

Análisis Topológico de la Economía Colombiana: 2008 - 2025

Daniel Posada, Juan Diego Naranjo

28 de noviembre de 2025

Resumen

Este estudio aplica técnicas de Análisis Topológico de Datos (TDA) para investigar la estructura y dinámica de la economía colombiana durante el período 2008–2025. Utilizando series temporales mensuales de siete indicadores macroeconómicos (precio del petróleo Brent, índice COLCAP, índice de precios al consumidor, índice de seguimiento económico, tasa de cambio real, tasa de desempleo y PIB), se construyen representaciones topológicas mediante dos enfoques complementarios: embeddings de Takens multivariados y univariados, y el algoritmo Mapper. Se calculan diagramas de persistencia, paisajes de persistencia y otras representaciones vectoriales para caracterizar la topología del sistema económico. Los resultados revelan patrones topológicos asociados con crisis económicas y cambios de régimen, proporcionando una perspectiva novedosa sobre la estructura de la economía colombiana y las interconexiones entre variables macroeconómicas.

Palabras clave: análisis topológico de datos, economía colombiana, embeddings de Takens, homología persistente, Mapper, series temporales macroeconómicas

Introducción

La economía colombiana ha experimentado fluctuaciones significativas en las últimas dos décadas [gallego2025], marcadas por crisis financieras globales (como la de 2008), cambios en los precios de materias primas o petróleo (como la fuerte caída iniciada en junio de 2014) y los efectos de la pandemia COVID-19. Además, ha mostrado vulnerabilidad ante choques externos y la influencia de fenómenos locales, como el choque asociado a la incertidumbre política, el estallido social (2019, 2021) y las propuestas de reforma y regulación petrolera promovidas por el gobierno de Colombia en 2022 [gallego2025]. Comprender la estructura subyacente y las dinámicas de sistemas económicos complejos, como el mercado financiero colombiano requiere herramientas que vayan más allá de los métodos estadísticos tradicionales [shultz2023].

El desafío fundamental en el análisis de datos modernos es la identificación, extracción y explotación de la estructura subyacente en datos que a menudo son complejos, de alta dimensión y

ruidosos. Los métodos tradicionales suelen asumir linealidad o recurren a modelos paramétricos que pueden ser insuficientes para capturar relaciones no lineales [**shultz2023**].

El Análisis Topológico de Datos (ATD) es un campo reciente y de rápido crecimiento que surgió de la topología (algebraica) aplicada y la geometría computacional a principios de siglo [**carlsson2009**]. El ATD ofrece un marco matemático robusto para estudiar la "forma" de los datos, capturando características geométricas y topológicas que permanecen invariantes bajo transformaciones continuas. El ATD propone un conjunto de teorías y herramientas que pueden ser utilizadas de forma independiente o en combinación con otras técnicas de análisis de datos y aprendizaje estadístico [**bubenik2015**].

Para lograr esto, ATD usa una representación topológica de los datos, comúnmente una nube de puntos en un espacio métrico y construye aproximaciones de su forma continua, a menudo a través de un complejo simplicial y/o una filtración de complejos. Una de las herramientas clave del ATD, la Homología Persistente (HP), permite cuantificar la forma y conectividad de los datos [**carlsson2009**], rastreando características topológicas como:

- Componentes conexas (H_0 o β_0).
- Ciclos o agujeros (H_1 o β_1).
- Cavidades o vacíos (H_2 o β_2).

La Homología Persistente permite identificar estructuras persistentes, es decir, estructuras topológicas que son robustas a las perturbaciones o al ruido en los datos [**carlsson2009**].

En el contexto económico y financiero, el ATD ha demostrado ser una herramienta poderosa para el análisis de series temporales multivariadas [**gallego2025, majumdar2020**]. Específicamente, ha sido utilizado para la detección temprana de señales de transiciones críticas o crisis [**gallego2025**]. Al aplicar ATD a series de tiempo financieras, se utiliza la técnica de ventanas deslizantes (sliding windows) para embeber los datos en un espacio de dimensión superior (fase space) [**meinecke2025**].

Al mismo tiempo, el Análisis Topológico de Datos (ATD) ofrece herramientas complementarias para abordar los datos desde perspectivas visuales y estadísticas. Por un lado, para la exploración de datos y la visualización de relaciones en conjuntos de datos complejos y de alta dimensionalidad, se utiliza el Algoritmo de Mapper. El algoritmo de Mapper construye un grafo simplificado que resume la información topológica de los datos, facilitando la interpretación y la comprensión de las estructuras subyacentes, como clusters o conectividad [**vanveen2019**].

Por otro lado, para la cuantificación y el análisis estadístico, la información topológica mostrada por los Diagramas de Persistencia puede resumirse en los Paisajes de Persistencia (Persistence Landscapes) [**bubenik2017dlotko**].

Estos paisajes, introducidos por Bubenik en 2015 [**bubenik2017dlotko**], transforman los diagramas de persistencia en una secuencia de funciones continuas y lineales por partes. Esta representación funcional permite mapear los diagramas a un espacio de Hilbert, lo que facilita la aplicación de análisis estadístico riguroso, incluyendo el cálculo de medias y varianzas. De

hecho, se ha demostrado que las normas L_p de los paisajes de persistencia exhiben un fuerte crecimiento antes de las crisis, lo que sugiere que el ATD puede generar señales de alerta temprana de colapsos inminentes en el mercado [gallego2025].

Estos hallazgos refuerzan la utilidad del ATD como una nueva herramienta econométrica, complementaria a los métodos tradicionales, para el monitoreo de riesgo sistémico [shultz2023].

Planteamiento del Problema

La economía colombiana, como la mayoría de sistemas macroeconómicos complejos, puede entenderse como un sistema dinámico multivariado, donde numerosos indicadores interactúan entre sí de manera no lineal, generando comportamientos emergentes que no siempre son evidentes mediante técnicas estadísticas tradicionales. Esta complejidad plantea la necesidad de enfoques capaces de capturar la estructura global de los datos y describir la evolución de su forma y estructura a lo largo del tiempo [shultz2023].

En este trabajo surge la motivación de construir una representación topológica razonable, manejable y robusta al ruido de la economía colombiana entre 2008 y 2025, con el fin de explorar si herramientas del Análisis Topológico de Datos (en particular, Diagramas de Persistencia, Paisajes de Persistencia y Grafo Mapper) pueden detectar cambios significativos en períodos de transición, alteraciones de régimen o episodios de crisis. Este interés se centra especialmente en dos momentos críticos recientes: la crisis financiera global de 2008 y la crisis derivada de la pandemia de 2020, eventos que afectaron profundamente la dinámica macroeconómica nacional [gallego2025].

Bajo esta perspectiva, el problema fundamental consiste en determinar si la topología subyacente del sistema económico colombiano contiene firmas estructurales, patrones geométricos persistentes o cambios abruptos en su forma global que puedan asociarse con estos episodios críticos.

