\mathbf{P}_{sa}$  is an  $N \times T$  submatrix containing the transition probabilities from a structural node  $v_i$  to an attribute node  $v_q$ ,

$$p_{v_i, v_q} = \begin{cases} \frac{w_q}{w_i + \sum_{i=1}^T w_{a_i}}, & \text{if } (v_i, v_q) \in E_a, \\ 0, & \text{otherwise.} \end{cases}$$

$\mathbf{P}_{as}$  is a  $T \times N$  transition probability submatrix for an attribute node  $v_p$  to reach a structural node  $v_j$ ,

$$p_{v_p, v_j} = \begin{cases} \frac{1}{N(v_p)}, & \text{if } (v_p, v_j) \in E_a, \\ 0, & \text{otherwise,} \end{cases}$$

and finally a  $T \times T$  submatrix  $\mathbf{0}$  for the transition probability from an attribute node  $v_p$  to an attribute node  $v_q$ ,

$$p_{v_p, v_q} = 0, \quad \forall v_p, v_q \in V_a,$$

with  $w_i$  the sum of edge weights for structural node  $v_i$ ,  $N(v_p)$  the number of neighbours to node  $v_p$  and  $p, q \in \{a_1, \dots, a_T\}$  as defined in Lin et al. (2021). Transition probabilities determine the likelihood of a label propagating from one node to another in a graph-based label propagation algorithm. They define how information (such as POI assignments) spreads through the network based on structural and attribute connections and their weights.

Consider a simple example illustrated in Figure 1 with structural nodes  $V = \{v_1, \dots, v_9\}$  such that  $N = 9$ . Each of the nodes represents a geographical grid with location based attributes captured in  $\Lambda$ . For this example, let the nodes be associated with only one attribute  $\lambda_1$  which consists of two levels  $a_1$  and  $a_2$ , i.e.  $A = 1, k_1 = 2$  and  $T = 2$ . Assume all structure and attribute weights in  $\mathbf{W}_a^*$  are 1. In this case then  $\mathbf{W}$  will simply be an adjacency matrix that defines structural neighbouring nodes using shared lines between grids. The set of indicator nodes for whether a node is associated with  $a_1$  or  $a_2$  is contained in the set  $V_a$ . The transition probability matrix  $\mathbf{P}_a$  for  $G_a$  is an  $11 \times 11$  matrix and can be illustrated in (2).

$$\mathbf{P}_a = \begin{pmatrix} 0 & 1/3 & 0 & 1/3 & \cdots & 0 & | & 0 & 1/3 \\ 1/4 & 0 & 1/4 & 0 & \cdots & 0 & | & 1/4 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & | & 1/3 & 0 \\ \text{---} & \text{---} \\ 0 & 1/7 & 1/7 & 1/7 & \cdots & 1/7 & | & 0 & 0 \\ 1/2 & 0 & 0 & 0 & \cdots & 0 & | & 0 & 0 \end{pmatrix}. \quad (2)$$

The transition from a structural node to a structural node is contained in the submatrix  $\mathbf{P}_s$  and from (2) one can note that since node  $v_1$  is only structural neighbours with nodes  $v_2$  and  $v_4$ , there is an assigned probability of  $1/3$  for moving from  $v_1$  to  $v_2$  or  $v_4$ . There is an additional transition probability of  $1/3$  associated with node  $v_1$  contained in the submatrix  $\mathbf{P}_{sa}$  which is the probability for structural node  $v_1$  to move to attribute node  $v_{a_2}$ , hence the sum over all rows still adds up to one and  $\mathbf{P}_a$  retains its stochastic matrix properties. The transition from the newly created attribute nodes, namely  $v_{a_1}$  and  $v_{a_2}$ , can be observed in the submatrix  $\mathbf{P}_{as}$  where the attribute node  $v_{a_1}$  shares a transition probability of  $1/7$  with the structural nodes  $v_2, v_3, v_4, v_5, v_6, v_8$  and  $v_9$ . The attribute node  $v_{a_2}$  shares a transition probability of  $1/2$  with structural nodes  $v_1$  and  $v_7$ . The probability to transition from an attribute node to an attribute node is 0 and is captured in the  $2 \times 2$  submatrix  $\mathbf{0}$ .

## 2.2 Attribute Based Fuzzy Lattice Data

In any network, if a set of labels is known for some nodes, label propagation allows us to predict labels for unmarked nodes by using their connections to the labelled ones. This helps assign values, to nodes where data are missing, ensuring that all nodes are considered based on their network links. It is a powerful way to fill missing information and make informed decisions across the entire network.

Using a semi-supervised learning approach to propagate labels to neighbouring nodes, the full geographical region can be represented as an attribute based fuzzy lattice structure, with each node having a set of probabilities associated with each label. As discussed in de Klerk and Fabris-Rotelli (2024), a fuzzy lattice data structure uses a random walk approach and label propagation to propagate labels to all nodes. A fuzzy lattice data structure only considers the graph's link structure to propagate labels, focusing on how nodes are connected and share common neighbours. However, nodes often have various attributes such as shared interests when examining social networks, or shared geographical features like weather patterns, topography, etc., when analysing geographical data (Zhou et al., 2009). An attribute based fuzzy lattice data structure is proposed herein using both the link structure and attributes associated with nodes to propagate labels. An unlabelled node can hold multiple labels received from neighbouring nodes, allowing for an overlap of catchment areas. The degree of overlap is captured by a probability and hence creates a fuzzy lattice data structure (de Klerk and Fabris-Rotelli, 2024).

Let  $\mathcal{Y}$  be the full set of  $M$  possible labels that is initially applied to the labelled nodes in the network. The foundation of label propagation is to propagate the labels in  $\mathcal{Y}$  to the remaining unlabelled nodes based on how nodes are connected to each other. A label can indicate a list of interests (books, movies, etc.) and the shared connections propagate these labels to the unlabelled nodes in the graph, predicting who shares similar interests. In a geographical application a node

represents a region with an edge indicating neighbouring areas and a label is a point of interest (POI) which an area is most likely to be associated with.

