OPERATIONS AND SUPPLY CHAIN MANAGEMENT
Vol. 16, No. 4, 2023, pp. 473 - 487
ISSN 1979-3561 | EISSN 2759-9363

Hybrid Clustering Strategy for Micro-hubs Location in Newspaper Distribution

Karla C. Alvarez-Uribe
Institute for Intelligent Systems Research and Innovation (IISRI)
Deakin University, Geelong, VIC 3216, Australia
Department of Production Engineering, Instituto Tecnológico Metropolitano, Medellín, Colombia
Email: kcalvarezuribe@deakin.edu.au (*Corresponding Author*)

Eduard Gañan-Cardenas
Department of Production Engineering, Instituto Tecnológico Metropolitano, Medellín, Colombia
Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia

Diego Perez-Montoya
Department of Production Engineering, Instituto Tecnológico Metropolitano, Medellín, Colombia

ABSTRACT

Facility location is one of the most critical factors in urban logistics planning, and the newspaper industry is no exception. Given the time-sensitive nature of newspapers and their narrow delivery time windows, efficient distribution network planning becomes essential. This research addresses the micro-hub location problem within the context of newspaper distribution across the Área Metropolitana del Valle de Aburrá in Medellin, Colombia, by developing a novel hybrid clustering strategy. We compare five clustering techniques: K-means, K-medians, K-medoids, Agglomerative Nesting (AGNES), and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Our strategy first uses AGNES (with single-linkage) to identify high-density regions and subsequently applies K-medoids within these identified areas to form compact clusters. Results demonstrated the superiority of the hybrid clustering strategy over both K-means and the individual clustering techniques. The hybrid approach generates more cohesive clusters, as evidenced by superior silhouette coefficients and within-cluster variance. The clustering proposal allowed 90% of customers to be located within 1.6 kilometers of a micro-hub, improving the distribution of newspapers in the urban areas.

Keywords: *clustering methods, facility location, micro-hubs, newspaper distribution, urban logistics*

1. INTRODUCTION

Population growth in urban areas has had a deep impact on urban freight distribution operations. The increasing demand for goods and services has boosted economic activity and the number of distribution logistics processes, resulting in more trips, traffic congestion, air pollution, and higher distribution costs (Browne *et al.*, 2012). Additionally, the growing popularity of e-commerce has favored demand fragmentation, thus leading to smaller, customized orders and more complex logistics (Singh and Gupta, 2020).

According to previous studies, delivery operations in urban areas are often inefficient and account for up to 28% of the total cost of the entire supply chain (Bergmann *et al.*, 2020). In addition to this, delivery operations generate pollutant emissions, which significantly exceed those from other transport activities (Katsela *et al.*, 2022), contributing to the creation of additional barriers to the city's sustainable logistics operations (Adetiloye and Pervez, 2015). Hence, the need to find solutions to optimize the urban distribution process. In this regard, some authors have proposed solutions or properly locating distribution centers and grouping customers (Lau *et al.*, 2010). Their goal has been to design and allocate a set of logistics facilities to better meet customer demands (Taaffe *et al.*, 2010). Some studies have reported improvements in profitability, with savings ranging from 5% to 10% in operating costs (Goetschalckx *et al.*, 2002).

Other solutions have focused on the distribution system design, with distribution center's location being the most critical factor in logistics planning and distribution network control (Saragih *et al.*, 2022). It has been proved that a proper location of distribution centers increases economic efficiency and reduces costs and environmental impacts while meeting customer expectations (Melkonyan *et al.*, 2020). Linear programming methods are often used to solve the facility location problem. However, despite their remarkable advantages in solving this type of problem, the larger the size of the network, the more difficult they are to solve. Therefore, resorting to clustering techniques can be an efficient and flexible alternative (Jarrah and Bard, 2012).

Several authors have employed clustering methods to solve problems associated with facility location, either by applying constraints on cluster formation or by clusters with sufficient and unlimited capacity. Clustering techniques have been widely used to find optimal locations. They have been used either as a starting point to develop other models (Faezy Razi, 2019; Sharma *et al.*, 2021) or as final models (Sahraeian and Kaveh, 2010; Varghese and Gladston, 2016; Wang *et al.*, 2018). In particular, K-means is one of the most extensively used clustering methods, even though emerging algorithms or possible combinations between them could produce better results.

K-means algorithm have been used to determine the optimal location of distribution centers, assigning customers to the nearest facility, and then create appropriate delivery routes (Varghese and Gladston, 2016; Wang *et al.*, 2018). Using the Euclidean distance, geographically neighboring customers are integrated into the same cluster, creating distinguishable spatial groups where customers belonging to the same cluster are close to each other and to their center, as well as clearly distant or differentiated from other customer groups. This solution is used as a preliminary stage or as an initial solution to the facility location and routing problem. However, K-means is not appropriate for identifying clusters with nonconvex shapes and varied sizes. Additionally, as K-means uses the mean value as the centroid, it makes its solution too sensitive to outliers (Han *et al.*, 2022). In spatial distribution, it could be very common to find some customers far away from the main areas of high customer density. This small group of customer-outliers would cause the centroid to move from the high-density area to a less optimal point, implying a larger trip to serve the area with the main concentration of customers. While a limited number of authors have explored alternative clustering methods, such as those conducted by Esnaf and Küçükdeniz (2009) and Sharma *et al.* (2017), their application remains somewhat constrained. Notably, we are not aware of any thorough assessment of hybrid strategies that make use of each clustering technique's advantages in the context of facility location in urban logistics.

This study aims to solve the problem of locating micro-hubs in the newspaper distribution industry, using various clustering algorithms. In this industry, planning and distribution activities are closely related. Since newspapers, as perishable products, must be distributed immediately once the news comes to inform customers as soon as possible (Boonkleaw *et al.*, 2009), production and delivery operations take place within very tight time windows, and their planning must consider geographic constraints and delivery deadlines (Cunha *et al.*, 2021). Driven by this need, we propose a solution that allows to determine the number and location of the micro-hubs, as well as the customers that must be assigned to each of them. Through the generation of compact clusters, we reduce the total distance traveled in the newspaper's distribution from the micro-hubs. The contribution of this study to the existing literature is that it addresses the problem of locating micro-hubs (as an intermediate facility distribution logistics) using a combination of different clustering algorithms. Besides K-means, we explore other clustering methods such as K-medians, K-medoids, Agglomerative Nesting (AGNES), and Density-Based Spatial Clustering of Applications with Noise (DBSCAN), which encompass partitioning, hierarchical, and density-based models, thus covering a broad spectrum of clustering alternatives. To validate the quality of the resulting clustering and determine the number of clusters, we use within-cluster variance and the silhouette coefficient. In addition, we propose a hybrid clustering strategy that exploits the strengths of each clustering technique involved.

The rest of this paper is organized as follows. Section 2 reviews the literature in the field. Section 3 describes the problem of joint newspaper distribution. Section 4 presents the strategy proposed in this study. Section 5 discusses the results. Finally, Section 6 concludes and recommends future lines of work.

2. LITERATURE REVIEW

Data clustering, as an unsupervised learning process, is often used as a preliminary step for data analysis or the structuring of other macro processes. It is employed to solve problems associated with facility location, route scheduling, and other configurations in the supply chain. Clustering methods have also been used as final models in the solution of the facility location problem. For instance, Varghese and Gladston (2016) applied the K-means algorithm to determine the optimal location of distribution centers in the United States and used the Euclidean distance as a function. Wang *et al.* (2018) developed a clustering-based approach to find the optimal locations for distribution centers in a set of potential facilities, allocate customers to the nearest facility, and design appropriate delivery routes. In their study, the authors segmented customers using the proximity coefficient and product preferences and employed the K-means method for customer clustering. Since their main challenge was to define the number of clusters (k), they used the silhouette coefficient to measure the performance of each possible value of k. Sabarish and Vidhya (2019) analyzed the location of schools, hospitals, and police stations in a delimited geographical region (Coimbatore city). For such purpose, they proposed an algorithm that used a dominating set and K-means to choose the facility and its corresponding cluster in the region. They validated each cluster using a series of metrics, including the Davies–Bouldin Index (DBI) and Dunn's index.

Clustering methods have also been used as a preliminary stage or as an initial solution to the facility location and routing problem. Oudouar *et al.* (2019) first determined the optimal location of customers using K-means and then planned routes from selected depots to a set of customers using Clarke and Wright's savings algorithm. In the same way, Santoso *et al.* (2021) applied center-based clustering as an initial stage, following which they formulated and solved individual vehicle routing problems for each resulting cluster. Sharma *et al.* (2017) proposed a hybrid approach that combines clustering and Mixed Integer Linear Programming (MILP). Their proposed method consists of two stages: the first stage uses K-means, and the second stage employs a MILP technique for each cluster to find the facility that produces the maximum profit. The results of their analysis show that, due to clustering, the average distance between the facility and the customer decreases significantly. Brimberg and Drezner (2019) used clustering to partition demand into a given number of subsets or groups that can be treated as smaller, independent subproblems. After clustering, the authors employed dynamic programming to determine the location of customers at each facility. Their proposed model assumes that each cluster has at least one facility exclusively assigned to it. Moreover, de Gusmão *et al.* (2020) developed a facility location model based on two methods: the first one uses K-means to find potential locations, and the second one employs the Maximum Covering Location Problem to select optimal locations. The authors used the model to determine the optimal placement of police stations in a Brazilian city based on crime occurrence. Faezy Razi (2019) clustered the maintenance stations of an oil refinery using K-means. In addition, the optimal number of clusters was calculated using

the silhouette coefficient, and the efficiency of each group of stations was estimated using Data Envelopment Analysis.

