



**UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA**  
DEPARTAMENTO DE MATEMÁTICA

## **Aplicación del análisis topológico de datos a patrones puntuales**

Memoria de Título presentada por:

**Sofía Catalina Madrid Mellado**

como requisito parcial para optar al título de

**Ingeniería Civil Matemática**

Otorgado por la Universidad Técnica Federico Santa María

Profesor Guía

Dr. Francisco Cuevas Pacheco

Profesor Coguía

Dr. Pedro Montero Silva

Marzo 2025.



TÍTULO DE LA MEMORIA:

**Aplicación del análisis topológico de datos a patrones puntuales**

AUTOR:

**SOFÍA CATALINA MADRID MELLADO**

TRABAJO DE MEMORIA, presentado en cumplimiento parcial de los requisitos para el título de Ingeniería Civil Matemática de la Universidad Técnica Federico Santa María.

Dr. Profesor Guía

---

Dr. Revisor 1

---

Dr. Revisor 2

---

Valparaíso, Chile, Marzo 2025.

---

---

# AGRADECIMIENTOS

A lo largo de estos años, muchas personas han sido fundamentales en mi camino académico y personal. Este trabajo no habría sido posible sin su apoyo, compañía y aliento en los momentos más desafiantes.

Quiero agradecer a mis amigos, tanto a aquellos que estuvieron desde el inicio de la carrera como a quienes llegaron en la etapa final del proceso. Cada uno, de distintas maneras, aportó con su compañía, apoyo y consejos, haciendo que este camino fuera mucho más llevadero.

A mis profesores guía y coguía, por su orientación y motivación en el área de la investigación. Trabajar con ellos en este último tiempo ha sido una experiencia enriquecedora que me permitió profundizar en temas que hoy en día me apasionan.

También quiero hacer una mención especial a mi gata Yayita, quien con su compañía estuvo presente en mis noches de estudio y en los momentos difíciles. Su presencia fue un gran consuelo y una fuente de tranquilidad en este proceso.

A mi familia, por estar presente a lo largo de estos años y acompañarme en este camino de formación.

Finalmente, agradezco a todas las personas que, de una u otra manera, me han apoyado en este camino. Cada palabra de aliento y cada gesto de apoyo han sido fundamentales para llegar hasta aquí.

**¡Gracias!**



---

---

# RESUMEN

En esta memoria se estudian patrones puntuales utilizando el enfoque del Análisis Topológico de Datos (TDA), una metodología derivada de la topología algebraica que permite analizar la estructura de datos espaciales desde una perspectiva geométrica y topológica. El objetivo principal es evaluar la aplicabilidad del TDA en el análisis de la densidad y distribución de puntos, explorando su capacidad para identificar patrones de agrupamiento en datos espaciales.

Dado que la combinación del TDA con el análisis de patrones puntuales aún es un área en desarrollo, se realizaron simulaciones para evaluar su efectividad. Para ello, se generaron configuraciones de puntos mediante un proceso de Poisson homogéneo, permitiendo la representación de diversas figuras geométricas, tanto simples como modificadas, con el propósito de analizar cómo el TDA captura y representa sus características topológicas.

En primer lugar, se presentan los fundamentos teóricos de los patrones puntuales y del TDA, proporcionando el contexto necesario para su aplicación en el análisis espacial. Posteriormente, se describen las simulaciones realizadas y se analizan los resultados obtenidos a través de distintos diagramas de persistencia, los cuales permiten identificar componentes conexas, ciclos y cavidades en las configuraciones de puntos.

Finalmente, la metodología se aplica a datos reales del Bosque de Barro Colorado (BCI), con el objetivo de estudiar la distribución espacial de diversas especies. Para complementar este análisis, se integraron técnicas de clustering mediante el método k-medoids, lo que permitió evaluar la capacidad del TDA para diferenciar agrupaciones

de puntos y revelar patrones ecológicos en la organización espacial de las especies. Los resultados obtenidos indican que el TDA permite caracterizar diferencias topológicas en los datos, proporcionando información adicional sobre la estructura de los patrones espaciales. Sin embargo, se identificaron desafíos computacionales y la necesidad de considerar métricas que incorporen la topografía del terreno. Estos hallazgos resaltan el potencial del TDA como herramienta para el análisis de datos espaciales y abren nuevas oportunidades de investigación en ecología y estadística espacial.

