International Journal of Advanced Engineering and Management Research

Vol. 11, No. 05; 2026

ISSN: 2456-3676

Multidimensional Analysis of Poverty Through Topological Data Analysis

[1]Oscar A. Jiménez, [2]Felipe A. Llaugel
[1]Instituto Tecnológico de Santo Domingo
Dominican Republic
https://orcid.org/0009-0000-7963-0790
[2]Universidad Dominicana O&M
Dominican Republic
https://orcid.org/0000-0001-7633-5219

doi.org/10.51505/ijaemr.2026.11501 URL: http://dx.doi.org/10.51505/ijaemr.2026.11501

Received: Aug 27, 2026 Accepted: Sep 08, 2026 Online Published: Sep 24, 2026

Abstract
This study addresses the phenomenon of poverty in low income community in Dominican Republic, from a multidimensional perspective, moving beyond the limitations of traditional income-only approaches. The central objective was to characterize the empirical structure of social deprivation by applying unsupervised learning techniques and Topological Data Analysis (TDA). Data from 591 households were processed. Using the Mapper algorithm together with DBSCAN, two candidate metric spaces were evaluated and compared to capture the "shape" of the data: a prevalence-weighted semimetric and the standard Euclidean metric. Unlike linear aggregation methods, the topology built on the Euclidean metric proved better suited to this micro-territorial context: it revealed that poverty in this sector does not operate in isolated clusters but instead forms an uninterrupted structural continuum that closely tracks the Quality of Life Index (ICV). These findings show that the absence of certain items responds strongly to sociocultural priorities and local consumption patterns.

Keywords: Multidimensional poverty, Topological Data Analysis (TDA), Mapper algorithm, Quality of Life Index (ICV), Urban vulnerability.

1. Introduction

The conceptualization of poverty has evolved from a strictly unidimensional perspective grounded in income levels toward a multidimensional paradigm that recognizes deprivation across multiple spheres of human well-being. Following the capability approach proposed by Amartya Sen (1999), poverty is defined not only as a lack of financial resources but as the deprivation of fundamental freedoms that prevent individuals from achieving the states of being and doing what they value.

Vol. 11, No. 05; 2026

ISSN: 2456-3676

In the Dominican Republic, despite sustained economic growth, significant structural gaps persist. The Mata Hambre sector, located in the National District, presents itself as a critical case study. Its strategic urban location contrasts with internal dynamics of vulnerability and inequality that are often diluted within municipal or national level statistical aggregations.

In response to these challenges, this study proposes a disruptive approach using Topological Data Analysis (TDA). This methodology, situated at the intersection of algebraic topology and data science, makes it possible to extract the intrinsic geometric structure of complex data sets.

The technical core of this work relies on the Mapper algorithm. Formally, given a data set X (in this case, the binary vectors representing asset and equipment ownership of Mata Hambre households) and a filter function $f : X \rightarrow Rd$, the algorithm builds a one-dimensional simplicial complex that summarizes the topology of the data space. The resulting graph makes it possible to identify clusters, branches, and loops that represent distinct vulnerability profiles, offering an analytical granularity that traditional descriptive statistics cannot achieve. Mapper has been used with great effectiveness in diverse fields such as medicine, genomics, neuroscience, chemistry, remote sensing, soil science, agriculture, sports, and economics. However, we believe that using TDA to study living standards and wealth inequality is novel (Anderson et al., 2022).

The use of TDA to study poverty in Mata Hambre not only represents an original methodological contribution but also seeks to provide a robust scientific basis for designing precisely targeted public policy. By uncovering the multidimensional structure of deprivation, it becomes possible to move from generic interventions toward human-development strategies grounded in the geometric evidence of the data.

1.1. Background

Poverty has ceased to be understood exclusively as income insufficiency and is now approached as a multidimensional phenomenon involving simultaneous deprivations across different spheres of well-being. Along these lines, the approach of Alkire and Foster (2011) consolidated a widely accepted methodology for identifying and measuring multidimensional poverty by jointly considering the various deprivations affecting households. In a complementary manner, Alkire and Santos (2010) strengthened the international discussion on the need to measure poverty beyond income, laying the conceptual foundations of the Multidimensional Poverty Index. To measure multidimensional poverty, Llaugel et al. (2024) presented an alternative approach for calculating the index, incorporating Principal Component Analysis (PCA) to derive the weights of the deprivation indicators and thereby use endogenous parameters for a municipality of Santo Domingo in the Dominican Republic.

Despite these advances, the literature has pointed out that traditional measurement methods tend to summarize information into aggregate indices or linear classifications, which makes it difficult to capture the internal heterogeneity of households and the coexistence of distinct deprivation profiles. For this reason, some recent studies have incorporated unsupervised approaches to

Vol. 11, No. 05; 2026

ISSN: 2456-3676

explore the structure of poverty and detect subgroups with differentiated patterns. In this regard, Abdul Rahman et al. (2021) use clustering techniques to identify multidimensional poverty indicators, while Trento Oliveira et al. (2023) demonstrate the potential of unsupervised learning and open data to identify marginalized urban areas.

In parallel, Topological Data Analysis (TDA) has emerged as a methodological alternative capable of studying the shape of data and detecting complex structures in high-dimensional spaces. The work of Singh et al. (2007) introduced the Mapper algorithm as a tool for representing complex data through a topological network while simultaneously preserving local and global information. Subsequently, Munch (2017), Wasserman (2018), and Chazal and Michel (2021) systematized the foundations of TDA, highlighting its usefulness for exploring nonlinear relationships, latent structures, and patterns that are not easily detectable through conventional statistical techniques.

A particularly relevant precedent for this line of research is the work of Anderson et al. (2022), who applied Mapper to UNICEF MICS survey data to study the relationship between wealth, asset possession, and urban-rural differences. Their findings show that the topological approach reveals gradients, groupings, and nonlinear configurations of well-being that extend the understanding offered by aggregate indicators. Likewise, Godwin et al. (2019) highlight that the use of TDA in the social sciences constitutes a highly promising avenue for studying complex, heterogeneous, and structurally fragmented phenomena.

In the context of the Dominican Republic, the Single Beneficiary System (SIUBEN, in Spanish), through the Universal Household Social Registry (RSUH), constitutes the most robust official source for identifying and characterizing households in situations of socioeconomic vulnerability. It is within this framework that advances in the international TDA literature converge with the local need for targeting, using the database corresponding to the data as of May 2024 for the Mata Hambre sector, in the National District. By representing each household as a point in a high-dimensional topological space, built exclusively from dichotomous variables on asset ownership, connectivity, and technological equipment, it becomes possible to translate the analytical state of the art into the deconstruction of urban poverty at a local level.

1.2. Problem Statement

Urban poverty is a complex, heterogeneous, and multidimensional phenomenon. In sectors such as Mata Hambre, households may share certain levels of deprivation while, at the same time, differing markedly in aspects related to housing, access to services, education, and asset ownership.

Conventional methods, based on aggregate indices, socioeconomic classifications, or linear parametric models, are useful for summarizing information and establishing general comparisons; however, they present critical limitations for identifying complex configurations, latent subpopulations, and patterns of structural fragmentation. In many cases, these approaches

Vol. 11, No. 05; 2026

ISSN: 2456-3676

reduce poverty to a synthetic measure that, although informative, can conceal nonlinear relationships between variables and substantive differences between households that, from a traditional standpoint, share the same vulnerability category (Alkire & Foster, 2011; Alkire & Santos, 2010).

Considering these methodological barriers, it becomes imperative to adopt an exploratory, analytical-computational approach capable of capturing the complexity, heterogeneity, and fragmentation of structural poverty while considering multiple dimensions simultaneously.

From an analytical standpoint, the Mapper algorithm offers a robust alternative for addressing this limitation, as it summarizes the geometry of high-dimensional data into a topological network in which nodes represent local clusters and edges represent connections or transitions between them (Chazal & Michel, 2021; Singh et al., 2007). Applied to the study of poverty, this approach has the potential to reveal gradients, isolated subgroups, and fragmented structures that remain invisible to classical methods.

1.3. Rationale

On the theoretical level, the study broadens the discussion on multidimensional poverty by proposing a structural and relational reading of the phenomenon. While classical literature has focused on identifying and aggregating deprivations, this research seeks to understand how such deprivations are organized within the data space and how they give rise to differentiated poverty configurations. In this sense, the work does not replace traditional approaches but rather complements them with a perspective oriented toward the shape and connectivity of the data (Alkire & Foster, 2011).

On the methodological level, the research is justified by the incorporation of Topological Data Analysis (TDA) into the study of urban poverty. The methodological design articulates a rigorous strategy that includes variable selection and curation, local clustering with DBSCAN, and the construction of a Mapper network to represent the geometric structure of poverty. This design constitutes a relevant innovation in the context of social science, where the use of TDA is still incipient (Godwin et al., 2019).

On the empirical level, the research relies on the SIUBEN Universal Household Social Registry database, corresponding to May 2024, which constitutes the main official source for the socioeconomic characterization of households in the Dominican Republic.

2. Literature Review

2.1. Multidimensional Poverty

Table 2.1 details the methodological structure of the 2025 Global Multidimensional Poverty Index (MPI), which organizes deprivations into three fundamental dimensions: health, education, and standard of living.