The K-means algorithm has been one of the most widely used clustering methods although it has undergone some modifications. Geetha et. al. (2009) added priority to the K-means algorithm as a measure for assigning customers to clusters. In their proposed approach, customers are assigned to the nearest cluster based on maximum demand and minimum distance. Hence, the customer with higher demand is assigned to the cluster first, and that with lower demand can be easily assigned to other clusters. Sahraeian and Kaveh (2010) developed a hybrid method that combines K-means and the Fixed Neighborhood Search (FNS) algorithm to solve a variant of the facility location problem known as the capacitated P-median problem. In this proposed hybrid method, since FNS is a local search algorithm, and it cannot provide an initial solution, the K-means algorithm is used for this task, and then, the FNS algorithm improves the quality of the obtained solutions. According to the author, the advantages of the proposed method are that it (i) omits unsuitable candidate sites, (ii) reduces the number of solutions, and (iii) avoids re-evaluating repeated solutions, which results in lower computational costs. Liao and Guo (2008) designed a clustering-based approach to solve a special version of the Capacitated Facility Location Problem and tested it in several scenarios. Their study evaluated performance using the average distance from site locations to their assigned facilities. The results of their experiments show that the proposed clustering-based approach can lead to near-optimal configurations of facility locations with fast convergences, regardless of whether the capacity of the facilities is sufficient or insufficient to cover the total demands. Sutanto *et al.* (2018) used the K-means algorithm to allocate customers to their designated facilities. In their proposed method, if a facility serves more customers than its capacity, reallocation takes place, which is done based on the average distance between customers and the available facilities. To evaluate the quality of the resulting clustering, the authors considered three aspects: connectedness, compactness, and spatial separation.

Although in a minor proportion, some studies have employed clustering methods other than K-means. For example, Sharma *et al.* (2021) proposed an optimization model in which facilities do not fulfill their service based on a radius. Instead, their proposed model considers a possible scenario in which topographic barriers must be handled. Their model consists of two stages: in the first stage, DBSCAN is used as the clustering algorithm, and, in the second stage, affinity propagation is employed to find the best location for a given facility. The overall goal was to reduce the connection distance between two points. Moreover, the model was validated on synthetic datasets of various sizes. Esnaf and Küçükdeniz (2009) developed a multi-facility location method aimed at minimizing distribution costs. Customers were first clustered using the fuzzy c-means algorithm. Each cluster was treated as an independent problem and solved using the center of gravity algorithm to locate the facilities. In a later work, Esnaf and Küçükdeniz (2013) presented a weighted fuzzy c-means algorithm for locating multiple facilities. Their proposed algorithm provides accurate solutions, eliminating the need to use a single-facility location method after clustering. It was tested on various datasets and compared to other methods (fuzzy c-means, center of gravity, and particle swarm optimization).

Table 1 shows different studies that employ clustering techniques to determine the location of facilities in different decision environments. Such studies were classified according to the method used by the author(s), the constraints, the objective function, and the environments in which the algorithms were tested.

Table 1 Classification of studies that use clustering techniques to determine the location of facilities

Author(s)	Type of decision		Method		Constraints				Objective function		Testing environment	
	Location	Location - assignment	Cluster techniques	Other techniques	No capacity	Capacity	Coverage	Others	Costs	Coverage	Real case	Test instances
Gülbay *et al.* (2021)	x		K-means	Integer Linear Programming		x		x	x		x	
Duong *et al.* (2021)		x	Fuzzy c-means K-means			x		x	x		x	
Rabbani *et al.* (2021)		x	K-means	Genetic Algorithms		x			x			x
Sharma *et al.* (2021)	x		DBSCAN	Density affinity propagation			x		x			x
Hidayat *et al.* (2020)	x		K-means			x			x		x	
Gülbay *et al.* (2021)	x		K-means	Integer Linear Programming		x		x	x		x	
Duong *et al.* (2021)		x	Fuzzy c-means K-means			x		x	x		x	
Rabbani *et al.* (2021)		x	K-means	Genetic Algorithms		x			x			x

Table 1 Classification of studies that use clustering techniques to determine the location of facilities (Con't)

Author(s)	Type of decision		Method		Constraints				Objective function		Testing environment	
	Location	Location - assignment	Cluster techniques	Other techniques	No capacity	Capacity	Coverage	Others	Costs	Coverage	Real case	Test instances
Sharma *et al.* (2021)	x		DBSCAN	Density affinity propagation			x		x			x
Hidayat *et al.* (2020)	x		K-means			x			x		x	
Hu *et al.* (2020)		x	Genetic Algorithms - Fuzzy c-means K-means	Genetic Algorithms	x			x	x		x	
Cai *et al.* (2020)		x	K-means	Center of gravity	x				x		x	
de Gusmão *et al.* (2020)	x		K-means	Maximum Covering Location Problem			x		x		x	
Gupta *et al.* (2019)		x	Fuzzy c-means K-means	Particle swarm optimization - Bat algorithm- Colony optimization		x			x			x
Sabarish and Vidhya (2019)		x	K-means	Domination set	x			x	x	x	x	
Faezy Razi (2019)		x	Fuzzy c-means	Integer Linear Programming		x				x		x
Wu *et al.* (2019)		x	K-medians									x
Simić *et al.* (2017)		x	Fuzzy c-means	Genetic Algorithms		x			x			x
Cabria and Gondra (2017)		x	K-means			x			x		x	
Varghese and Gladston, (2016)		x	K-means		x				x			x
Jiang *et al.* (2016)		x	K-harmonic means	Particle swarm optimization - Bat algorithm		x			x		x	
Esnaf and Küçükdeniz (2013)		x	Fuzzy c-means		x				x		x	
Küükdeniz *et al.* (2012)	x		Fuzzy c-means	Convex programming		x			x			x
Sahraeian and Kazemi (2011)		x	Fuzzy c-means	Fuzzy set	x				x			x
Sahraeian and Kaveh (2010)	x		K-means	FNS algorithm		x			x			x

The literature on logistics in the newspaper industry has been rather scarce. Most studies have focused on vehicle route planning and considered other aspects of production in the supply chain (e.g., Boonkleaw *et al.*, 2009; Chiang *et al.*, 2009; Kamble *et al.*, 2017; Osaba *et al.*, 2017; Ree & Yoon, 1996; Russell *et al.*, 2008). We found a few papers that addressed facility location in the newspaper industry. For instance, Cunha and Mutarelli (2007) investigated the production and distribution of a newsmagazine in Brazil. They developed a mixed-integer programming model to simultaneously determine the optimal number and location of the facilities, demand allocation, production sequencing, and modes of transport. Their proposed model achieved a cost reduction of about 7.1%, which is significant for this type of product. Cunha *et al.* (2021) studied how to optimize a three-echelon newspaper production and distribution network. Their goal was to determine the optimal number and location of distribution centers and synchronize production schedules with last-mile delivery. For facility location, they first used K-means to cluster customers and then employed a multi-layer algorithm to optimize the

location of the distribution centers. Their proposed approach yielded a 28% reduction in distribution costs.

In order to contribute to the existing literature in the field, this study presents a hybrid strategy based on clustering techniques to solve the problem of locating facilities without capacity constraints. For such purpose, different clustering techniques are compared and the goal to reduce the total travel distance in the distribution of newspapers in the Area Metropolitana del Valle de Aburrá in Medellín, Colombia. Importantly, besides proposing an approach for the location of micro-hubs in the newspaper industry, we also contribute by addressing a problem with real data while considering several practical and complex aspects when it comes to the search for related information.

3. PROBLEM DESCRIPTION

In Colombia, there are around 80 newspaper and magazine publishers and other information media companies that are part of the Asociación de Medios de Información (AMI by its Spanish acronym). According to a report by Dinero magazine, some of the country's most representative information media have experienced a sustained drop in their net profits (Garzón, 2019). The newspaper industry in Colombia derives 68% of its revenues from advertising, 27% from circulation, and the remaining 9% from services provided to third parties. Nevertheless, according to data from the Asociación Nacional de Medios de Comunicación (Asomedios by its Spanish acronym), advertising revenue amounted to COP 4.2 trillion in 2012 and decreased to COP 1.8 trillion in 2018. In addition to this, newsrooms have been reduced by 40% over the last five years due to low profitability margins. From 2011 to 2016, the number of employees decreased by 40%. In 2017, however, there was a slight stabilization, causing a 5.5% increase in employability in 2018 (Zambrano, 2020).

Although digital news distribution has had an important growth in Colombia, people prefer traditional media to be informed. In the case of newspapers, 73% of Colombians prefer to read the print edition even though digital versions are often free and easy to access (González, 2016). This is consistent with the results obtained in a 10-country study conducted by Two Sides (an international organization dedicated to promoting the responsible production, use, and recovery of paper and prints worldwide), which reported that consumers prefer to read the print version of books (72%), magazines (72%), and newspapers/news (55%) over digital alternatives (Two sides, 2019).

For several years, the regional press in Colombia has been experiencing an advertising crisis in traditional print media. This situation has been exacerbated by the COVID-19 pandemic, which has led to a 40% to 80% decline in advertising (according to estimates from the AMI) and has made it impossible to distribute newspapers door-to-door (Vita, 2020). The pandemic has indeed changed consumption habits, as consumers currently prefer to read the news digitally to avoid interaction with delivery personnel. This has aggravated the crisis and directly impacted newspaper consumption production, as production and distribution personnel must comply with the stringent confinement measures. Given the unfavorable forecasts of the print press and after the disastrous wave of the COVID-19 pandemic, the national press continues to keep high demand margins for paid and free subscribers (Franco, 2022).

As indicated in the studies by Chiang *et al.* (2009) and Wang *et al.* (2021), daily newspaper distribution is performed by combining truck and motorcycle routes that start at a single distribution center located at the production facility. Each delivery node must be served before the 4:00 a.m. delivery deadline to allow individual carrier (motorcycles) routes to complete their deliveries before the 6:00 a.m. delivery deadline. Newspapers are then transferred from the distribution center to an informal transit point – micro-hubs (a service station or a store's parking lot), where a cross-docking operation takes place. In this operation, which was studied by Cunha *et al.* (2021), distribution centers are usually located in public spaces; therefore, no fixed facility and location costs are incurred.