---

---

# CONTENIDO

<b>AGRADECIMIENTOS</b>	<b>I</b>
<b>RESUMEN</b>	<b>III</b>
<b>1. INTRODUCCIÓN</b>	<b>1</b>
1.1. Introducción	1
<b>2. PRELIMINARES</b>	<b>7</b>
2.1. Patrones puntuales	7
2.1.1. Tipos de Patrones Puntuales	8
2.1.2. Propiedades estadísticas	9
2.1.3. Estadísticos de Resumen	11
2.1.4. Modelos de Procesos Puntuales	14
2.2. Topología	17
2.2.1. Homología simplicial	17
2.2.2. Homología Singular	20
2.2.3. Números de Betti	21
2.2.4. Diagramas	24
<b>3. ALGORITMOS</b>	<b>27</b>
3.1. Aplicación del TDA a patrones puntuales	27
3.2. Herramientas computacionales	32
3.3. Metodología	33

<b>4. EXPERIMENTOS Y DISCUSIÓN</b>	<b>35</b>
4.1. Generación de puntos mediante procesos de Poisson	36
4.2. Simulaciones de figuras simples	36
4.2.1. Toro	37
4.2.2. Esfera	41
4.2.3. Cono	44
4.2.4. Cilindro	48
4.2.5. Paraboloide Positivo (sin agujero)	52
4.2.6. Paraboloide Negativo (sin agujero)	56
4.2.7. Plano	60
4.3. Simulaciones de figuras modificadas	65
4.3.1. Toro Entrelazado	66
4.3.2. Toro Plano	68
4.3.3. Paraboloide Positivo con Agujero	72
4.3.4. Paraboloide Negativo con Agujero	75
4.3.5. Plano con Agujero	78
4.4. Cluster de superficies	82
4.4.1. Construcción de Descriptores Topológicos	82
4.4.2. Algoritmo k-medoids	83
4.4.3. Procedimiento de Clusterización	84
4.5. Clasificación y comparación de resultados	85
4.5.1. Clusterización con todas las figuras	86
4.5.2. Clusterización de subconjuntos específicos	89
4.5.3. Impacto de las modificaciones estructurales	98
4.5.4. Análisis de Clustering con el Toro, Esfera y Plano	99
4.5.5. Análisis de Clustering con 100 Repeticiones y 11 Figuras Simuladas	101
<b>5. APLICACIÓN DEL ANÁLISIS A LOS DATOS DE BCI</b>	<b>104</b>
5.1. Datos usados en el análisis	105
5.2. Metodología	106

---

5.3. Análisis de la distribución espacial de especies a lo largo de los censos	108
5.3.1. Distribución de Posola	108
5.3.2. Distribución de Guazul	109
5.3.3. Distribución de Pourbi	110
5.4. Análisis de agrupaciones de especies mediante k-medoids	112
5.4.1. Resultados del clustering	114
<b>6. COMENTARIOS FINALES Y TRABAJOS A FUTURO</b>	<b>116</b>
<b>ANEXO: TABLAS</b>	<b>118</b>
6.1. Resultados de Clustering simulaciones	118
6.1.1. Clusterización con todas las figuras	118
6.1.2. Clusterización de subconjuntos específicos	119
6.2. Resultados de Clustering para los datos BCI	123
<b>BIBLIOGRAFÍA</b>	<b>126</b>