Consider geographical spatial POIs  $P = \{P_1, P_2, \dots, P_M\}$  as a known set of  $M$  locations. The centre of each grid cell  $v_n$ ,  $n = 1, \dots, N$  will represent the structural nodes in graph  $G_a$ . The grid size should be determined based on the specific geographical region of the study. Let  $V_m = \{v_1, \dots, v_M\}$  be the subset of the  $M$  labelled structural nodes which contains a POI and let  $V_u = \{v_{M+1}, \dots, v_N\}$  be the subset of remaining  $N - M$  initially unlabelled structural nodes which doesn't contain a POI.

Let all attribute nodes contained in set  $V_a$  initially be unlabelled nodes. Then the set of all unlabelled (both structural and attribute) nodes are contained in the ordered set  $V_u^* = V_u \cup V_a = \{v_{M+1}, \dots, v_N, v_{a_1}, \dots, v_{a_T}\}$ . Let  $V$  be ordered such that the first  $M$  rows are the cells from  $V_m$  and the remaining  $N - M + T$  rows are the nodes from  $V_u^*$  such that  $V = V_m \cup V_u^* = \{v_1, \dots, v_M, v_{M+1}, \dots, v_N, v_{a_1}, \dots, v_{a_T}\}$ .

Let  $\mathcal{Y}$  be the full set of  $M$  possible labels and  $\mathbf{Y}_m$  be an  $M \times M$  indicator matrix carrying the initial  $M$  multi-class labels indicating the POI associated with the corresponding node in set  $V_m$ . Similarly let  $\mathbf{Y}_u^*$  be an  $(N - M + T) \times M$  matrix indicating the POI associated with corresponding nodes in  $V_u^*$ . The initial label matrix  $\mathbf{Y}$  is an  $(N + T) \times M$  matrix with the first  $M$  rows as  $\mathbf{Y}_m$  and the remaining  $N - M + T$  rows as  $\mathbf{Y}_u^*$  or 0 (as the nodes in  $V_u^*$  are initially unlabelled).

Assume that all unlabelled nodes can reach a labelled node in a finite number of steps and will therefore have an associated label at the end of the iteration process, i.e. graph  $G$  is label connected (Azran, 2007). A graph is considered non-label connected when some nodes remain unlabelled after the iteration process. This can occur when certain nodes are isolated, having no direct edges linking to the rest of graph  $G$ . In a social network, this can happen if there are individuals who do not share any common interests with the rest of the network. When considering geographical applications, a graph may be non-label connected when some grid cells are disconnected from the rest, separated by natural or man-made barriers such as rivers, mountains, or boundaries like regional or state borders.

Consider transition matrix  $\mathbf{P}_a$  as defined in (1). Using a random walk approach with  $\mathbf{P}_a^t$  the corresponding matrix at time  $t$ , then  $\mathbf{P}_a^{t \rightarrow \infty}$  indicates the steady-state distribution matrix at which all nodes are associated with label  $c \in \mathcal{Y}$  as  $t \rightarrow \infty$  as shown in Bhagat et al. (2011). In matrix form

$$\hat{\mathbf{Y}} = \mathbf{P}_a^{t \rightarrow \infty} \mathbf{Y}, \quad (3)$$

where

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{\mathbf{Y}}_m \\ \hat{\mathbf{Y}}_u^* \end{bmatrix} \text{ and } \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_m \\ \mathbf{0} \end{bmatrix}, \quad (4)$$

and  $\hat{\mathbf{Y}}$  contains the output labels from the converged iteration process for all nodes in set  $V$ .

Labelled nodes in set  $V_m$  are classified as absorbent states such that they exhibit probability 1 of staying in the same node and probability 0 of leaving the node. Therefore the labels of all nodes  $v_n \in V_m$  do not change. Since the nodes are ordered in such a way that the first  $M$  rows in  $V$  are  $V_m$  and the last  $N - M + T$  rows are  $V_u^*$ , the transition matrix can be split into 4 submatrices, indicating the probability to move between states, namely

$$\mathbf{P}_a = \begin{pmatrix} \mathbf{P}_{mm} & \mathbf{P}_{mu} \\ \mathbf{P}_{um} & \mathbf{P}_{uu} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{P}_{um} & \mathbf{P}_{uu} \end{pmatrix}.$$

The probabilities of transitioning from the unlabelled states to the labelled states are captured in the  $(N - M + T) \times M$  matrix  $\mathbf{P}_{um}$ , and from the unlabelled states to the unlabelled states are captured in the  $(N - M + T) \times (N - M + T)$  matrix  $\mathbf{P}_{uu}$ . Since all labelled nodes are defined as absorbent states, the probability of staying in a labelled node is 1 and the probability of exiting a labelled node is 0. This simplifies  $\mathbf{P}_{mm}$  to an  $M \times M$  identity matrix  $\mathbf{I}$ , and  $\mathbf{P}_{mu}$  to an  $M \times (N - M + T)$  zero matrix  $\mathbf{0}$ .

Since the graph is label connected and there are  $M$  absorbent states, the limiting distribution of  $\lim_{t \rightarrow \infty} \mathbf{P}_a^t$  is

$$\mathbf{P}_a^{t \rightarrow \infty} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ (1 - \mathbf{P}_{uu})^{-1} \mathbf{P}_{um} & \mathbf{0} \end{pmatrix} \quad (5)$$

as shown in Bhagat et al. (2011) and Azran (2007). Substituting (5) and (4) into (3), the labels can be computed as

$$\begin{bmatrix} \hat{\mathbf{Y}}_m \\ \hat{\mathbf{Y}}_u^* \end{bmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ (1 - \mathbf{P}_{uu})^{-1} \mathbf{P}_{um} & \mathbf{0} \end{pmatrix} \begin{bmatrix} \mathbf{Y}_m \\ \mathbf{0} \end{bmatrix}. \quad (6)$$

From (6), the labels for the absorbent state nodes contained in  $V_m$  remain unchanged with  $\hat{\mathbf{Y}}_m = \mathbf{Y}_m$ . The labels of the unlabelled nodes are obtained by computing

$$\hat{\mathbf{Y}}_u^* = (1 - \mathbf{P}_{uu})^{-1} \mathbf{P}_{um} \mathbf{Y}_m. \quad (7)$$

The  $(N - M + T) \times M$  matrix  $\hat{\mathbf{Y}}_u^*$  will contain the probability of each label assigned to the nodes in  $V_u^*$  as outlined in Algorithm 1, with the sum over each row adding to 1.

