Hospital Accessibility Catchment Areas as a Fuzzy Lattice Data Structure

Michelle de Klerk[1], Inger Fabris-Rotelli[1]

[1] Department of Statistics, University of Pretoria, Pretoria, South Africa - dklerkm@gmail.com, inger.fabris-rotelli@up.ac.za

Keywords: Catchment Areas, Fuzzy Lattice Data, Semi-supervised, Label Propagation, Label Connected

Abstract

The accessibility to basic facilities and services plays a pivotal role in every society and city planning. Spatial accessibility can vary between cities and countries and is mainly defined by the ease at which facilities can be accessed by communities. Facilities can provide essential services and/or products such as pharmacies, clinics, schools, universities, etc. Spatial accessibility is dependent on the spatial impedance between a facility and the target population and can be illustrated with catchment areas. We propose a fuzzy lattice catchment area method which uses a semi-supervised learning algorithm to create overlapping catchment areas. This methodology is applied to determine the accessibility to hospitals in South Africa and provides an illustration on the difference for regions with high accessibility compared to low accessibility. The application can easily be adapted in a variety of fields based on industry type, drive-time thresholds, supply capacity and the target population.

1. Introduction

1.1 Background

Ease of accessibility to essential facilities is an important component of any society. This is important for several reasons including equality in provision of resources and services (Wang, 2014). Unfortunately, due to geographic, and non-geographic, barriers not all communities receive equal accessibility to basic facilities (Rader et al., 2022). This is especially true for disadvantaged population groups considering factors like income and minority groups. Globally there is also a disparity between countries when considering basic services such as health care, for which low- and middle-income countries tend to have poorer accessibility than first world countries (Peters et al., 2008).

Access can be classified into one of four categories namely: potential spatial, potential aspatial, realised spatial and realised aspatial access as defined by (Khan, 1992). Aspatial access focusses on social disparities between communities and identifies nongeographic barriers such as wages, gender, race, childcare services, educational attainment, linguistic barriers etc. as discussed in (Wang, 2014; Rader et al., 2022). Potential and realised spatial access is however based on geographic distance as well as the distribution and size of facilities. Services or products offered within geographic accessibility of communities are potential spatial access and only when it is utilised, is it referred to as realised spatial access.

Potential spatial accessibility can be defined by the ease at which a facility, or any point of interest, POI, is arrived at from a demand location (Wang, 2014). POIs are facilities which provide services and/or products and can be grouped by sub-industry level (laboratories, pharmacies, clinics, schools, universities, etc.) or industry level (health care, education etc.). Spatial accessibility is dependent on the spatial impedance (drive-time or Euclidean distance) between the POI and demand location and on the capacity of the POI in question (Tao et al., 2018; Luo and Wang, 2003; Shao and Luo, 2022).

This paper proposes a method to identify fuzzy lattice catchment areas, which represent potential spatially accessible regions. Fuzzy in this context refers to assigning degrees of membership as a probability ranging from 0 to 1. The method employs a semi-supervised approach to classify overlapping catchment areas. The fuzzy lattice catchment areas approach is compared to existing methods to highlight how the distribution in resources and demand is affected when using a weighted approach as opposed to assigning equal weights. Finally it is used as an application to create fuzzy lattice catchment areas for hospitals in South-Africa and illustrates how areas with high and low accessibility can be influenced by the weighted distribution of resources.

1.2 Literature review

A catchment area is the geographical area from which a POI attracts communities that utilises its services or products. It is usually defined by the maximum traveling distance users are willing to travel to a POI and can either be created naturally to a point that people are drawn to, by natural geographic boundaries or as a predefined establishment (Luan et al., 2020).

There are an array of methods available to determine catchment areas in spatially accessible areas. These methods include the circular buffer approach (Andersen and Landex, 2008) and Euclidean buffer-based method (Lin et al., 2020) to more specialised methods such as the floating catchment, two-step floating catchment area (2SFCA) (Radke and Mu, 2000; Luo and Wang, 2003), gravity based methods (Wang, 2014), distance decay functions and variable catchment areas as summarised in (Tao et al., 2018). As highlighted in (Wang, 2014), spatial accessibility is dependent on the capacity of a POI and/or the spatial impedance between a POI and demand location.

Catchment areas formed around POIs can be classified into one of the following three categories: in potential spatial accessible areas it can either be i) overlapping or ii) non-overlapping (Challen et al., 2022) and beyond spatial accessible areas, iii) spatially inaccessible (Wang, 2014). Overlapping catchment areas indicates a choice for the POI being visited and is a more

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-2024-99-2024

realistic approach than imposing non-overlapping spatial boundaries (Challen et al., 2022). Non-overlapping catchment areas provide a simplification to modelling techniques. There is however the disadvantage that demand could be incorrectly assigned. If a community is only assigned to one catchment area, when more than one is available, the result will tend to be spatially uneven (Challen et al., 2022). Spatially inaccessible areas, also referred to as off-network areas (Lin et al., 2020), occur when the distance or time impedance falls beyond the distance or drive-time threshold of the spatially accessible areas (Wang, 2014). Communities which falls in the spatially inaccessible areas are typically excluded from any further analysis (Lin et al., 2020).