Vol. 11, No. 05; 2026

ISSN: 2456-3676

Table 2.1: Indicators of the Multidimensional Poverty Index (MPI)

Poverty dimensions	Indicator	Deprived if living in a household where...	Weight
Health	Nutrition	Any adult under 70 years of age or any child with nutritional information shows undernutrition.	1/6
Health	Child mortality	Any child under 18 has died in the household in the past five years.	1/6
Education	Years of schooling	No household member of school-entry age plus six years or more has completed at least six years of schooling.	1/6
Education	School attendance	Any school-age child is not attending school up to the age corresponding to completion of eighth grade.	1/6
Standard of living	Cooking fuel	The household cooks with dung, wood, charcoal, or coal.	1/18
Standard of living	Sanitation	Sanitation facilities are not improved, or if improved, are shared with other households.	1/18
Standard of living	Drinking water	The household does not have access to improved drinking water, or it is more than a 30-minute round trip away.	1/18
Standard of living	Electricity	The household has no electricity.	1/18
Standard of living	Housing	The floor, roof, or walls are inadequate, being made of natural or rudimentary materials.	1/18
Standard of living	Assets	The household owns at most one of the following: radio, television, telephone, computer, cart, bicycle, motorcycle, or refrigerator, and does not own a car or truck.	1/18

Source: UNDP, 2025.

As can be observed, the analysis goes beyond mere income availability to assess simultaneous deprivations across ten specific indicators, ranging from school attendance and nutrition to

Vol. 11, No. 05; 2026

ISSN: 2456-3676

access to basic services and asset ownership. It is precisely this multiplicity of concurrent factors and their respective weights that generates a high-dimensional data space. To model the way households experience the combination of these deprivations, the use of advanced analytical tools such as Topological Data Analysis (TDA) becomes indispensable, as it can unravel how these deprivations interact and structurally overlap across the territory (UNDP, 2025).

One of the most influential contributions in this field is that of Alkire and Santos (2010), who propose a multidimensional poverty index oriented toward developing countries, integrating different dimensions of human well-being into a single synthetic measure. Subsequently, Alkire and Foster (2011) develop a counting methodology that identifies the multidimensionally poor and aggregates information on the intensity of their deprivations. This approach has been widely valued for its conceptual clarity, flexibility, and capacity to support targeted public policy formulation.

However, although these methods have represented an important advance in the measurement of poverty, their analytical structure continues to rely on aggregation and classification procedures that can conceal complex relationships between indicators. In practice, two households may present similar levels of aggregate poverty yet differ substantially in the specific combination of deprivations they face.

Along these lines, Abdul Rahman et al. (2021) apply clustering techniques to identify multidimensional poverty indicators in vulnerable populations, showing that unsupervised machine-learning methods can complement conventional measurements by revealing differentiated deprivation profiles. Similarly, Trento Oliveira et al. (2023) use unsupervised learning and open data to detect deprived areas of well-being in São Paulo, highlighting the value of flexible approaches for capturing complex spatial and social configurations.

These contributions support the claim that multidimensional poverty requires not only comprehensive indicators but also analytical tools capable of representing the internal heterogeneity of households and territories. In this context, Topological Data Analysis offers a promising alternative, as it allows the relational structure of the data to be studied rather than only its aggregate values.

2.2. Topological Data Analysis (TDA)

Origin of the topological intuition: the Seven Bridges of Königsberg problem

Vol. 11, No. 05; 2026

ISSN: 2456-3676

Figure 2.1: Representation of the Seven Bridges of Königsberg problem and its abstraction as a graph. Source: Murtagh, 2024

One of the most important historical antecedents of topology and graph theory is the Seven Bridges of Königsberg problem, studied by Leonhard Euler in 1736. The relevance of this problem lies not only in its solution but in the conceptual shift it introduced: Euler observed that, in order to determine whether it was possible to cross the city by traversing each bridge exactly once, it was not necessary to preserve every detail of the map, such as distances, angles, sizes, or shapes. Instead, it was sufficient to represent each landmass as a node and each bridge as an edge.

In other words, Euler discarded non-essential geometric information and preserved only the relational shape of the system. This step was decisive because it showed that certain problems can be analyzed through properties that are invariant under deformation, an idea that would later become central to topology.

Topological Data Analysis (TDA) is a recent approach within data science that seeks to extract information from the shape or geometric structure of data. Rather than focusing solely on linear relationships, averages, or rigid partitions, TDA makes it possible to identify connectivity, groupings, gaps, cycles, and other global features that emerge in complex data sets. Chazal and Michel (2021) note that TDA provides mathematical, statistical, and computational tools for inferring and exploiting the underlying topological and geometric structures in highly complex data.

From an introductory perspective, Munch (2017) highlights that the value of TDA lies in its ability to reveal patterns that may go unnoticed by other exploration techniques. Complementarily, Wasserman (2018) emphasizes its importance as a bridge between topology, statistics, and inference, especially in contexts where data present noise, high dimensionality, and non-trivial structures.

A fundamental aspect of TDA is that it does not constitute a direct replacement for classical methods but rather a complementary tool. According to Chazal and Michel (2021), its main strength lies in providing new descriptions of the data that can be used both for visualization and for subsequent analysis, in combination with other methodologies.

2.2.1. Simplicial Complexes

Simplicial complexes constitute one of the mathematical foundations of TDA. Since observed data are usually presented as a finite set of points, they do not reveal topological properties on their own. To overcome this limitation, combinatorial structures are built that connect nearby points and approximate an underlying continuous shape. Chazal and Michel (2021) explain that simplicial complexes can be understood as generalizations of graphs to higher dimensions: they consider not only connections between pairs of points but also among triples, quadruples, and

Vol. 11, No. 05; 2026

ISSN: 2456-3676

larger sets of nearby observations.

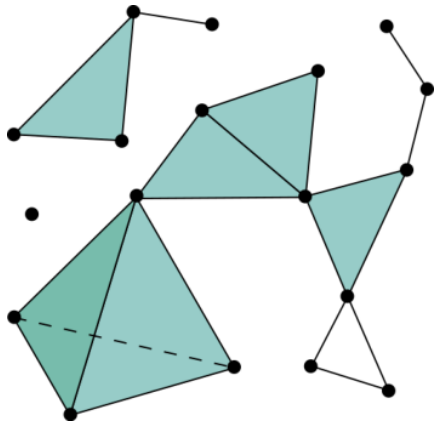

Figure 2.2: An example of a simplicial complex. Source: Arenas et al., 2023

As shown in Figure 2.2, a simplicial complex can contain simplices of different dimensions. This one consists of 17 points (0-simplices), 22 edges (1-simplices), 8 triangles (2-simplices), and 1 tetrahedron (3-simplex).

Formally, a zero-dimensional simplex corresponds to a vertex, a one-dimensional simplex to an edge, a two-dimensional simplex to a triangle, and a three-dimensional simplex to a tetrahedron. A simplicial complex is then a collection of these simplices that satisfy certain compatibility conditions, particularly that every face of a simplex also belongs to the complex. This structure makes it possible to represent neighborhoods, connectivity, and proximity relationships in a richer way than an ordinary graph.

In TDA, the construction of simplicial complexes is fundamental because it provides a mathematical foundation on which topological invariants can be computed. Among the best-known complexes are the Čech complex and the Vietoris–Rips complex, both built from distances between points and a scale parameter. Their practical usefulness lies in translating a set of discrete observations into a structure suitable for topological analysis.

For applied research such as the study of socioeconomic vulnerability, simplicial complexes are relevant because they make it possible to represent similarity configurations between households without imposing a prior classification.

2.2.2. Persistent Homology

Persistent homology is one of the most powerful tools of TDA, as it allows the evolution of the topological features of data to be studied across different scales. Intuitively, homology seeks to identify elements such as connected components, cycles, or voids within a structure; persistence, in turn, observes which of these features remain across a sequence of simplicial complexes built at different resolution levels.

Vol. 11, No. 05; 2026

ISSN: 2456-3676

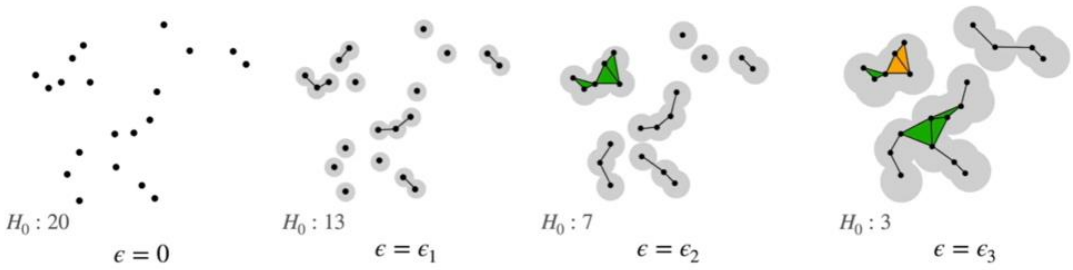

Figure 2.3: Evolution of a point cloud as the scale parameter ϵ increases. As ϵ grows, connections form between nearby points and the number of connected components decreases, illustrating the idea of topological persistence in dimension zero (H0). Source: Chazal and Michel (2021).

Figure 2.3 intuitively illustrates the concept of persistent homology in dimension zero. Initially, when ϵ = 0, each point in the cloud constitutes an independent connected component. As the scale parameter increases, the neighborhoods (balls of radius ϵ) around the points begin to overlap, forming connections between nearby observations. As a result, isolated components progressively merge, reducing the total number of connected components (H0). This continuous filtration process makes it possible to identify which topological structures persist across different scales (representing underlying global features) and which disappear quickly (associated with local variation or noise in the data) (Chazal & Michel, 2021).