Consider the simple example illustrated in Figure 1 with added labelled nodes. In this example seven structural nodes  $V_u = \{v_1, v_3, v_4, v_5, v_6, v_8, v_9\}$  and two attribute nodes  $V_a = \{v_{a_1}, v_{a_2}\}$  are unlabelled and contained in the set  $V_u^* = V_u \cup V_a$ . Two structural nodes  $V_m = \{v_2, v_7\}$  are labelled with associated POIs  $P_1$  and  $P_2$ , respectively. Let  $V$  be the ordered set, contained in  $G_a$ , consisting of  $N + T = 11$  nodes with the first  $M = 2$  rows the labelled nodes and the remaining  $N - M + T = 9$  rows the unlabelled nodes, i.e.  $V = \{v_2, v_7, v_1, v_3, v_4, v_5, v_6, v_8, v_9, v_{a_1}, v_{a_2}\}$ . Let  $\mathbf{Y}_m$  be an indicator matrix with the first column representing label  $P_1$  and the second column representing label  $P_2$ . Applying the methodology as described above, the probability of labels  $P_1$  and  $P_2$  to be assigned to nodes in set  $V$  is calculated in (8).

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{\mathbf{Y}}_m \\ \hat{\mathbf{Y}}_u^* \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0.59 & 0.79 & 0.48 & 0.67 & 0.69 & 0.49 & 0.62 & 0.68 & 0.30 \\ 0 & 1 & 0.41 & 0.21 & 0.52 & 0.33 & 0.31 & 0.51 & 0.38 & 0.32 & 0.70 \end{bmatrix}^T. \quad (8)$$

From (8) it can be noted that all nodes in  $V_m$  remained in the initial node and label associated with it with probability 1. The nodes contained in set  $V_u^*$ , however, received a probability associated with labels  $P_1$  and  $P_2$ . If ignoring the attribute nodes as done in de Klerk and Fabris-Rotelli (2024), the probability for node  $v_1$  to be associated with label  $P_1$  is 0.72 and with  $P_2$  is 0.28. In this example, however, nodes  $v_1$  and node  $v_7$ , which is the node associated with label  $P_2$ , share a similar attribute  $a_2$ , hence even if they are not as structurally close as  $v_1$  and  $v_2$ , the node associated with label  $P_2$ , they now share a connected link through node  $v_{a_2}$ . This can also be observed in the increased probability for node  $v_1$  to be associated with  $P_2$  from 0.28 to 0.41.

The attribute nodes also carry a probability to be associated with labels  $P_1$  and  $P_2$ . In a label propagation model, the labels are typically spread across the graph using the connections between

---

**Algorithm 1** Attribute Based Fuzzy Lattice Data
 

---

**Require:**  $V \cup V_a$ : Combined node set (size  $N + T$ ) of structural nodes  $V$  and attribute nodes  $V_a$ .

**Require:**  $V_m$ : Set of  $M$  labelled nodes (absorbent states).

**Require:**  $E \cup E_a$ : Combined edge set (size  $N + T$ ) of structure- ( $E$ ) and attribute-edge weights ( $E_a$ ).

**Require:**  $\mathbf{W}_a^*$ : Weights matrix of size  $(N + T) \times (N + T)$ .

**Require:**  $\mathbf{X}$ : Attribute matrix of size  $N \times T$ .

**Require:**  $\mathbf{Y}_m$ : Indicator matrix for all labelled states of size  $M \times M$ .

**Ensure:**  $\hat{\mathbf{Y}}$ : Matrix with output labels from the converged iteration process for all nodes in combined set  $V \cup V_a$  of size  $(N + T) \times M$ .

- 1: **Initialise graph:** Define  $G_a = (V \cup V_a, E \cup E_a, \mathbf{W}_a^*)$ .
- 2: **Compute transition probability matrix  $\mathbf{P}_a$ :** by computing submatrices  $\mathbf{P}_s, \mathbf{P}_{sa}, \mathbf{P}_{as}$  with structural weights  $w_{ij} \in \mathbf{W}$  and attribute weights  $w_{ai} \in \mathbf{W}_a$  as outlined in (1):

$$\mathbf{P}_a = \begin{bmatrix} \mathbf{P}_s & \mathbf{P}_{sa} \\ \mathbf{P}_{as} & \mathbf{0} \end{bmatrix}.$$

- 3: **Subdivide matrix  $\mathbf{P}_a$  into labelled and unlabelled sub matrices:**

$$\mathbf{P}_a = \begin{bmatrix} \mathbf{P}_{mm} & \mathbf{P}_{mu} \\ \mathbf{P}_{um} & \mathbf{P}_{uu} \end{bmatrix},$$

$\mathbf{P}_{mm} = \{\mathbf{P}_a[i, j] \mid 1 \leq i \leq M; 1 \leq j \leq M\} = \mathbf{I}$  as an  $M \times M$  identity matrix,

$\mathbf{P}_{um} = \{\mathbf{P}_a[i, j] \mid M + 1 \leq i \leq N + T; 1 \leq j \leq M\}$  as an  $(N - M + T) \times M$  matrix,

$\mathbf{P}_{mu} = \{\mathbf{P}_a[i, j] \mid 1 \leq i \leq M; M + 1 \leq j \leq N + T\} = \mathbf{0}$  as an  $M \times (N - M + T)$  zero matrix,

$\mathbf{P}_{uu} = \{\mathbf{P}_a[i, j] \mid M + 1 \leq i \leq N + T; M + 1 \leq j \leq N + T\}$

as an  $(N - M + T) \times (N - M + T)$  matrix.

