



Análisis topológico de datos y su aplicación en mercados financieros

EDISSON GALLEG^a, DANNY A. J. GÓMEZ-RAMÍREZ^b, JAIBERTH PORRAS^a

^a Institución Universitaria Tecnológico de Antioquia, Medellín, Colombia.

^b Visión Real Cognitiva (Cognivisión) S.A.S., Itagüí, Colombia.

Resumen. Presentamos una introducción a los fundamentos del análisis topológico de datos, con especial énfasis en la homología persistente. Para facilitar la comprensión de los conceptos principales, algunos se abordan desde una perspectiva gráfica. Asimismo, examinamos aplicaciones recientes de esta disciplina en el ámbito económico, particularmente en la detección temprana de crisis en los mercados financieros. En este contexto, empleamos una metodología previamente desarrollada y la aplicamos al mercado financiero colombiano, representado por la Bolsa de Valores de Colombia.

Palabras clave: Homología persistente, paisaje de persistencia, normas p , mercados financieros, detección de crisis.

MSC2010: 55N31, 14F35, 14F45.

Topological data analysis and its application in financial markets

Abstract. We present an introduction to the fundamentals of topological data analysis, with a particular emphasis on persistent homology. To facilitate the understanding of the main concepts, some are approached from a graphical perspective. Likewise, we examine recent applications of this discipline in the economic field, particularly in the early detection of crises in financial markets. In this context, we employ a previously developed methodology and apply it to the Colombian financial market, represented by the Colombian Stock Exchange.

Keywords: Persistent homology, persistence landscape, p norms, financial markets, crisis detection.

E-mail: egalleg@gmail.com^a, daj.gomezramirez@gmail.com^b, lunovane10@gmail.com^c .

Recibido: 27 Junio 2025. Aceptado: 26 Septiembre 2025.

Para citar este artículo: E. Gallego, D. Gómez-Ramírez y J. Porras, Análisis topológico de datos y su aplicación en mercados financieros, *Rev. Integr. Temas Mat.*, 43 (2025), No. 2, 16–26. doi: 10.18273/revint.v43n2-2025002

1. Introducción

El análisis topológico de datos (TDA, por sus siglas en inglés) es un enfoque que utiliza herramientas de la topología algebraica para estudiar la forma o estructura global de los datos. Esta disciplina ha generado gran interés en los últimos años, gracias a su riqueza matemática, así como a sus múltiples aplicaciones. El TDA integra conceptos de topología, álgebra, geometría, análisis y estadística. Las aplicaciones del TDA incluyen la biología, las redes sociales y neuronales, y el análisis de la volatilidad en los mercados financieros.

Una de las áreas más prominentes del TDA es la homología persistente. Desde su introducción en [5], la homología persistente ha suscitado gran interés y se han escrito numerosos artículos en el área, tanto de carácter teórico como aplicado. La homología persistente permite identificar características estructurales de los datos que persisten a través de múltiples escalas de un filtrado. La homología persistente se ha utilizado para clasificar tipos de tumores mamarios [2], para el análisis de redes biológicas y sociales [1], y en la detección temprana de crisis en los mercados financieros [7].

Existen varias formas de codificar la homología persistente, entre las que se destacan el diagrama de persistencia, el código de barras y el paisaje de persistencia. En cuanto a la parte computacional, uno de los algoritmos más eficientes en la actualidad para el cálculo de las diferentes codificaciones de la homología persistente es el algoritmo de Ripser [3].

Este trabajo tiene como primer objetivo ofrecer una introducción clara y comprensiva a los conceptos esenciales del TDA, con especial atención en la homología persistente. Se parte de la definición de complejos simpliciales, se describe el procedimiento de construcción del complejo de Vietoris-Rips y se introducen las nociones de filtraciones, los grupos de homología persistente y sus diferentes formas de codificación.

Como segundo objetivo, se examinan aplicaciones recientes de la homología persistente en la detección temprana de crisis en los mercados bursátiles. En particular, se consideran los aportes de Gidea y Katz [6], quienes analizaron índices del mercado estadounidense, y los de Leaverton [7], centrados en mercados bursátiles europeos.

2. Fundamentos del análisis topológico de datos

En esta sección se ofrece una introducción a los conceptos fundamentales del TDA, con énfasis en la homología persistente. Se abordan los siguientes temas:

- Complejos simpliciales – complejo de Vietoris-Rips;
- Filtraciones de complejos simpliciales;
- Homologías y persistencia;
- Código de barras, diagrama de persistencia y paisaje de persistencia.

Definición 2.1. Considere un conjunto finito de puntos $X = \{x_0, x_1, \dots, x_m\}$ en el espacio n -dimensional \mathbb{R}^n . Un complejo simplicial $\mathbb{K} \subset \mathcal{P}(X)$ sobre X es un subconjunto no vacío de partes de X que satisface:

- $x_i \in \mathbb{K}$ para $i = 0, 1, 2, \dots, m$;
- Si $\sigma \in \mathbb{K}$ y $l \subset \sigma$, entonces $l \in \mathbb{K}$.