In distribution logistics, transportation costs are one of the main sources of total facility operating expenses and are directly affected by distance (Sitek *et al.*, 2021). Hence, the efficiency of the distribution network can be improved by reducing the distance between customers (Wang *et al.*, 2015). As demonstrated by the existing literature in the field, customer clustering is a viable strategy for optimizing the configuration of logistics networks (Wang, Zhang, *et al.*, 2018). In this study, we consider a medium-sized newspaper that circulates in Medellín city and parts of the department of Antioquia in Colombia with the purpose to seek the location of micro-hubs and thus reduce the distance between distribution points and potential customers.

Inspired by previous studies, particularly that of Cunha *et al.* (2021) and Wang *et al.* (2015) for this specific application, we highlight the following characteristics of the problem of locating micro-hubs:

- Informal transshipment points are assumed to have no capacity constraints and can, therefore, meet all the demand allocated to them.
- Since micro-hubs are typically located in public spaces (e.g., a public area used for free parking or a service station), no costs are incurred for their use, and they can be easily relocated.
- Micro-hubs can be located at any continuous point on the city map.
- Demand is deterministic.
- Each demand point has a fixed location, while hub locations can be rearranged to reduce travel distance.
- Cost is defined as the average distance between each demand point and its designated facility.
- The total capacity of the facilities is assumed to be the same and high enough to meet all demands.

4. METHODOLOGY

Unsupervised machine learning methods have demonstrated their ability to identify underlying clusters in data. The strategy we propose here is to use clustering techniques to find groups of customers that are spatially close to each other but distant from the other groups. In this way, we expect to determine the number and location of micro-hubs whose position reduces the distance to the final customer. The number and location of the micro-hubs are determined based on the validation properties of the clusters, such as compactness and spatial separation. In addition, as

mentioned above, in this study, micro-hubs do not have large physical requirements and their costs are low. Therefore, we do not define restrictions on the number of locations, letting them be defined only from the cluster properties. **Figure 1** describes the steps followed to compare the different clustering techniques used here and determine the location of the micro-hubs. The proposed strategy begins with data preparation, followed by the selection of clustering methods and metrics for cluster validation and comparison. Then, based on the obtained results, the most appropriate model(s) for the intended purpose is selected, or multi-stage or hybrid techniques are evaluated.

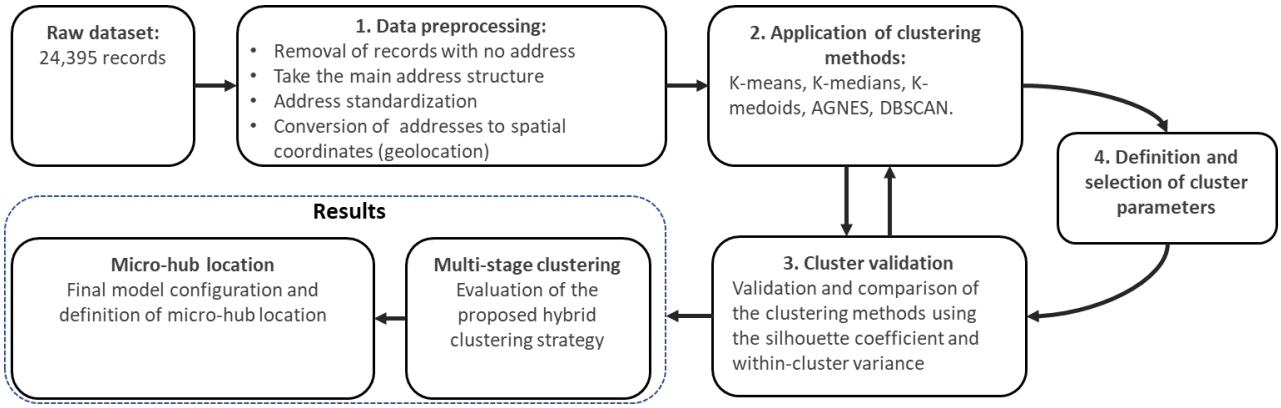

Figure 1 Proposed clustering strategy

4.1 Data Preprocessing

Historical data on customers' geographic locations were collected. For our research and due to the limited availability of data, we only considered the newspaper distribution network in Medellín and its neighboring suburban areas. As a result, we obtained a total of 24,395 records (customers). To optimize the geolocation process, we verified the uniqueness of customer addresses and found that multiple customers could have the same address (without including their additional information) because they resided in buildings. Thus, by considering unique addresses, we obtained a total of 14,080 records. These unique addresses were standardized and geolocated using the Google Maps API to obtain their geographic coordinates (latitude and longitude). We were able to geolocate 92% of the records, which resulted in a database containing 12,954 complete records. **Figure 2** shows the geographic distribution of the customer population. Importantly, just the latitude and longitude variables of each customer will be used for the proposed location exercise.

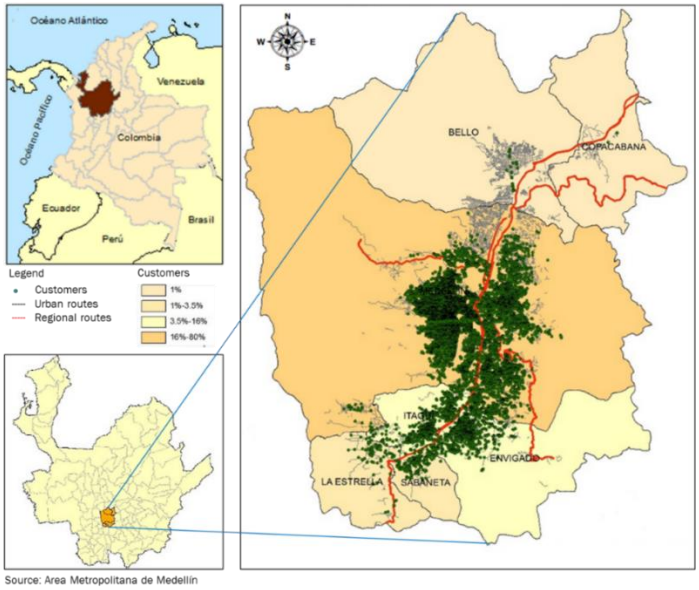

Figure 2 Geographic distribution of the customer population

4.2 Clustering Methods

In this study, we employed five clustering methods (K-means, K-medians, K-medoids, AGNES, and DBSCAN), which encompass partitioning, hierarchical, and density-based models, thus covering a broad spectrum of clustering alternatives. As shown below, each method has advantages and disadvantages in terms of clustering. Therefore, besides comparing them, we combined some of them to make up for their drawbacks. The methods under analysis consider exclusive allocation, i.e., when a customer is assigned to a cluster, it will be served exclusively by such micro-hub and will not belong to any other or multiple facilities.

K-means: It is a partitioning method and one of the most widely used clustering algorithms (Han *et al*., 2022; Manoharan *et al*., 2016). It is characterized by its simplicity and speed (Celebi *et al*., 2013) and seeks to minimize the

sum of the squared Euclidean distances between the objects in a cluster and their centroid, which corresponds to the mean value of the objects in the cluster. Simplicity, efficiency, and stability are some of its main features; however, it requires the number of clusters to be specified in advance (although this may not be a disadvantage if the number of facilities has been previously defined). It is not appropriate for identifying clusters with nonconvex shapes and of varied sizes (Han *et al*., 2022). As it uses the mean value as the centroid, it is sensitive to outliers. Also, it may be highly sensitive to the initial choice of cluster centers, which may lead to unstable results (Cardot *et al*., 2012).

K-medians: It is a variant of the K-means method. It uses the median, instead of the mean, as the centroid of a cluster, as well as the Manhattan distance (also known as the taxicab metric). Contrary to K-means, which uses the squared differences between a point and its centroid, K-medians consider their absolute difference. These characteristics make K-medians more resistant to outliers although it shares the disadvantages of partitioning methods (Cardot *et al*., 2012).

K-medoids: It is also a partitioning method and another variant of the K-means algorithm. In this method, the centroid of each cluster is an object from the current dataset. This object is known as the medoid, and it is a representative object of a given cluster. The remaining objects are assigned to the most similar medoid. This similarity is determined by the absolute difference between each object and its corresponding medoid (Han *et al*., 2022). The most popular example of this method is the Partition Around Medoids (PAM) algorithm, which is effective on small datasets but inefficient on large datasets (Jinyin *et al*., 2017). Like K-medians and while maintaining the other characteristics of K-means, K-medoids have the advantage of being resistant to outliers and can be used with other dissimilarity measures (Maechler *et al*., 2018).

AGNES (Agglomerative Nesting): It is an agglomerative hierarchical method (Maechler *et al*., 2015), in which each observation forms its cluster first, and then pairs of clusters are merged until all the observations are part of the same cluster (Boongoen and Iam-On, 2018). A particular feature of this way of forming new clusters is that if two clusters are merged at a given level, they are already hierarchically grouped in the rest of the levels. The number of clusters is obtained by cutting the final tree structure (dendrogram) where the tree levels are highly dissimilar. In this study, we used AGNES along with the single-linkage criterion (nearest neighbor) and the Euclidean distance, which allows for the formation of clusters under the density-based approach (Handl *et al*., 2005). Besides not requiring the number of clusters to be defined in advance, this method has the advantage of finding individuals with high local proximity and clusters of various sizes. This characteristic, however, makes it sensitive to noise at higher levels in the hierarchy (Han *et al*., 2022).