- 4: **Compute labels for absorbent and unlabelled states:**

$$\hat{\mathbf{Y}}_m = \mathbf{Y}_m,$$

$$\hat{\mathbf{Y}}_u = (\mathbf{I} - \mathbf{P}_{uu})^{-1} \mathbf{P}_{um} \mathbf{Y}_m.$$

- 5: **Return computed labels:** Return  $\hat{\mathbf{Y}} = \begin{bmatrix} \hat{\mathbf{Y}}_m \\ \hat{\mathbf{Y}}_u \end{bmatrix}$ .
-

structural nodes. When attribute nodes are added they act as additional connections related to the structural nodes. By associating a probability with each attribute node to a label, the model is indicating how likely that attribute is to be connected to a particular label through the structural nodes. This probability reflects the attribute's influence or relationship with the label, as mediated by its connections to the structural nodes. Attribute node  $v_{a_2}$  carries a higher probability to be associated with label  $P_2$  than with  $P_1$  since the nodes which are associated with attribute  $a_2$  are closer in connectivity to node  $v_7$ , which is associated with label  $P_2$ , and similar for attribute node  $v_{a_1}$  to label  $P_1$ .

### 2.3 Seasonal Catchment Areas

Assuming that  $G$  is a labelled connected graph, each node receives a set of probabilities corresponding to each label in the set  $\mathcal{Y}$ . As discussed in de Klerk and Fabris-Rotelli (2024), the proposed fuzzy lattice data structure enables the identification of realistic catchment areas by applying drive-time thresholds to the structural nodes. This approach involves overlaying polygons representing POIs and ensuring that only nodes which fall within the drive-time polygons are considered when assigning probabilities of association with a POI.

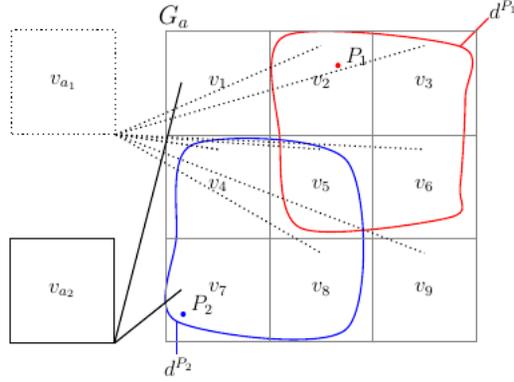
Consider the set of POIs  $P = \{P_1, P_2, \dots, P_M\}$  with an associated drive-time threshold distance  $d^{P_m}$  assigned to each  $P_m, m = 1, \dots, M$ . The threshold drive-time distance  $d^{P_m}$  depends on the order of the POIs and will be denoted by  $d_{low}^{P_m}, d_{mid}^{P_m}$  or  $d_{high}^{P_m}$  for low-, middle- or high-order POIs, respectively, as discussed in Green et al. (2016). The highest drive-time threshold considered is at 15 minutes based on the fifteen-minute city concept (Pozoukidou and Angelidou, 2022). For each POI, we identify all structural nodes  $v_n \in V$  which are within a threshold drive-time distance  $d^{P_m}$  of  $P_m$ .

Let  $G_{P_m}$  define the region such that  $G_{P_m} = \bigcup_n v_n \ni d_{v_n, P_m} \leq d^{P_m}, G_{P_m} \subseteq G$  with  $d_{v_n, P_m}$  the drive-time distance between the nearest network point of structural nodes  $v_n \in V$  and  $P_m$ . Let  $O_{v_n} = \{m : v_n \in G_{P_m}\}$  define the indices of POI  $P_m$ , that is snapped to the same network as  $v_n \in V$ . Nodes that are contained in more than one  $G_{P_m}$  have an overlap of accessibility to different POIs. There can, however, exist nodes which are not contained in any set of  $G_{P_m}$ , resulting in spatially disjoint areas.

Since catchment areas can only be applied to structural nodes,  $v_n \in V$ , only the first  $N - M$  rows in  $\hat{\mathbf{Y}}_u^*$ , i.e.  $\hat{\mathbf{Y}}_u = \{\hat{\mathbf{Y}}_u^*[i, :] \mid 1 \leq i \leq N - M\}$ , are used in the drive-time threshold calculations. Let all structural nodes which fall beyond the threshold drive-time distance  $d^{P_m}$  for each  $P_m$  be nullified in  $\hat{\mathbf{Y}}_u$  i.e. the probability for a node to be assigned to a label which falls beyond the drive-time threshold is 0. If a region falls beyond the drive-time threshold for all POIs, the region is spatially inaccessible (disjoint) and will be assigned a probability of 0 for all labels  $c \in \mathcal{Y}$ . The matrix  $\hat{\mathbf{Y}}_u$  is row standardised to ensure that all rows adds up to 1.

Consider the example provided in Figure 2 with drive-time thresholds for POIs  $P_1$  and  $P_2$ . Suppose the values for  $d^{P_1}$  and  $d^{P_2}$  are such that  $G_{P_1} = \{v_2, v_3, v_5, v_6\}, G_{P_2} = \{v_4, v_5, v_7, v_8\}, V_m = \{v_2, v_7\}$  and regions  $v_1$  and  $v_9$  are spatially disjoint. Let all structural nodes which fall beyond drive-time threshold  $d^{P_1}$  for catchment area of  $P_1$  carry a 0 probability to be assigned to label  $P_1$ , and similarly  $P_2$ . After the matrix  $\hat{\mathbf{Y}}_u$  has been row standardised, the results from (8) with drive-time thresholds applied are

$$\hat{\mathbf{Y}}_u = \begin{bmatrix} 0 & 1 & 0 & 0.67 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.33 & 0 & 1 & 0 \end{bmatrix}^T. \quad (9)$$



**Figure 2.** Graph  $G_a$  with structural nodes  $v_1, \dots, v_9$  and attribute nodes  $v_{a_1}$  and  $v_{a_2}$ , POIs  $P_1$  and  $P_2$ , and corresponding drive-time thresholds  $d^{P_1}$  and  $d^{P_2}$ .