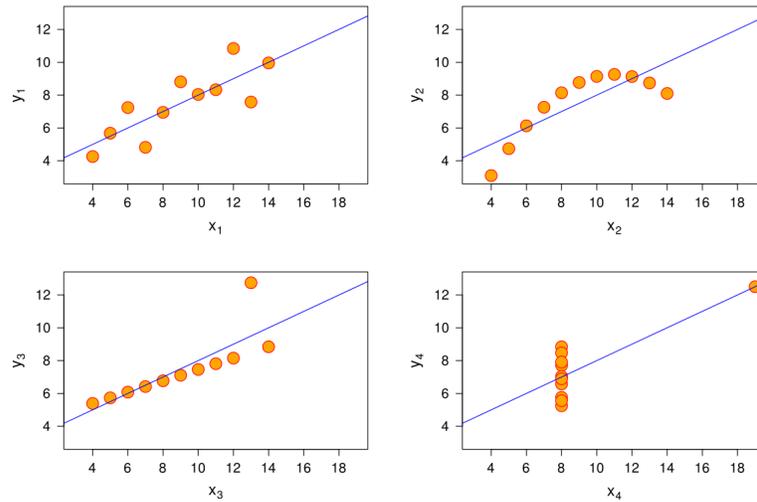


# INTRODUCCIÓN

### 1.1. Introducción

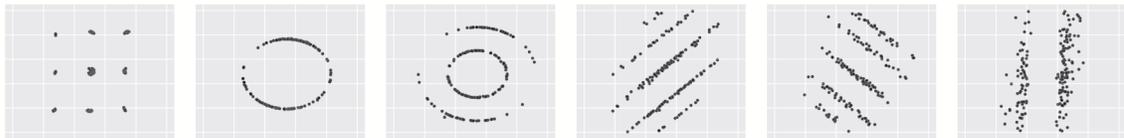
En los últimos años, se ha visto un incremento exponencial en la cantidad de datos, los cuales han llegado a tener diferentes formas, tamaños y dimensiones, dificultando cada vez más su estudio y análisis. Esto ha llevado a muchos investigadores a explorar metodologías para comprender mejor la estructura de los datos.

Un problema recurrente es que los descriptores estadísticos tradicionales, tales como media, varianza, desviación estándar, no siempre son suficientes para la comprensión y descripción de los datos. Un claro ejemplo de lo expuesto es lo visto en el trabajo de [Anscombe \(1973\)](#) conocido como 'El cuarteto de Anscombe' en el que se presentan cuatro conjuntos de datos con media, varianza, correlación y ecuación de regresión lineal idénticas, pero con distribuciones gráficas completamente diferentes, las cuales se ilustran en la Figura 1.1 presentada a continuación.



**Figura 1.1:** Cuarteto de Anscombe.

Para reforzar esta idea de buscar nuevas formas de analizar los datos y la importancia de tomar en cuenta la apariencia gráfica para complementar el análisis estadístico, en el estudio presentado por [Matejka and Fitzmaurice \(2017\)](#) se presenta un método para generar distintos conjuntos de datos que cumplen con las mismas estadísticas descriptivas básicas, como la media, la desviación estándar y la correlación de Pearson, pero que presentan distribuciones gráficas completamente diferentes. Estos ejemplos refuerzan la importancia del análisis visual en combinación con herramientas estadísticas. A continuación, se presenta otro ejemplo en la Figura 1.2.



**Figura 1.2:** Colección de conjuntos de datos que poseen apariencias diferentes, pero que a su vez todos poseen los mismos valores en sus estadísticos de resumen.

Para abordar este problema es posible utilizar las ideas de topología en el análisis de datos. La topología es una rama de las matemáticas que estudia las propiedades de las figuras que permanecen invariantes ante deformaciones continuas, como esti-

ramientos y torsiones. Este enfoque permite capturar características de los datos que no son evidentes mediante técnicas tradicionales del análisis de datos.

La conexión entre el análisis de datos y la topología se basa en la homología persistente, esto permite analizar características topológicas de los datos en diferentes escalas, lo que ha dado origen al campo del Análisis de Datos Topológicos (conocido por sus siglas en inglés como TDA). Este enfoque fue formalmente introducido por primera vez en el trabajo de [Delfinado and Edelsbrunner \(1993\)](#), donde se presentó un algoritmo incremental para calcular invariantes topológicos, como los números de Betti que contienen la información sobre la cantidad de estructuras topológicas presentes en un espacio según dimensión. Este avance sentó las bases para el desarrollo de metodologías más avanzadas relacionadas con la homología persistente. Posteriormente, [Edelsbrunner et al. \(2002\)](#) consolidaron las bases teóricas y computacionales de la persistencia topológica, posicionándose como los principales pioneros en el área. Su trabajo permitió comprender cómo las características topológicas, tales como componentes conexas, agujeros y cavidades, emergen, persisten y desaparecen a medida que varía un parámetro, ofreciendo una herramienta poderosa para el análisis estructural de los datos.