DBSCAN (Density-Based Spatial Clustering of Applications with Noise): The DBSCAN algorithm is used to identify non-spherical-shaped clusters or clusters of arbitrary shapes. It looks for regions with a high concentration of points and separated by low-density or scattered areas (Han *et al.*, 2022). It requires two user-defined parameters: (i) a radius (ε), which determines the neighborhood around a point (o_i), and (ii) minPoints, which defines the minimum number of points within the neighborhood of o_i for it to be considered a core or high-density point (Han *et al*., 2022; Tran *et al.*, 2013). The cluster of points within the neighborhood is expanded by checking all new points and verifying if they also have more minPoints at epsilon, extending the cluster recursively if so (Dudik *et al.*, 2015). Hence, all the points that are part of the same cluster are densely connected (Kumar & Reddy, 2016).

4.3 Performance Metrics

To evaluate the quality of the clustering results, a cluster validation strategy that considers the various methods employed here must be defined. Clusters can be validated intrinsically or extrinsically (Han *et al.*, 2022). Extrinsic cluster validation is considered a supervised process in which the cluster to which the object belongs is known in advance. In intrinsic cluster validation, such information is not available, and the validation depends on the resulting clustering and the underlying information of the dataset. In this study, since there is no prior labeling of each observation and because this would not be either the case in a typical location process, we used intrinsic cluster validation methods.

According to the classification made by (Handl *et al.*, 2005), cluster validation methods are closely related to the objectives of each clustering method; hence, they can be classified considering the following three aspects: compactness, connectedness, and spatial separation. Compactness-based methods use within-cluster variance to assess homogeneity between points from the same cluster. Connectedness-based methods evaluate how well the formed cluster adheres to the idea that a point must be matched with its nearest neighborhood. Spatial separation-based methods employ within-cluster variance to assess how differentiated or separated clusters are.

Cohesion between objects in the same cluster and distinctive separation between clusters are two desirable characteristics of any clustering result. To measure the performance of all the clustering methods employed in this study, we used two internal evaluation metrics: the silhouette coefficient (Sutanto *et al*., 2018) and intra-cluster variance. Importantly, since each method is designed to produce better results on either of the two fronts, the best method will be the one with a greater balance (between the two metrics) in solving the location problem.

The silhouette coefficient: It has the advantage of simultaneously measuring compactness and spatial separation. Equation (1) shows how this coefficient is calculated. In the equation, a(i) is the average distance between an observation (i) and the other observations in the same cluster, and b(i) denotes the average distance between an observation (i) and the observations in the nearest cluster to which it does not belong (Rousseeuw, 1987). The silhouette coefficient can take a value between -1 and 1. A value close to 1 indicates that the observation is closer to the observations of its own cluster than to those that do not belong to it, which means that it is in the right cluster.

$$S(i) = \frac{\{b(i) - a(i)\}}{max\{a(i) - b(i)\}} \quad (1)$$

Total within-cluster variance: Also known as within-cluster sum of squares, it is a measure of cluster compactness. It is calculated using Equation (2), where C is the number of clusters; n_k, is the number of elements in cluster k, and μ_k, is the cluster's centroid. First, the Euclidean distance from each element in a cluster to its centroid is computed, and then these values are totaled for each cluster. A cluster with a small within-cluster variance value is considered more compact than a cluster with a large value.

$$Within-cluster\ variance = \sum_{k=1}^{C} \sum_{i \in k}^{n_k} (x_i - \mu_k)^2 \quad (2)$$

4.4 Determination of The Number of Clusters

For K-means, K-medians, and K-medoids, which use the number of clusters as an input parameter, an iterative process must be previously performed, in which the number of clusters is changed from 2 to 30, and the silhouette coefficient and intra-cluster variance are calculated at each iteration. This process yields the cluster number configurations that will be used to compare the methods. For hierarchical methods, the number of clusters is defined by inspecting the dendrogram to identify the level in the hierarchy with a significant increase in the similarity measure between the grouped levels.

In cluster analysis, the elbow method is a technique used to analyze and test consistency and designed to help determine the appropriate number of clusters in a dataset. It consists in testing a range of values of a given parameter, graphically representing the results obtained with each value, and identifying the point on the curve where the improvement is no longer significant (the likelihood principle) (Yuan and Yang, 2019). This method calculates total within-cluster variance based on the number of clusters. It chooses, as optimal, the value that enables it to add more clusters as soon as it produces a minimum improvement. In the case of the silhouette coefficient, the points where this coefficient achieves the best performance are identified. The best cluster number configuration for all the methods is thus the point where the best balance between the two metrics is struck.

5. RESULTS

5.1 Comparison of The Individual Clustering Methods

5.1.1 Determination of The Number of Clusters

To define the number of clusters (k) for K-means, K-medians, and K-medoids, we plotted the estimated silhouette coefficient and intra-cluster variance for each value of k (from 2 to 30), as shown in **Figure 3**. Regarding the silhouette coefficient (**Figure 3a**), the highest performance peaks are observed when k=2, k=4, and k=9 in most methods. As for intra-cluster variance (**Figure 3b**), an inflection point is observed when k=9, and we may consider another one when k=26. After finding a balance between the results of the two metrics, 4 and 9 were chosen as the values of k with the best performance in both metrics.

Figure 4 presents the resulting dendrogram of the hierarchical clustering algorithm used in this study (AGNES with single-linkage). From it, we may conclude that the most suitable number of clusters is obtained by cutting it at an approximate height of 0.02, which results in 4 clusters. In the case of DBSCAN, which does not require a predefined number of clusters, the algorithm predicted 4 clusters as the optimal value.

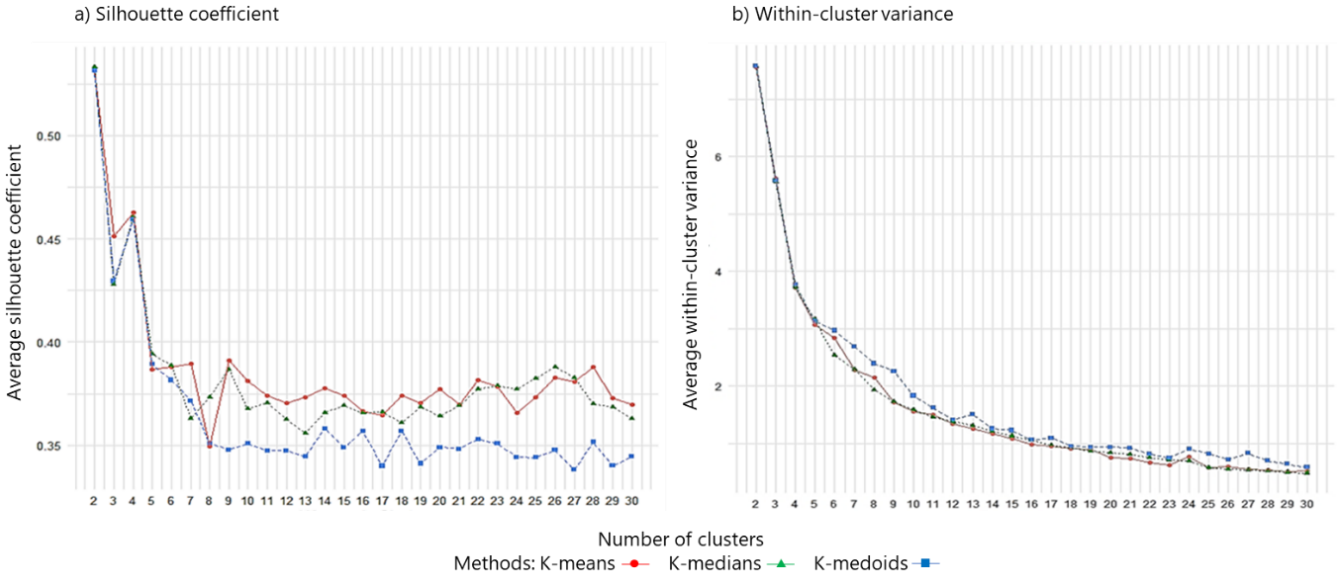

Figure 3 Evaluation of K-means, K-medians, and K-medoids at different cluster number configurations (From K= 2 to K=30)

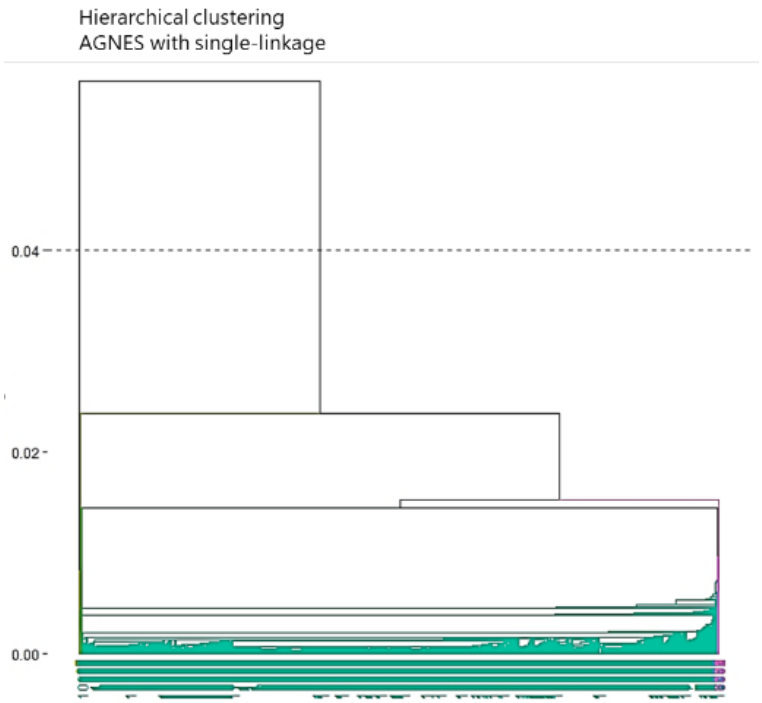

Figure 4 Results of the hierarchical clustering method (AGNES with single-linkage)

5.1.2 Cluster Comparison

Table 2 shows the estimated performance metrics of the different clustering methods compared in this study for k=4 and k=9. As observed, DBSCAN is the only method with no results for k=9 because its very design does not allow its performance to be evaluated in several clusters other than the predicted one (k=4, in this case). As for AGNES (the hierarchical method), which enables users to define the number of clusters based on their criteria, we also measured its performance for k=9 for consistency purposes.