The spatial impedance or threshold distance between a POI and demand location can be based on different methods such as a distance buffer, distance-decay, drive-times etc. (Wang, 2014). A welfare approach was taken by (Green et al., 2016) to define threshold distances by classifying different services and products by answering the fundamentals of “who gets what, where and how”. Using this approach essential services, such as basic education, healthcare etc., should be available to the majority of the population and should have a smaller threshold distance (and catchment area) than facilities providing inessential services. As discussed in (Green et al., 2016), low-order basic facilities define access to essential services such as schools and naturally form small catchment areas around communities. Middle-order POIs serve more communities than low-order POIs and can be located at a further distance such as 24-hour health care and Home Affairs/State Department offices. Finally high-order POIs have the largest possible catchment areas as people are willing to commute further to POIs that provide high value services or products such as higher education and leisure centres. According to (Green et al., 2016), middle-order POIs should ideally have a maximum distance of 30 km and should be accessible to 91% of the population. The fifteen-minute city concept specifies that all critical urban services and facilities should be within a 15 minute walking, cycling or driving distance for optimal and sustainable city planning and development (Pozoukidou and Angelidou, 2022). The distance and time associated with low-, middle- and high-order POIs can differ vastly between cities and each region should be analysed individually based on its road networks and infrastructure.

It is important to quantify the capacity or supply of a POI and the demand at surrounding communities for optimal resource and facility planning (Green et al., 2016). Supply is quantified based on whether an activity or algorithmic approach is used to model catchment areas (Challen et al., 2022). An activity based approach takes into account the daily movement of communities between different regions to access a specific service. An algorithmic approach uses static measures in an area at a specific point in time by considering the size, capacity and ease of access to quantify supply. Demand is calculated in a similar fashion by taking the population size of an area at a point in time. For our application in section 3 an algorithmic approach will be used to compare hospital bed capacity and demand at a specific point in time.

In this paper a similar approach is used as in (Challen et al., 2022) to define catchment areas using network structures, specifically considering label propagation (Raghavan et al., 2007). Network structures takes into account how nodes (or users) are linked and incorporates information from neighbours to classify unlabelled nodes. There are multiple methods for node classification in network structures available. These include community score, multi-objective optimisation, modularity optimisation, genetic algorithm, label propagation, game theory, clustering etc. as summarised in (Bedi and Sharma, 2016).

The fuzzy lattice catchment areas methodology that is proposed herein will however differ from the approach in (Challen et al., 2022) where the final result is overlapping catchment areas rather than non-overlapping areas. Non-overlapping catchment areas occur when a node is associated with only one label at the end of the iteration process. In our method a probabilistic approach is followed where each node will have a probability to be assigned to each label in the network structure. Drive-time thresholds are applied to ensure that the probability structure also takes into account the geographical boundaries of the POIs and population.

2. METHODOLOGY

In this section the initial data inputs and the algorithmic approach to creating fuzzy lattice catchment areas will be covered. This is achieved by integrating the foundational work in (Bhagat et al., 2011) on node classification in network structures with model-based label propagation techniques for resource allocation as done in (Challen et al., 2022). The methodology in (Challen et al., 2022) focuses on how propagation rates adjust as demand outstrips supply in non-overlapping catchment areas. In contrast, our method suggests overlapping catchment areas where each label has a set of probabilities associated with each node at different drive-time intervals.

A label can be a binary, single or multi-label which can represent any characteristic of the associated node and an edge represents a similarity between two nodes. If a node is an individual in a social network, a label can represent age, marital status etc. and an edge between two individuals can represent a shared characteristic such as a shared family connection, interest etc. In this application, let the network graph be the full geographical region which is subdivided into smaller grid cells (whose centres then represent the nodes) and an edge between two nodes is a shared line segment between two grid cells. Labels according to a chosen connectivity will be the POI associated or classified with each grid cell. Initially only the nodes which spatially overlaps with a POI will be labelled, the remaining nodes will be unlabelled.

Using a semi-supervised learning approach to propagate labels to neighbouring grid cells, the full geographical region can be represented as a fuzzy lattice structure, with each grid cell having a set of probabilities associated with each label, as outlined in section 2.1. Network threshold distances for low-, middle- and high-order POIs will then be overlayed on the fuzzy lattice data structure to create fuzzy lattice catchment areas in section 2.2.

2.1 Fuzzy lattice data

Fuzzy lattice catchment areas uses a random walk approach and label propagation to propagate labels to all nodes by only using the link structure of the graph. Label propagation (Raghavan et al., 2007) relies on an iterative process where each node considers its neighbours. Each unlabelled node joins the community to which the maximum number of its neighbours belong to. As discussed in (Raghavan et al., 2007) when a tie between two labelled nodes occur, the label is determined by selecting one of the neighbouring labels based on a uniform random approach. In this research we will consider the approach in (Xie and Szymanski, 2013) where a node can keep multiple labels received from neighbouring nodes, allowing for an overlap of catchment areas. The degree of overlap is captured by a probability and hence create a fuzzy lattice data structure.