Chazal and Michel (2021) define persistent homology as a tool that computes, studies, and encodes multiscale topological features in nested families of simplicial complexes. They further emphasize that it not only computes Betti numbers at each scale but also describes the evolution of homology groups along a filtration. This idea of filtration is central: a filtration is a nested family of sub complexes in which, as a distance parameter increases, topological features appear and disappear.

The main advantage of persistent homology is that it distinguishes between stable topological features and artifacts produced by noise or poorly chosen scales. Wasserman (2018) specifically highlights the relevance of persistence for giving TDA a more solid inferential foundation.

2.2.3. The Mapper Algorithm

The Mapper algorithm occupies a central place in the applied TDA literature, as it translates high-dimensional data into an interpretable graphical representation. It was introduced by Singh et al. (2007) as a tool for analyzing complex data sets while preserving relevant topological information. Its general purpose is to reduce dimensionality without completely losing the global structure of the data set, which distinguishes it from other classical reduction methods (such as PCA).

Vol. 11, No. 05; 2026

ISSN: 2456-3676

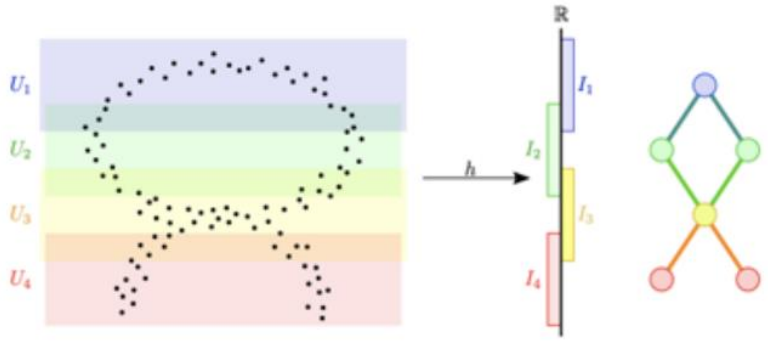

Figure 2.4: Illustration of the Mapper algorithm. The filter function h : R2 → R projects the data onto the vertical coordinate. The image of h is covered with overlapping intervals I1, . . . , I4, whose preimages in the point cloud induce subsets that are then clustered to form the nodes of the graph. Source: Belchí Guillamón et al. (2019).

Figure 2.4 schematically shows how the Mapper algorithm works. In this example, the filter function projects the point cloud onto a single coordinate (vertical axis), and this projection is covered with overlapping intervals. For each interval, its preimages in the original data space are then extracted and a local clustering algorithm is applied. Each detected group or cluster becomes a node of the final graph, while connections (edges) between nodes are established when there is an intersection of points between adjacent groups. This topological construction makes it possible to summarize the global structure of the data while preserving locally relevant analytical information (Belchí Guillamón et al., 2019).

Anderson et al. (2022) describe Mapper as an unsupervised learning algorithm that combines dimensionality reduction and clustering to produce a graph from a data cloud. In this graph, nodes represent clusters of nearby observations, while edges represent overlaps between those clusters.

The Mapper procedure can be summarized in several stages. First, a data set X in a high-dimensional space is taken as the starting point. Next, a filter function h : X → Rd is defined, whose purpose is to highlight some relevant features of the data. The image of that function is then covered with overlapping intervals. For each region of that cover, its preimage in the original space is considered and a local clustering algorithm is applied. Each resulting group becomes a node of the final graph. Finally, when two groups from overlapping regions share observations, the corresponding nodes are connected by an edge. A crucial strength of Mapper is its flexibility. Its construction depends on key methodological decisions: the filter function, the distance metric, the number of intervals, the degree of overlap, and the clustering algorithm.

On the applied level, the most direct precedent for this research is the work of Anderson et al. (2022), who use Mapper to analyze UNICEF MICS survey data in Serbia. Their study shows that the resulting graph makes it possible to examine the relationship between the wealth index, the

Vol. 11, No. 05; 2026

ISSN: 2456-3676

urban-rural setting, and asset ownership. The authors conclude that this approach provides invaluable information for public policy, revealing household profiles that are not evident through PCA. They note that Mapper offers a nuanced perspective, making it possible to study branches (flares) that reflect differentiated deprivation patterns.

Several free implementations of the Mapper algorithm exist, such as Python Mapper, TDAmapper (an R implementation), and KeplerMapper for Python (KeplerMapper, 2024). The latter is the one used in this analysis.

Mapper is a data-reduction method and, in that sense, is like principal component analysis. However, while nodes are created from local information, the edges of the graph retain some knowledge of the global topological features of the data as a subspace. This capacity to extrapolate from the local to the global is what makes Mapper attractive and useful.

It should be noted that the user must make several decisions when implementing Mapper, such as the filter function, the number of sets in the cover, the size of the overlap, the metric, and the clustering procedure, and variation in these parameters can produce very different graphs. In summary, the Mapper algorithm represents one of the most solid TDA tools for the study of social phenomena, serving as a useful instrument for hypothesis generation.

3. Methodology

The following section details the methodological route followed for the analysis of multidimensional poverty in households of the Mata Hambre sector. The research is framed as a statistical and computational data analysis aimed at exploring and describing how poverty is structured. To this end, the situation of households is analyzed as reported in the data-collection instruments, without modifying their environment. Taking as its main methodological reference the work of Anderson, Memić, and Volić (2022) on the application of Topological Data Analysis (TDA) to UNICEF's Multiple Indicator Cluster Surveys (MICS), the central pillar of this design lies in the convergence between data science and computational topology.

3.1. Population and Sample: Mata Hambre Sector

The spatial unit of analysis for this study is the Mata Hambre sector, located in the National District, Dominican Republic. Since access was available to the targeted census of the Single Beneficiary System (SIUBEN), no probabilistic sampling formula was applied; instead, the full universe of data available for this polygon at the selected time point was used. The effective study population consisted of 591 registered households that had complete or imputable information across the dimensions required for the topological analysis.

3.2. Data Collection

The data used comes from secondary sources, extracted directly from the Universal Household Social Registry (RSUH) administered by SIUBEN, corresponding to May 2024. Because the

Vol. 11, No. 05; 2026

ISSN: 2456-3676

original data were at the level of individual or disaggregated records, a structural transformation was carried out to establish the household as the primary unit of analysis. This process was divided into three fundamental stages:

3.2.1. Binarization of Asset Ownership Variables

Eighteen critical variables related to asset and service ownership were identified (television, internet, appliances, and transportation). The original responses, which included "No" and "Don't know" categories, were transformed through a binarization process:

A value of 1 was assigned only to confirmed presence of the good or service.
A value of 0 was assigned to any other condition (absence, "don't know" responses, or null values).

3.2.2. Aggregation and Consolidation Rules

To consolidate information from multiple household members into a single record per household, an aggregation logic based on the unique household identifier (IdHogar) was applied:

Asset Variables: The maximum-value function (max) was used. This ensures that if at least one household member reports owning the good, the household is counted as the owner of it.

Context Variables: For indicators that are invariant within the household unit (such as the Quality of Life Index, ICV), the first record found was retained.

3.2.3. Derivation of New Indicators

As part of enriching the dataset, the calculated variable "Number of Members" was generated by counting the records associated with each household identifier. Finally, all processed dimensions were merged (integrated) to obtain a single, consistent dataset.

Table 3.1: Result of the data-processing pipeline.

Metric	Description
Unit of Analysis	Single Household
Variables Processed	18 asset indicators + ICV + Household size
Cleaning Logic	Conversion of nulls and "don't know" responses to 0 (Absence)

3.2.4. Variables

The variables selected for this analysis correspond to dichotomous questions on household asset ownership, access to information technologies, and a basic capacity associated with human capital. This type of question facilitates coding and subsequent analysis, since each affirmative

Vol. 11, No. 05; 2026

ISSN: 2456-3676

response can be represented by the value 1 and each negative response by the value 0. Consequently, each household is described by an 18-dimensional binary vector, in which each component indicates the presence or absence of a specific good, service, or attribute.

$$x = (x_1, \ldots, x_{18})$$

Our analysis focuses on the following household possession (yes or no): television, cable television, computer, computer with internet access, tablet, cell phone, cell phone with internet access, landline telephone, radio, stove, microwave, air conditioning, refrigerator, washing machine, power inverter, electric generator, private, automobile, private motorcycle.

As a reference variable for interpreting the resulting topological structure, SIUBEN's Quality of Life Index (ICV in Spanish) will be used. According to Llaugel and Llaugel (2023), the ICV, under the SIUBEN 2A model, classifies Dominican households into four categories: ICV-1 (extreme poverty), which groups households with the greatest deprivations; ICV-2 (moderate poverty), which includes households with unmet basic needs; ICV-3 (poor/vulnerable), which defines a lower-middle stratum with greater stability but at risk of falling into poverty in the event of a crisis; and ICV-4 (non-poor), which represents households with the best socioeconomic conditions within the classification.

Much of our analysis involves overlaying the Quality of Life Index (ICV) data onto the Mapper graph resulting from the questions above. The ICV questions for Mata Hambre are broader than these and, in addition to material possessions, collect data on access to water, housing characteristics, and so on.