Los elementos de \mathbb{K} se llaman *simplejos* o *caras* de \mathbb{K} . Si $\sigma = \{x_{i_0}, x_{i_1}, \dots, x_{i_t}\}$ es un elemento de \mathbb{K} , decimos que σ es una cara t -dimensional de \mathbb{K} o que σ es un t -simplejo de \mathbb{K} . Los 0-simplejos se llaman *vértices* de \mathbb{K} y los 1-simplejos se denominan *aristas* de \mathbb{K} . Decimos que \mathbb{K} es un complejo simplicial n -dimensional si contiene al menos un n -simplejo.

Ejemplo 2.2. Sea $X = \{v_1 = (1, 3), v_2 = (2, 4), v_3 = (7, 6), v_4 = (2, -4), v_5 = (4, 7)\}$. En la siguiente definición de un complejo simplicial \mathbb{K} a partir de X , la primera fila contiene los vértices; la segunda, las aristas; la tercera, los 2-simplejos, y la cuarta fila, el 3-simplejo.

$$\mathbb{K} = \left\{ \begin{array}{l} \{\{v_1\}, \{v_2\}, \{v_3\}, \{v_4\}, \{v_5\}, \\ l_1 = \{v_1, v_2\}, l_2 = \{v_1, v_3\}, l_3 = \{v_1, v_4\}, l_4 = \{v_2, v_4\}, l_5 = \{v_2, v_5\}, \\ \sigma_1 = \{v_1, v_2, v_3\}, \sigma_2 = \{v_1, v_2, v_4\}, \sigma_3 = \{v_1, v_4, v_3\}, \sigma_4 = \{v_2, v_4, v_3\}, \\ \eta = \{v_1, v_2, v_3, v_4\} \end{array} \right\}.$$

Como el complejo simplicial \mathbb{K} tiene dimensión 3, entonces podemos asociarle un espacio topológico en \mathbb{R}^3 , que llamaremos la *realización geométrica* de \mathbb{K} , y que se observa en la Figura 1.

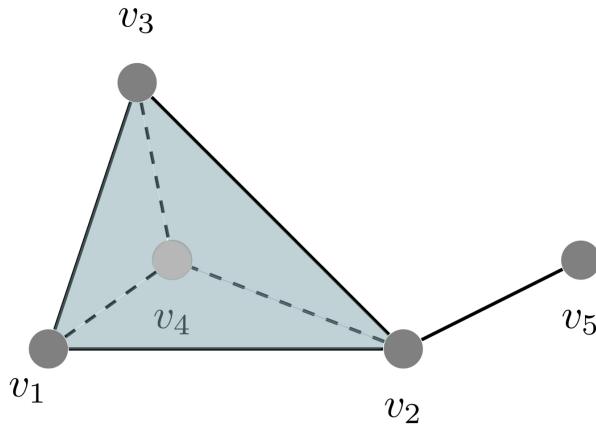


Figura 1. Realización geométrica del complejo simplicial \mathbb{K} .

Aquí, cada vértice $\{v_i\}$ se identifica con un punto (en este ejemplo, un punto de \mathbb{R}^3), que usualmente se denota con el mismo nombre v_i del vértice. Cada arista $l_i = \{v_r, v_s\}$ se identifica con el segmento de recta que va de v_r a v_s . Por ejemplo, la arista l_1 se identifica con el segmento de recta que va de v_1 a v_2 . Cada 2-simplejo σ se identifica con el triángulo relleno cuyos vértices son aquellos que aparecen en σ . En general, todo n -simplejo $\sigma = \{v_0, v_1, \dots, v_n\}$ se identifica con el conjunto de puntos $\bar{X} = \sum_{i=0}^n \lambda_i v_i$, donde cada $\lambda_i \geq 0$ y $\sum_{i=0}^n \lambda_i = 1$.

Definición 2.3. Dados un conjunto finito de puntos $X = \{x_0, x_1, \dots, x_m\}$ en un espacio métrico (M, d) y un número real positivo r , definimos el complejo de Vietoris-Rips de X con parámetro r , como el complejo simplicial denotado $\mathbb{K}_r(X)$ que satisface:

- Los vértices de $\mathbb{K}_r(X)$ son los puntos de X .
- $\sigma = \{y_0, y_1, \dots, y_t\} \in \mathbb{K}_r(X)$ si y solo si $d(y_i, y_j) \leq r$ para todo $i, j \in \{0, 1, \dots, t\}$.