It can be noted that, after standardisation, each node is only associated with the POI falling within the same network. Node  $v_5$  has an overlap of drive-time thresholds  $d^{P_1}$  and  $d^{P_2}$  and hence retains the probability of being assigned both label  $P_1$  and  $P_2$ . Nodes  $v_1$  and  $v_9$  fall beyond the drive-time thresholds for all POIs and are therefore spatially disjoint with probability 0 to be assigned to label  $c \in \mathcal{Y}$ .

## 2.4 Measures of Spatial Accessibility

In this section, we apply supply-demand ratios and accessibility measures to demonstrate the distribution of resources when using attribute based fuzzy lattice data. Supply may refer to a provider's capacity, such as the number of available hospital beds (Challen et al., 2022), the quantity of providers like primary care physicians in a given region (Luo and Wang, 2003; Wang, 2014), or the resources offered by a provider, such as a doctor's available services or stock availability at a store, as in Shao and Luo (2022). Demand typically represents the population and communities in need of the specified services or products.

Let all POIs  $P_m \in P$  have supply size (capacity) of  $S(P_m)$  and all nodes  $v_n \in V$  have an associated demand (population size) of  $D(v_n)$  as formulated in Challen et al. (2022). Then the supply-demand ratio for supplier  $m$  in region  $G_{P_m}$  can be expressed by

$$R_{P_m} = \frac{S(P_m)}{D(G_{P_m})}, \quad (10)$$

and the accessibility for all structural nodes  $v_n \in V$  is

$$A_{v_n} = \sum_{m \in O_{v_n}} R_{P_m} = \sum_{m \in O_{v_n}} \frac{S(P_m)}{D(G_{P_m})}, \quad (11)$$

as illustrated by Luo and Wang (2003) and Wang (2014).

Consider the example provided in Figure 2 with  $N = 9$  grids and POIs  $P_1$  and  $P_2$ . Suppose the demand for all regions are 1, i.e.  $D(v_n) = 1, n = 1, \dots, 9$ , and supply for all suppliers are 4,  $S(P_m) = 4, m = 1, 2$ . Suppose the values for  $d^{P_1}$  and  $d^{P_2}$  are such that  $G_{P_1} = \{v_2, v_3, v_5, v_6\}$ ,

$G_{P_2} = \{v_4, v_5, v_7, v_8\}$ ,  $V_m = \{v_2, v_7\}$  and regions  $v_1$  and  $v_9$  are spatially disjoint. Using attribute based fuzzy lattice catchment areas to determine accessibility and supply-demand ratio using a probabilistic approach, the demand on catchment area level will change according to the weight assigned by  $\hat{Y}_u$ . Let the demand be proportionally assigned to each supplier's catchment area using the probabilities in (9). Then  $D(G_{P_1}) = 3.67$  and  $D(G_{P_2}) = 3.33$  which results in the supply-demand ratio for  $P_1$  to change to  $R_{P_1} = \frac{S(P_1)}{D(G_{P_1})} = 1.09$  and  $P_2$  to  $R_{P_2} = \frac{S(P_2)}{D(G_{P_2})} = 1.2$ . Since region  $v_5$  still has access to both  $P_1$  and  $P_2$ , the accessibility of  $v_5$  is  $A_{v_5} = R_{P_1} + R_{P_2} = 2.288$ .

Comparing this to the same example in de Klerk and Fabris-Rotelli (2024), the demand in region  $G_{P_1}$  has increased which in turn has decreased the supply-demand ratio for  $P_1$ . The only overlapping node  $v_5$  has a higher probability to be associated with label  $P_1$  because of the shared attribute node  $v_{a_1}$  with structural nodes  $v_2, v_3, v_4, v_5, v_6, v_8$  and  $v_9$ . The accessibility for node  $v_5$  remains unchanged whether fuzzy lattice catchment areas or attribute based fuzzy lattice catchment areas are used because even though the demand is proportionally different, the sum over supply-demand ratios remains the same. The decreased supply-demand ratio for  $P_1$  indicates that supplier  $P_1$  is under greater pressure to provide the same level of supply while facing increased demand. The attribute based fuzzy lattice data offers the added advantage of incorporating factors that influence both demand and supply, rather than relying solely on the proximity of a community to a POI. This allows for a more comprehensive analysis, accounting for attributes beyond just geographical distance. This will be illustrated with an example in the next section by not only considering the link structure but also underlying attributes such as ease of accessibility and seasonal characteristics.

### 3. Application

To test the proposed methodology on real-world applications, the population size and sales of one of South Africa's leading pharmaceutical chains (with close to 880 stores located in South Africa) were considered. The supply allocated to each of the stores is determined by sales and demand by population size at an EA<sup>1</sup> level.

Let region  $G$  be the district municipality City of Cape Town<sup>2</sup> located in the Western Cape province in South Africa. Region  $G$  is subdivided into 592 square grids of size 2km  $\times$  2km, which is the optimal grid size to ensure adequate variation in landscape features and changes across the nodes.

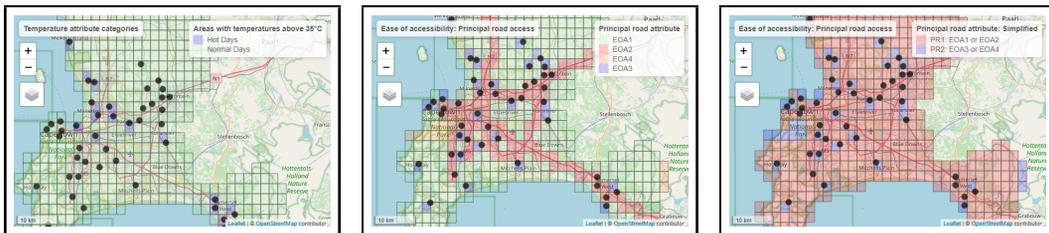
In addition to the link structure, a set of geographical attributes assigned to all nodes in  $G$  will be considered. Seasonal attributes such as extreme temperature and rainfall patterns are considered since gastrointestinal illnesses increase at higher temperatures or rainfall shortages as highlighted by the Department of Health South Africa<sup>3</sup>. Research from Musengimana et al. (2016) also indicates that high temperatures can promote the growth of bacteria, viruses, and parasites in food and water which increases the risk of waterborne diseases like diarrhoea.