Table 2 Performance of the clustering algorithms

Method	Average silhouette coefficient		Total within-cluster variance	
	K=4	K=9	K=4	K=9
K-means	0.4624	0.3572	3.72	1.775
K-medoids	0.4606	0.3865	3.738	1.725
K-medians	0.45	0.3546	3.77	1.765
AGNES	0.76	0.379	17.9	17.85
DBSCAN	0.2198	-	17.85	-

According to the information in **Table 2**, AGNES obtained the best silhouette coefficient with k=4. The best within-cluster variance values, however, were obtained by the partitioning methods with k=9, with K-medoids showing the best performance but no significant differences from the other methods. AGNES (with single-linkage) and DBSCAN produced the worst results in terms of within-cluster variance, which is not surprising given that these methods are not designed to form compact clusters.

Figure 5 presents the spatial distribution of the clusters formed by the partitioning methods for k=4. As observed, the clusters appear to be similar in size in the region with a high concentration of points and to be altered by the inclusion of outliers lying beyond the boundary formed by the mass of points. **Figure 6** shows the spatial distribution of the clusters formed by AGNES (with single-linkage) and DBSCAN for k=4. As can be seen, both methods generated a huge cluster that groups together all the points of the central region that comprise the main mass of points. The other clusters (in both methods) correspond to outliers from the north and south that move away from the high-density region. This result is consistent with the characteristics of both methods, which struggle to identify clusters when the groups are not separated by a significant distance and thus end up grouping all the points into the same cluster. This scenario would not be a feasible or practical solution for our case study because all customers in the Área Metropolitana del Valle de Aburrá would be served from a single distribution center.

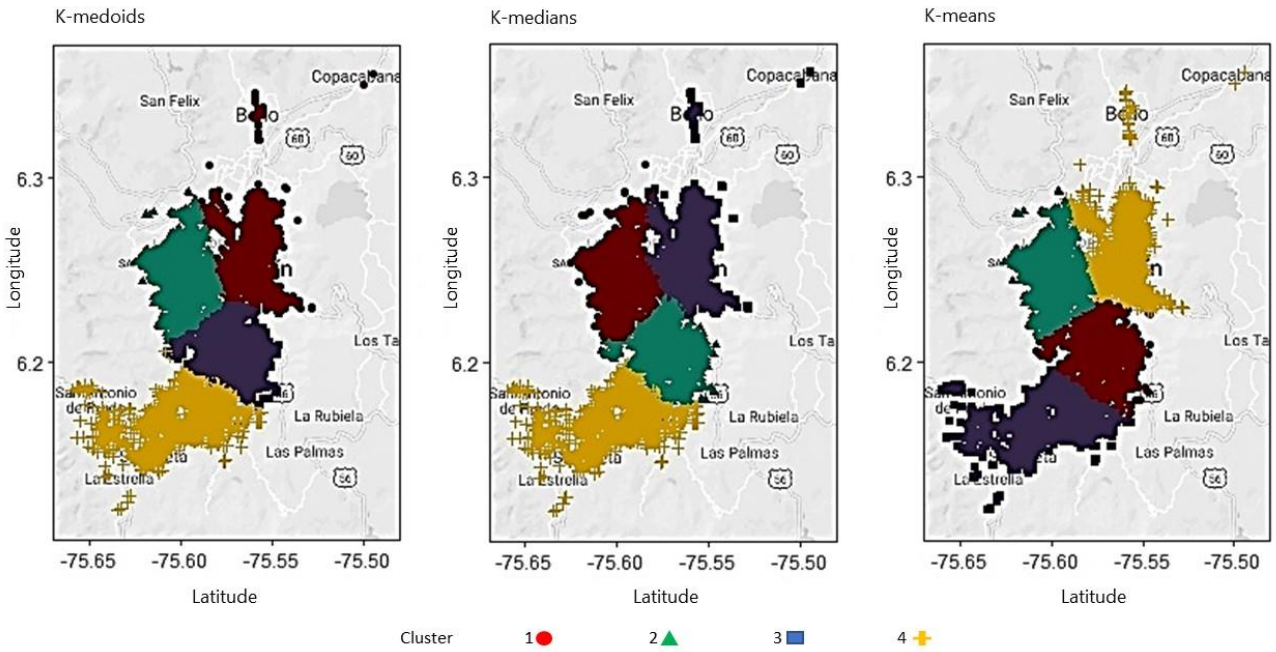

Figure 5 Distribution of the clusters formed by the partitioning algorithms

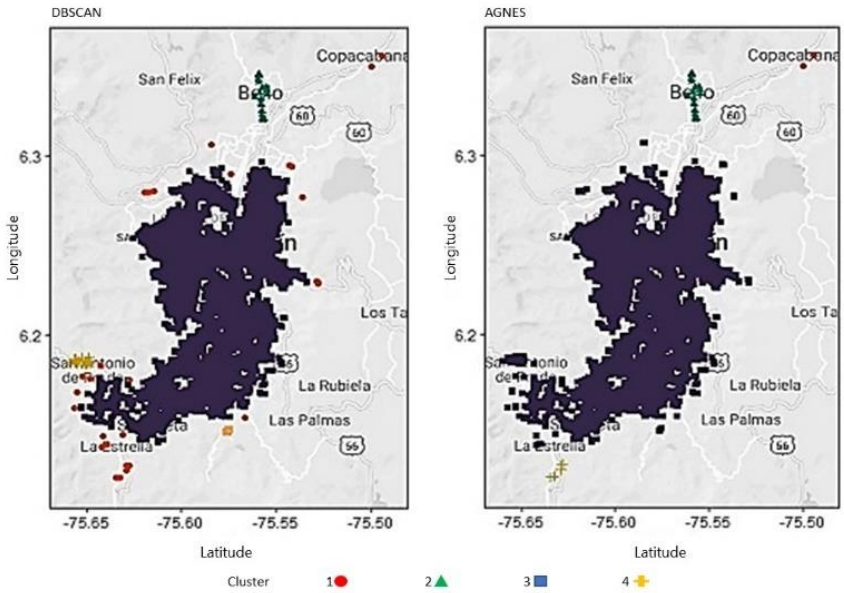

Figure 6 Distribution of the clusters formed by the DBSCAN and AGNES algorithms

5.2 Hybrid Clustering Strategy

According to the results presented, no method alone was able to produce satisfactory results in the two-performance metrics used in this study. In some methods, the resulting clustering can be affected by outliers lying too far from the main area of the scope of the cluster; in the other methods, all the points are assigned to the same cluster due to their high density, and outliers are assigned to independent clusters. We propose a hybrid clustering strategy (see **Figure 7**) that exploits the advantages of density-based and partitioning methods. In this proposed strategy, (i) the single-linkage method is first used to identify scattered areas and high-density ones, and then (ii) K-medoids is used to cluster high-density regions once the optimal number of clusters has been determined.

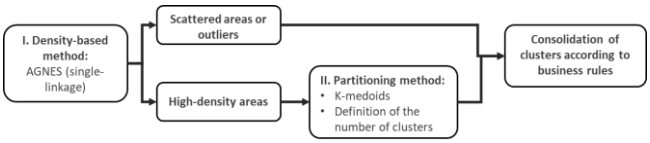

Figure 7 Proposed hybrid two-stage clustering

Table 3 shows the results of the proposed hybrid model. First, according to the results in **Table 2**, AGNES obtained the best silhouette with a configuration of k=4 clusters (**Figure 6**-AGNES): A high-density cluster (Cluster #3), two scattered clusters (Clusters #2 and #4), and one outlier (Cluster #1). Then, once Cluster #3 was identified as the high-density cluster, K-medoids was used over this area, resulting in an optimal partition of 9 clusters (k=9). In addition, as a business rule, a service center must have a minimum of 30 customers to be served. Thus, among the dispersed and atypical clusters, only one met this condition;

hence, it remained as an independent cluster, while the others were reassigned to the nearest cluster, resulting in a total number of 10 clusters. As expected, this affected the compactness of the receiving cluster as shown in **Table 3**.

For comparison purposes, the other partitioning methods were also evaluated individually for k=10. The results showed that the proposed model outperformed the other techniques, as it obtained the highest silhouette coefficient (0.3874) and one of the lowest intra-cluster variance values (1.614). Therefore, we may say that more compact and connected clusters were obtained using the approach proposed in this study, whose superiority is based on its ability to produce compact classes and isolate scattered sets of customers. **Figure 8** illustrates the final spatial distribution of the ten clusters, which, for our case study, will result in 10 micro-hubs that will meet the total demand.

Table 3 Performance of the proposed hybrid model and its comparison with the other partitioning methods for k=10

Metric	Silhouette coefficient	Within-cluster variance
	k=10	
Hybrid method	0.3874	1.614
K-means	0.3636	1.6337
K-medoids	0.3678	1.5815
K-medians	0.3671	1.6495

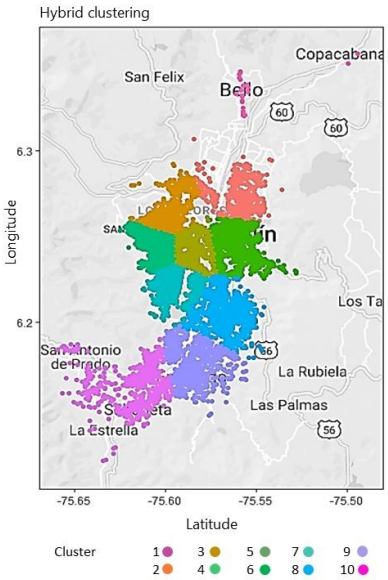

Figure 8 Spatial distribution of the clusters formed by the proposed hybrid clustering

To validate the quality of the clusters obtained with the proposed method, the distances between customers' locations and the centroid of the cluster to which they are assigned were analyzed using the haversine formula (Sinnott, 1984). For better visualization and analysis, **Figure 9** presents the frequency histogram of the distances in kilometers. The average distance in all the clusters is approximately 861 meters (i.e., less than 1 kilometer), which means that customers are, on average, less than nine blocks from their distribution center (the centroid). In logistics terms, this distance can be quickly travelled by the vehicles delivering the product at their average speed. Only Cluster 1 presented a high intra-cluster variance because it received more scattered points. Also, 90% of the observations were at a maximum distance of 1600 meters from their corresponding centroid, and only 5% were at a distance between 2200 and 9300 meters approximately. This 5% of observations were mostly grouped in Cluster 1 and, to a lesser extent, in Cluster 10.