En los ejemplos que se presentarán posteriormente, X será un conjunto finito de puntos de \mathbb{R}^m . Sin embargo, puede ocurrir que $\dim(\mathbb{K}_r(X)) > m$ para algunos valores de r . Por lo tanto, la realización geométrica de $\mathbb{K}_r(X)$ puede no ser un subespacio topológico de \mathbb{R}^m (en el Ejemplo 2.2 la nube de puntos X es un subconjunto de \mathbb{R}^2 pero $\mathbb{K} \not\subset \mathbb{R}^2$).

Notemos que si $r < r'$, entonces $\mathbb{K}_r(X) \subset \mathbb{K}_{r'}(X)$. Esta contención induce un homomorfismo entre las p -ésimas homologías simpliciales $H_p(\mathbb{K}_r(X)) \rightarrow H_p(\mathbb{K}_{r'}(X))$, para todo p .

Definición 2.4. Una colección de subcomplejos simpliciales $\{\mathbb{K}_r\}_{r \in J}$ de \mathbb{K} tales que $\mathbb{K}_r \subset \mathbb{K}_{r'}$ si $r < r'$, se denomina una *filtración* de \mathbb{K} si $\bigcup \mathbb{K}_r = \mathbb{K}$.

Ejemplo 2.5. A continuación presentamos una filtración del complejo simplicial \mathbb{K} dado en el Ejemplo 2.2. La filtración se observa en la Figura 2. Por facilidad, solo presentamos la representación gráfica de cada subcomplejo. \mathbb{K}_0 está conformado por todos los 0-simplejos (vértices) de \mathbb{K} ; \mathbb{K}_1 consta de todos los vértices y aristas de \mathbb{K} ; \mathbb{K}_2 está compuesto por todos los simplejos de dimensión menor o igual que 2 de \mathbb{K} . Finalmente, $\mathbb{K}_3 = \mathbb{K}$ contiene además el interior del tetraedro con vértices v_1, v_2, v_3 y v_4 . De esta forma, $\mathbb{K}_0 \subset \mathbb{K}_1 \subset \mathbb{K}_2 \subset \mathbb{K}_3$.

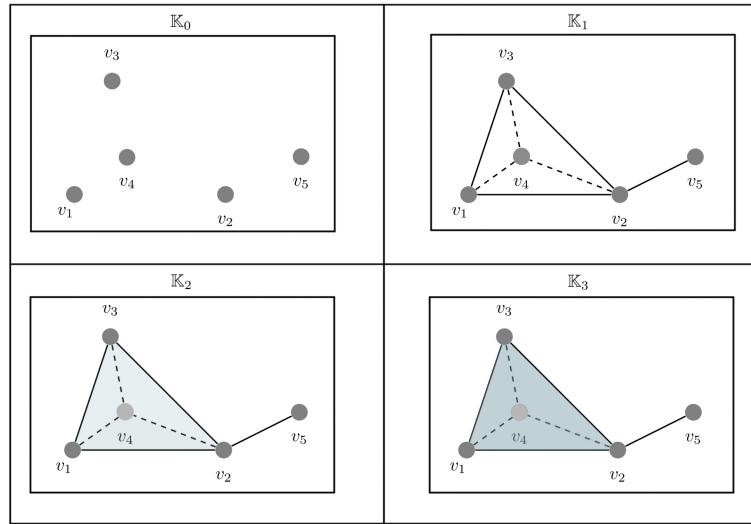


Figura 2. Filtración del complejo simplicial \mathbb{K} .

Dada una filtración $\{\mathbb{K}_i\}$ de complejos simpliciales

$$\mathbb{K}_0 \stackrel{f^0}{\subset} \mathbb{K}_1 \stackrel{f^1}{\subset} \cdots \stackrel{f^{n-2}}{\subset} \mathbb{K}_{n-1} \stackrel{f^{n-1}}{\subset} \mathbb{K}_n,$$

donde f^i es la inclusión, y dado un entero $p \geq 0$, podemos inducir una sucesión de mapeos lineales en las homologías de la siguiente forma:

$$H_p(\mathbb{K}_0) \xrightarrow{f_p^0} H_p(\mathbb{K}_1) \xrightarrow{f_p^1} \cdots \xrightarrow{f_p^{n-2}} H_p(\mathbb{K}_{n-1}) \xrightarrow{f_p^{n-1}} H_p(\mathbb{K}_n).$$

Para $i < j$, denotemos $f_p^{ij} = f_p^j \circ \cdots \circ f_p^i$. Definimos los *grupos de homología de persistencia* p -dimensionales de la filtración $\{\mathbb{K}_i\}$ como

$$H_p^{i,j} = \frac{Z_p(\mathbb{K}_i)}{B_p(\mathbb{K}_j) \cap Z_p(\mathbb{K}_i)},$$

donde $Z_p(\mathbb{K}_i)$ representa los p -ciclos de \mathbb{K}_i y $B_p(\mathbb{K}_j)$ las p -fronteras de \mathbb{K}_j .