Using node classification formulation as illustrated in (Bhagat et al., 2011), consider graph $G = G(V, E)$ that is subdivided into $N, n = 1, ..., N$ nodes $V = \{v_1, v_2, ..., v_N\}$ and let an edge $(i, j) \in E$ represent a shared line segment between nodes v_i and v_j as defined by the Rook's contiguity (Wang, 2014). Let $\mathbf{D}$ be a $N \times N$ diagonal degree matrix indicating the number of neighbouring edges to node $v_n, n = 1, ..., N$ and weights matrix $\mathbf{A}$ be an $N \times N$ adjacency matrix defining neighbouring nodes using common edges between grids.

Consider geographical spatial POIs, $P = \{P_1, P_2, ..., P_M\}$, $m = 1, ..., M$ as a known set of M locations. The centre of each grid cell $v_n, n = 1, ..., N$ will represent the nodes in graph G. The grid size should be determined based on the specific geographical region of the study. In addition every node is assumed to contain at most one POI, thus the grid size should be chosen to maintain this at a minimum. Let V_m be the subset of the M labelled nodes which contains a POI and V_u be the subset of remaining $N - M$, initially unlabelled nodes, which doesn't contain a POI. Let V be ordered such that the first M rows are the cells from V_m and the remaining $N - M$ rows are the nodes from V_u such that $V = V_m \cup V_u = \{v_1, ..., v_M, v_{M+1}, ..., v_N\}$ (Bhagat et al., 2011).

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 $(N-M) \times M$ matrix indicating the POI associated with corresponding nodes in V_u. The initial label matrix, $\mathbf{Y}$, is an $N \times M$ matrix with the first M rows as $\mathbf{Y}_m$ and the remaining $N - M$ 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 not label connected when there remains unlabelled nodes at the end of the iteration process. This will occur when there exists unlabelled node/s which share no direct edges connecting them with the remaining structure in graph G. When considering a social network structure this can happen when there are people who share no similar interest with the remaining individuals in the network. For a geographical application, a graph can be non label-connected when there are grid cells which doesn't share a direct edge with the remaining grid cells due to natural/man-made boundaries such as rivers, mountains, region/state border boundaries etc.

Consider transition matrix $\mathbf{P}$, with $\mathbf{P}^t$ the corresponding matrix at time t where p_{ij}^t is the probability of reaching node v_j from v_i in t steps. As $t \to \infty$, $p_{ij}^{t\to\infty}$ will indicate the $(i, j)^{th}$ probability in $\mathbf{P}^{t\to\infty}$ at which the process reached node v_j, with associated label $c \in \mathcal{Y}$, from v_i in convergence with $0 \le p_{ij} \le 1$ and $\sum_j p_{ij} = 1$. In matrix form

$$\hat{\mathbf{Y}} = \mathbf{P}^{t\to\infty}\mathbf{Y} \tag{1}$$

where

$$\hat{\mathbf{Y}} = \begin{bmatrix}\hat{\mathbf{Y}}_m \\ \hat{\mathbf{Y}}_u\end{bmatrix} \text{ and } \mathbf{Y} = \begin{bmatrix}\mathbf{Y}_m \\ \mathbf{0}\end{bmatrix} \tag{2}$$

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

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 $N - M$ rows are V_u, the transition matrix can be split into 4 sub-matrices, indicating the probability to move between states, namely $\mathbf{P} = \begin{pmatrix}\mathbf{P}_{mm} & \mathbf{P}_{mu} \\ \mathbf{P}_{um} & \mathbf{P}_{uu}\end{pmatrix} = \begin{pmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{P}_{um} & \mathbf{P}_{uu}\end{pmatrix}$.

The probability of transition from the unlabelled states to the labelled states are captured in the $(N - M) \times M$ matrix $\mathbf{P}_{um}$ and from the unlabelled states to the unlabelled states are captured in the $(N-M)\times(N-M)$ 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 $M \times (N - M)$ zero matrix $\mathbf{0}$.

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

$$\mathbf{P}^{t\to\infty} = \begin{pmatrix}\mathbf{I} & \mathbf{0} \\ (1 - \mathbf{P}_{uu})^{-1}\mathbf{P}_{um} & \mathbf{0}\end{pmatrix} \tag{3}$$

as shown in (Bhagat et al., 2011; Azran, 2007). Substituting equations (3) and (2) into (1) the labels can be computed by

$$\begin{bmatrix}\hat{\mathbf{Y}}_m \\ \hat{\mathbf{Y}}_u\end{bmatrix} = \begin{pmatrix}\mathbf{I} & \mathbf{0} \\ (1 - \mathbf{P}_{uu})^{-1}\mathbf{P}_{um} & \mathbf{0}\end{pmatrix}\begin{bmatrix}\mathbf{Y}_m \\ \mathbf{0}.\end{bmatrix}. \tag{4}$$

From (4), the labels for the absorbent state nodes contained in V_m remains unchanged with $\hat{\mathbf{Y}}_m = \mathbf{Y}_m$. The labels of the unlabelled nodes are obtained by computing $\hat{\mathbf{Y}}_u = (1 - \mathbf{P}_{uu})^{-1}\mathbf{P}_{um}\mathbf{Y}_m$. The $(N - M) \times M$ matrix $\hat{\mathbf{Y}}_u$ will contain the probability of each label assigned to grid $v_n \in V_u$ with the sum over each row adding to 1. The process will iterate over all nodes defined in G.