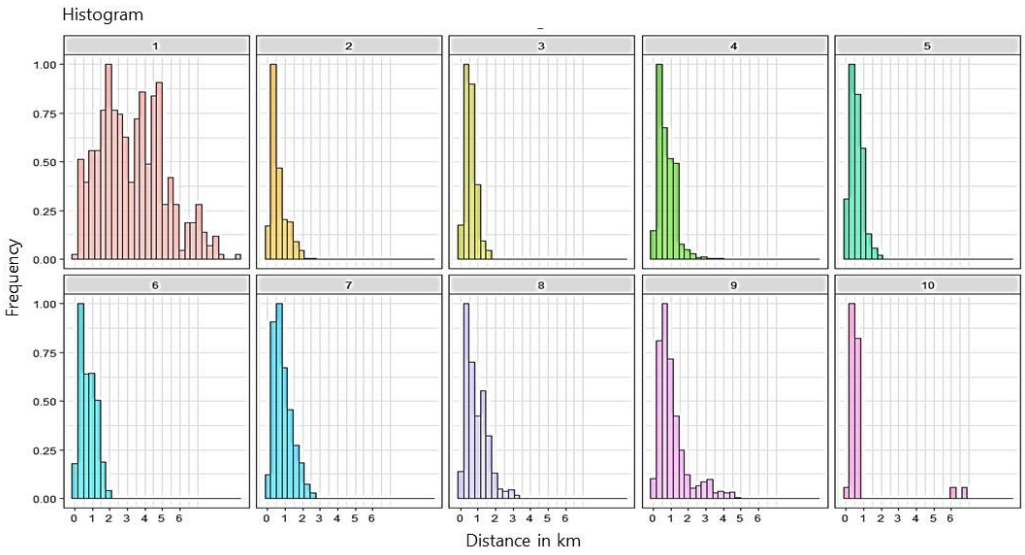

Figure 9 Histogram of the distances between customers' locations and the micro-hubs obtained with the proposed hybrid strategy

6. CONCLUSIONS

In this study, we presented a clustering-based approach to determine the number and location of newspaper micro-hubs using a real dataset containing the spatial location of customers. We considered, as a case study, the distribution of printed newspapers, which, according to our literature review, has been little studied using the clustering methods we employed here. Although different clustering techniques have been employed as a starting point to develop other models or as final models in the facility location problem, K-means has been the most widely used without sufficient justification. To select the best clustering results, we compared three partitioning methods (K-means, K-medians, K-medoids), a hierarchical method (AGNES), and a density-based method (DBSCAN). According to our results, we found that K-means is not properly the best-performing clustering method. The method failed to create a connected neighborhood and can be greatly affected by the presence of outliers. In general, it can be said that none of the clustering methods managed to simultaneously produce clusters with compact and connected points.

According to our findings, AGNES (with single-linkage), which is mostly a density-based method, provided the best silhouette coefficient. However, it had a poor performance in terms of cluster compactness. For their part, the three partitioning methods, especially K-medoids, produced the best results in terms of compactness, but their performance was hampered by outliers or widely scattered regions. In general, these results are consistent with the performance expected from each clustering technique. For instance, the density-based method identified outliers, as well as scattered and distant regions, but grouped the entire high-density area into a single cluster. The partitioning methods, on the contrary, created a uniform and convex-shaped clusters but included points from scattered areas lying far from the main mass of points. Therefore, considering that no method alone provided a satisfactory solution, we proposed a hybrid clustering model to solve the facility location problem. This proposed hybrid model combines the qualities of density-based and partitioning methods and consists of two stages. The first stage uses the AGNES (with single-linkage) algorithm, and the second stage employs the K-medoids algorithm to cluster high-density regions.

When validating the quality of the clusters, the proposed hybrid method was found to outperform each clustering method. It managed to create more compact clusters, with 90% of customers being within 1.6 kilometers from the centroid. This centroid corresponds to the proposed location for the micro-hubs, which would minimize the travel distance between customers and their distribution center. The solution provided by our proposed method would result in 10 micro-hubs throughout the Área Metropolitana del Valle de Aburrá, which is 1,157 km2 in size. As a result, the distribution network could be better managed, and future improvements could be implemented, considering that the product must be distributed in a short time window in the morning for it to arrive at the customers' location on time.

Compared to existing studies into facility location for newspaper distribution, the benefits of our proposed approach can be reformulated to simultaneously solve facility location and demand allocation problems. Importantly, the proposed approach has limitations and provides opportunities for further research. Future studies, for instance, could focus on analyzing the time-space distribution of customers, considering their dynamic behavior, as well as predicting their future spatial distribution to make the facility location process proactive rather than reactive. The methodology outlined in the context of this study exhibits the capacity for expansion and application across a variety of product distribution scenarios. Furthermore, its versatility extends to the incorporation of variables like customer purchase frequency and service reliability levels, while also accounting for the uncertainty associated with delivery fleet travel time.

Given the anticipated reduction in print newspaper readership attributable to the growing prevalence of digital media, it becomes imperative to examine alternative strategies for mitigating distribution expenses within forthcoming distribution models. One viable approach may involve the integration of consolidation and delivery schemes, such as the promotion of mobile depots, facilitating the customer assignment and vehicle fleet synchronization. Additionally, the exploration of collaborative delivery schemes, including crowdsourcing and Mobility as a Service (MaaS), has the potential to yield cost-saving benefits.

Acknowledgements: This research was partially funded by Instituto Tecnológico Metropolitano (project P20239) and iMOVE CRC and supported by the Cooperative Research Centres program, an Australian Government initiative (Project 5-036).

REFERENCES

Adetiloye, T., & Pervez, G. (2015). A Macro and Micro-Level Evaluation of Stakeholders' Collaboration for Sustainable City Logistics Operations. *Operations and Supply Chain Management: An International Journal*, 8(2), pp. 90–100. https://doi.org/10.31387/oscm0200147.

Bergmann, F., Wagner, S., & Winkenbach, M. (2020). Integrating First-Mile Pickup and Last-Mile Delivery on Shared Vehicle Routes for Efficient Urban E-Commerce Distribution. *Transportation Research Part B: Methodological*, 131, pp. 26–62. https://doi.org/10.1016/j.trb.2019.09.013.

Boongoen, T., & Iam-On, N. (2018). Cluster Ensembles: A Survey of Approaches with Recent Extensions and Applications. *In Computer Science Review*, 28, pp. 1–25. https://doi.org/10.1016/j.cosrev.2018.01.003.

Boonkleaw, A., Suthikarnnarunai, N., & Srinon, R. (2009). Strategic Planning and Vehicle Routing Algorithm for Newspaper Delivery Problem: Case study of Morning Newspaper, Bangkok, Thailand. *Lecture Notes in Engineering and Computer Science*, 2179(1), pp. 1067–1071.

Brimberg, J., & Drezner, Z. (2019). Solving Multiple Facilities Location Problems with Separated Clusters. *Operations Research Letters*, 47(5), pp. 386–390. https://doi.org/10.1016/j.orl.2019.07.007.

Browne, M., Allen, J., Nemoto, T., Patier, D., & Visser, J. (2012). Reducing Social and Environmental Impacts of Urban Freight Transport: A Review of Some Major Cities. *Procedia - Social and Behavioral Sciences*, 39, pp. 19–33. https://doi.org/10.1016/j.sbspro.2012.03.088.

Cabria, I., & Gondra, I. (2017). Potential-K-Means for Load Balancing and Cost Minimization in Mobile Recycling Network. *IEEE Systems Journal*, 11(1), pp. 242–249. https://doi.org/10.1109/JSYST.2014.2363156.

Cai, C., Luo, Y., Cui, Y., & Chen, F. (2020). Solving Multiple Distribution Center Location Allocation Problem Using K-Means Algorithm and Center of Gravity Method Take Jinjiang District of Chengdu as An Example. *IOP Conference*

Series: Earth and Environmental Science, 587(1), pp. 1–6 https://doi.org/10.1088/1755-1315/587/1/012120.

Cardot, H., Cénac, P., & Monnez, J. (2012). A Fast and Recursive Algorithm for Clustering Large Datasets with K-Medians. *Computational Statistics and Data Analysis*, 56(6), pp. 1434–1449. https://doi.org/10.1016/j.csda.2011.11.019.

Celebi, M., Kingravi, H., & Vela, P. (2013). A Comparative Study of Efficient Initialization Methods for the K-Means Clustering Algorithm. *Expert Systems with Applications*, 40(1), pp. 200–210. https://doi.org/10.1016/j.eswa.2012.07.021.

Chiang, W., Russell, R., Xu, X., & Zepeda, D. (2009). A Simulation/Metaheuristic Approach To Newspaper Production and Distribution Supply Chain Problems. *International Journal of Production Economics*, 121(2), pp. 752–767. https://doi.org/10.1016/j.ijpe.2009.03.001.

Cunha, C., Guazzelli, C., Yoshizaki, H., Masteguim, R., Colacioppo, D., & Ajzenberg, M. (2021). A Multi-Stage Approach for Optimizing the Three-Echelon Joint Newspaper Distribution Network of Two Major Publishers in São Paulo, Brazil. *Case Studies on Transport Policy*, 9(3), pp. 1073–1083. https://doi.org/10.1016/j.cstp.2021.05.008.