A partir de ello, las preguntas que guían este estudio son:

- ¿Qué estructura topológica subyace al espacio de estados de la economía colombiana?
- ¿Cómo se relacionan topológicamente las diferentes variables macroeconómicas a lo largo del tiempo?
- ¿Presentan los métodos del ATD (Diagramas de Persistencia, Paisajes de Persistencia, Grafos Mapper) cambios significativos durante períodos de transición o inestabilidad?

Objetivos

Objetivo General:

Caracterizar la estructura topológica de la economía colombiana durante 2008-2025 mediante técnicas de análisis topológico de datos.

Objetivos Específicos:

1. Construir embeddings de Takens multivariados y univariados de series temporales macroeconómicas colombianas.
2. Calcular diagramas de persistencia y paisajes de persistencia en múltiples dimensiones homológicas.
3. Aplicar el algoritmo Mapper para visualizar la estructura global del espacio de estados económicos.
4. Identificar patrones topológicos asociados con eventos económicos significativos.

Marco Teórico

Análisis Topológico de Datos

El Análisis Topológico de Datos (ATD) es un campo de las matemáticas aplicadas que estudia la forma de los datos utilizando conceptos de topología algebraica. A diferencia de métodos estadísticos tradicionales, el ATD es capaz de capturar características globales y multiescala de conjuntos de datos complejos.

El pipeline usual en Análisis Topológico de Datos (ATD) consta de las siguientes etapas:

1. **Adquisición de datos:** Se obtiene un conjunto de puntos $X = \{x_1, \dots, x_n\}$ en un espacio métrico, típicamente \mathbb{R}^d , que representa las observaciones o mediciones del fenómeno de interés.
2. **Construcción de la filtración:** Se construye una secuencia anidada de complejos simpliciales

$$\emptyset = K_0 \subseteq K_1 \subseteq K_2 \subseteq \dots \subseteq K_n$$

parametrizada por una escala ϵ o radio. Construcciones comunes incluyen el complejo de Vietoris-Rips, el complejo de Čech, o el complejo Alpha.

3. **Cálculo de homología persistente:** Se aplica homología simplicial a cada complejo en la filtración y se rastrean los generadores de homología a través de las escalas, identificando cuándo nacen y mueren las características topológicas.
4. **Generación de diagramas de persistencia:** Se visualizan las características topológicas como puntos (b, d) en el plano de persistencia, donde b es el parámetro de nacimiento y d el parámetro de muerte.
5. **Ánalysis:** Se extraen características relevantes de los diagramas mediante descriptores estadísticos (landscape de persistencia, imágenes de persistencia, curvas de Betti, Mapper) para su uso en tareas de aprendizaje automático o análisis estadístico.

Embedding de Takens

Dado que el pipeline del ATD requiere una nube de puntos (o un objeto que permita la generación de complejos simpliciales), es necesario embeber las series temporales en un espacio apropiado para su análisis topológico. Para ello, es usual usar embebimientos de ventanas deslizantes que permiten la reconstrucción del espacio de fases de la serie temporal.

El teorema de embedding de Takens [takens1981] establece que, bajo condiciones adecuadas, la dinámica de un atractor de un sistema dinámico puede reconstruirse a partir de observaciones escalares mediante embeddings de retardo temporal.

Embedding Univariado

Para una serie temporal univariada $\{x_t\}$, el embedding de Takens con dimensión d y retardo τ produce vectores:

$$\mathbf{y}_t = (x_t, x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-(d-1)\tau})$$

Los parámetros óptimos τ y d pueden determinarse mediante información mutua y el método de falsos vecinos más cercanos, respectivamente.

Embedding Multivariado

Para un sistema multivariado con n variables $\{x_t^{(1)}, \dots, x_t^{(n)}\}$, el embedding multivariado captura la dinámica conjunta:

$$\mathbf{y}_t = (x_t^{(1)}, \dots, x_t^{(n)}, x_{t-\tau}^{(1)}, \dots, x_{t-\tau}^{(n)}, \dots, x_{t-(d-1)\tau}^{(1)}, \dots, x_{t-(d-1)\tau}^{(n)})$$

Este enfoque es particularmente apropiado para sistemas económicos donde múltiples variables interactúan.

Complejo Vietoris-Rips

Dado un conjunto de puntos X en un espacio métrico (X, d) y un parámetro de escala $\epsilon > 0$, el complejo de Vietoris-Rips $\text{VR}_\epsilon(X)$ es el complejo simplicial definido por:

$$\text{VR}_\epsilon(X) = \{\sigma \subseteq X : d(x_i, x_j) \leq \epsilon \text{ para todo } x_i, x_j \in \sigma\}$$

Es decir, un conjunto de puntos $\{x_0, x_1, \dots, x_k\}$ forma un k -símplice en $\text{VR}_\epsilon(X)$ si y solo si la distancia entre cualquier par de puntos es a lo sumo ϵ .

Equivalentemente, $\text{VR}_\epsilon(X)$ es la bandera del grafo cuyos vértices son los puntos de X y cuyas aristas conectan pares de puntos a distancia a lo sumo ϵ . Esta construcción tiene la ventaja computacional de depender únicamente de las distancias por pares entre puntos, y satisface:

$$\text{VR}_{\epsilon_1}(X) \subseteq \text{VR}_{\epsilon_2}(X) \quad \text{si } \epsilon_1 \leq \epsilon_2$$

generando así una filtración natural al variar ϵ .

Homología Persistente

La homología persistente es la herramienta fundamental de ATD que rastrea la evolución de características topológicas (componentes conectadas, ciclos, cavidades) a través de múltiples escalas. Dado un conjunto de datos, se construye una filtración, es decir, una secuencia anidada de complejos simpliciales indexados por un parámetro de escala.

Formalmente, sea X un conjunto de puntos y consideremos una filtración de complejos simpliciales:

$$\emptyset = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n$$

donde cada K_i es un complejo simplicial parametrizado por valores $t_0 < t_1 < \cdots < t_n$.

Al incrementar el parámetro de filtración, las características topológicas “nacen” y “mueren”, generando un diagrama de persistencia que registra estos eventos como pares (nacimiento, muerte) en el plano. La persistencia de una característica, definida como muerte – nacimiento, cuantifica su importancia topológica.

Estos diagramas son una representación visual de las características topológicas detectadas por homología persistente. Consiste en un multiconjunto de tuplas, donde cada punto (b, d) representa una característica topológica que nace en el parámetro b y muere en el parámetro d , con $b < d$.

Los puntos cercanos a la diagonal $y = x$ representan características de corta duración (ruído), mientras que los puntos alejados de la diagonal corresponden a características persistentes (señal topológica significativa). La distancia vertical $d - b$ de un punto a la diagonal cuantifica su persistencia.