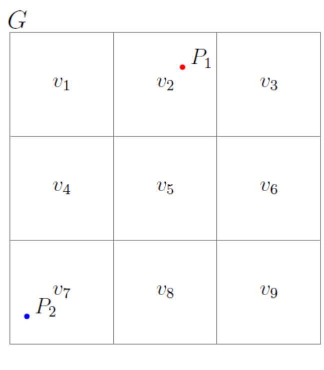

Figure 1. Demand for grids $v_n, n = 1, ..., 9$ with POIs P_1 and P_2.

Consider the simple example illustrated in Figure 1. In this example seven regions, $V_u = \{v_1, v_3, v_4, v_5, v_6, v_8, v_9\}$, are unlabelled and two regions, $V_m = \{v_2, v_7\}$, are labelled with associated POIs P_1 and P_2 respectively. Let V be the ordered set, contained in G, consisting of $N = 9$ nodes with the first $M = 2$ rows the labelled nodes and the remaining $N - M = 7$ rows the unlabelled nodes, i.e. $V = \{v_2, v_7, v_1, v_3, v_4, v_5, v_6, v_8, v_9\}$. Let $\mathbf{Y}_m$ 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 $v_n \in V, n = 1, ..., 9$ is calculated in (5).

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{\mathbf{Y}}_m \\ \hat{\mathbf{Y}}_u \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0.72 & 0.28 \\ 0.83 & 0.17 \\ 0.45 & 0.55 \\ 0.62 & 0.38 \\ 0.66 & 0.34 \\ 0.38 & 0.62 \\ 0.52 & 0.48 \end{bmatrix}. \tag{5}$$

From the results in (5) 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. It can be expected that a unlabelled grid will have a higher probability associated with a labelled grid that is closer in the graph structure than labelled grids that is further away. Node v_3 has the highest probability of being associated with label $P_1(0.83)$ and v_8 with label $P_2(0.62)$ as can be verified by the graph structure of G as illustrated in Figure 1.

2.2 Fuzzy lattice catchment areas

Using the proposed fuzzy lattice data structure, realistic catchment areas can be obtained on this structure using drive-time thresholds. When assuming G is labelled connected, all nodes will receive a set of probabilities assigned to each label in set $\mathcal{Y}$. This is however not practical when considering geographic data and the threshold drive-time distance a person is willing to travel to a given POI.

Consider the set of M POIs, $P = \{P_1, P_2, ..., P_M\}$, $m = 1, ..., M$, with an associated drive-time threshold distance d^{P_m} assigned to each $P_m, m = 1, ..., M$. The threshold drive-time distance d^{P_m} depends on the order of the POIs and will be denoted by $d^{P_m}_{low}, d^{P_m}_{mid}$ or $d^{P_m}_{high}$ for low-, middle- or high-order POIs respectively as discussed in (Green et al., 2016). For each POI, we identify all nodes v_n 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 v_n 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. 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.

Let all 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 $N = 9$ grids and 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 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 (5) with drive-time thresholds applied is provided in (6).

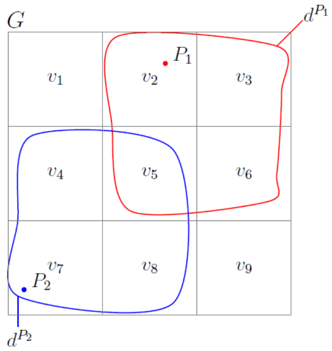

Figure 2. Demand for grids $v_n, n = 1, ..., 9$ with POIs P_1 and P_2 and corresponding drive-time thresholds d^{P_1} and d^{P_2}.

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{\mathbf{Y}}_m \\ \hat{\mathbf{Y}}_u \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0.62 & 0.38 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{6}$$

It can be noted that after standardisation that 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 falls beyond the drive-time thresholds for all POIs and is therefore spatially disjoint with probability 0 to be assigned to label $c \in \mathcal{Y}$.

2.3 Comparing measures of spatial accessibility

Spatial accessibility can be measured by considering how the distribution of supply and demand is connected in space (Wang, 2014). Supply can represent the capacity of a supplier such as the number of beds available in a hospital (Challen et al., 2022), the number of suppliers such as the number of primary care physicians in an area (Wang, 2014; Luo and Wang, 2003) or resources that are available such as a doctor's resources as done in (Shao and Luo, 2022). Demand usually represents the population and communities that requires the specified services or products. In this section the difference in spatial accessibility when comparing fuzzy lattice catchment areas to 2SFCA (Radke and Mu, 2000; Luo and Wang, 2003) will be illustrated.

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

$$R_{P_m} = \frac{S(P_m)}{D(G_{P_m})} \tag{7}$$

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

$$A_{v_n} = \sum_{m \in O_{v_n}} R_{P_m} = \sum_{m \in O_{v_n}} \frac{S(P_m)}{D(G_{P_m})}. \tag{8}$$

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

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., $D(v_n) = 1, n = 1, ..., 9$ 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 2SFCA, the supply-demand ratio for P_1 and P_2 are the same as both have the same units of demand and supply i.e. $R_{P_1} = R_{P_2} = 1$. Since region v_5 has access to both P_1 and P_2 the accessibility of v_5 is $A_{v_5} = R_{P_1} + R_{P_2} = 2$. The supply-demand ratio and accessibility is a good indication on how resources are distributed. This approach however doesn’t take into account the structure of the data and all demand locations within a catchment area are given equal weights.