Las técnicas de TDA son útiles para analizar la estructura subyacente de los datos, basándose en características topológicas como la presencia de agujeros, cavidades y componentes conexas, las cuales se calculan utilizando los números de Betti. Esto proporciona información sobre la organización y distribución de los datos en el espacio, además de distinguir figuras y entender la morfología de los datos.

El objetivo de este trabajo es incorporar esta nueva técnica al análisis de patrones puntuales. El análisis de patrones puntuales es una rama de la estadística espacial que se enfoca en el estudio de la distribución y las relaciones espaciales entre puntos, dentro de un espacio definido. Estos puntos pueden representar ubicaciones geográficas, eventos en un plano, posiciones de árboles en un bosque, entre otros ejemplos.

Cabe destacar que no es primera vez que se busca aplicar TDA en el análisis de patrones puntuales, pues en el trabajo expuesto por [Biscio et al. \(2020\)](#) se propone una metodología para realizar pruebas de ajustes en procesos puntuales, lo que implica determinar si un conjunto de datos puede ser razonablemente modelado por un proceso puntual teórico dado. Para ello proponen un enfoque basado en diagramas de persistencia como estadísticos clave para la prueba de ajuste. Entre los principales resultados obtenidos se muestra que el TDA es más sensible a la estructura espacial, siendo útil en modelos donde los datos exhiben dependencias espaciales complejas. Se destaca que es robusto frente al ruido y que no requiere suposiciones estrictas sobre la naturaleza exacta de las dependencias espaciales.

A diferencia de dicho estudio, en este trabajo se explorará el reconocimiento de agujeros y otras estructuras en la superficie, junto con el análisis de la distribución espacial de los puntos.

Para ello se realizarán simulaciones de distintas figuras con el fin de identificar casos borde y verificar si esta técnica es capaz de diferenciarlas, y detectar la estructura solo a partir de una nube de puntos. Posteriormente, se aplicará esta metodología a un conjunto de datos de ecosistemas forestales de Barro Colorado, Panamá. Este conjunto de datos utilizado proviene de un extenso estudio realizado en esta isla, la cual es considerada un laboratorio natural, que cuenta con la información de más de 500 especies y cerca de 424.000 puntos distribuidos en el espacio. A partir de este análisis, se explorará el uso de TDA en combinación con técnicas de clustering y se compararán sus resultados con los obtenidos mediante estadísticos de resumen tradicionales en patrones puntuales. Estas comparaciones permitirán evaluar la eficiencia del TDA, identificar sus ventajas y limitaciones en la aplicación a datos ecológicos y analizar cómo las estructuras topológicas detectadas se correlacionan con los patrones ecológicos identificados con otros métodos.

En particular, se espera que el TDA proporcione una comprensión más detallada de la organización espacial y la dinámica ecológica en el ecosistema forestal de Barro Colorado. Este enfoque busca no solo validar la efectividad del TDA en patrones puntuales, sino también ofrecer nuevas perspectivas para el análisis espacial de datos

complejos.

Esta tesis se estructura en seis capítulos que abordan de manera progresiva el análisis topológico de datos aplicado a patrones puntuales. En el Capítulo 1, se presenta la Introducción, donde se expone el contexto y la motivación del trabajo, junto con una descripción general de las herramientas utilizadas. El Capítulo 2 está dedicado a los Preliminares, donde se establecen las bases teóricas necesarias para comprender tanto los patrones puntuales como los conceptos fundamentales de la topología algebraica. En el Capítulo 3, se detallan los Algoritmos, explicando cómo el TDA se aplica al estudio de patrones puntuales, además de describir las librerías y métodos computacionales empleados. El Capítulo 4 está centrado en los Experimentos y Discusión, presentando los resultados obtenidos a partir de figuras simuladas simples y modificadas, así como análisis de clustering para evaluar la capacidad del TDA de diferenciar o identificar estructuras topológicas al combinar distintas figuras. El Capítulo 5 aborda la Aplicación del Análisis a los Datos de BCI, donde se implementan las técnicas desarrolladas sobre datos reales del Bosque de Barro Colorado. Finalmente, el Capítulo 6 presenta los Comentarios Finales y Trabajos Futuros, donde se resumen las conclusiones del estudio y se proponen posibles líneas de investigación para continuar explorando el potencial del TDA en el análisis de patrones espaciales.