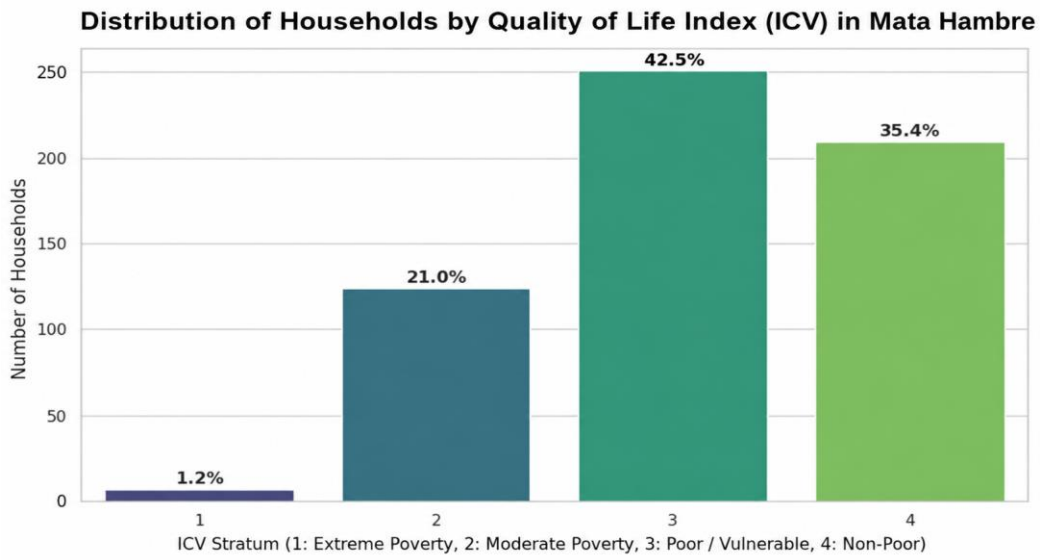

Figure 3.1: Distribution of households by Quality of Life Index (ICV) in Mata Hambre.

Vol. 11, No. 05; 2026

ISSN: 2456-3676

The distribution of households by ICV category in Mata Hambre, shown in Figure 3.1, shows a significant concentration in the intermediate and upper levels of the classification. In particular, 42.5% of households fall under ICV-3, corresponding to the poor/vulnerable group, while 35.4% are classified as ICV-4, associated with the non-poor stratum. Meanwhile, 21.0% of households belong to ICV-2, i.e., moderate poverty, and only 1.2% fall under ICV-1, corresponding to extreme poverty.

This distribution suggests that, although the incidence of households in extreme poverty is small within the analyzed set, there is still a relevant proportion of households in vulnerable conditions, especially within the moderate-poverty category. Likewise, the high concentration in ICV-3 indicates the presence of a broad segment of households classified as poor/vulnerable, which makes the use of TDA especially pertinent for exploring internal heterogeneities that are not visible in an aggregate classification.

3.5. TDA / Mapper Data Processing

3.5.1. Metrics

As explained in Section 2.2.3, in order to apply the Mapper algorithm to the SIUBEN data set corresponding to Mata Hambre households (denoted as X), it is essential to rigorously define the distance measure d(x, y) between the response vectors of two households x = (x1, . . . , xn) and y = (y1, . . . , yn), where n represents the number of selected socioeconomic variables. To address the complexity of these deprivations, this research adopts the probability-based semimetric function developed by Anderson, Memić, and Volić (2022). According to the authors, for a distance function to be effective in household surveys, it must satisfy two fundamental, intuitive properties:

1. The accumulation of shared characteristics must reduce distance. That is, if households x and z agree on all the variables on which x and y already agree, and additionally share other characteristics, then the distance between x and z cannot be greater than the distance between x and y. Formally, if A is the set of variables on which x and y agree, and B is the set on which x and z agree, such that $A \subseteq B$, then the following inequality must hold:

$$d(x, z) \leq d(x, y)$$

2. Statistically unusual similarities must exert a greater influence on closeness. For example, if most households in a sector have a mobile phone, but very few have access to their own vehicle, the fact that two households own a vehicle represents a much stronger multidimensional link than the fact that both have a mobile phone.

To mathematically satisfy these conditions, we implement the semimetric proposed by Anderson et al. (2022). The procedure begins by calculating the population proportion pi of households that present an affirmative condition on variable i. This proportion empirically represents the

Vol. 11, No. 05; 2026

ISSN: 2456-3676

probability that a randomly selected household shares that condition. The topological distance between two households is defined as zero if all their responses are identical. If they differ ($x \neq y$), the variables on which they do agree are identified, and the distance is defined as the probability that two randomly generated vectors would present the same agreements. Formally, letting $p = (p1, \ldots, pn)$ be the vector of proportions for the n items, the distance function is defined as follows:

$$d : X \times X \to \mathcal{R}$$
$$(x, y) \to d(x, y) = \prod_{1 \leq i \leq n : x_i = y_i} \left(p_i^2 + (1 - p_i)^2\right)$$

As shown in the original work, this function satisfies the basic axioms of a semimetric:

- *Positivity: $d(x, y) \geq 0$*
- *Identity of indiscernibles: $d(x, y) = 0 \Rightarrow x = y$*
- *Symmetry: $d(x, y) = d(y, x)$*

It should be noted that this function is strictly classified as a semimetric (and not a true metric) because it is susceptible to violating the triangle inequality; that is, in certain scenarios $d(x, y) > d(x, z) + d(y, z)$ can occur. Despite this technical limitation, the semimetric $d(x, y)$ is well suited for processing with Mapper because it guarantees the two desired structural properties. First, since pi represents a probability, its value is bounded within the interval [0, 1], which ensures that the expression $p^2i + (1 - pi)^2 \in [0, 1]$. Thus, if x and y agree on the subset of variables A, and x and z agree on the subset B (where $A \subseteq B$), the distance decomposes multiplicatively as:

$$d(x, z) = \prod_{i \in B} \left(p_i^2 + (1 - p_i)^2\right) = \prod_{i \in A} \left(p_i^2 + (1 - p_i)^2\right) * \prod_{i \in B \setminus A} \left(p_i^2 + (1 - p_i)^2\right)$$
$$d(x, z) = d(x, y) * \prod_{i \in B \setminus A} \left(p_i^2 + (1 - p_i)^2\right) \leq d(x, y)$$

Second, the function $p^2i + (1 - pi)^2$ traces a parabola whose global minimum is located at 0.5. Consequently, infrequent similarity contributes a smaller multiplicative factor to the computation of $d(x, y)$, which reduces the dissimilarity between two households with a greater impact than a common similarity would, thereby making it possible to effectively group together the most critical vulnerabilities in the sector studied.

3.5.2. Filter Function

To configure the Mapper algorithm, it is essential to define a filter function (or lens). In line with the methodological strategy of the reference literature, the filter was set as the sum of the components of each vector $x \in X$. In practical terms, this function counts the total number of goods or affirmative conditions reported by each household in SIUBEN. In this way, the lens

Vol. 11, No. 05; 2026

ISSN: 2456-3676

initially groups together, within the same intervals, those families that report a similar numerical amount of assets.

Table 3.2 details the empirical distribution of Mata Hambre households according to the sum of reported possessions. The analysis reveals a marked concentration in the intermediate values of the distribution: specifically, between 5 and 9 assets, 453 households are grouped, representing 76.6% of the total studied. The modal value of the sample is 7 assets (115 households), closely followed by the 6-asset stratum (112 households) and the 8-asset stratum (89 households).

In contrast, observations of the extreme values of the distribution are atypical. At the threshold of greatest vulnerability, only 2 households reported lacking all the assessed goods and 3 households reported owning only one; at the upper end of accumulation, only a single household registered 14 goods. This frequency structure suggests that the bulk of the sector's population maintains intermediate levels of material provision, with scenarios of absolute deprivation or extraordinary accumulation being statistically marginal. The table below summarizes these descriptive findings:

Table 3.2: Distribution of households by number of assets owned

Number of Assets Owned	Number of Households
0	2
1	3
2	13
3	27
4	46
5	78
6	112
7	115
8	89
9	59
10	31
11	11
12	4
14	1

Vol. 11, No. 05; 2026

ISSN: 2456-3676

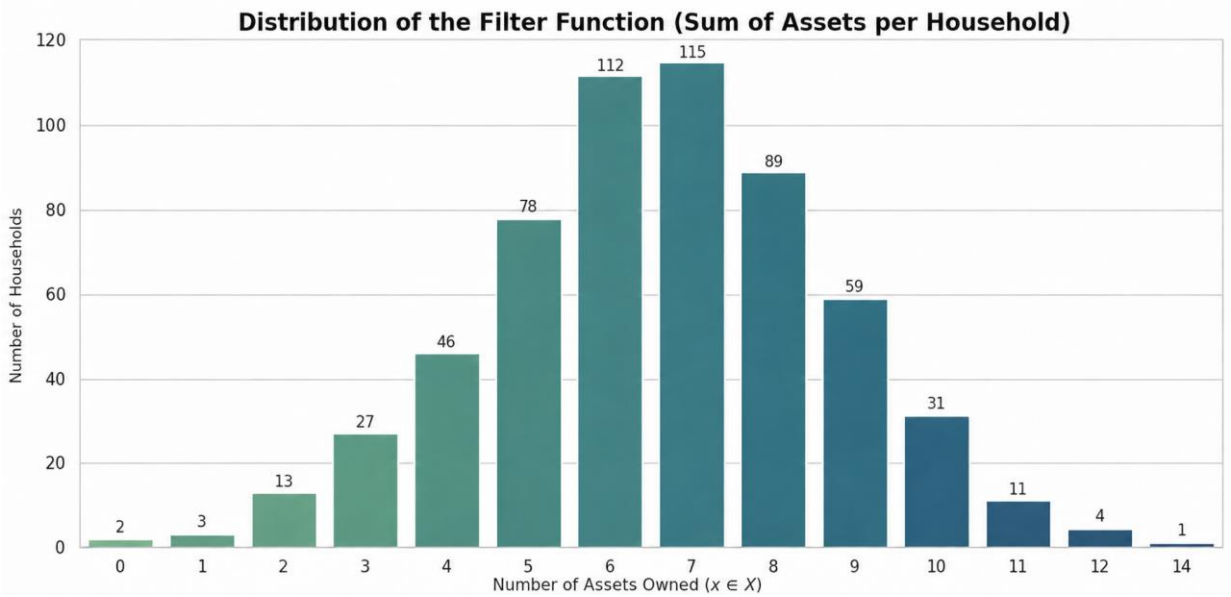

Figure 3.2: Distribution of the filter function, defined as the sum of assets per household.