Sea $\sigma \in H_p(\mathbb{K}_i)$ una clase de equivalencia de un p -simplejo. Diremos que σ *nace* en \mathbb{K}_i si $\sigma \notin H_p^{i-1,i}$. Ahora, supongamos que el simplejo σ nace en \mathbb{K}_i . Diremos que σ *muere* en \mathbb{K}_j si este simplejo se identifica con otra clase más antigua mediante el correspondiente homomorfismo de \mathbb{K}_{j-1} a \mathbb{K}_j , es decir, si $f_p^{i,j-1}(\sigma) \notin H_p^{i-1,j-1}$ pero $f_p^{i,j}(\sigma) \in H_p^{i-1,j}$.

Sea $X = \{x_i\}$ un conjunto finito de puntos en \mathbb{R}^n y $\{\mathbb{K}_{r_i}(X) = \mathbb{K}_{r_i}\}$ una filtración de complejos de Vietoris-Rips de la cubierta convexa de X , ordenados de tal forma que $\mathbb{K}_{r_i} \subset \mathbb{K}_{r_j}$ si $r_i < r_j$. Consideremos $\sigma \in H_p(\mathbb{K}_{r_i})$ un p -simplejo que nace \mathbb{K}_{r_i} y muere en \mathbb{K}_{r_j} . Asociamos a σ la pareja ordenada (r_i, r_j) que indica su nacimiento y muerte. A partir de estas parejas definimos tres formas de representar o codificar la homología de persistencia: el código de barras, el diagrama de persistencia y el paisaje de persistencia.

Código de Barras. El código de barras de la filtración $\{\mathbb{K}_{r_i}\}$ en dimensión p es el conjunto de líneas horizontales que parten en r_i y finalizan en r_j , una línea por cada p -simplejo que nazca en algún r_i y muera en algún r_j .

Diagrama de persistencia. El diagrama de persistencia es otra forma de representar la información recopilada de los nacimientos y muertes de los p -simplejos que van apareciendo y desapareciendo a medida que se avanza en la filtración $\{\mathbb{K}_{r_i}\}$. El diagrama de persistencia en dimensión p asociado a $\{\mathbb{K}_{r_i}\}$ es el conjunto de parejas ordenadas

$$D_p = \{(r_i, r_j) \mid \exists \sigma \text{ tal que } \sigma \text{ nace en } r_i \text{ y muere en } r_j\}.$$

Paisaje de persistencia. El paisaje de persistencia fue introducido en el 2015 por Bubenik en [4], como una forma funcional de representar los diagramas de persistencia. Esta transformación convierte cada pareja $(a, b) \in D_p$ en una función $f_{(a,b)}(x)$ cuya gráfica tiene la forma de triángulo isósceles y que se define como

$$f_{(a,b)}(x) = \begin{cases} x - a, & \text{si } a < x \leq \frac{a+b}{2}; \\ b - x, & \text{si } \frac{a+b}{2} < x < b; \\ 0, & \text{en otro caso.} \end{cases}$$

Para cada $k \in \mathbb{N}$ definimos la función $f_k(x)$ como el k -ésimo mayor valor del conjunto $\{f_{(a,b)}(x)\}$, variando (a, b) sobre todos los elementos de D_p . La función $f : \mathbb{N} \times \mathbb{R} \rightarrow \mathbb{R}$ dada por $f(k, x) = f_k(x)$ se denomina paisaje de persistencia en dimensión p asociada a $\{\mathbb{K}_{r_i}\}$.

De este modo, el paisaje de persistencia es una secuencia de funciones que capturan la estructura topológica del conjunto de datos y que forman parte de un espacio de Banach, lo cual permite aplicar técnicas estadísticas como medias, normas, distancias y pruebas de hipótesis. Por lo tanto, los paisajes de persistencia combinan topología algebraica con análisis estadístico riguroso.

Ejemplo 2.6. Consideremos en \mathbb{R}^3 el conjunto de puntos

$$X = \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (0, 1, 1)\}.$$

Para este conjunto de puntos X tenemos el código de barras, el diagrama de persistencia y el paisaje de persistencia que se muestran en la Figura 3. Como es habitual, se presentan los códigos de barras en un mismo gráfico, con un color diferente para cada dimensión (en este ejemplo, asignamos barras negras para dimensión cero y rojas para dimensión 1). De manera similar, el diagrama y paisaje de persistencia se presentan con los mismos colores.