## Capítulo 2

---

# PRELIMINARES

En este capítulo se presenta una breve introducción a patrones puntuales y topología para proveer los conceptos fundamentales necesarios para entender y contextualizar el análisis de datos topológicos (TDA) aplicado a los patrones espaciales. Se describen las bases teóricas de los patrones puntuales, incluyendo su clasificación y los procesos subyacentes, así como las nociones básicas de topología, como homología simplicial, números de Betti y diferentes tipos de diagramas, entre los que se incluyen los diagramas de persistencia, los diagramas de código de barras y los landscapes o paisajes de persistencia. Este marco teórico permitirá comprender las herramientas utilizadas en los experimentos y su interpretación en el contexto del análisis de datos.

### 2.1. Patrones puntuales

En esta sección, se presentan los conceptos necesarios para la comprensión de los patrones puntuales, este enfoque es clave para analizar cómo los eventos, objetos o entidades se distribuyen espacialmente en un área determinada. Estos conceptos son fundamentales porque trabajaremos con datos espaciales, donde la disposición y las relaciones entre los puntos ofrecen información crucial para comprender las

dinámicas subyacentes de los sistemas estudiados.

Para establecer los fundamentos necesarios, utilizaremos [Baddeley et al. \(2015\)](#); [Daley et al. \(2003\)](#); [Moller and Waagepetersen \(2003\)](#), los cuales proporcionan una base sólida tanto teórica como práctica. Este marco será esencial para desarrollar las herramientas y metodologías requeridas para estudiar patrones puntuales, abriendo paso a su análisis avanzado en escenarios más complejos.

Sea  $S$  un espacio métrico equipado con una medida de referencia  $\nu$ . Se dice que  $Y$  es un proceso puntual si es un subconjunto aleatorio, numerable y localmente finito. Específicamente, se asumirá que  $Y$  es un proceso puntual simple, es decir, con probabilidad uno, todos los puntos ocurren en ubicaciones diferentes. En la práctica, únicamente se observan los puntos contenidos en una ventana de observación acotada  $W \subseteq S$ . Sea  $n(A)$  la cardinalidad de  $A$ . Luego, se define  $N(A)$  como la cantidad de elementos de  $Y$  que caen en  $A$ , es decir,

$$N(A) = n(Y \cap A).$$

Para un conjunto  $A \subseteq S$ , el número de puntos en  $A$  se denota por  $N(A)$ , y  $Y_A = Y \cap A$  representa la restricción de  $Y$  a  $A$ . Si el proceso incluye variables aleatorias adicionales, estas forman un proceso puntual marcado, donde los elementos se describen como pares  $(Y, M_Y)$ , siendo  $M_Y$  una marca asociada al punto  $Y$ .

### 2.1.1. Tipos de Patrones Puntuales

Los procesos puntuales pueden clasificarse en función de la distribución espacial de sus puntos. Existen tres tipos principales de patrones:

- Aleatorio: No presenta estructura aparente, los puntos se distribuyen de manera independiente sin interacción.
- Regular: Exhibe repulsión entre puntos, lo que implica una disposición más uniforme en el espacio.

- Agregado: Presenta agrupamientos o clusters, lo que indica dependencia entre los puntos y posibles interacciones subyacentes.

Esta clasificación es fundamental para el análisis de datos espaciales, ya que permite distinguir entre diferentes estructuras subyacentes en los procesos observados y seleccionar las herramientas adecuadas para su estudio.

### 2.1.2. Propiedades estadísticas

El estudio de las propiedades estadísticas de un proceso puntual  $Y$  es fundamental para describir su comportamiento espacial. Estas propiedades se dividen en dos niveles principales: primer orden y segundo orden.