Dimensiones Homológicas

En este estudio consideramos tres dimensiones homológicas:

- H_0 : Componentes conectadas (mide fragmentación del espacio)
- H_1 : Ciclos unidimensionales (captura comportamiento cíclico/periódico)
- H_2 : Cavidades bidimensionales (detecta vacíos o agujeros en el espacio de estados)

Representaciones Topológicas

Paisajes de Persistencia

Los diagramas de persistencia no residen en un espacio vectorial, lo que complica análisis estadísticos estándar. Diversas representaciones vectoriales permiten trabajar con estos diagramas estadísticamente y en machine learning. Una de ellas son los paisajes de persistencia:

Los paisajes de persistencia son una representación funcional de los diagramas de persistencia que permite su vectorización y el uso de herramientas estadísticas estándar. Introducidos

por Bubenik (2015), transforman el diagrama de persistencia en una secuencia de funciones continuas y lineales a trozos.

Dado un diagrama de persistencia $\text{Dgm} = \{(b_i, d_i)\}$, para cada punto (b_i, d_i) se define la función piramidal:

$$f_i(t) = \begin{cases} t - b_i & \text{si } b_i \leq t \leq \frac{b_i + d_i}{2} \\ d_i - t & \text{si } \frac{b_i + d_i}{2} \leq t \leq d_i \\ 0 & \text{en otro caso} \end{cases}$$

El k -ésimo paisaje de persistencia $\lambda_k : \mathbb{R} \rightarrow \mathbb{R}$ se define como:

$$\lambda_k(t) = \text{k-max}\{f_i(t) : i = 1, \dots, n\}$$

donde k-max denota el k -ésimo valor más grande.

Una de las características más importantes de los paisajes de persistencia es que son estables bajo perturbaciones de los datos, lo cual permite confiabilidad en su interpretación y uso. Además, admiten operaciones vectoriales y estadísticas directamente en su espacio funcional, y pueden discretizarse mediante muestreo para obtener vectores de características de dimensión fija. Estas propiedades los hacen útiles para pruebas de hipótesis, análisis de clasificación y regresión sobre datos topológicos.

Medidas Estadísticas y Normas en Paisajes de Persistencia.

Medida	Definición	Interpretación
Normas		
L^∞	$\sup_{k,t} \lambda_k(t) $	Captura la persistencia máxima, identificando la característica más prominente.
L^p	$\left(\sum_k \int \lambda_k(t) ^p dt \right)^{1/p}$	Pondera características según persistencia. $p = 1$: suma áreas; $p = 2$: enfatiza las más persistentes.
Estadísticas		
Media	$\bar{\lambda} = \frac{1}{n} \sum_{i=1}^n \lambda^{(i)}$	Identifica características topológicas consistentes en la población.
Varianza	$\frac{1}{n} \sum_{i=1}^n \ \lambda^{(i)} - \bar{\lambda}\ _2^2$	Cuantifica la variabilidad topológica del conjunto de datos.
Distancia	$d(\lambda, \lambda') = \ \lambda - \lambda'\ _p$	Mide similitud topológica entre conjuntos de datos.

Algoritmo Mapper

El algoritmo Mapper [singh2007] construye un grafo simplificado que captura la estructura topológica de datos de alta dimensión. El proceso consta de tres pasos:

1. **Filtro:** Proyectar los datos a un espacio de menor dimensión mediante una función $f : X \rightarrow \mathbb{R}^m$
2. **Cobertura:** Cubrir la imagen de f con conjuntos solapados
3. **Clustering:** Para cada conjunto en la cobertura, agrupar los puntos en su preimagen y crear un nodo por cada cluster.

Los nodos se conectan si sus clusters correspondientes comparten puntos, resultando en un grafo que preserva propiedades topológicas del espacio original mientras proporciona una visualización interpretable desde las técnicas de análisis usual en grafos no dirigidos (componentes, puentes y distancias).

Aplicaciones de TDA en Economía

El Análisis Topológico de Datos (TDA) ha demostrado ser útil para estudiar dinámicas complejas en mercados financieros y sistemas macroeconómicos. Gidea y Katz (2017, 2018) mostraron que las normas L^p de los paisajes de persistencia aumentan de manera sistemática antes de crisis bursátiles, identificando señales tempranas previas al colapso de las puntocom en 2000 y a la quiebra de Lehman Brothers en 2008.

En el ámbito macroeconómico, Majumdar y Laha [majumdar2020] emplearon el algoritmo Mapper para analizar datos de India, identificando transiciones entre regímenes económicos y períodos de volatilidad elevada. Su estudio evidencia el potencial de TDA para segmentar configuraciones dinámicas en sistemas multivariados.

Shultz (2023) recopiló diversas aplicaciones de TDA en economía, destacando su uso en la detección de shocks financieros, la caracterización de ciclos no lineales y el análisis de co-movimientos entre activos. Asimismo, investigaciones recientes han aplicado homología persistente al estudio de criptomonedas y curvas de rendimiento, mostrando que los cambios topológicos reflejan episodios de tensión financiera.

Estos trabajos posicionan al TDA como una herramienta emergente para detectar transiciones, analizar interdependencias y caracterizar dinámicas complejas en datos económicos y financieros.

Metodología

A continuación, se describe la metodología utilizada para el análisis topológico de las series macroeconómicas, incluyendo las etapas de extracción de datos, preprocessamiento, transformación, construcción de estructuras topológicas y visualización.

Datos

Variables y Fuentes

Se utilizaron series temporales mensuales de siete indicadores macroeconómicos colombianos para el período febrero 2008 - enero 2025, con un total de 204 observaciones:

1. **Precio del petróleo Brent (USD por barril)**: Indicador de shocks externos en commodities.
2. **Índice COLCAP**: Principal índice bursátil colombiano.
3. **Índice de Precios al Consumidor (IPC)**: Medida de inflación.
4. **Índice de Seguimiento a la Economía (ISE)**: Proxy mensual del PIB.
5. **Índice de Tasa de Cambio Real**: Medida de competitividad externa.
6. **Tasa de Desempleo**: Indicador del mercado laboral.
7. **Producto Interno Bruto (PIB)**: Valor total de bienes y servicios, interpolado mensualmente para esta serie.

Los datos fueron obtenidos del Banco de la República de Colombia, DANE (Departamento Administrativo Nacional de Estadística), y fuentes internacionales para el precio del petróleo. Se realizó un proceso de extracción, ajuste y limpieza mensual, incluyendo interpolación para el PIB, abarcando el intervalo temporal máximo común a todas las fuentes disponibles.

Preprocesamiento

Todas las series de datos fueron normalizadas mediante escalamiento Min-Max al rango para garantizar la comparabilidad entre variables con diferentes unidades y magnitudes. La normalización se realizó con la siguiente fórmula:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Esta transformación preserva las propiedades dinámicas de las series, y además permite intuir una topología más rica en relaciones de crecimiento y decaimiento, al evitar que variables con mayor varianza dominen el análisis multivariado.

Pipeline Computacional

El análisis se implementó en Python utilizando librerías especializadas para Análisis Topológico de Datos: `giotto-tda` [tauzin2021], que ofrece herramientas para calcular diagramas de persistencia; `kmapper` [vanveen2019], la implementación de Mapper para construir grafos Mapper y `Persim`, utilizada para generar Paisajes de Persistencia.