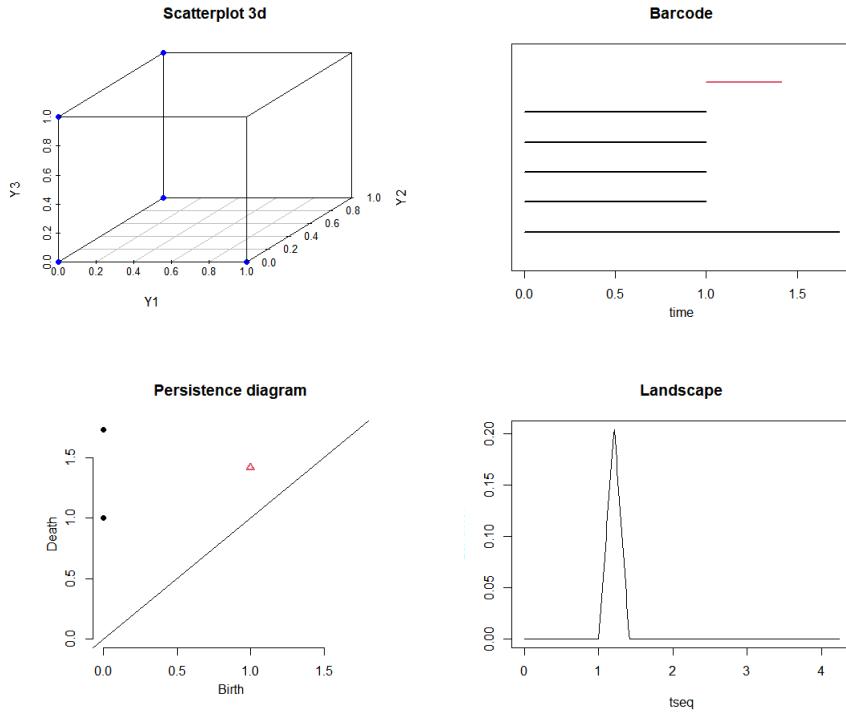


Figura 3. Representaciones de la homología persistente.

Observamos que en el código de barras solo aparece una barra en dimensión 1 (línea roja), que nace en $a = 1$ y muere en $b = \sqrt{2}$, por lo que $D_1 = \{(1, \sqrt{2})\}$. Esta única barra

da lugar al paisaje de persistencia $f(k, x)$ definido como $f(k, x) = 0$ si $k \neq 1$ y

$$f(1, x) = \begin{cases} x - 1, & \text{si } 1 < x \leq \frac{1+\sqrt{2}}{2}; \\ \sqrt{2} - x, & \text{si } \frac{1+\sqrt{2}}{2} < x < \sqrt{2}; \\ 0, & \text{si } x \notin (1, \sqrt{2}). \end{cases}$$

3. Metodología

Gidea y Katz [6] estudiaron los cuatro principales índices bursátiles estadounidenses en los períodos previos al colapso del sector tecnológico, en el año 2000, y a la crisis financiera asociada con los créditos hipotecarios en 2008. En su análisis identificaron que las normas p de los paisajes de persistencia asociados a series de tiempo de estos índices presentan un fuerte crecimiento en momentos previos y durante una crisis financiera. Estos hallazgos fueron confirmados por Leaverton [7] para índices bursátiles europeos. Las series de tiempo sobre las cuales trabajaron Gidea y Katz se formaron a partir de *ventanas deslizantes* que permiten embeber los datos en un espacio de dimensión superior. Antes de examinar el método aplicado por Gidea y Katz, recordaremos la definición de la norma p y revisaremos el concepto de ventanas deslizantes.

Sea $f: X \rightarrow \mathbb{R}$ una función medible definida sobre un espacio de medida (X, \mathcal{M}, μ) , y sea $p \in [1, \infty)$. Entonces, la *norma p* de f en el espacio $L^p(X)$ se define como:

$$\|f\|_p = \left(\int_X |f|^p d\mu \right)^{\frac{1}{p}}.$$

Ejemplo 3.1. Sea $f: \mathbb{N} \times [0, 1] \rightarrow \mathbb{R}$ definida por $f(k, x) = \frac{x+1}{k^2}$. Consideremos la medida μ en $\mathbb{N} \times [0, 1]$ como el producto entre la medida de conteo en \mathbb{N} con la medida de Lebesgue en $[0, 1]$. Para $p \geq 1$, por el teorema de Fubini tenemos que:

$$\|f\|_p^p = \int_{\mathbb{N} \times [0, 1]} |f|^p d\mu = \sum_{k=1}^{\infty} \int_0^1 \left(\frac{x+1}{k^2} \right)^p dx = \sum_{k=1}^{\infty} \frac{2^{p+1} - 1}{(p+1)k^{2p}} < \infty.$$

Ventanas deslizantes. El método de las ventanas deslizantes, también conocido como reconstrucción por retardos temporales, tiene su origen en el estudio de sistemas dinámicos no lineales, y fue formalizado en la década de 1980 gracias al trabajo del matemático holandés Florence Takens. En su teorema, publicado en 1981 [8], Takens demostró que es posible reconstruir el espacio de fases de un sistema dinámico complejo utilizando únicamente una serie temporal y sus versiones retardadas en el tiempo. Hoy en día, esta herramienta se aplica en campos como el análisis de datos, la predicción de fenómenos, el procesamiento de señales y el aprendizaje automático.