#### Propiedades de Primer Orden

El primer orden se describe mediante la medida de intensidad  $\alpha$ , definida como:

$$\alpha(A) = \mathbb{E}[N(A)], \quad A \subseteq S,$$

donde  $N(A)$  es el número de puntos del proceso  $Y$  contenidos en el subconjunto  $A$  del espacio  $S$ . La medida  $\alpha$  representa el valor esperado del conteo de puntos en la región  $A$ .

Si existe una función  $\rho(u)$ , denominada función de intensidad, tal que:

$$\alpha(A) = \int_A \rho(u) d\vartheta(u),$$

entonces  $\rho(u)$  indica la densidad promedio esperada de puntos en la vecindad del punto  $u$ , en relación con la medida de referencia  $\vartheta$ . Para procesos homogéneos,  $\rho(u)$  es constante en todo  $S$ , mientras que para procesos inhomogéneos,  $\rho(u)$  varía dependiendo de  $u$ .

### Propiedades de Segundo Orden

El comportamiento de segundo orden, que captura las interacciones espaciales entre puntos, se describe mediante la medida factorial de segundo orden  $\alpha^{(2)}$ , definida como:

$$\alpha^{(2)}(B) = \mathbb{E} \left[ \sum_{u \neq v \in Y} I[(u, v) \in B] \right], \quad B \subseteq S \times S,$$

donde  $I[\cdot]$  es la función indicadora que verifica si el par  $(u, v)$  pertenece al subconjunto  $B$ . Si existe una función de densidad producto de segundo orden  $\rho^{(2)}(u, v)$ , tal que:

$$\alpha^{(2)}(B) = \int \int I[(u, v) \in B] \rho^{(2)}(u, v) d\vartheta(u) d\vartheta(v),$$

entonces  $\rho^{(2)}(u, v)$  describe la probabilidad conjunta de observar un par de puntos en las vecindades de  $u$  y  $v$ .

### Interacciones Espaciales y Cardinalidad

La función de densidad producto  $\rho^{(2)}(u, v)$  es especialmente útil para analizar las interacciones espaciales entre puntos. Si  $\rho^{(2)}(u, v) > \rho(u)\rho(v)$ , esto indica agrupamiento, mientras que  $\rho^{(2)}(u, v) < \rho(u)\rho(v)$  sugiere repulsión. Para procesos donde no hay interacción espacial, como en el proceso de Poisson homogéneo, se cumple que  $\rho^{(2)}(u, v) = \rho(u)\rho(v)$ .

La cardinalidad de un proceso puntual se refiere al número total de puntos en una región dada, lo que está directamente relacionado con la intensidad del proceso en esa región. Por ejemplo, en un proceso homogéneo, el valor esperado de la cardinalidad en una región  $A$  es proporcional al tamaño de  $A$  y está dado por:

$$\mathbb{E}[N(A)] = \rho\vartheta(A),$$

donde  $\rho$  es la intensidad constante del proceso y  $\vartheta(A)$  es la medida de referencia del conjunto  $A$ .

Relación entre Propiedades de Primer y Segundo Orden

La medida factorial de segundo orden  $\alpha^{(2)}$  también está relacionada con los momentos de  $N(A)$ , el número de puntos en  $A$ . En particular, para dos subconjuntos disjuntos  $A, B \subseteq S$ , se tiene:

$$\mathbb{E}[N(A)N(B)] = \alpha^{(2)}(A \times B) + \alpha(A \cap B).$$

### 2.1.3. Estadísticos de Resumen

Las estadísticas funcionales resumen se han convertido en una herramienta muy útil para describir patrones puntuales espaciales. De hecho, el análisis exploratorio, el ajuste de modelos y la validación de modelos se basan en estimaciones de estas estadísticas. Las estadísticas resumen más populares se fundamentan en propiedades de segundo orden de  $N(\cdot)$ , como la función de correlación pareada y la función  $K$ , o en la distribución de distancias entre puntos, como las funciones  $F$ ,  $G$  y  $J$ . A continuación, se describen las estadísticas resumen más comúnmente utilizadas tanto para casos homogéneos como inhomogéneos.