Since the range of this data set runs from 0 to 14 assets owned, a methodologically appropriate choice is to divide it into 5 intervals (the n_cubes parameter), maintaining an overlap level between 30% and 35%, consistent with the criterion used in the reference study. In that work, the total range was 34 assets and was divided into 10 intervals, so that each covered approximately 4 to 5 assets. To preserve a similar resolution for the Mata Hambre case, a proportional partition is adopted over the observed range.

Thus, if 5 intervals with 35% overlap are used, the cover of the filter function for the Mata Hambre data set can be represented as shown in Table 3.3.

Table 3.3: Cover of the filter function for the Mata Hambre data set

	0	**1**	**2**	**3**	**4**
Interval elements	0.0–3.9	2.5–6.4	5.1–8.9	7.6–11.5	10.1–14.0
Number of households	45	263	316	190	16

The Interval Elements row will show decimal values (for example, 0.0–3.9 or 2.5–6.4). This is normal and consistent with the TDA approach, since the algorithms build the cover over continuous spaces, even though the variable analyzed here—the sum of assets owned—takes integer values. As a result, a household with 3 assets can simultaneously belong to two intervals, since that value falls within both overlapping ranges.

Vol. 11, No. 05; 2026

ISSN: 2456-3676

3.5.3. Clustering

For the local clustering stage within the intervals defined by the Mapper algorithm, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) was implemented, in line with the methodological strategy of the reference literature. This general-purpose algorithm, introduced by Ester et al. (1996), is widely recognized in the literature for its ability to identify arbitrarily shaped clusters and detect outliers in noisy data spaces.

Unlike other methods that require the number of clusters to be predefined, DBSCAN operates on the notion of local density and requires setting two fundamental parameters: the neighborhood radius (ε) and the minimum number of points (minPts). Operationally, the algorithm classifies as a core point any observation with at least minPts neighbors within a maximum distance ε. From these core points, DBSCAN recursively groups densely connected points.

The parameterization of minPts in the literature is usually guided by the heuristic rule based on the order of magnitude of the natural logarithm of the sample, i.e., ln(n). In the context of this research, the data space consists of n = 591 households that completed the 18 selected SIUBEN socioeconomic variables. Since ln(591) ≈ 6.38, and under the principle of maintaining a parsimonious parameterization that avoids excessive fragmentation of the clusters, the value minPts = 5 was adopted.

For the empirical determination of the ε parameter, a k-distance graph was constructed, set to k = 5 (the distance from each household to its fifth nearest neighbor). Applying the probability-based semimetric developed for this study, the resulting graph (see Figure 3.3) shows a gradually increasing curve that culminates in a steep slope at the upper tail. This inflection point, known as the "elbow," suggests the optimal density threshold. Consequently, based on the onset of that curvature, the parameter $\varepsilon \approx 2.2 \times 10^{-2}$ was set for the final topological clustering.

Vol. 11, No. 05; 2026

ISSN: 2456-3676

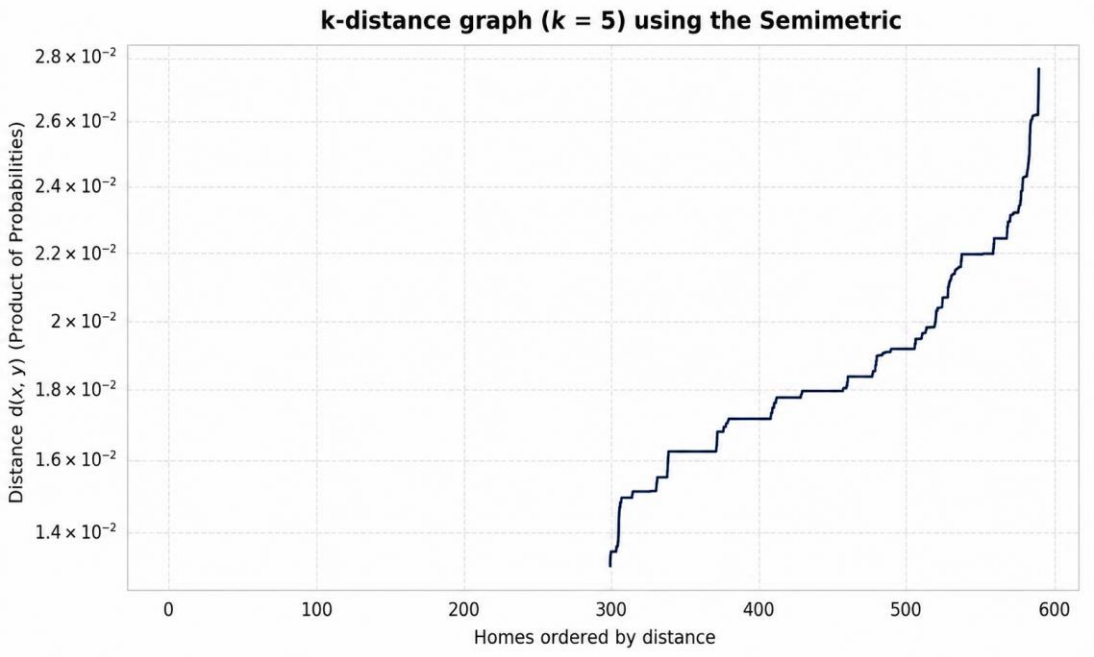

Figure 3.3: k-distance graph (k = 5) using the semimetric.

For the Euclidean metric, we repeated the same procedure using the k-distance graph again, with k = 5, presented in Figure 3.4. In this case, the graph's step-like structure reflects the fact that the data are binary, so the possible Euclidean distances between observations take on a discrete set of values. Although an "elbow" as smooth as that seen with continuous data is not observed, a significant change between the main distance levels is nonetheless apparent.

For this reason, we selected $\varepsilon = 1.5$, a value that captures sufficiently wide neighborhoods without excessively merging groups with distinct response patterns.

In summary, the parameters used for the DBSCAN algorithm were minPts = 5, $\varepsilon \approx 2.2 \times 10^{-2}$ for the new semimetric, and $\varepsilon = 1.5$ for the Euclidean metric.

Vol. 11, No. 05; 2026

ISSN: 2456-3676

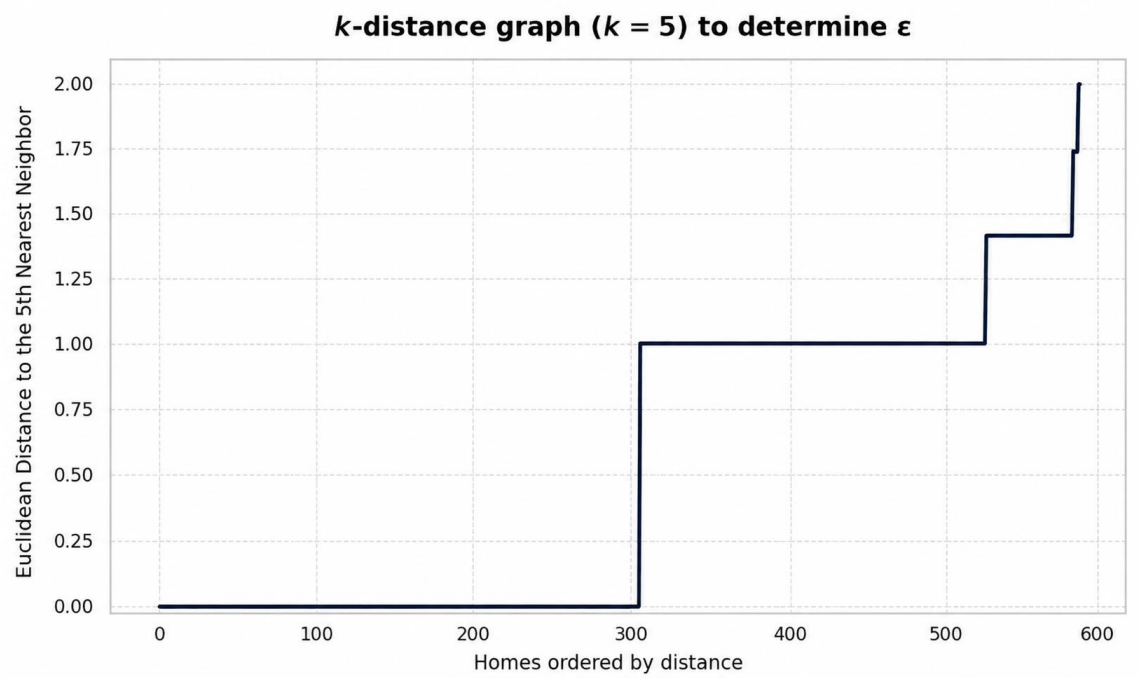

Figure 3.4: k-distance graph (k = 5) for the Euclidean metric.