La idea de la ventana deslizante consiste en desplazar una ventana de tamaño fijo sobre los datos con el fin de embeber los datos en un espacio de dimensión superior. Poste-

riormente, se analizan los datos embebidos, por ejemplo, a partir de la construcción del complejo simplicial de Vietoris-Rips.

Ejemplo 3.2. Supongamos que tenemos como nube de datos la colección $X = \{5, 8, 12, 14, 16, 18, 20, 25, 30\}$. Si queremos formar ventanas deslizantes de tamaño 3, con un paso de desplazamiento de una unidad, entonces la primera ventana es $w_1 = (5, 8, 12)$; la segunda ventana es $w_2 = (8, 12, 14)$; la tercera ventana es $w_3 = (12, 14, 16)$, y así sucesivamente hasta llegar a la ventana $w_7 = (20, 25, 30)$. Es sencillo comprobar que, dado un conjunto de N datos, al generar ventanas deslizantes de tamaño $w < N$, se obtienen $N - w + 1$ ventanas posibles.

El método aplicado por Guidea y Katz [6] es el siguiente. A partir de un conjunto de datos relacionados con el cierre diario de un conjunto de índices del mercado accionario, se toman ventanas de tamaño fijo que se desplazan una unidad de tiempo (un día). Para la nube de puntos que conforma cada ventana se construye un filtrado de complejos simpliciales de Vietoris-Rips. Luego se aplican algoritmos establecidos en la literatura (como por ejemplo el algoritmo de Ripser [3]) para determinar los grupos de homología, los grupos de homología persistente y el paisaje de persistencia asociados al filtrado. Se calculan las normas p de los diagramas de persistencia de la nube de puntos de cada ventana y se analizan sus cambios en el tiempo. Finalmente, se monitorea el comportamiento de las normas p en momentos previos de estrés o inestabilidad de los índices accionarios.

A continuación, describimos el método de forma más precisa. Considere d series temporales $\{x_n^k\}$, con $k \in \{1, \dots, d\}$, donde el vector

$$x(n) = (x_n^1, x_n^2, \dots, x_n^d)$$

tiene por coordenadas los valores de las series en el día n . Considere un tamaño de ventana w . El resumen del método es el siguiente:

- Para cada día n , se construye el conjunto (la ventana) X_n de puntos en \mathbb{R}^d dado por:

$$\begin{aligned} X_n &= \{x(n), x(n+1), \dots, x(n+w-1)\} \\ &= \{(x_n^1, x_n^2, \dots, x_n^d), \dots, (x_{n+w-1}^1, x_{n+w-1}^2, \dots, x_{n+w-1}^d)\}. \end{aligned}$$

- Sobre cada conjunto de puntos X_n , se calcula la homología persistente de la filtración de Vietoris-Rips, y se obtienen diagramas de persistencia que describen las estructuras topológicas, como ciclos.
- Las características topológicas se resumen en *paisajes de persistencia*, que son funciones reales en un espacio de Banach.
- Se computan las *normas* en L^p (por ejemplo, $p \in \{1, 2\}$) de estos paisajes de persistencia para cada ventana, generando una serie temporal de estas normas.

- La evolución de estas normas refleja cambios en la topología de los datos; permite detectar patrones que pueden indicar próximas crisis o cambios importantes en el mercado financiero.

Guidea y Katz aplicaron el método a series de tiempo correspondientes a cuatro índices bursátiles: S&P 500, Dow Jones Industrial Average (DJIA), NASDAQ Composite y Russell 2000, lo cual equivale a tomar $d = 4$. Para el análisis, se utilizaron ventanas deslizantes de tamaño $w = 50$ y $w = 100$. En lugar de emplear los precios de cierre diarios de los índices, se trabajó con las series de tiempo de los retornos logarítmicos $r_{i,j}$, definidos como:

$$r_{i,j} = \ln\left(\frac{P_{i,j}}{P_{i-1,j}}\right),$$

donde $P_{i,j}$ es el precio ajustado del índice j en el día i . Esto significa que el punto $x(k)$ de la ventana X_n es $x(k) = (r_{k,1}, r_{k,2}, \dots, r_{k,d})$, para $k \in \{n, n+1, \dots, n+w-1\}$.

En este trabajo aplicamos el método propuesto por Guidea y Katz al mercado accionario colombiano. Con el propósito de obtener una representación integral de dicho mercado, se consideró el índice COLCAP junto con las tres acciones más representativas y líquidas de la Bolsa de Valores de Colombia (BVC): la acción ordinaria de Bancolombia, Ecopetrol e ISA. En consecuencia, el número de series temporales analizadas es $d = 4$. Se emplearon ventanas de tamaño $w = 30$, diferenciándose así del tamaño utilizado por Guidea y Katz, con el objetivo de examinar el efecto que produce la variación en este parámetro.