El pipeline comprende las siguientes etapas:

Embeddings de Takens

Se implementaron dos enfoques paralelos:

Enfoque Multivariado:

- Input: Matriz (204×7) transpuesta a $(1 \times 7 \times 204)$ para formato $(n_{\text{samples}}, n_{\text{variables}}, n_{\text{timestamps}})$
- Parámetros: $\tau = 3$ (retardo trimestral), $d = 5$ (dimensión de embedding)
- Output: Nube de puntos en \mathbb{R}^{35} (5 retardos \times 7 variables)
- Interpretación: Sistema económico completo como una trayectoria única

Enfoque Univariado:

- Input: (7×204) tratando cada variable independientemente
- Parámetros: $\tau = 3, d = 5$
- Output: Siete nubes de puntos separadas en \mathbb{R}^5
- Interpretación: Dinámicas individuales de cada indicador

La elección de $\tau = 3$ meses refleja ciclos económicos trimestrales, y $d = 5$ captura aproximadamente 1.5 años de historia ($3 \times 5 = 15$ meses).

Generación de Diagramas Óptimos: Con objetivos de visualización, para cada indicador univariado se determinaron parámetros óptimos de retardo τ y dimensión d con métodos implementados en la librería Giotto-tda, tales como la información mutua y el método de falsos vecinos más cercanos. Esto permitió construir diagramas de persistencia individuales que reflejan la estructura topológica óptima de cada serie.

Homología Persistente

Se aplicó el algoritmo de Vietoris-Rips (giotto-tda) para calcular diagramas de persistencia:

- Dimensiones homológicas: H_0, H_1, H_2
- Parámetro `max_edge_length`: ∞ (filtración completa)
- Coeficiente: \mathbb{Z}_2 (aritmética módulo 2)
- Procesamiento paralelo: `n_jobs=-1` (todos los núcleos)

Paisajes de Persistencia

Los diagramas se vectorizaron mediante paisajes de persistencia usando la librería `Persim`:

- Número de capas: $k = 5$
- Resolución: $n_{bins} = 100$ valores de muestreo
- Dimensiones homológicas: H_0 y H_1 (componentes conexas y ciclos)
- Output multivariado: $(1, 10, 100)$ donde $10 = 2$ dimensiones \times 5 capas
- Output univariado: $(7, 10, 100)$ para comparaciones entre variables

Para cada dimensión homológica y capa, se calcularon las siguientes normas:

Normas L_p Calculadas:

- **Norma L^1 :** $\|\lambda\|_1 = \int |\lambda(t)|dt$ (área total bajo el paisaje, magnitud acumulada)
- **Norma L^2 :** $\|\lambda\|_2 = (\int |\lambda(t)|^2 dt)^{1/2}$ (energía del paisaje)
- **Norma L^∞ :** $\|\lambda\|_\infty = \sup_t |\lambda(t)|$ (altura máxima, característica más prominente)

Estas normas se calcularon de forma agregada por dimensión homológica (promediando sobre las 5 capas), permitiendo cuantificar:

1. La intensidad de características topológicas (H_0 : fragmentación del sistema; H_1 : ciclos económicos)
2. La dominancia relativa entre regímenes económicos y periodicidades.

Algoritmo Mapper

Para complementar el análisis local de homología persistente, se aplicó el algoritmo Mapper para una visualización global de la topología:

- **Función filtro:** Análisis de Componentes Principales (PCA) proyectando a las dos primeras componentes.
- **Cobertura:** Intervalos solapados con $n_{\text{intervals}} = 15$ y $\text{overlap} = 30\%$.
- **Clustering:** DBSCAN con parámetros adaptativos por cubrimiento.
- **Coloración:** Número de fila (tiempo).

El grafo generado con Mapper permitió detectar componentes conexos que se interpretan como cambios de régimen económico. También identificó anomalías, crisis y transiciones, entendidas a partir de nodos con una sola conexión, sin conexiones o que funcionan como nodos puente en la estructura. Esta capacidad de revelar ramificaciones, lazos y estructuras de red facilita la interpretación de la dinámica económica subyacente.

El código completo está disponible en el siguiente repositorio:

- <https://github.com/Daniel-PosadaPUJ/Economic-Cycles-TDA-.git>

para más información de librerías usadas, consultar el archivo `requirements.txt` incluido en el repositorio.

Resultados y Análisis

En esta sección se presentan los resultados obtenidos mediante la aplicación de embeddings de Takens, diagramas de persistencia, paisajes de persistencia y el algoritmo Mapper. El análisis se organiza en tres partes: (i) diagramas de persistencia; (ii) paisajes de persistencia; y (iii) algoritmo mapper.

Diagramas de Persistencia

Los diagramas de persistencia constituyen una herramienta esencial para identificar la estructura topológica presente en las reconstrucciones del sistema. En esta sección se presentan dos grupos de diagramas: (i) el correspondiente al embedding multivariado concatenado, que resume la dinámica conjunta del sistema económico, y (ii) los diagramas individuales para cada una de las variables macroeconómicas, construidos empleando sus parámetros óptimos de reconstrucción.

Diagrama del embedding Multivariado Concatenado

A partir del embedding multivariado previamente descrito, se calculó la homología persistente del complejo de Vietoris–Rips asociado usando la librería *Giotto-TDA*. El diagrama resultante muestra una estructura topológica relativamente limpia: la mayoría de las características en H_0 persisten construyendo poco a poco una única componente conexa, lo cual sugiere una distribución coherente del embedding sin separación excesiva entre regiones. Más notablemente, emergen cuatro ciclos persistentes en H_1 , indicativos de patrones recurrentes y trayectorias cíclicas presentes cuando las variables se analizan de manera conjunta. La ausencia de cavidades complejas y la baja presencia de ruido topológico reflejan que el sistema multivariado mantiene una geometría consistente en alta dimensión.