Consider the same application, but applied to fuzzy lattice catchment areas to determine accessibility and supply-demand ratio using a probabilistic approach. The supply and demand associated with each grid and supplier will remain the same i.e. $D(v_n) = 1, n = 1, .., 9$ and $S(P_m) = 4, m = 1, 2$ but the demand on catchment area level will change according to the weight assigned by $\hat{\mathbf{Y}}_u$. Let the demand be proportionally assigned to each supplier’s catchment area using the probabilities in (6). Then $D(G_{P_1}) = 3.62$ and $D(G_{P_2}) = 3.38$ which results in the supply-demand ratio for P_1 to change to $R_{P_1} = \frac{S(P_1)}{D(G_{P_1})} = 1.105$ and P_2 to $R_{P_2} = \frac{S(P_2)}{D(G_{P_2})} = 1.183$. Since region v_5 still has access to both P_1 and P_2 the accessibility of v_5 is $A_{v_5} = R_{P_1} + R_{P_2} = 2.288$.

Since semi-supervised learning takes into account the structure of the data, a higher weight is placed on label P_1 than for P_2 for region v_5 due to the placement of the supplier. This results in a lower supply-demand ratio for P_1 compared to P_2 since the same level of supply needs to be provided to a slightly higher level of demand.

When comparing fuzzy lattice catchment areas to other overlapping catchment area techniques, such as 2SFCA, the advantage of a probabilistic approach rather than assigning equal weights can be demonstrated when comparing how the population size (or demand in an area) and capacity at a given POI (or supply) is distributed. This provides a better illustration of how demand and supply is allocated based on the structure of the data rather than assigning equal weights to each region. This will also be illustrated in the next section with application to hospital bed capacity and population size in South Africa.

3. Application

Access to hospitals for different parts in South Africa will be investigated to test the proposed methodology. The facilities which will be considered are district, central, military, mining, regional and tertiary public hospitals as well as private hospitals. All hospitals which provide basic healthcare services and trauma related emergencies are considered as the POIs, where specialised facilities such as rehabilitation and psychiatric centres are excluded.

To illustrate the difference between areas which are highly accessible and areas with limited accessibility, two district municipalities[1] in Gauteng and Eastern Cape province (as illustrated in Figure 3) and their surrounding areas in South-Africa will be considered.

Figure 3. South Africa with hospital accessibility catchment areas application in Gauteng and Eastern Cape provinces. (OpenStreetMap contributors, 2024).

The first district municipality is City of Johannesburg (JHB) which is situated in the Gauteng province and is categorised as a metropolitan municipality. This is the largest city in South Africa with 4.8 million people of which 44% have completed their high school certificate and 15% holds a higher education qualification. The housing in this district is predominantly formal dwellings (90%) where 95% have access to running water on site as reported in the 2022 South Africa census[2]. The provincial road network of Gauteng consists of majority paved roads (76%) and 24% unpaved roads as reported by the National and provincial Departments of Transport[3].

The second district municipality is O.R. Tambo which is located in the Eastern Cape province and isn’t categorised as one of the six metropolitan municipalities in South Africa. This district has a population of 1.5 million people of which 22% have completed their high school certificate and 8% hold a higher education qualification. The housing in this district was predominantly informal, with only 43.4% formal housing recorded in 2011, but this value has increased significantly to 77% in 2022. Access to running water on site in this district is however still only 43% as reported in the 2022 South Africa census[2].

Even though the Eastern Cape is geographically larger than Gauteng, this province has limited accessibility where majority of the provincial road network is unpaved at 90% and only

[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)

[2] Statistics South Africa, *Census 2022 Municipal Factsheet*, census.statssa.gov.za (accessed February 23, 2024)

[3] National Treasury, *Chapter 7 - Roads and Transport*, www.treasury.gov.za (accessed February 23, 2024)

https://doi.org/10.5194/isprs-annals-X-4-2024-99-2024 | © Author(s) 2024. CC BY 4.0 License.