4. Recolección de datos y descripción de softwares utilizados

Los datos empleados en este estudio fueron obtenidos del portal Investing.com y comprenden todos los días hábiles de negociación del mercado entre el 15 de enero de 2008 y el 23 de julio de 2025. El procesamiento y análisis de la información se llevaron a cabo en el entorno Google Colab, utilizando el lenguaje de programación Python y las librerías Pandas, NumPy, Gudhi y matplotlib.pyplot.

En particular, Pandas se empleó para la lectura, organización y manipulación de los datos; NumPy, para la ejecución de operaciones de cálculo numérico; Gudhi, como herramienta principal para el cálculo de diagramas de persistencia y la obtención de sus normas p ; y matplotlib.pyplot, para la generación de representaciones gráficas y visualizaciones.

5. Análisis de resultados

La Figura 4 presenta los paisajes de persistencia correspondientes a la homología en dimensiones 1 (ciclos o túneles) y 2 (cavidades o vacíos), construidos a partir de series

temporales con dimensión de embebimiento $d = 4$, generadas con los retornos logarítmicos del índice COLCAP y las tres acciones más representativas y líquidas de la BVC, empleando ventanas de tamaño $w = 30$ mediante un esquema de retardos temporales. En la Figura 5 se ilustra la norma L^2 (esto es, la norma p con $p = 2$) de dichos paisajes en ambas dimensiones homológicas. En esta figura se observa un marcado crecimiento de la norma p en momentos previos y durante cuatro períodos específicos, que coinciden con momentos de turbulencia o inestabilidad en la BVC.

El primer episodio corresponde a la crisis financiera global de 2008, originada por la quiebra del banco de inversión estadounidense Lehman Brothers, el 15 de septiembre de ese año. El segundo, de carácter local, estuvo asociado con la fuerte caída de los precios del petróleo, iniciada en junio de 2014 por un exceso de oferta. Este fenómeno se extendió por cerca de dos años y, dada la alta dependencia de la economía colombiana de la exportación de materias primas, así como la fuerte ponderación de Ecopetrol en el COLCAP, arrastró hacia abajo significativamente al índice.

El tercer episodio corresponde a la crisis bursátil global desencadenada por la pandemia del COVID-19 a partir de marzo de 2020. En los períodos previos y durante esta crisis se observa el incremento más pronunciado de la norma p , el cual coincide con la caída más severa de los mercados en comparación con las demás crisis analizadas. Finalmente, el cuarto evento obedece a un choque local en 2022, asociado con la incertidumbre política y las propuestas de reforma y regulación petrolera promovidas por el gobierno de Colombia, liderado por el presidente Gustavo Petro.

En las últimas ventanas analizadas, correspondientes al año 2025, se observa un incremento en la norma p . Si se asume el carácter predictivo del método, esta tendencia, de consolidarse con mayor claridad, podría estar sugiriendo señales tempranas de una eventual caída en las valoraciones del índice COLCAP y, en general, de las principales acciones de la BVC.

Estos resultados refuerzan la utilidad del TDA como herramienta para la detección de crisis en los mercados bursátiles, incluso en contextos de baja liquidez como el de la BVC.

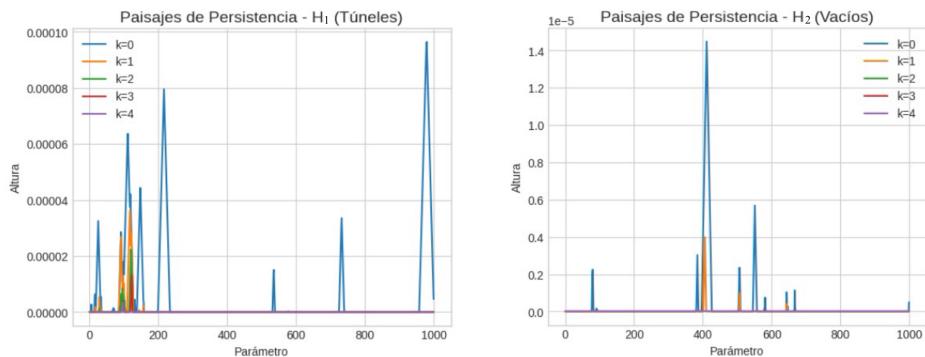


Figura 4. Paisaje de persistencia de la serie de tiempo de dimensión $d = 4$ con ventanas de tamaño $w = 30$.