Daily minimum and maximum temperatures as well as rainfall for 86 weather stations in the

<sup>1</sup>An enumerations area (EA) is the smallest geographical unit (piece of land) into which the country is divided for enumeration purposes. Enumeration areas contain between 100 to 250 households. Source: Stats SA.

<sup>2</sup>District municipalities consist of multiple local municipalities. They are administrative divisions which are accountable for providing basic services within the area. Source: Education and Training Unit (ETU), [www.etu.org.za](http://www.etu.org.za) (accessed February 15, 2024)

<sup>3</sup>Department of Health South Africa, *NATIONAL HEAT HEALTH ACTION GUIDELINES: Guide to extreme heat planning in South Africa for the human health sector*, <https://www.health.gov.za/> (accessed August 20, 2024)



**Figure 3.** Temperature attribute indicating "hot days" and ease of access and accessibility to principal roads. Map data © OpenStreetMap contributors, under the Open Database License (ODbL).

Western Cape were provided by SAWS<sup>4</sup>. The average rainfall, minimum and maximum temperatures from 2016 to 2019 for the four different seasons were calculated. In the grids which have no overlap with a SAWS weather station, ordinary Kriging was used to estimate the average rainfall, minimum and maximum temperatures for the different seasons.

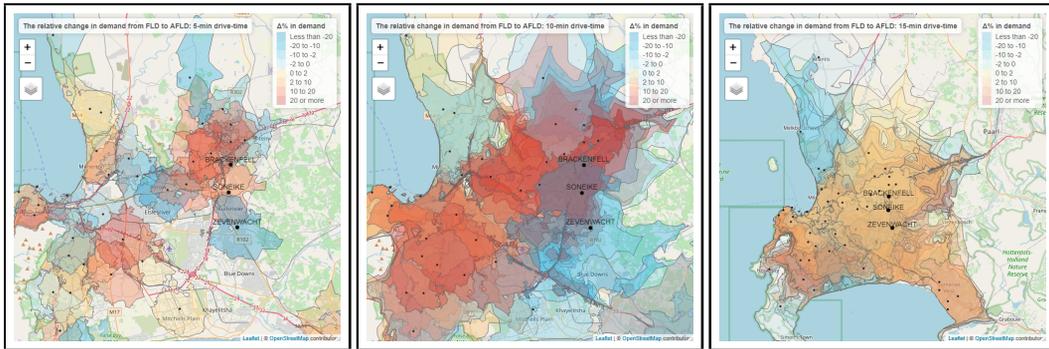
The seasonal relationship between total seasonal sales and weather variables was tested using a simple linear regression, with average seasonal sales (from 2016 to 2019) as the dependent variable and average maximum temperature, average minimum temperature, and average rainfall (all averaged from 2016 to 2019) as the independent variables. These variables were captured in 2km × 2km grids for which sales and weather data were available, which were 46 distinct areas in the City of Cape Town. The results indicate that the average minimum temperature has the most significant positive effect on average sales ( $p = 0.0295$ ) and average maximum temperature somewhat of an effect ( $p = 0.0791$ ) which aligns with findings from Musengimana et al. (2016). However, rainfall does not appear to significantly influence sales ( $p = 0.9453$ ). It should be noted that Cape Town experienced a water crisis from 2017 to 2019 and 574 of the 592 nodes indicated rainfall at the 10th percentile or lower. Hence, very low rainfall measures were captured over the majority of Cape Town. The multiple R-squared value of 0.1628 suggests that only about 16.3% of the variability in sales can be explained by the model, indicating other factors may play a role. For the continuation of this research we will only be using temperature and ease of access as attributes.

Let the attribute temperature have two levels, namely "hot days" and "normal days". All structural nodes in  $G$  will be associated with either one of these two levels. The first, "hot days", will include all structural nodes, overlapping with at least one POI and experiences maximum temperatures above 35°C in summer. The structural nodes associated with "hot days" correspond with the same areas identified by Heat Watch Cape Town<sup>5</sup>. All other remaining structural nodes are assigned to "normal days", reflecting the tendency to select areas that are less impacted by extreme weather, as illustrated in Figure 3. This also accounts for potential shortages of essential medicines for gastrointestinal illnesses in heat-affected regions. In addition to temperature, ease of access to a POI will be considered as an attribute. Each structural node will have an indication on whether it has easy access to a principal road or not. A principal road can be either a major road, such as a highway, or a main road, typically serving as a key route for local traffic within a town or city<sup>6</sup>. This designation signifies

<sup>4</sup>South African Weather Services, <https://www.weathersa.co.za/>

<sup>5</sup> 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)

<sup>6</sup> Department of Health South Africa, *TRH 26: South African Road Classification and Access Management Manual*,



**Figure 4.** The relative change in demand from FLD to AFLD for 5, 10, and 15-minute drive-time thresholds. Map data © OpenStreetMap contributors, under the Open Database License (ODbL).

whether the node is connected to key transportation routes that facilitate faster and more direct travel between nodes. The principal road attribute will have four levels that each structural node can be associated with. The first level, “EOA1”, indicates that the structural node has demand but no easy access to a principal road. The second level, “EOA2”, indicates that a structural node has easy access to a principal road. The third level, “EOA3”, is associated with structural nodes that are overlapping with a POI, but not within easy access of a principal road, and level four, “EOA4”, includes all other structural nodes not within easy access to principal roads. Given the assumption that people will prefer POIs that are easily accessible, the nodes associated with “EOA1” and “EOA2” should share a connection through the same attribute. For this reason, the principal road attribute will be simplified into two levels: “PR1,” which includes all nodes from levels “EOA1” and “EOA2”, and “PR2”, which includes all nodes from levels “EOA3” and “EOA4”, as illustrated in Figure 3.