3.5.4. Construction of the Topological Graph (Mapper Algorithm)

To model the household network, the Mapper algorithm was implemented following these technical parameters:

1. Filter / Lens (Filter): Instead of using generic dimensionality-reduction methods, a one-dimensional filter function was defined based on the sum of possessions (physical and technological assets) over the 18-dimensional binary matrix. This made it possible to project household data into an interpretable space based on their accumulation volume, in which households are ordered according to the number of goods they own.

2. Cover: The projected one-dimensional space was divided using overlapping continuous intervals. Based on the distribution of the filter function, the cover was parameterized with 5 intervals (n_cubes = 5) and 35% overlap (perc_overlap = 0.35). This overlap ensures that the transition between different levels of resilience and poverty is captured as a continuum, allowing "bridge" households to connect different structures of the graph.

3. Local Clustering (Clustering): Within each interval of the cover, the density-based spatial clustering algorithm DBSCAN was applied. To adequately capture the nature of the data, both the Euclidean distance and the probability-based semimetric were applied. This semimetric assigns greater weight to uncommon similarities (rare assets) and reduces the importance of widespread deprivations. This made it possible to group households with genuinely similar

Vol. 11, No. 05; 2026

ISSN: 2456-3676

vulnerability profiles and to isolate outliers (noise), thereby forming the nodes and edges of the final simplicial complex.

4. Results Analysis

Before the analytical interpretation of the topological networks obtained for the Mata Hambre sector, it is essential to define the notational convention that will govern the reading of the graphs generated by the Mapper algorithm.

Formally, each vertex or node n of the topological network is identified by the ordered pair (i, c). In this notation, i represents the index of the interval (derived from the filter function) in which the households are located, ordered ascendingly according to the total number of assets or goods reported in SIUBEN. In turn, c designates the label of the specific cluster formed internally within that interval i after applying the DBSCAN algorithm. It should be noted that the numbering of c is strictly categorical and arbitrary.

As an illustration, a node labeled (3, 0) encapsulates exclusively those households belonging to interval 3 that, based on their density, were grouped into cluster 0. Following this logic, node (3, 1) encompasses the households from that same interval 3 that form a different cluster, cluster 1.
Regarding network connectivity, the topology dictates that an edge is drawn between two nodes if, and only if, there is a non-empty intersection between the sets of households they represent. Since the filter function used projects the data space onto a linear dimension (the sum of assets), the overlap configuration determines that an interval can only intersect with its immediate neighbors. For example, households captured in interval 1 can only share membership with clusters originating in intervals 0 or 2.

This mathematical architecture implies that the graph's connections are strictly restricted to nodes from contiguous intervals. Additionally, because DBSCAN performs an exclusive partition within each block, the internal clusters of the same interval are disjoint by definition. Empirically, this means that a Mata Hambre household cannot belong simultaneously to nodes (3, 0) and (3, 1). However, a household can indeed be present at the same time in node (2, 0) and node (3, 0), and it is precisely this overlap that generates the edge and reveals the hidden continuities in the structure of vulnerability.

4.1. Graph — Probability-Based Semimetric

Figure 4.1 presents the Mapper graph generated from the SIUBEN data for the Mata Hambre sector, using the probability-based semimetric and the sum-of-components filter function. In this representation, each node encapsulates a cluster of households grouped within the same interval by the DBSCAN algorithm.

Vol. 11, No. 05; 2026

ISSN: 2456-3676

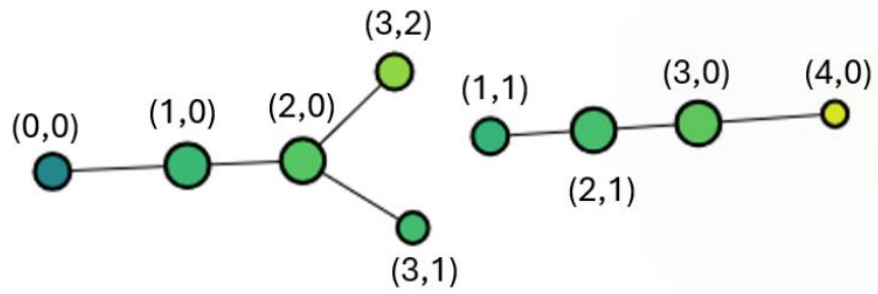

Figure 4.1: Mapper graph built from the probability-based semimetric. Each node is labeled (i, c), where i indicates the interval of the filter function and c the cluster identified within that interval.

A central methodological aspect of this visualization is the color-scale configuration: the color of each node reflects the average Quality of Life Index (ICV) of the households that comprise it. This projection allows the discovered geometric structure to be directly contrasted with SIUBEN's official vulnerability metric. Darker tones (green/bluish) indicate a lower ICV (conditions of greater poverty), while lighter tones (yellow) correspond to higher average ICV values. Likewise, the diameter of each node is proportional to the number of households represented in that grouping.

Unlike a continuous topological structure, the graph reveals a clear fragmentation into two connected components. On one hand, a main structure is observed that begins at node (0, 0) and branches from (2, 0) into the branches (3, 1) and (3, 2). On the other hand, there is a topologically isolated subgroup that evolves linearly from node (1, 1) to (4, 0).

This disconnection is a fundamental analytical finding. It indicates that, although households from both groups may fall within the same intervals of the filter function, the specific nature of their sociodemographic characteristics is radically different. Since the semimetric used assigns greater weight to statistically unusual similarities, the isolated subgroup (1,1)–(4,0) groups together households that share atypical deprivation or possession patterns relative to the rest of the Mata Hambre sector. The semimetric amplifies this multidimensional distance, fracturing the socioeconomic space and isolating these households from the main distribution, regardless of whether their average ICV values are similar to those of the majority group. As will be shown in the comparative analysis, this high sensitivity of the semimetric to marginal probabilities tends to hyper-fragment the continuity of the social fabric.

Vol. 11, No. 05; 2026

ISSN: 2456-3676

4.2. Graph — Euclidean Metric

Figure 4.2 presents the Mapper graph generated using the standard Euclidean metric. As in the previous model, the color scale of the nodes reflects the average Quality of Life Index (ICV) of the households that comprise them.

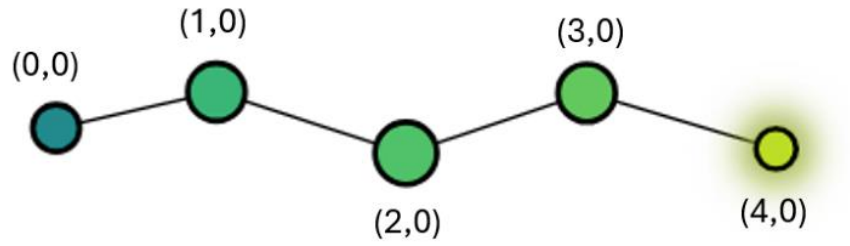

Figure 4.2: Mapper graph built using the Euclidean metric. Each node is labeled (i, c), where i represents the interval of the filter function and c the corresponding cluster within that interval.

Unlike the level of fragmentation observed with the semimetric, the Euclidean metric reveals a perfectly continuous and linear topological structure. The graph is composed of five consecutive nodes: (0,0), (1,0), (2,0), (3,0), and (4,0). This indicates that the DBSCAN algorithm, operating under the notion of Euclidean distance, identified exactly one cohesive density cluster for each interval of the filter function. Moreover, the uninterrupted presence of edges between all adjacent nodes confirms that there are a smooth transition and a genuine overlap of households throughout the entire distribution.

In reference literature, Anderson et al. (2022) considered the consolidation of a single node per interval under the Euclidean metric to be a limitation, arguing that it omitted atypical branches in national-scale studies. The probability-based semimetric outweighed marginal deprivations, artificially fracturing groups of households that shared similar living conditions. The Euclidean metric, by contrast, treats the feature space in a more balanced way. In doing so, it succeeds in modeling urban poverty not as a set of disconnected anomalies but as a structural continuum.

As can be seen in the visual progression of the graph, there is a clear transition, free of breaks or isolation, from the levels of greatest vulnerability and low ICV (node (0,0), dark tones) to the strata of least deprivation (node (4, 0), yellow tone). To empirically support this geometric configuration, Table 4.1 details the descriptive statistics of the Quality of Life Index within each cluster.

Vol. 11, No. 05; 2026

ISSN: 2456-3676

Table 4.1: Summary of average ICV score by topological node

Node	Average ICV	ICV σ	Minimum ICV	Maximum ICV	# households
(0, 0)	2.422	0.753	1	4	45
(1, 0)	3.015	0.751	1	4	263
(2, 0)	3.169	0.729	1	4	451
(3, 0)	3.273	0.744	2	4	187
(4, 0)	3.700	0.675	2	4	10

The statistical data confirms a monotonically increasing progression: the average ICV rises steadily from 2.422 at the base to 3.700 at the apex of the network. Particularly notable is the structural change beginning at node (3,0), where the recorded minimum ICV rises from 1 to 2, indicating the statistical disappearance of extreme poverty in that section of the graph.