Cunha, C., & Mutarelli, F. (2007). A Spreadsheet-Based Optimization Model for The Integrated Problem of Producing and Distributing A Major Weekly Newsmagazine. *European Journal of Operational Research*, 176(2), pp. 925–940. https://doi.org/10.1016/j.ejor.2005.06.065.

de Gusmão, A., da Costa Borba, B., & Clemente, T. (2020). Management Information System for Police Facility Location. In: Moreno-Jiménez, J., Linden, I., Dargam, F., Jayawickrama, U. (eds) Decision Support Systems X: Cognitive Decision Support Systems and Technologies. ICDSST 2020. *Lecture Notes in Business Information Processing*, 384, pp. 86–98. https://doi.org/10.1007/978-3-030-46224-6_7.

Dudik, J., Kurosu, A., Coyle, J., & Sejdić, E. (2015). A Comparative Analysis of DBSCAN, K-Means, and Quadratic Variation Algorithms for Automatic Identification of Swallows from Swallowing Accelerometry Signals. *Computers in Biology and Medicine*, 59, pp. 10–18. https://doi.org/10.1016/j.compbiomed.2015.01.007.

Duong, Q., Nguyen, D., & Nguyen, Q. (2021). Hub and Spoke Logistics Network Design for Urban Region with Clustering-Based Approach. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-79457-6_51.

Duong, Q., Nguyen, D., & Nguyen, Q. (2021). Hub and Spoke Logistics Network Design for Urban Region with Clustering-Based Approach. In: Fujita, H., Selamat, A., Lin, J.CW., Ali, M. (eds) Advances and Trends in Artificial Intelligence. Artificial Intelligence Practices. IEA/AIE 2021. *Lecture Notes in Computer Science*, pp. 598–605. https://doi.org/10.1007/978-3-030-79457-6_51

Esnaf, Ş., & Küçükdeniz, T. (2009). A Fuzzy Clustering-Based Hybrid Method for a Multi-Facility Location Problem. *Journal of Intelligent Manufacturing*, 20(2), pp. 259–265. https://doi.org/10.1007/s10845-008-0233-y.

Esnaf, Ş., & Küçükdeniz, T. (2013). Solving Uncapacitated Planar Multi-Facility Location Problems by A Revised Weighted Fuzzy C-Means Clustering Algorithm. *Journal of Multiple-Valued Logic and Soft Computing*, 21(1–2), pp. 147–164.

Faezy Razi, F. (2019). A Hybrid DEA-based K-means and Invasive Weed Optimization for Facility Location Problem. *Journal of Industrial Engineering International*, 15(3), pp. 499–511. https://doi.org/10.1007/s40092-018-0283-5.

Franco, G. (2022, November 22). ¿Cuándo desaparecerá el último lector del diario impreso? La República. Available in: https://www.larepublica.co/empresas/cuando-desaparecera-el-ultimo-lector-del-diario-impreso-3493210.

Garzón, J. (2019, September 24). Crisis en la industria de la comunicación en Colombia ¿Oportunidad para reinventarse?? TV Noticias, 1–2. Available in: https://www.tvnoticias.com.co/crisis-en-la-industria-de-la-comunicacion-en-colombia-oportunidad-para-reinventarse.

Geetha, S., Poonthalir, G., & Vanathi, P. T. (2009). Improved K-Means Algorithm for Capacitated Clustering Problem. *Journal of Computer Science*, 8(4), pp. 52-59.

Goetschalckx, M., Vidal, C., & Dogan, K. (2002). Modeling and Design of Global Logistics Systems: A Review of Integrated Strategic and Tactical Models and Design Algorithms. *European Journal of Operational*, 143(1), pp. 1–18. https://doi.org/10.1016/S0377-2217(02)00142-X.

González, C. (2016, April 1). Siete de cada 10 colombianos aún prefiere leer los periódicos impresos. La República, 1. Available in: https://www.larepublica.co/internet-economy/siete-de-cada-10-colombianos-aun-prefiere-leer-los-periodicos-impresos-2364096.

Gülbay, İ. *et al.* (2021). Location Optimization of Receivers for IoT- Based Infrastructures. In: Durakbasa, N.M., Gençyılmaz, M.G. (eds) Digital Conversion on the Way to Industry 4.0. ISPR 2020. *Lecture Notes in Mechanical Engineering*, pp. 930–942. https://doi.org/10.1007/978-3-030-62784-3_77.

Gupta, R., Muttoo, S., & Pal, S. (2019). Meta-Heuristic Algorithms to Improve Fuzzy C-Means and K-Means Clustering for Location Allocation of Telecenters Under E-Governance in Developing Nations. *International Journal of Fuzzy Logic and Intelligent Systems*, 19(4), pp. 290–298. https://doi.org/10.5391/IJFIS.2019.19.4.290

Hajiaghaei-Keshteli, M. (2011). The Allocation of Customers to Potential Distribution Centers in Supply Chain Networks: GA and AIA Approaches. *Applied Soft Computing Journal*, 11(2), pp. 2069–2078. https://doi.org/10.1016/j.asoc.2010.07.004.

Han, J., Pei, J., & Tong, H. (2022). Data Mining: Concepts and Techniques (Morgan kaufmann., Ed.; 4th ed.).

Handl, J., Knowles, J., & Kell, D. (2005). Computational Cluster Validation in Post-Genomic Data Analysis. Bioinformatics, 21(15), pp. 3201–3212. https://doi.org/10.1093/bioinformatics/bti517

Hidayat, R., Akhmad, S., Winarso, K., & Arendra, A. (2020). K-Means Method for Determining Location of Facilities and Development of Supply Chain Network for Salt Commodities in Sumenep District. *6th Information Technology International Seminar*, ITIS 2020, pp. 193–197. https://doi: 10.1109/ITIS50118.2020.9321040.

Hu, W., Dong, J., Hwang, B., Ren, R., & Chen, Z. (2020). Network Planning of Urban Underground Logistics System with Hub-And-Spoke Layout: Two Phase Cluster-Based Approach. *Engineering, Construction and Architectural Management*, 27(8), pp. 2079–2105. https://doi.org/10.1108/ECAM-06-2019-0296.

Jarrah, A., & Bard, J. (2012). Large-Scale Pickup and Delivery Work Area Design. *Computers and Operations Research*, 39(12), pp. 3102–3118. https://doi.org/10.1016/j.cor.2012.03.014.

Jiang, Y., Yeh, W., Lai, C., Liu, H., Yeh, C., Chung, Y., & Lin, J. (2016). Integrated Use of Soft Computing and Clustering for Capacitated Clustering Single-Facility Location Problem with One-Time Delivery. *IEEE Congress on Evolutionary Computation*, CEC 2016, pp. 2701–2705. https://doi.org/10.1109/CEC.2016.7744128.

Jinyin, C., Xiang, L., Haibing, Z., & Xintong, B. (2017). A Novel Cluster Center Fast Determination Clustering Algorithm. *Applied Soft Computing Journal*, 57, pp. 539–555. https://doi.org/10.1016/j.asoc.2017.04.031.

Kamble, S., Raut, R., & Gawankar, S. (2017). Optimizing the Newspaper Distribution Scenarios using Genetic Algorithm: A Case Study of India. *American Journal of Applied*

Sciences, 14(4), pp. 478–495. https://doi.org/10.3844/ajassp.2017.478.495.

Katsela, K., Güneş, Ş., Fried, T., Goodchild, A., & Browne, M. (2022). Defining Urban Freight Microhubs: A Case Study Analysis. *Sustainability (Switzerland)*, 14(532), pp. 1–27. https://doi.org/10.3390/su14010532.

Klose, A., & Drexl, A. (2005). Facility Location Models for Distribution System Design. *European Journal of Operational Research*, 162(1), pp. 4–29. https://doi.org/10.1016/j.ejor.2003.10.03.

Kumar, M., & Reddy, R. (2016). A Fast Dbscan Clustering Algorithm by Accelerating Neighbor Searching using Groups Method. *Pattern Recognition*, 58, pp. 39–48. https://doi.org/10.1016/j.patcog.2016.03.008.

Küükdeniz, T., Baray, A., Ecerkale, K., & Esnaf, Ş. (2012). Integrated Use of Fuzzy C-Means and Convex Programming for Capacitated Multi-Facility Location Problem. *Expert Systems with Applications*, 39(4), pp. 4306–4314. https://doi.org/10.1016/j.eswa.2011.09.102.

Lau, H., Jiang, Z., Ip, W., & Wang, D. (2010). A Credibility-Based Fuzzy Location Model with Hurwicz Criteria for the design of distribution systems in B2C e-commerce. *Computers and Industrial Engineering*, 59(4), pp. 873–886. https://doi.org/10.1016/j.cie.2010.08.018.

Liao, K., & Guo, D. (2008). A Clustering-Based Approach to The Capacitated Facility Location Problem. *Transactions in GIS*, 12(3), pp. 323-339. https://doi.org/10.1111/j.1467-9671.2008.01105.x.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., Studer, M., Roudier, P., Gonzalez, J., & Kozlowski, K. (2018). *Cluster: Cluster Analysis Basics and Extensions*. R Package, Version 2.0. 7–1.

Maechler, M., Struyf, A., Hubert, M., Hornik, K., Studer, M., & Roudier, P. (2015). *Package 'cluster': Cluster Analysis Basics and Extensions*. R Topics Documented, Version 2.0. 7–1.

Manoharan, J., Ganesh, S., & Sathiaseelan, J. (2016). Outlier Detection Using Enhanced K-Means Clustering Algorithm and Weight Based Center Approach. *International Journal of Computer Science and Mobile Computing*, 5(4), pp. 453–464.