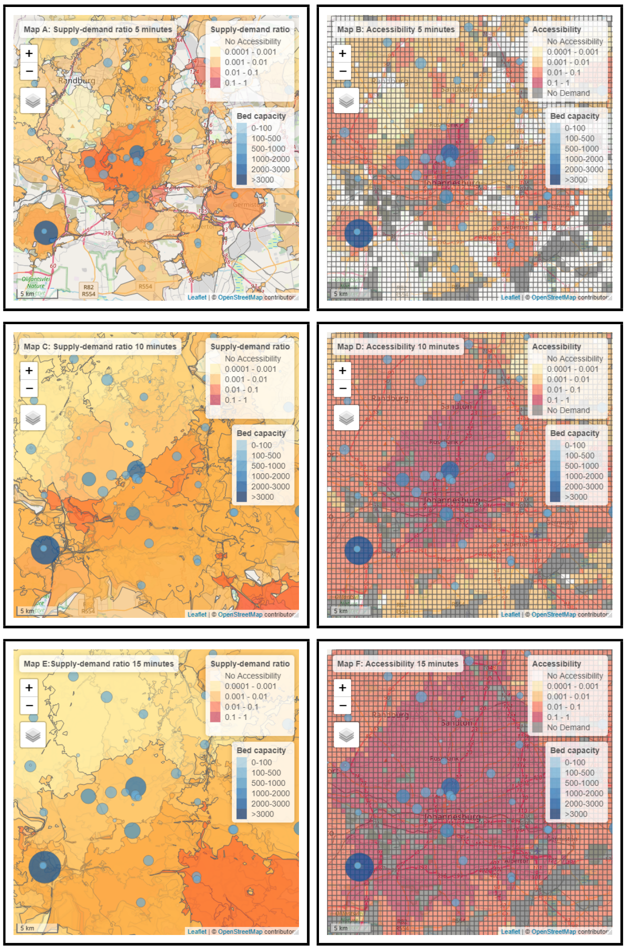

Figure 4. High Accessibility: Illustration of supply-demand ratio (Map A) and accessibility (Map B) at 5 minute drive-time threshold, supply-demand ratio (Map C) and accessibility (Map D) at 10 minute drive-time threshold and supply-demand ratio (Map E) and accessibility (Map F) at 15 minute drive-time threshold of City of JHB. (OpenStreetMap contributors, 2024).

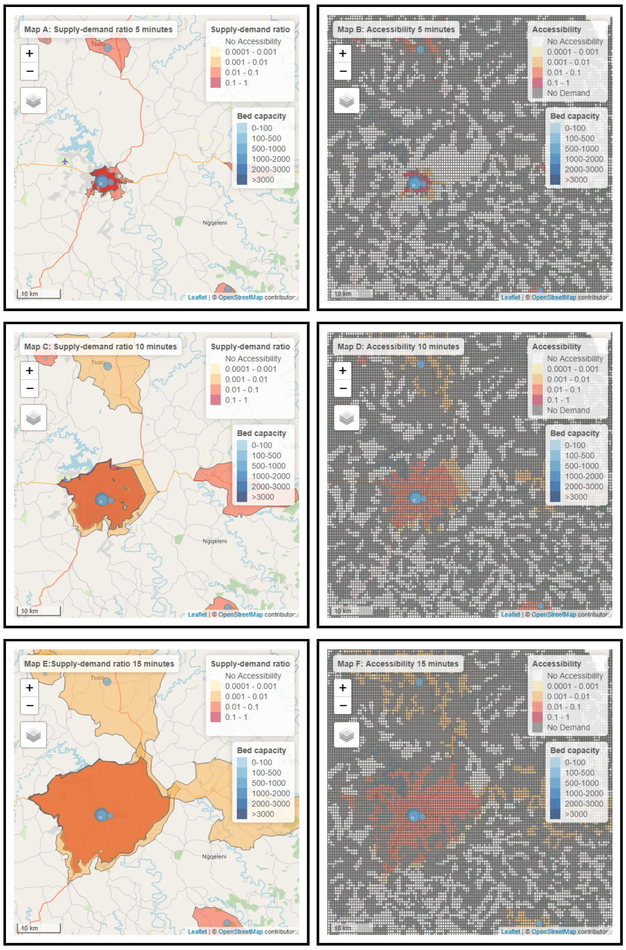

Figure 5. Low Accessibility: Illustration of supply-demand ratio (Map A) and accessibility (Map B) at 5 minute drive-time threshold, supply-demand ratio (Map C) and accessibility (Map D) at 10 minute drive-time threshold and supply-demand ratio (Map E) and accessibility (Map F) at 15 minute drive-time threshold of O.R. Tambo. (OpenStreetMap contributors, 2024).

10% is paved as reported by the National and provincial Departments of Transport[3].

Using the proposed fuzzy lattice catchment areas approach let graph G be South Africa which is subdivided into square grids of size $500m \times 500m$. This is the optimal grid size to ensure that only one POI is allocated to each node. The demand allocated to each grid, $D(v_n)$, is population size which is obtained from the overlaying census and deeds office data as captured by Lightstone (Pty) Ltd Ltd[4] for 2022. The supply for each POI, $S(P_m)$, is the bed capacity for each hospital. Drive-times thresholds of 5, 10 and 15 minutes for all the POIs will be used as the value for ($d^{P_m}_{low}$) low-, ($d^{P_m}_{mid}$) middle- and ($d^{P_m}_{high}$) high-order facility drive-time thresholds. The highest drive-time threshold considered for this example is at 15 minutes based on the fifteen-minute city concept (Pozoukidou and Angelidou, 2022).

Consider Figure 4 which illustrates supply-demand ratio and accessibility for POIs situated in the City of JHB, and surrounding areas. The capacity of each POI is illustrated by the size of the dot. The hospital with the largest capacity in South Africa is Chris Hani Baragwanath public hospital[5] with a capacity of 3,200 beds and approximately 6,760 staff members. It is the third largest hospital in the world and is located in the bottom left hand corner for all maps in Figure 4.