Likewise, the household-count column reflects the mathematical nature of the Mapper algorithm. The total sum of observations in the network (956) exceeds the 591 households in the original SIUBEN sample. This difference occurs because the overlap parameter allows families located in multidimensional transition zones to belong simultaneously to two adjacent nodes. The bulk of the population is concentrated in the intermediate strata (1,0) and (2,0), showing that most of the sector shares a common space of vulnerability. It is therefore concluded that the Euclidean metric offers a substantially more robust, faithful, and interpretable geometric representation of the sector's socioeconomic reality.

4.2.1. Projection of Socioeconomic Indicators onto the Topological Network

In this section, complementary multidimensional information from the SIUBEN microdata is integrated, overlaying it onto the continuous structure of the Mapper graph generated with the Euclidean metric (see Figure 4.2). By reconfiguring the node color scale as a function of specific variables, it becomes possible to visually reveal the distribution of different vulnerabilities throughout the network.

This analytical approach allows deep conclusions to be drawn about the direct relationship between the Quality of Life Index (ICV) and the concentration of particular deprivations. By projecting these indicators onto the “shape” of the data, it becomes easier to identify which goods, services, or deprivations are the true catalysts that position a household at the extremes of greater or lesser poverty within the urban fabric of the Mata Hambre sector.

Relationship Between the Level of Vulnerability and Asset Ownership

To further examine the poverty structure of the Mata Hambre sector, the relationship between the network structure and the ownership of specific goods was analyzed. To this end, the continuous Euclidean graph (previously validated) was colored according to the percentage of households

Vol. 11, No. 05; 2026

ISSN: 2456-3676

that reported owning each item. In the resulting visualizations, dark purple tones indicate low prevalence of the good at that node, while light yellow tones represent high ownership rates.

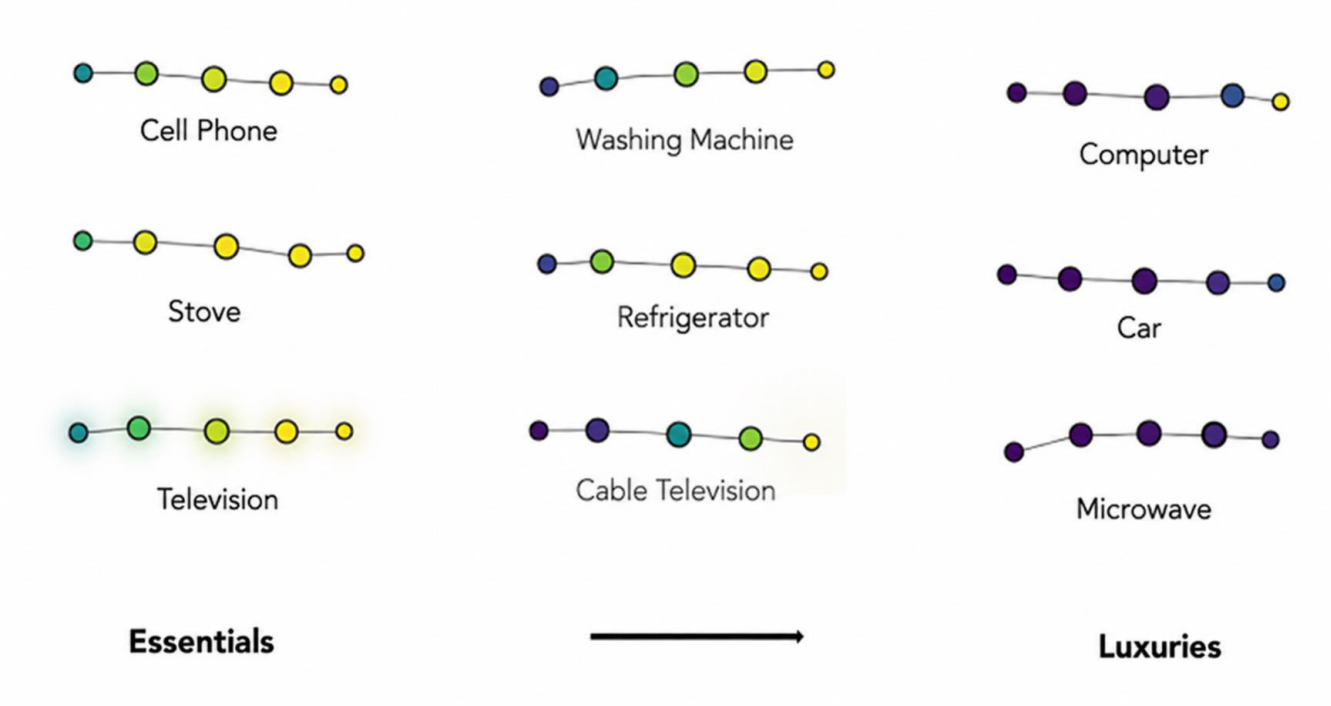

Figure 4.3: Overlay of asset ownership on the Mapper graph.

Figure 4.3 shows this projection for a representative selection of goods, empirically organized into three main categories based on their topological behavior.

First, the essential goods are identified. Items such as the cell phone, stove, and television show high ownership rates (yellow and light-green tones) across nearly the entire network, with the sole exception of node (0, 0), which concentrates extreme vulnerability. This suggests that, in the current Dominican context, basic connectivity and food preparation are prioritized regardless of income constraints.

Second, the transitional goods emerge. Items such as the washing machine and cable television show a very pronounced color progression. They present nearly zero ownership rates in the most vulnerable strata (nodes 0, 0 and 1, 0), but experience a significant jump toward lighter tones from node (2,0) and (3,0) onward. These goods act as topological markers of the transition out of critical poverty and into an intermediate standard of living.

Finally, the relatively privileged goods. Surprisingly, possessions such as the computer (with or without internet) and the automobile remain very dark in tone across almost the entire graph.

Vol. 11, No. 05; 2026

ISSN: 2456-3676

Their ownership is restricted exclusively to the last nodes at the right-hand end, and even there, at moderate rates.

This last finding allows for important sociocultural reflections on the measurement of poverty. For example, comparing the television graph with that of the microwave, the microwave behaves almost like a luxury item (very few own one), while the television permeates almost the entire network. From a strict market standpoint, a microwave is usually cheaper than a television; however, Mata Hambre families clearly prioritize access to information and entertainment over technological convenience in the kitchen.

This divergence makes even more sense when the sociocultural factor is introduced: whereas in countries such as the United States the microwave is considered an essential, ubiquitous everyday tool, in the Dominican context it is a much less desirable appliance due to marked differences in culinary customs and traditional food-preparation methods.

To empirically support the visual observations derived from the colored graph, Table 4.2 details the exact proportion of households that own each good within the five topological nodes. Analysis of these frequencies reveals that the poverty transition in Mata Hambre is not linear but instead occurs through well-defined structural "jumps" between clusters.

Table 4.2: Proportion of households owning each good, by node

Variable / Good	**(0, 0)**	**(1, 0)**	**(2, 0)**	**(3, 0)**	**(4, 0)**
HasTelevision	0.511	0.753	0.907	0.984	1.0
HasCableTelevision	0.044	0.141	0.472	0.829	1.0
HasComputer	0.0	0.03	0.084	0.267	1.0
HasComputerWithInternet	0.0	0.019	0.062	0.241	0.9
HasTablet	0.022	0.038	0.067	0.16	0.4
HasCellPhone	0.489	0.844	0.925	1.0	1.0
HasCellPhoneWithInternet	0.244	0.631	0.827	0.936	0.9
HasLandline	0.0	0.099	0.293	0.599	1.0
HasRadio	0.067	0.125	0.279	0.487	0.5
HasStove	0.711	0.966	0.991	0.989	1.0
HasMicrowave	0.0	0.008	0.053	0.123	0.1
HasAirConditioning	0.0	0.0	0.018	0.037	0.1
HasRefrigerator	0.2	0.81	0.967	0.989	1.0
HasWashingMachine	0.156	0.559	0.845	0.963	1.0
HasInverter	0.0	0.0	0.013	0.043	0.0
HasElectricGenerator	0.0	0.0	0.002	0.005	0.0
HasPrivateAutomobile	0.0	0.015	0.035	0.112	0.3
HasPrivateMotorcycle	0.0	0.008	0.024	0.037	0.0

Vol. 11, No. 05; 2026

ISSN: 2456-3676

A detailed inspection of the proportions confirms the topological stratification previously proposed and yields three fundamental quantitative findings:

- **The basic-subsistence gap (Node 0 to 1): Node (0,0)** shows critical precariousness. Items that the rest of the sector take for granted are luxuries here: only 20% own a refrigerator and barely 15.6% have a washing machine. However, upon moving to node (1,0) (with which topological overlap exists), an abrupt jump occurs: refrigerator ownership soars to 81% and washing-machine ownership to 55.9%.
- **The consolidation of connectivity (Nodes 2 and 3):** The transition to node (2,0) marks the democratization of basic digital access. It is in this cluster where cell phones with internet surpass 80% penetration. Upon reaching node (3,0), intermediate-comfort goods such as cable television experience their largest jump, rising from 47.2% at node 2 to 82.9% at node 3.
- **The exclusivity barrier (Node 3 to 4):** The jump to node (4,0) shows how the algorithm mathematically separated the sector's non-poor household group. Computer ownership, which reached only 26.7% in the immediately preceding stratum, jumps sharply to 100% at node (4,0). Likewise, private-automobile ownership triples, rising from 11.2% to 30%. These drastic changes confirm that node (4,0) encapsulates a non-poor socioeconomic profile with a qualitatively different purchasing power from the rest of Mata Hambre.