Melkonyan, A., Gruchmann, T., Lohmar, F., Kamath, V., & Spinler, S. (2020). Sustainability assessment of last-mile logistics and distribution strategies: The case of local food networks. *International Journal of Production Economics*, 228, pp. 1–17. https://doi.org/10.1016/j.ijpe.2020.107746

Osaba, E., Yang, X., Diaz, F., Onieva, E., Masegosa, A., & Perallos, A. (2017). A Discrete Firefly Algorithm to Solve A Rich Vehicle Routing Problem Modelling A Newspaper Distribution System With Recycling Policy. *Soft Computing*, 21(18), pp. 5295–5308. https://doi.org/10.1007/s00500-016-2114-1.

Oudouar, F., el Fallahi, A., & Zaoui, E. (2019). An Improved Heuristic Based on Clustering AND Genetic Algorithm for Solving the Multi-Depot Vehicle Routing Problem. *International Journal of Recent Technology and Engineering*, 8(3), pp. 6535–6540. https://doi.org/10.35940/ijrte.C5256.098319.

Rabbani, M., Mokhtarzadeh, M., & Manavizadeh, N. (2021). A Constraint Programming Approach and A Hybrid of Genetic and K-Means Algorithms to Solve the P-Hub Location-Allocation Problems. *International Journal of Management Science and Engineering Management*, 16(2), pp. 123–133. https://doi.org/10.1080/17509653.2021.1905096

Ree, S., & Yoon, B. (1996). A Two-Stage Heuristic Approach for the Newspaper Delivery Problem. *Computers and Industrial Engineering*, 30(3), pp. 501–509 https://doi.org/10.1016/0360-8352(96)00013-7

Rousseeuw, P. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, 20, pp. 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Russell, R., Chiang, W., & Zepeda, D. (2008). Integrating Multi-Product production and Distribution in Newspaper Logistics. *Computers and Operations Research*, 35(5), pp. 1576–1588. https://doi.org/10.1016/j.cor.2006.09.002.

Sabarish, B., & Vidhya, S. (2019). Facility Recommendation System Using Domination Set Theory in Graph. *International Journal of Innovative Technology and Exploring Engineering*, 8(9), pp. 313–317. https://doi.org/10.35940/ijitee.h7419.078919.

Sahraeian, R., & Kaveh, P. (2010). Solving Capacitated P-Median Problem by Hybrid K-Means Clustering and Fixed Neighborhood Search Algorithm. *International Conference on Industrial Engineering and Operation Managemen*t 2010, pp. 1–6.

Sahraeian, R., & Kazemi, M. (2011). A Fuzzy Set Covering-Clustering Algorithm for Facility Location Problem. *IEEE International Conference on Industrial Engineering and Engineering Management 2011*, pp. 1098–1102. https://doi.org/10.1109/IEEM.2011.6118085.

Santoso, A., Parung, J., Prayogo, D. N., & Winardi. (2021). Designing Clusters of Distribution Area and Delivery Route for Maximizing the Vehicle Utilization and Minimizing the Workload Gap and Transportation Cost. *Operations and Supply Chain Management: An International Journal*, 14(4), pp. 536–544. https://doi.org/10.31387/oscm0470322.

Saragih, N. I., Bahagia, S. N., Suprayogi, & Syabri, I. (2022). Location-Inventory-Routing Problem in A Context of City Logistics: A Case Study of Jakarta. *Operations and Supply Chain Management: An International Journal*, 15(2), pp. 218–227. https://doi.org/10.31387/oscm0490342.

Sharma, A., Sharma, A., & Jalal, A. (2017). Clustering Based Hybrid Approach for Facility Location Problem. *Management Science Letters*, 7(12), pp. 577–584. https://doi.org/10.5267/j.msl.2017.8.007.

Sharma, A., Sharma, A., Jalal, A., & Kant, K. (2021). A Two Step Clustering Method for Facility Location Problem. *International Journal of Advanced Intelligence Paradigms*, 10(1), pp. 337–355. https://doi.org/10.1504/ijaip.2018.10011478.

Simić, D., Ilin, V., Svircevic, V., & Simic, S. (2017). A Hybrid Clustering and Ranking Method for Best Positioned Logistics Distribution Centre in Balkan Peninsulaetia. *Logic Journal of the IGPL*, 25(6), pp. 991–1005. https://doi.org/10.1093/jigpal/jzx047.

Singh, M., & Gupta, S. (2020). Urban Rail System for Freight Distribution in A Mega City: Case study of Delhi, India. *Transportation Research Procedia*, 48, pp. 452–466. https://doi.org/10.1016/j.trpro.2020.08.052.

Sinnott, R. (1984). Virtues of the Haversine. *Sky and Telescope, 68(2):158.*

Sitek, P., Wikarek, J., Rutczyńska-Wdowiak, K., Bocewicz, G., & Banaszak, Z. (2021). Optimization of Capacitated Vehicle Routing Problem with Alternative Delivery, Pick-Up and Time Windows: A Modified Hybrid Approach. *Neurocomputing*, 423, pp. 670–678. https://doi.org/10.1016/j.neucom.2020.02.126.

Sutanto, G., Kim, S., Kim, D., & Sutanto, H. (2018). A Heuristic Approach to Handle Capacitated Facility Location Problem Evaluated Using Clustering Internal Evaluation. *IOP Conference Series: Materials Science and Engineering 2018*, 332, pp. 1–8. https://doi.org/10.1088/1757-899X/332/1/012023.

Taaffe, K., Geunes, J., & Edwin, E. (2010). Supply Capacity Acquisition and Allocation with Uncertain Customer Demands. *European Journal of Operational Research*, 204(2), pp. 263–273. https://doi.org/10.1016/j.ejor.2009.10.030.

Tran, T., Drab, K., & Daszykowski, M. (2013). Revised DBSCAN Algorithm to Cluster Data with Dense Adjacent Clusters.

Chemometrics and Intelligent Laboratory Systems, 120, pp. 92–96. https://doi.org/10.1016/j.chemolab.2012.11.006

Two sides. (2019). Print and Paper in A Digital World - An International Survey of Consumer Preferences, Attitudes, and Trust. Available in: https://www.midlandpaper.com/wp-content/uploads/2019/07/Two_Sides_Print_and_Paper_In_A_Digital_World.pdf.

Varghese, S., & Gladston, R. (2016). Clustering Based Model for Facility Location in Logistic Network Using K-Means. *International Journal of Scientific Inventions and Innovations*, 1(1), pp. 26–32.

Vita, L. (2020, April 22). La pandemia del Covid-19 ¿una prueba de fuego para los periódicos o su estocada? La República. Available in: https://www.larepublica.co/empresas/la-pandemia-del-covid-19-una-prueba-de-fuego-para-los-periodicos-o-su-estocada-2994991.

Wang, Y., Assogba, K., Liu, Y., Ma, X., Xu, M., & Wang, Y. (2018). Two-Echelon Location-Routing Optimization with Time Windows Based on Customer Clustering. *Expert Systems with Applications*, 104, pp. 244–260. https://doi.org/10.1016/j.eswa.2018.03.018.

Wang, Y., Ma, X., Xu, M., Liu, Y., & Wang, Y. (2015). Two-Echelon Logistics Distribution Region Partitioning Problem Based on A Hybrid Particle Swarm Optimization-Genetic Algorithm. *Expert Systems with Applications*, 42(12), pp. 5019–5031. https://doi.org/10.1016/j.eswa.2015.02.058.

Wang, Y., Sun, Y., Guan, X., Fan, J., Xu, M., & Wang, H. (2021). Two-Echelon Multi-Period Location Routing Problem with Shared Transportation Resource. *Knowledge-Based Systems*, 226, pp. 1–22. https://doi.org/10.1016/j.knosys.2021.107168

Wang, Y., Zhang, S., Assogba, K., Fan, J., Xu, M., & Wang, Y. (2018). Economic and Environmental Evaluations in The Two-Echelon Collaborative Multiple Centers Vehicle Routing Optimization. *Journal of Cleaner Production*, 197, pp. 443–461. https://doi.org/10.1016/j.jclepro.2018.06.208.

Wu, L., Chen, H., Yu, X., Chao, S., Yu, Z., Dou, R. (2019). Grid Partition and Agglomeration for Bidirectional Hierarchical Clustering. In: Li, J., Liu, Z., Peng, H. (eds) Security and Privacy in New Computing Environments. SPNCE 2019. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, 284, pp. 707–722. https://doi.org/10.1007/978-3-030-21373-2_60.

Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. J, 2(2), pp. 226–235. https://doi.org/10.3390/j2020016.

Zambrano, W. (2020). Reinvención de la prensa y la radio colombiana: un nuevo ecosistema comunicativo (Fondo de Publicaciones Universidad Sergio Arboleda, Ed.). Universidad Sergio Arboleda, Bogotá, Colombia, 218. https://doi.org/10.22518/book/9789585158054.

Karla C. Alvarez-Uribe received the B.Sc. in industrial engineering and M.Sc degree in engineering from Universidad Nacional de Colombia, Medellin, Colombia. Currently, she serves as a full professor at the Instituto Tecnológico Metropolitano – ITM, Medellín, Colombia and PhD candidate at the Institute for Intelligent Systems Research and Innovation (IISRI) at Deakin University, Australia. Her research interest includes the development of optimization and simulation models applied to urban logistics.

Eduard Gañan-Cárdenas has a B.Sc. in Industrial engineering M.Sc. in statics from National University of Colombia, Medellin, Colombia. Currently, he is a PhD candidate at the same university and holds a position as a research professor at the Instituto Tecnológico Metropolitano – ITM Medellín, Colombia. His research interests are data mining, machine learning, and decision support systems.

Diego Perez-Montoya completed his Bachelor of Science degree in Production Engineering at the Instituto Tecnológico Metropolitano – ITM Medellín, Colombia. He began his research journey as a young researcher affiliated with the Quality and Production Research Group at the same university. Currently, he is employed as a specialist in business intelligence.