In Figure 4, maps A, C and E illustrate the supply-demand ratio and maps B, D and F the accessibility at drive-time thresholds of 5, 10 and 15 minutes respectively. One can note that as the drive-times per POI increases that the supply-demand ratio of each POI decreases as a higher level of demand is expected at the same level of supply. At the lowest drive-time threshold of 5 minutes, the ratio of supply to demand is very high in areas where the POI capacity is high. Maps B, D, and F demonstrate accessibility, and as more POIs are included within the drive-time thresholds, it becomes evident that nodes intersecting multiple POIs have greater accessibility compared to nodes connected to just one POI. When considering accessibility, it can be noted that at the lowest drive-time threshold, high accessibility is only evident where multiple POIs are located in the same vicinity. At the highest drive-time threshold (map F), all nodes have high accessibility, which illustrates that in highly

[4] 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.

[5] Chris Hani Baragwanath Academic Hospital, chrishanibaragwanathhospital.co.za (accessed February 20, 2024).

https://doi.org/10.5194/isprs-annals-X-4-2024-99-2024 | © Author(s) 2024. CC BY 4.0 License.

accessible cities, majority of people will have access to all basic services within a 15 minute drive-time threshold.

When considering an area with low accessibility a different view is observed when analysing how resources and demand is distributed. Consider the district municipality O.R. Tambo in Figure 5 where maps A, C and E illustrates the supply-demand ratio and maps B, D and F the accessibility at 5, 10 and 15 minutes drive-time thresholds respectively. In the low accessibility region one can note that there isn't a significant difference between the supply-demand ratio at 10 and 15 minutes drive-time respectively. This is due to the fact that the number of additional demand that is captured in extending the drive-time isn't as much as in a metropolitan area, since the number of people in a square meter is less than in City of JHB. One can note however that the supply-demand ratio for this region is lower than in the highly accessible region for most POIs and this is due to the supply capacity available at each of these POIs.

The capital of O.R. Tambo is the town Mthatha and there are multiple towns and communities that are formed around it. From Figure 5, Mthatha (situated close to the midpoint, but slightly off-centre to the left in maps A to F) has one of the highest health service per capita in its district, but since the district O.R. Tambo is largely rural and has a limited budget capacity (as reported by COGTA[6]) one can note that the accessibility to hospitals for other towns in this district is very low. The accessibility for this district, illustrated in maps B, D and F, has the opposite effect as in the highly accessible district. In City of JHB, additional POIs are captured as the drive-time increases, therefore the nodes have access to more hospitals and accessibility increases as drive-time threshold increases. In O.R. Tambo however, the POIs remain the only POIs in the 15 minute catchment area. Therefore, as the drive-time threshold increases, additional demand is added but the supply remains at the same level, which will lead to a decrease in accessibility for nodes that is in a 5 or 10 minute catchment area of the POIs. When considering map F in Figure 5 it is noted that some regions have no accessibility even at a 15 minute drive-time threshold. The grayed out blocks indicate open areas with no demand. No demand is identified with census and Lightstone (Pty) Ltd data and occurs where there is vacant land, water bodies, commercial or industrial structures and no residential homes.

On inspection when comparing the high accessible district to the district with lower accessibility it can be noted how the distribution of resources is affected as demand increases and whether the supply can increase to match the level of demand.

4. Discussion

We introduce fuzzy lattice catchment areas which are overlapping catchment areas and ensure a spatially even distribution of demand and supply. Fuzzy lattice catchment areas are created using the fundamentals of label propagation and is based on a probabilistic approach. Applying fuzzy lattice catchment areas has the advantage of creating more realistic, overlapping catchment areas, allowing a person to choose from multiple POIs within their accessibility range. This method ensures that demand is proportionally assigned rather than concentrated in a single region, which can lead to spatially uneven catchment areas. Together with this, drive-time thresholds are considered to ensure that the final catchment areas are representative of services/or goods being provided by POIs and the target population in that area.

The fuzzy lattice data structure is created by assuming the area is label connected, therefore that each unlabelled node will receive a probability to be associated with labelled node in a finite number of steps. This can be limiting however if there exists a natural barrier in the geographical data where a labelled node can't be reached and it is not labelled connected. Future research includes adjusting the edge weights (Bhagat et al., 2011) between nodes which will enable the specification of barriers in the graph structure. Fuzzy lattice catchment areas defines each labelled node as an absorbent state which implies that the node will assume the label with a probability of 1. This assumption can however be relaxed as proposed by (Szummer and Jaakkola, 2001) where a labelled node is not required to be an absorbent state. The relaxation of this assumption could be applied to cases where the population that is placed in the same node as a POI has a probability to be assigned to all POIs. In this application the grid sizes were small enough that this didn't have an effect on creating spatially uneven areas, but if the nodes are chosen at a larger size, relaxing this assumption could be beneficial.