The demand allocated to each grid,  $D(v_n)$ , is the population size which is obtained from the overlaying census and deeds office data as captured by Lightstone Pty Ltd<sup>7</sup> from 2016 to 2019. The supply capacity at each node is the total sales over summer (December, January and February) of an affordable over-the-counter diarrhoea medication, averaged over 2016 to 2019.

The accessibility and supply-demand ratio will be computed using both an attribute based fuzzy lattice data structure and a fuzzy lattice data structure to compare how the probabilities assigned to each POI differ when ease of access and temperature are taken into account. The supply considered here includes only one pharmaceutical chain in South Africa, which is not an exhaustive list of all suppliers providing diarrhoea medication. As a result, the supply-demand ratio values are very small due to the under-representation of supply. Therefore, when comparing the 5, 10, and 15-minute drive-time thresholds visually, the differences in demand are mapped to illustrate how demand changes when using the fuzzy lattice data structure (FLD) compared to the attribute based fuzzy lattice data structure (AFLD). Figure 4 shows the proportionate difference between demand using AFLD and demand using FLD, relative to the demand using FLD, indicating how much demand has increased

<https://wayleave.tshwane.gov.za/> (accessed August 15, 2024)

<sup>7</sup>Lightstone (Pty) Ltd procures its data directly from the Deeds Office and is comprised of a snapshot of all South African property ownership as at 1993, with a full history of all transactions to augment with census data. Ethics clearance number: NAS003/2023.

**Table 1.** Supply, demand and supply-demand ratios for Zevenwacht ( $P_1$ ), Soneike ( $P_2$ ) and Brackenfell ( $P_3$ ) at 5, 10 and 15 minute drive-time thresholds using a fuzzy lattice data (FLD) and attribute based fuzzy lattice data (AFLD) structure .

$S(P)$	Drive-time	FLD		AFLD	
		Demand	SD-Ratio	Demand	SD-Ratio
$S(P_1) = 1\,252$	5min	93\,638	0.013	93\,638	0.013
	10min	252\,558	0.005	211\,542	0.006
	15min	225\,203	0.006	118\,533	0.011
$S(P_2) = 1\,654$	5min	57\,993	0.029	57\,238	0.029
	10min	69\,743	0.024	81\,338	0.020
	15min	86\,922	0.019	108\,908	0.015
$S(P_3) = 674$	5min	30\,950	0.022	31\,253	0.022
	10min	45\,415	0.015	64\,683	0.010
	15min	45\,468	0.015	79\,372	0.009

or decreased as a percentage when using AFLD instead of FLD.

Areas with easy access to principal roads and normal temperature fluctuations tend to see an increase in demand, while areas further from main routes and experiencing temperatures above 35°C typically experience a decrease in demand. Note that in this example, both attributes were given equal weights of 1 to indicate that a link between the levels exists, though these weights can be adjusted to reflect differing levels of importance.

When comparing Brackenfell, Soneike, and Zevenwacht (indicated with labels in Figure 4), as done in Table 1, it's important to note their proximity and how their 5-minute drive-time catchment areas overlap slightly, becoming fully overlapping at the 15-minute drive-time threshold catchment areas. Brackenfell benefits from easy access to a principal road, an attribute that Soneike and Zevenwacht lack. Additionally, Zevenwacht is associated with the "hot days" category in the temperature attributes. As a result, Brackenfell holds a stronger connection with the surrounding population due to its shared attributes, while Zevenwacht has attributes that aren't shared by the surrounding population and demand will be allocated only on the link structure.

When examining Zevenwacht's demand across different drive-time thresholds, we see that as we move from fuzzy lattice data to attribute based fuzzy lattice data, demand decreases significantly with increased drive-time, which causes the supply-demand ratio to rise. Soneike, although not linked to a principal road, still experiences a demand increase because it shares the temperature attribute, absorbing some of the demand that would have shifted to Zevenwacht.

The greatest increase in demand, however, occurs in Brackenfell. Demand nearly doubles when using the AFLD, driven by the two shared attributes: normal temperature and ease of access. As Brackenfell's population grows, the supply-demand ratio decreases, indicating that the store will need to adjust its supply to meet the increased demand.

The comparison between the attribute based fuzzy lattice data structure (AFLD) and the fuzzy lattice data structure (FLD) highlights the impact of incorporating attribute nodes on demand alloca-

tion. This helps to capture more realistic demand allocation to areas based on additional factors other than the link structure. Adjusting the weights of these attributes can further refine the calculation of accessibility and supply-demand ratios, allowing for more accurate predictions to varying conditions and priorities.

#### **4. Discussion**

The benefits of an attribute based fuzzy lattice data structure are evident in its ability to propagate labels more effectively within a network. By combining both structural connections and attribute similarities, nodes can share a connection through a similar attribute node. Traditional methods typically rely on link structures, but many real-world networks contain shared attributes between nodes that are not captured by structural connectivity alone. This approach leverages both aspects, enabling labels to propagate based on structural connections and shared characteristics such as common interests, similar topography, etc.

In this application, edge weights were equally assigned across different attributes and structural edges; it can however be adjusted. If there are natural or man-made barriers, structural weights can be reduced or even set to zero. Similarly, if certain attributes hold varying levels of importance, weights can be adjusted accordingly to ensure that these attributes are prioritised.

Only temperature and ease of access were considered as factors influencing the sales of diarrhoea medication. The low R-squared value when testing sales against temperature suggests that other factors, which were not investigated, likely play a significant role in these sales. Factors such as socioeconomic conditions, economic considerations, marketing strategies, etc., could also be pivotal. Additionally, the model only accounted for the supply of one easily accessible over-the-counter diarrhoea medication from one pharmaceutical chain store in South Africa. As multiple other pharmacies and medications are available, the supply figures indicated per region are likely an underestimate.