Diagrama de Persistencia - Embedding Concatenado (7 variables, 35D)

Complejo Vietoris-Rips | 171 características topológicas | Persistencia máx: 0.6671

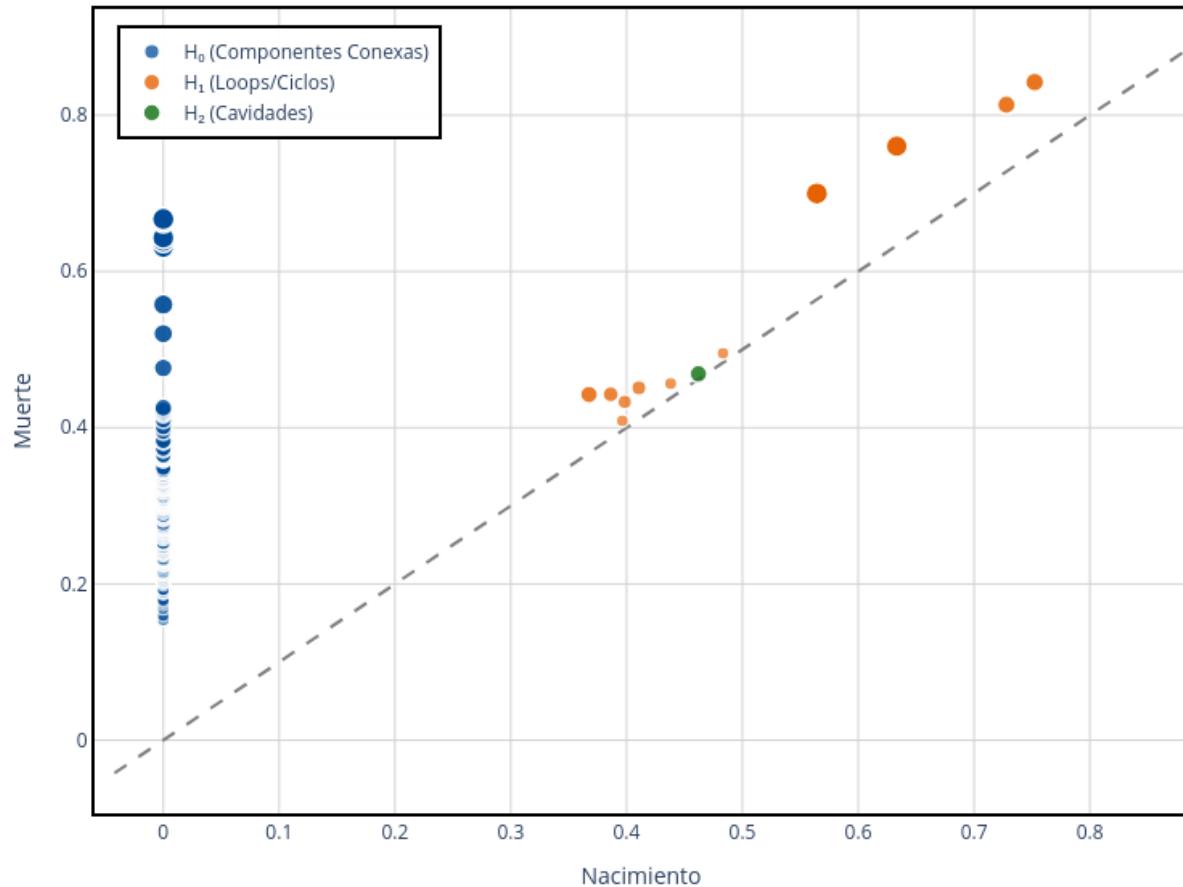


Figura 1: Diagrama de persistencia del Embedding Concatenado.

Diagramas Individuales por Indicador

De forma complementaria, se calcularon los diagramas de persistencia para cada variable macroeconómica utilizando la dimensión y el retardo óptimo identificados en su reconstrucción individual. Estos diagramas se incluyen con fines comparativos y de visualización, permitiendo observar las diferencias dinámicas entre series.

Las variables muestran comportamientos topológicos heterogéneos: algunas presentan ciclos claros y persistentes en H_1 (Petróleo Brent, Colcap), lo que indica patrones cíclicos más marcados; mientras que otras exhiben estructuras dominadas por componentes conexas, reflejando

dinámicas más estables, suavizadas o menos complejas (PIB, ISE). Esta comparación permite identificar qué series aportan mayor complejidad topológica al embedding conjunto y cuáles presentan patrones más simples o lineales.

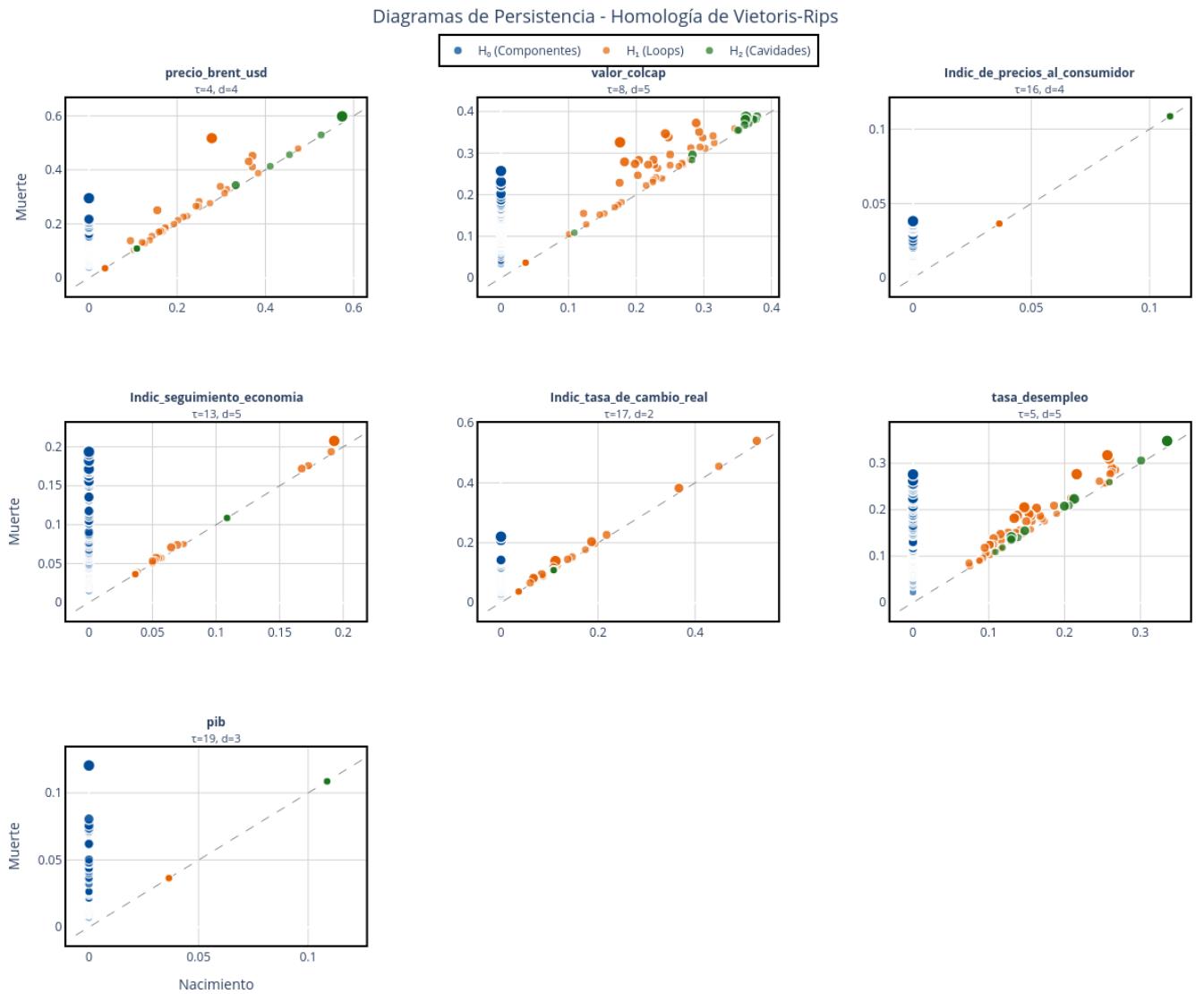


Figura 2: Diagramas de persistencia de cada indicador macroeconómico, junto con sus parámetros de construcción (τ, d).