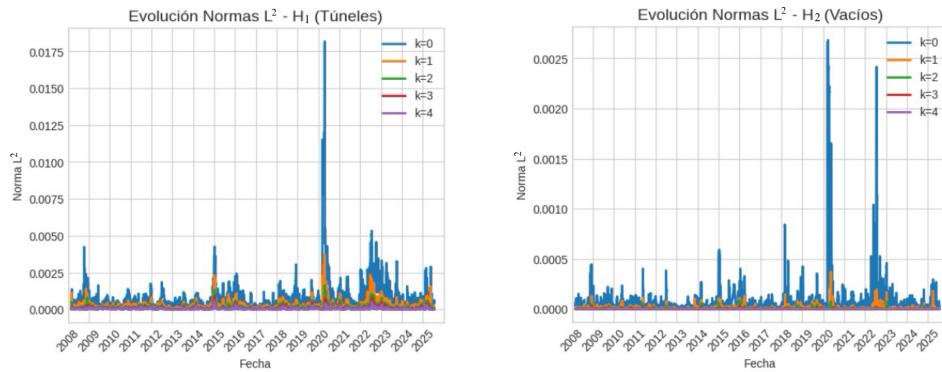


Figura 5. Norma L^2 del paisaje de persistencia en dimensión homológica 1 (izquierda) y en dimensión homológica 2 (derecha).

6. Conclusiones y perspectivas de trabajo futuro

En este trabajo se ofreció una introducción clara y concisa al TDA, dirigida a lectores que no necesariamente poseen formación especializada en topología algebraica. A través de distintas referencias, se expuso cómo el TDA ha emergido como una herramienta novedosa y prometedora para el análisis de grandes volúmenes de información, con el fin de identificar patrones relevantes en contextos complejos.

Dentro del TDA, destaca la homología persistente, una técnica capaz de capturar características estructurales y relaciones no lineales en los datos. Esta herramienta ha demostrado ser útil en diversas aplicaciones, especialmente en el ámbito financiero, donde ha mostrado capacidades predictivas al ser aplicada a datos de mercados financieros globales previos a crisis financieras. Por esta razón, se recomienda su utilización durante períodos de estrés en los mercados, con el fin de anticipar posibles crisis.

La aplicación del TDA al principal índice de la Bolsa de Valores de Colombia, en conjunto con las tres acciones más líquidas de este mercado, evidencia la vulnerabilidad del mercado colombiano frente a choques externos, así como la capacidad del TDA para anticipar posibles crisis bursátiles. Los resultados obtenidos son consistentes con los hallazgos reportados por Gidea y Katz [6] para los cuatro principales índices bursátiles de Estados Unidos, tanto en la crisis financiera global de 2008 como en la crisis bursátil generada por la pandemia del COVID-19 a partir de marzo de 2020. Asimismo, el TDA mostró capacidad predictiva para detectar crisis de carácter local, como la provocada por el desplome de los precios del petróleo, iniciado en junio de 2014, y el choque político de 2022 asociado con la llegada del nuevo gobierno en Colombia. Finalmente, la menor liquidez del mercado colombiano, en comparación con los índices estadounidenses, resalta el potencial de este método para la detección temprana de crisis financieras.

Como línea de investigación futura, se propone aplicar la homología persistente a conjuntos de datos provenientes de ámbitos distintos al financiero. Un ejemplo de particular interés sería la utilización del TDA en el análisis de información derivada del

monitoreo de condiciones ambientales en las principales ciudades de Colombia, con el propósito de generar alertas tempranas que contribuyan a la prevención de crisis en el ámbito de la salud pública.

Referencias

- [1] Aktas M.E., Akbas E. and Fatmaoui A.E., “Persistence Homology of Networks: Methods and Applications”, *Appl. Netw. Sci.*, 4 (2019), No. 1, 1–28. doi: 10.1007/s41109-019-0179-3
- [2] Asaad A., Ali D., Majeed T. and Rashid R., “Persistent Homology for Breast Tumor Classification Using Mammogram Scans”, *Mathematics*, 10 (2022), No. 21, 1–18. doi: 10.3390/math10214039
- [3] Bauer U., “Ripser: Efficient Computation of Vietoris–Rips Persistence Barcodes”, *J. Appl. Comput. Topol.*, 5 (2021), No. 3, 391–423. doi: 10.1007/s41468-021-00071-5
- [4] Bubenik P., “Statistical Topological Data Analysis Using Persistence Landscapes”, *J. Mach. Learn. Res.*, 16 (2015), No. 1, 77–102
- [5] Edelsbrunner H., Letscher D. and Zomorodian A., “Topological Persistence and Simplification”, *Discrete Comput. Geom.*, 28 (2002), No. 4, 511–533. doi: 10.1007/s00454-002-2885-2
- [6] Gidea M. and Katz Y., “Topological Data Analysis of Financial Time Series: Landscapes of Crashes”, *Physica A*, 491 (2018), 820–834. doi: 10.1016/j.physa.2017.09.028
- [7] Leaverton L., “Analysis of Financial Time Series Using TDA: Theoretical and Empirical Results”, *Thesis (M.Sc.)*, Universitat de Barcelona, Barcelona, 2020, p. 48.
- [8] Takens F., “Detecting Strange Attractors in Turbulence”, in *Dynamical Systems and Turbulence, Warwick 1980* (eds. Rand D.A. and Young L.S.), Springer, 1981, 366–381. doi: 10.1007/BFb0091924.