Typology of Households Along the Socioeconomic Continuum

The analysis of asset distribution over the Euclidean graph makes it possible to sketch an empirical typology of Mata Hambre households based on their topological position.

Nodes (0,0) and (1,0) group together families facing the most severe levels of vulnerability (the lowest average ICV values). In this base stratum, ownership of middle-class amenities is virtually nonexistent, and in the most critical cases at node (0,0), even access to the full set of essential goods is not guaranteed. These are households operating at the margins of urban material subsistence.

Node (2,0) acts as the graph's inflection point, representing a stratum of “moderate vulnerability” or transition. In this cluster, households already have secured basic connectivity and cooking (essential goods) and begin to show the first significant rates of acquisition of items that facilitate household logistics, such as the refrigerator or washing machine.

For their part, nodes (3,0) and (4,0) encapsulate the households with the greatest relative consolidation within the sector. Node (3,0) marks the point at which transitional amenities, such as cable television, reach most families, defining the profile of poor/vulnerable households. Finally, node (4,0) represents the apex of the local distribution and is associated with the non-poor stratum. It is the only region of topological space where relatively privileged goods, such as the personal computer or the automobile, begin to have a representative presence, outlining a profile that resembles the city's most stable socioeconomic stratum.

Vol. 11, No. 05; 2026

ISSN: 2456-3676

It is imperative to note, however, that traditional sociological labels such as “extreme poverty,” “moderate poverty,” “poor/vulnerable,” or “non-poor” are imprecise analytical categories with fuzzy boundaries. The main strength of the topological model over linear parametric classification is that it mathematically evidences this fluidity: the uninterrupted existence of edges connecting the sequence from node (0,0) to (4,0) demonstrates that there is a genuine overlap of households between each level. Poverty within the urban fabric of Mata Hambre does not operate in watertight compartments but rather as a structural continuum in which transitions of accumulation and deprivation intertwine.

5. Conclusions

At a general level, it has been empirically shown that Topological Data Analysis (TDA) constitutes a highly effective computational tool for investigating the multidimensional “shape” of urban poverty. By applying the Mapper algorithm to the SIUBEN data, it was possible to move beyond the limitations of traditional aggregate indices, revealing the underlying geometric structure of vulnerability in the Mata Hambre sector.

First, it was found that the choice of distance metric is a determining factor in the topological modeling of socioeconomic data. While prior literature suggested the use of probability-based semimetrics for household surveys, in the micro-territorial context of Mata Hambre this approach tended to hyper-fragment the network. By contrast, the application of the standard Euclidean metric proved superior for this urban ecosystem, revealing that poverty does not operate as a set of isolated anomalies but as a structural continuum. The overlap between clusters (nodes) confirmed that stratum divisions are fluid and that households transition gradually through different levels of accumulation and deprivation.

Second, overlaying specific variables onto the Euclidean graph proved to be an excellent method for hypothesis generation and for understanding household behavior. This technique made it possible to classify goods into three clear topological categories: essential, transitional, and relatively privileged. Analysis of these projections showed that Mata Hambre families make consumption decisions strongly influenced by cultural patterns and logistical needs specific to the Dominican urban environment (such as prioritizing communication and entertainment devices over certain kitchen appliances), factors that the Quality of Life Index (ICV) alone fails to capture.

Ultimately, the integration of unsupervised learning algorithms (DBSCAN) and TDA provides a much more granular and interpretable view of inequality, establishing topology as a robust analytical framework for data science applied to the social sciences in the Dominican Republic.
This study has several methodological limitations that should be considered when interpreting its findings and that outline priorities for future refinement.

First, although the comparison between the prevalence-weighted semimetric and the Euclidean metric (Sections 4.1 and 4.2) shows that the latter yields a more cohesive and interpretable graph

Vol. 11, No. 05; 2026

ISSN: 2456-3676

for this micro-territory, this comparison was carried out qualitatively, through visual inspection of graph connectivity and ICV coloring.

Second, the alignment between the topological continuum and the Quality of Life Index (ICV) is, at present, a visual and descriptive result rather than a statistically quantified one.

Third, missing and "no sabe" (don't know) responses were coded as 0 (absence) during binarization (Section 3.2.1). While this decision is conservative and transparent, it can bias deprivation estimates upward for items with high non-response.

References

Abdul Rahman, M., Sani, N. S., Hamdan, R., Ali Othman, Z., & Abu Bakar, A. (2021). A Clustering Approach to Identify Multidimensional Poverty Indicators for the Bottom 40 Percent Group. PLOS ONE, 16(8), e0255312. https://doi.org/10.1371/journal.pone.0255312

Alkire, S., & Foster, J. (2011). Counting and Multidimensional Poverty Measurement. Journal of Public Economics, 95(7–8), 476-487. https://doi.org/10.1016/j.jpubeco.2010.11.006

Alkire, S., & Santos, M. E. (2010). Acute Multidimensional Poverty: A New Index for Developing Countries (Tech. Rep. No. 2010/11). Human Development Report Office, United Nations Development Programme. https://hdr.undp.org/system/files/documents/hdrp201011.pdf

Anderson, J. R., Memić, F., & Volić, I. (2022). Topological Data Analysis and UNICEF Multiple Indicator Cluster Surveys. Journal of Quantitative Economics, 20(2), 281-309. https://doi.org/10.1007/s40953-022-00288-w

Arenas, R. D. M., Baltazar, M. P. D., Aguilar, J. A. F., Vera, E. J. Z., Arriola, G. C. P., Espinoza, H. J. G., & Castaneda, H. A. (2023). Applications of simplicial complexes in the calculation of the homotopy group of the sphere and in systems engineering. Proceedings of the 21st LACCEI International Multi-Conference for Engineering, Education and Technology. https://doi.org/10.18687/LACCEI2023.1.1.1079

Belchí Guillamón, J., Brodzki, J., Burfitt, M., & Niranjan, M. (2019). A Numerical Measure of the Stability of Mapper-Type Algorithms. arXiv preprint arXiv:1906.01507. https://arxiv.org/abs/1906.01507

Chazal, F., & Michel, B. (2021). An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. Frontiers in Artificial Intelligence, 4, 667963. https://doi.org/10.3389/frai.2021.667963

Dłotko, P. (2019). Ball Mapper: A Shape Summary for Topological Data Analysis. arXiv preprint arXiv:1901.07410. https://arxiv.org/abs/1901.07410

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. kdd, 96(34), 226-231.

Godwin, A., Thielmeyer, A. R. H., Rohde, J. A., Verdín, D., Benedict, B. S., Baker, R. A., & Doyle, J. (2019). Using Topological Data Analysis in Social Science Research: Unpacking

Decisions and Opportunities for a New Method. ASEE Annual Conference and Exposition, Conference Proceedings. https://doi.org/10.18260/1-2--33522

IBM. (n.d.). CRISP-DM help overview [IBM Documentation]. Retrieved March 24, 2026, from https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview

KeplerMapper. (2024). KeplerMapper 2.1.0 documentation. Retrieved March 24, 2026, from https://kepler-mapper.scikit-tda.org/en/latest/index.html

Llaugel, F., & Llaugel, M. (2023). Análisis del impacto de los subsidios sociales: propuesta para mejor asignación en República Dominicana. Ciencia, Ingenierías y Aplicaciones, 6(1), 81-106. https://doi.org/10.22206/cyap.2023.v6i1.pp81-106

Llaugel, F. A., Perez, E., & Llaugel, M. (2024). Aplicación del método de análisis de componentes principales al índice de pobreza multidimensional: caso Los Alcarrizos. Ciencia y Sociedad, 49(1), 45–70. https://doi.org/10.22206/cys.2024.v49i1.2981

Munch, E. (2017). A User's Guide to Topological Data Analysis. Journal of Learning Analytics, 4(2), 47-61. https://doi.org/10.18608/jla.2017.42.6

Murtagh, J. (2024, March). How a Classic Bridge-Crossing Puzzle Inspired New Math. Scientific American. Retrieved April 6, 2026, from https://www.scientificamerican.com/article/how-the-seven-bridges-of-koenigsberg-spawned-new-math/

Sen, A. (1999). *Development as Freedom*. New York: Alfred A. Knopf / Oxford University Press.

Singh, G., Mémoli, F., & Carlsson, G. (2007). Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. Symposium on Point Based Graphics, 91-100. https://research.math.osu.edu/tgda/mapperPBG.pdf

Trento Oliveira, L., Kuffer, M., Schwarz, N., & Pedrassoli, J. C. (2023). Capturing Deprived Areas Using Unsupervised Machine Learning and Open Data: A Case Study in São Paulo, Brazil. European Journal of Remote Sensing, 56(1), 2214690. https://doi.org/10.1080/22797254.2023.2214690

UNDP (United Nations Development Programme). 2025 Global Multidimensional Poverty Index (MPI): Overlapping Hardships: Poverty and Climate Hazards. Retrieved April 6, 2026, from https://hdr.undp.org/content/2025-global-multidimensional-poverty-index-mpi

Wasserman, L. (2018). Topological Data Analysis. Annual Review of Statistics and Its Application, 5(1), 501-532. https://doi.org/10.1146/annurev-statistics-031017-100045