En conjunto, el análisis del embedding multivariado y los diagramas individuales proporciona una visión más rica de la estructura global del sistema económico, así como las contribuciones particulares de cada variable a la geometría dinámica observada.

Paisajes de Persistencia

La Figura 3 muestra los paisajes de persistencia para H_0 y H_1 obtenidos a partir del embedding multivariado.

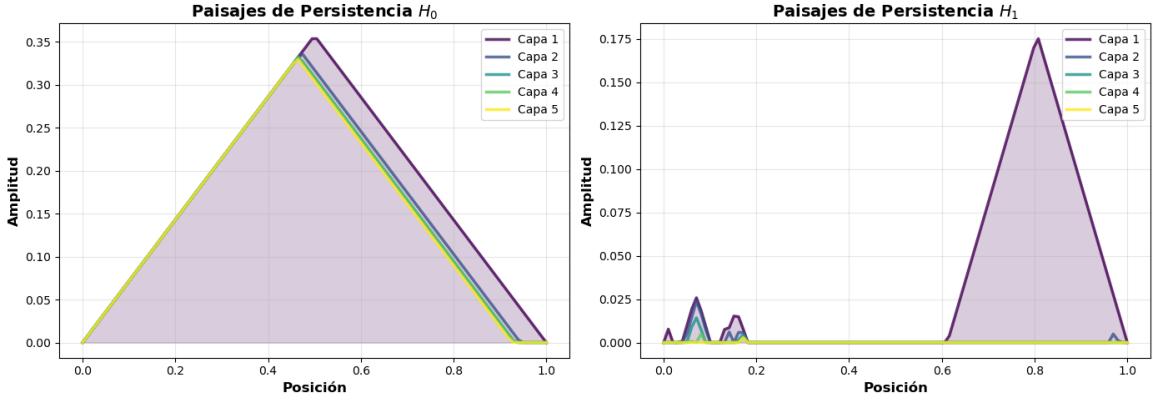


Figura 3: Paisajes de persistencia para H_0 y H_1 (5 capas).

Los paisajes revelan una dominancia de H_0 : las curvas presentan mayor amplitud y un decaimiento progresivo, indicando la presencia de una estructura global conectada. Por el contrario, H_1 muestra amplitudes muy reducidas y un decaimiento abrupto, señalando que los ciclos detectados son débiles y poco persistentes. Las capas superiores no presentan picos relevantes, lo que confirma que la estructura significativa se concentra principalmente en la primera capa.

Normas Agregadas

La Figura 4 presenta las normas agregadas L^1 , L^2 y L^∞ para cada dimensión homológica. Estas permiten cuantificar la magnitud relativa de la persistencia en cada tipo de característica topológica.

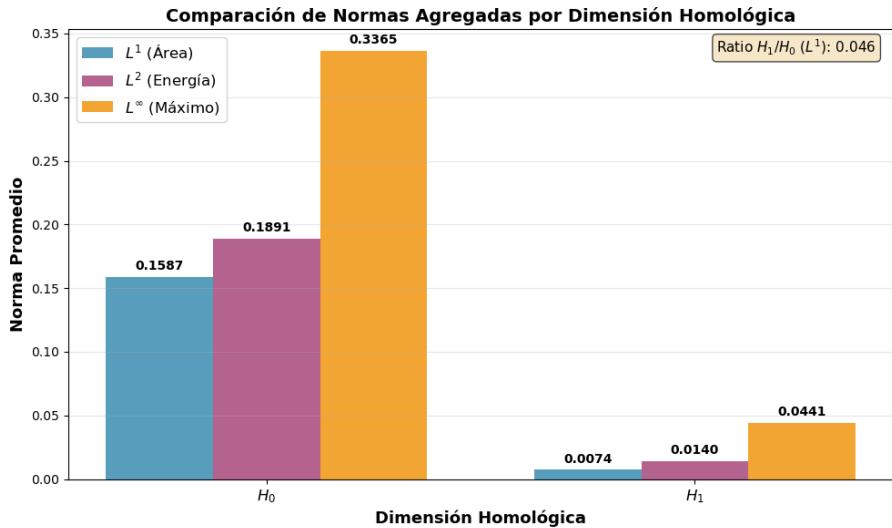


Figura 4: Normas L^1 , L^2 y L^∞ agregadas para H_0 y H_1 .

A partir de los valores agregados se observa lo siguiente:

- **Norma L^1 :** H_0 presenta un área sustancialmente mayor que H_1 , indicando que la mayor parte de la características topológicas del sistema provienen de componentes conectadas persistentes. El valor bajo en H_1 confirma que los ciclos son marginales en la dinámica del embedding.
- **Norma L^2 :** La energía topológica es muy superior en H_0 , lo que refuerza la idea de una conectividad dominante y robusta. En H_1 , la energía es casi nula, lo cual implica que no existen fluctuaciones cíclicas suficientemente fuertes como para influir en la estructura global.
- **Norma L^∞ :** El máximo de persistencia en H_0 supera ampliamente al de H_1 , evidenciando que la conectividad es una característica importante, mientras que el valor reducido de H_1 sugiere que los ciclos, cuando aparecen, tienen muy baja estabilidad y mueren rápidamente.

En conjunto, estas métricas muestran que la topología significativa en los datos está dominada casi exclusivamente por H_0 , mientras que los ciclos (H_1) tienen contribución mínima.

Interpretación Económica

Desde una perspectiva económica, los resultados permiten inferir que:

- Las variables macroeconómicas analizadas presentan una dinámica conjunta coherente y conectada, reflejada en la fuerte persistencia de la conectividad (H_0).

- No se observan ciclos económicos robustos en las escalas capturadas por el análisis; las oscilaciones detectadas en H_1 son débiles y no persistentes.

Algoritmo Mapper

El grafo obtenido mediante el algoritmo *Mapper*, construido sobre el *embedding multivariado*, se presenta en la Figura 5. Este grafo resume la estructura topológica de los datos y permite identificar patrones a partir de la segmentación y agrupamiento inducidos por la función de filtro (PCA) y el algoritmo de clustering empleado (DBSCAN).