#### **5. Conclusion**

Traditional community detection methods focus primarily on a network's topology, but integrating attribute information can enhance these methods by improving the understanding of node similarities and capturing relationships that go beyond structural connectivity. Incorporating attributes into the fuzzy lattice data structure allows for a more accurate estimate of demand distribution. In networks where nodes share common attributes, such as similar accessibility or environmental conditions, the attribute based fuzzy lattice data structure can propagate labels across the network more accurately. This approach overcomes limitations of purely structural connectivity by accounting for both the link structure and attributes, resulting in a more detailed and accurate depiction of the network's interactions. The attribute based fuzzy lattice data structure offers significant advantages for modelling and analysing networks. By considering both attribute and structural links, the attribute-based fuzzy lattice data structure provides a balance between proximity and shared attributes.

#### **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

- ANDERSEN, J. L. E. AND LANDEX, A. (2008). Catchment areas for public transport. *WIT Transactions on the Built Environment*, **101**, 175–184.
- ANYAMBA, A., SMALL, J. L., BRITCH, S. C., TUCKER, C. J., PAK, E. W., REYNOLDS, C. A., CRUTCHFIELD, J., AND LINTHICUM, K. J. (2014). Recent weather extremes and impacts on agricultural production and vector-borne disease outbreak patterns. *PLoS One*, **9**, e92538.
- AZRAN, A. (2007). The rendezvous algorithm: Multiclass semi-supervised learning with Markov random walks. In *Proceedings of the 24th International Conference on Machine Learning*. 49–56.
- BHAGAT, S., CORMODE, G., AND MUTHUKRISHNAN, S. (2011). Node classification in social networks. *Social Network Data Analytics*, 115–148.
- CHALLEN, R. J., GRIFFITH, G. J., LACASA, L., AND TSANEVA-ATANASOVA, K. (2022). Algorithmic hospital catchment area estimation using label propagation. *BMC Health Services Research*, **22**, 1–12.
- CORLUKA, G. (2019). Tourism seasonality – An overview. *Journal of Business Paradigms*, **4**, 21–43.
- DE KLERK, M. AND FABRIS-ROTELLI, I. (2024). Hospital accessibility catchment areas as a fuzzy lattice data structure. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.
- GREEN, C. A., MANS, G. G., NGIDI, M., SOGONI, Z., AND MARTIZ, J. (2016). Using catchment area analysis and GIS based spatial analysis for prioritising spatial investment in non-metro South Africa. In *Proceedings of the 52nd ISOCARP Conference*. International Society of City and Regional Planners, Durban, South Africa.  
URL: <http://hdl.handle.net/10204/12066>
- KHAN, A. A. (1992). An integrated approach to measuring potential spatial access to health care services. *Socio-Economic Planning Sciences*, **26**, 275–287.
- LIN, D., ZHU, R., YANG, J., AND MENG, L. (2020). An open-source framework of generating network-based transit catchment areas by walking. *ISPRS International Journal of Geo-Information*, **9**, 467.
- LIN, H., ZHAN, Y., ZHAO, Z., CHEN, Y., AND DONG, C. (2021). Overlapping community detection based on attribute augmented graph. *Entropy*, **23**, 680.
- LUAN, X., CHENG, L., SONG, Y., AND ZHAO, J. (2020). Better understanding the choice of travel mode by urban residents: New insights from the catchment areas of rail transit stations. *Sustainable Cities and Society*, **53**, 101968.
- LUO, W. AND WANG, F. (2003). Measures of spatial accessibility to health care in a GIS environment: Synthesis and a case study in the Chicago region. *Environment and Planning B: Planning and*

- Design*, **30**, 865–884.
- MACHARIA, P. M., RAY, N., GIORGI, E., OKIRO, E. A., AND SNOW, R. W. (2021). Defining service catchment areas in low-resource settings. *BMJ Global Health*, **6**.
- MUSENGIMANA, G., MUKINDA, F. K., MACHEKANO, R., AND MAHOMED, H. (2016). Temperature variability and occurrence of diarrhoea in children under five-years-old in Cape Town metropolitan sub-districts. *International Journal of Environmental Research and Public Health*, **13**, 859.
- PETERS, D. H., GARG, A., BLOOM, G., WALKER, D. G., BRIEGER, W. R., AND HAFIZUR RAHMAN, M. (2008). Poverty and access to health care in developing countries. *Annals of the New York Academy of Sciences*, **1136**, 161–171.
- POZOUKIDOU, G. AND ANGELIDOU, M. (2022). Urban planning in the 15-minute city: Revisited under sustainable and smart city developments until 2030. *Smart Cities*, **5**, 1356–1375.
- RADER, B., ASTLEY, C. M., SEWALK, K., DELAMATER, P. L., CORDIANO, K., WRONSKI, L., RIVERA, J. M., HALLBERG, K., PERA, M. F., CANTOR, J., ET AL. (2022). Spatial modeling of vaccine deserts as barriers to controlling SARS-CoV-2. *Communications Medicine*, **2**, 141.
- RADKE, J. AND MU, L. (2000). Spatial decompositions, modeling and mapping service regions to predict access to social programs. *Geographic Information Sciences*, **6**, 105–112.
- RAGHAVAN, U. N., ALBERT, R., AND KUMARA, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, **76**, 036106.
- SHAO, Y. AND LUO, W. (2022). Supply-demand adjusted two-steps floating catchment area (SDA-2SFCA) model for measuring spatial access to health care. *Social Science & Medicine*, **296**, 114727.
- TAO, Z., CHENG, Y., ZHENG, Q., AND LI, G. (2018). Measuring spatial accessibility to healthcare services with constraint of administrative boundary: A case study of Yanqing District, Beijing, China. *International Journal for Equity in Health*, **17**, 1–12.
- VER HOEF, J. M., PETERSON, E. E., HOOTEN, M. B., HANKS, E. M., AND FORTIN, M.-J. (2018). Spatial autoregressive models for statistical inference from ecological data. *Ecological Monographs*, **88**, 36–59.
- WANG, F. (2014). *Quantitative Methods and Socio-Economic Applications in GIS*. CRC Press, Boca Raton, FL.
- ZHOU, Y., CHENG, H., AND YU, J. X. (2009). Graph clustering based on structural/attribute similarities. *Proceedings of the VLDB Endowment*, **2**, 718–729.