To accurately capture the hospital bed capacity in a third world country such as South Africa can be challenging. How medical resources, staff and accessibility to the resources are reported can also be different between sources especially when comparing public and private hospitals. The list of all public and private[7] health facilities were obtained as reported by the hospital groups or by South African home affairs[8]. A few independent reports, including one from PMG[9], have however raised concerns that the capacity at some of the hospitals are not as reported and some having only half of the capacity available due to budget cuts and lack of health care professionals. Variation on bed capacity other than that officially reported could however not be taken into account and therefore the number of beds per capita that is indicated could be a very conservative estimate. In the approach taken by (Challen et al., 2022) caution was raised in creating overlapping hospital catchment areas when considering specialised hospitals that doesn't provide the same service but as explained in this application only hospitals that provide basic healthcare services were considered. Hence when applying this method all POIs considered needs to provide the same service or product. Another challenge when calculating catchment areas for low- and middle-income countries is the setup and maintenance of basic infrastructure. Depending on the accessibility of a region to main routes and the condition of the roads, the drive-time (given the same distance) can vary greatly between cities. The determination of the threshold for drive-time or distance should be made independently for each unique case, considering the type of service an industry offers and how essential it is to the surrounding communities. Future work will include investigation of a generalised application in the healthcare sector using a simulation study.

[6] Department of Cooperative Governance and Traditional Affairs, *Profile and Analysis: OR TAMBO District Municipality EC*, https://www.cogta.gov.za/ddm/wp-content/uploads/2020/11/ORTamnco-September-2020.pdf (accessed February 23, 2024)

[7] SA Private Hospitals, www.saprivatehospitals.com. (accessed February 02, 2024).

[8] Department of Home Affairs (DHA), *List of Connected Health Facilities per Province*, www.dha.gov.za (accessed February 15, 2024).

[9] PMG (People's Assembly Monitor), *SA Military Health Services briefing on improving capacity and capabilities at Military Hospitals*, https://pmg.org.za/committee-meeting/37138/ (accessed February 23, 2024)

Fuzzy lattice catchment areas is a very flexible approach for an array of different industry type services and target population scenarios. It can efficiently depict either an oversupply or undersupply of a services to the nearby population by considering the accessibility criteria based on drive-time or distance.

5. Conclusion

A realistic representation of facility catchment areas is crucial for any city resource planning. This will highlight any constraints in a city, such as limited accessibility to basic services. Fuzzy lattice catchment areas give a weighted and spatially even approach to understand the supply-demand ratio for any POI together with the accessibility available to communities.

This approach can be applied in various fields where the target population could only be a subset of the population for a specific industry type, for example only considering children in a local community and their access to parks, schools, or daycare. The algorithm has minimal assumptions and is mainly dependent on the link structure of the geographical region, supply per POI, demand at the desired nodes and drive-time thresholds.

Acknowledgments

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., Landex, A., 2008. Catchment areas for public transport. *WIT Transactions on the Built Environment*, 101, 175–184.

Azran, A., 2007. The rendezvous algorithm: Multiclass semi-supervised learning with Markov random walks. *Proceedings of the 24th International Conference on Machine Learning*, 49–56.

Bedi, P., Sharma, C., 2016. Community detection in social networks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 6(3), 115–135.

Bhagat, S., Cormode, G., Muthukrishnan, S., 2011. Node classification in social networks. *Social Network Data Analytics*, 115–148.

Challen, R. J., Griffith, G. J., Lacasa, L., Tsaneva-Atanasova, K., 2022. Algorithmic hospital catchment area estimation using label propagation. *BMC Health Services Research*, 22(1), 1–12.

Green, C. A., Mans, G. G., Ngidi, M., Sogoni, Z., Martiz, J., 2016. Using catchment area analysis and GIS based spatial analysis for prioritising spatial investment in non-metro south africa. *Proceedings of the 52nd ISOCARP Conference*, International Society of City and Regional Planners, Durban, South Africa.

Khan, A. A., 1992. An integrated approach to measuring potential spatial access to health care services. *Socio-Economic Planning Sciences*, 26(4), 275–287.

Lin, D., Zhu, R., Yang, J., Meng, L., 2020. An open-source framework of generating network-based transit catchment areas by walking. *ISPRS International Journal of Geo-Information*, 9(8), 467.

Luan, X., Cheng, L., Song, Y., 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., 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(6), 865–884.

OpenStreetMap contributors, 2024. OpenStreetMap. https://www.openstreetmap.org. Accessed: March 7, 2024.

Peters, D. H., Garg, A., Bloom, G., Walker, D. G., Brieger, W. R., Hafizur Rahman, M., 2008. Poverty and access to health care in developing countries. *Annals of the New York Academy of Sciences*, 1136(1), 161–171.

Pozoukidou, G., Angelidou, M., 2022. Urban Planning in the 15-Minute City: Revisited under Sustainable and Smart City Developments until 2030. *Smart Cities*, 5(4), 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(1), 141.

Radke, J., Mu, L., 2000. Spatial decompositions, modeling and mapping service regions to predict access to social programs. *Geographic Information Sciences*, 6(2), 105–112.

Raghavan, U. N., Albert, R., Kumara, S., 2007. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3), 036106.

Shao, Y., 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.

Szummer, M., Jaakkola, T., 2001. Partially labeled classification with Markov random walks. *Advances in Neural Information Processing Systems*, 14.

Tao, Z., Cheng, Y., Zheng, Q., 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), 1–12.

Wang, F., 2014. *Quantitative Methods and Socio-Economic Applications in GIS*. CRC Press, United States.

Xie, J., Szymanski, B. K., 2013. Labelrank: A stabilized label propagation algorithm for community detection in networks. *2013 IEEE 2nd Network Science Workshop (NSW)*, IEEE, 138–143.