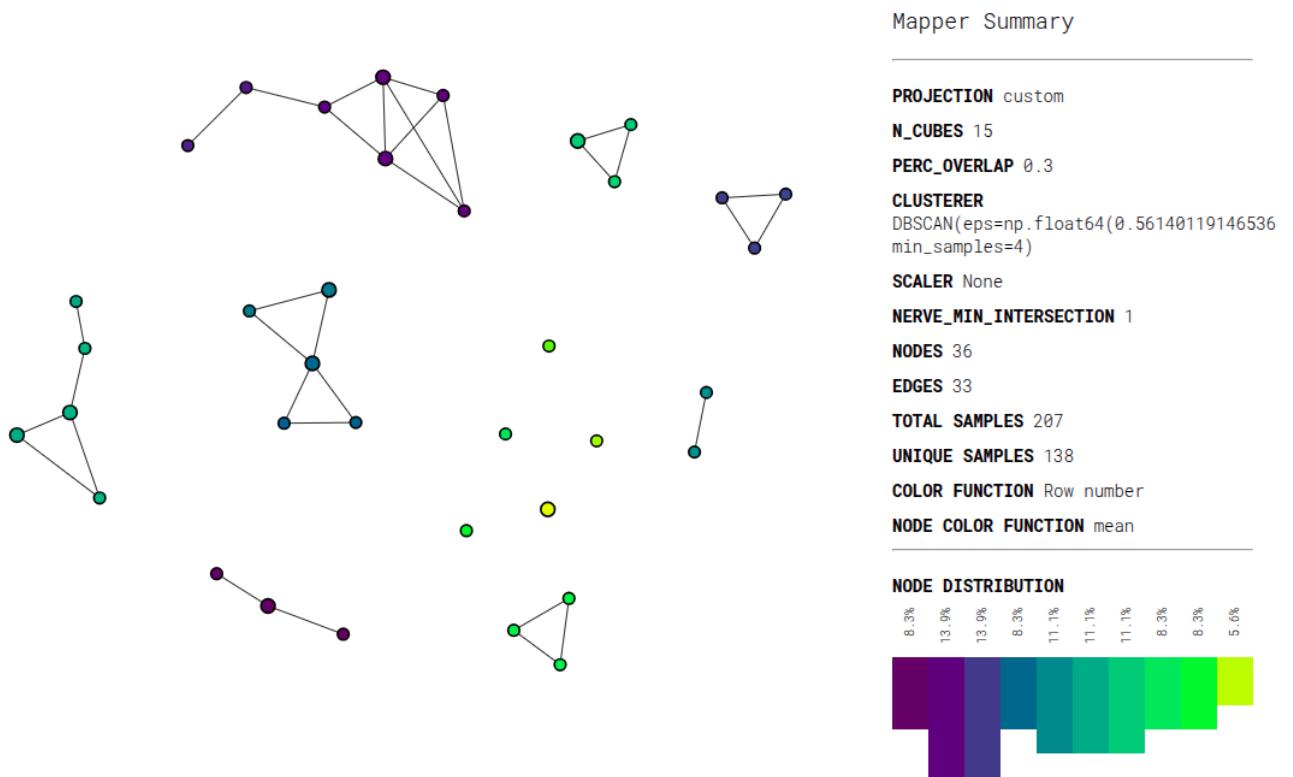


Figura 5: Grafo generado con la librería KMapper, acompañado de un resumen de sus características topológicas.

A partir del análisis detallado del grafo Mapper, se obtiene la siguiente clasificación de nodos según su grado $\delta(v)$, así como la identificación de puntos de articulación y componentes conexas:

- **Nodos aislados ($\delta(v) = 0$):** 5
- **Nodos periféricos ($\delta(v) = 1$):** 6

- **Nodos puente (puntos de articulación):** 6
- **Nodos internos o no críticos:** 19
- **Componentes conexas:** 13

La distribución anterior refleja la heterogeneidad estructural del espacio subyacente. En términos topológicos y, potencialmente, económicos, estos nodos pueden interpretarse como regiones asociadas a comportamientos particulares de los datos, tales como episodios de crisis, aparición de anomalías, transiciones entre regímenes o zonas de estabilidad relativa. Estas propiedades permiten utilizar el grafo Mapper como una herramienta descriptiva y exploratoria para el estudio de dinámicas complejas en sistemas financieros o económicos.

Interpretación Económica

La combinación de técnicas topológicas ofrece una lectura coherente del periodo 2008–2025:

- **El periodo comprendido entre finales de 2019 y 2020 muestra una agrupación notable en la topología,** coherente con la disrupción económica causada por la pandemia y sus efectos inmediatos.
- **A partir de 2022 se observan configuraciones topológicas más dispersas,** lo que podría asociarse a un entorno macroeconómico con mayor heterogeneidad en las dinámicas subyacentes.
- **Las series financieras (como Brent, COLCAP y TCR) concentran las variaciones topológicas más pronunciadas,** mientras que las variables reales (IPC, PIB, desempleo) exhiben estructuras más estables y menos fluctuantes.

Regímenes Económicos Separados (Componentes Conexas del Mapper)

13 regímenes distintos detectados

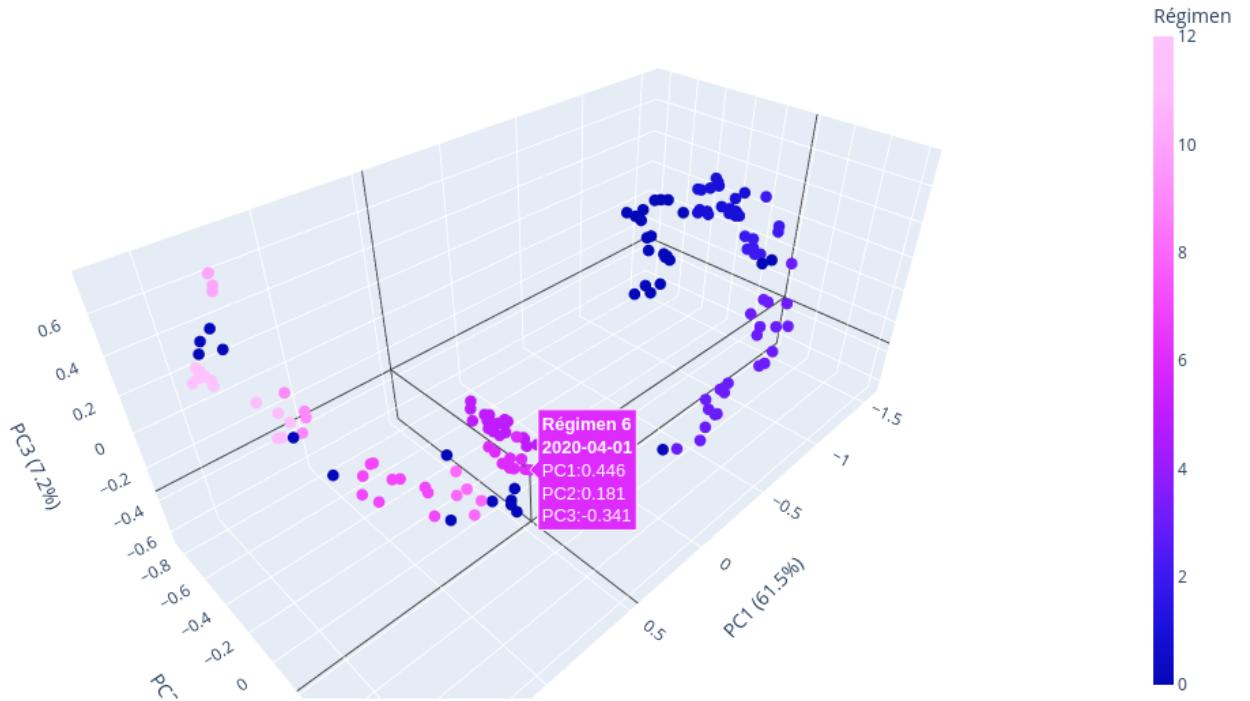


Figura 6: Embedding concatenado proyectado en dos dimensiones mediante PCA y coloreado según las componentes conexas del grafo Mapper.

La visualización sugiere que existen conexiones reconocibles entre distintos momentos de la economía colombiana. En particular, ciertos períodos, como el tránsito en todo el periodo 2008-2009, entre finales de 2019 e inicios de 2020, o los agrupamientos que aparecen desde mediados de 2020 hasta comienzos de 2021, forman conjuntos claramente diferenciados que se reflejan en las componentes conexas del grafo Mapper. Estas agrupaciones indican que el algoritmo es capaz de capturar similitudes en la dinámica conjunta de las variables durante fases específicas del ciclo económico.

Asimismo, dentro de cada componente es posible distinguir patrones más sutiles revisando la estructura interna del componente. La presencia de nodos aislados, nodos con pocas conexiones o puntos de articulación aporta información adicional sobre momentos de transición, episodios atípicos o configuraciones relativamente inestables dentro del sistema. En conjunto, estas características sugieren que el grafo Mapper no solo identifica agrupamientos globales, sino que también permite caracterizar la heterogeneidad interna de cada periodo y los cambios de dinámica a lo largo del tiempo.

Estos patrones son consistentes con trabajos previos como [shultz2023, gallego2025], y evidencian la capacidad del ATD para identificar y caracterizar transiciones económicas críticas.

Conclusiones

Los resultados obtenidos muestran que la economía colombiana presenta una dinámica heterogénea, con transiciones frecuentes entre distintos estados o configuraciones del sistema. El análisis topológico permite visualizar esta fragmentación a través de cambios en la estructura del embedding y en la organización del grafo Mapper, donde emergen componentes conexas diferenciadas, nodos puente y patrones internos que corresponden a fases económicas particulares.

Sin embargo, desde una perspectiva global, los paisajes de persistencia muestran que dicha fragmentación local coexiste con una estructura topológica conexa. Concretamente, la prevalencia de la homología en dimensión H_0 y la persistencia extremadamente baja de H_1 indican que, aunque existen variaciones y transiciones locales detectadas por Mapper, el sistema conserva una forma general conectada y sin ciclos persistentes. Esto sugiere que las variables macroeconómicas mantienen una tendencia conjunta y que no se observan dinámicas cíclicas profundas en las escalas capturadas por la homología persistente.

El algoritmo Mapper demostró ser especialmente útil para este estudio: ofrece un flujo de análisis claro, una representación visual intuitiva y una capacidad efectiva para destacar transiciones, agrupamientos y estructuras internas difíciles de detectar mediante técnicas tradicionales. Su complementariedad con la homología persistente resulta evidente, pues mientras esta identifica características topológicas globales —capturadas a través de los paisajes—, Mapper permite localizar y describir dónde y cómo se manifiestan dichas características a lo largo del tiempo y cómo se fragmenta localmente el embedding.

La correspondencia cualitativa entre las estructuras detectadas y eventos históricos relevantes (como episodios de estrés financiero o las disruptiones asociadas a la pandemia de 2020) refuerza el potencial de este enfoque para el estudio de sistemas económicos complejos. En conjunto, los resultados sugieren que Mapper constituye una herramienta valiosa para explorar, caracterizar y comunicar la evolución dinámica de sistemas multivariados en contextos económicos, mientras que los paisajes de persistencia aportan evidencia sobre la forma global y la coherencia estructural del sistema analizado.

1. Referencias

- [1] Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, 46(2), 255–308.
- [2] Edelsbrunner, H., Letscher, D., & Zomorodian, A. (2002). Topological persistence and simplification. *Discrete & Computational Geometry*, 28(4), 511–533.
- [3] Zomorodian, A., & Carlsson, G. (2005). Computing persistent homology. *Discrete & Computational Geometry*, 33(2), 249–274.

- [4] Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. (2007). Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1), 103–120.
- [5] Ghrist, R. (2008). Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1), 61–75.
- [6] Singh, G., Mémoli, F., & Carlsson, G. E. (2007). Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *Eurographics Symposium on Point-Based Graphics* (pp. 91–100).
- [7] Van Veen, H., Saul, N., Eargle, D., & Mangham, S. (2019). Kepler Mapper: A flexible Python implementation of the Mapper algorithm. *Journal of Open Source Software*, 4(42), 1315.
- [8] Tauzin, G., et al. (2021). giotto-tda: A topological data analysis toolkit for machine learning and data exploration. *Journal of Machine Learning Research*, 22(39), 1–6.
- [9] Bubenik, P. (2015). Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16(1), 77–102.
- [10] Bubenik, P., & Dłotko, P. (2017). A persistence landscapes toolbox for topological statistics. *Journal of Symbolic Computation*, 78, 91–114.
- [11] Meinecke, L. S. (2025). *Sliding Windows and Persistence: Topological Analysis of Time Series Data*. Master's thesis, ETH Zürich, Department of Mathematics, Zürich, Switzerland. Advisor: Dr. S. Kališnik Hintz.
- [12] Shultz, C. (2023). *Applications of Topological Data Analysis in Economics*. Technical Report, SSRN. Disponible en: <https://ssrn.com/abstract=4378151>
- [13] Gallego, E., Gómez-Ramírez, D. A. J., & Porras, J. (2025). Análisis topológico de datos y su aplicación en mercados financieros. *Revista Integración*, 43(2), 16–26. doi: 10.18273/revint.v43n2-2025002.
- [14] Gidea, M., & Katz, Y. (2018). Topological data analysis of financial time series: Landscapes of crashes. *Physica A: Statistical Mechanics and its Applications*, 491, 820–834.
- [15] Majumdar, S., & Laha, A. K. (2020). Clustering and classification of time series using topological data analysis with applications to finance. *Expert Systems with Applications*, 162, 113868.
- [16] Singh, G., Mémoli, F., & Carlsson, G. E. (2007). Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *Eurographics Symposium on Point-Based Graphics* (pp. 91–100).

- [17] Gidea, M., & Katz, Y. (2018). Topological data analysis of financial time series: Landscapes of crashes. *Physica A: Statistical Mechanics and its Applications*, 491, 820–834.