La primera estadística resumen es la función de correlación pareada, que se define como:

$$g(u, v) = \frac{\rho^{(2)}(u, v)}{\rho(u)\rho(v)}, \quad u, v \in S,$$

para  $\rho(u)\rho(v) > 0$ , siempre que existan  $\rho$  y  $\rho^{(2)}$ . Intuitivamente,  $g(u, v) > 1$  (respectivamente  $g(u, v) < 1$ ) significa que la presencia de un punto en  $u$  aumenta (respectivamente disminuye) la probabilidad de observar otro punto en  $v$ . Por esta razón,  $g$  se utiliza como estadística funcional resumen.

Otra estadística importante es la función  $K$ , definida como:

$$K(B) = \frac{1}{\vartheta(A)} \mathbb{E} \left[ \sum_{u \neq v \in Y} I[u \in A, d(u, v) \in B] \frac{1}{\rho(u)\rho(v)} \right], \quad A, B \subseteq S,$$

donde  $d(\cdot, \cdot)$  es una métrica en  $S$ . Si  $K(B)$  no depende de la elección de  $A \subseteq S$ , entonces  $Y$  se denomina estacionario de segundo orden reponderado por intensidad

(SOIRS, por sus siglas en inglés), y  $K$  se llama la medida de momento reducido de segundo orden. Si un proceso puntual satisface la condición SOIRS, entonces se tiene:

$$K(B) = \int_B g_0(h) \vartheta(h), \quad B \subseteq S.$$

Finalmente, la función  $K$  se define como:

$$K(r) = K(b_S(0, r)), \quad r \geq 0,$$

donde  $b_S(0, r) = \{v \in S : d(0, v) \leq r\}$  es la bola con centro 0 y radio  $r$ . Es importante notar que la función  $K$  se define bajo la condición SOIRS, que depende del espacio  $S$ .

Además de la función  $K$ , otras estadísticas funcionales resumen ampliamente utilizadas en el análisis de patrones puntuales son las funciones  $F$ ,  $G$  y  $J$ . Estas funciones se basan en la distribución de distancias entre puntos y ofrecen diferentes perspectivas para caracterizar patrones espaciales, especialmente en términos de agrupamiento, dispersión y aleatoriedad.

La función  $F$ , conocida como la función de distribución de distancias desde un punto arbitrario al punto más cercano del patrón, se define como:

$$F(r) = \mathbb{P}(d(u, Y) \leq r), \quad u \notin Y,$$

donde  $d(u, Y)$  es la distancia desde un punto arbitrario  $u$  hasta el punto más cercano del patrón  $Y$ . Intuitivamente,  $F(r)$  mide la probabilidad de que un punto seleccionado aleatoriamente en el espacio  $S$  esté a una distancia menor o igual a  $r$  de algún punto del patrón  $Y$ . Para procesos homogéneos,  $F(r)$  puede interpretarse como la proporción de puntos del espacio que se encuentran a una distancia menor o igual a  $r$  de al menos un punto del patrón.

La función  $G$ , denominada función de distribución de distancias desde un punto del patrón a su vecino más cercano, se define como:

$$G(r) = \mathbb{P}(d(u, v) \leq r \mid u \in Y, v \in Y, u \neq v),$$

donde  $d(u, v)$  es la distancia entre dos puntos  $u$  y  $v$  en  $Y$ . Esta función describe la probabilidad de que, dado un punto  $u$  perteneciente al patrón, su vecino más cercano esté a una distancia menor o igual a  $r$ . En procesos homogéneos,  $G(r)$  proporciona información sobre la distribución de las distancias mínimas entre puntos, lo que permite identificar patrones de agrupamiento o dispersión.

Finalmente, la función  $J$  combina las funciones  $F$  y  $G$  para proporcionar una medida de interacción espacial entre puntos. Se define como:

$$J(r) = \frac{1 - G(r)}{1 - F(r)}, \quad F(r) < 1.$$

La función  $J$  ofrece una perspectiva adicional para distinguir entre diferentes tipos de patrones espaciales. Sus principales interpretaciones son las siguientes:

- Si  $J(r) > 1$ , existe evidencia de repulsión espacial entre los puntos del patrón. Esto ocurre, por ejemplo, en procesos inhibidos donde los puntos tienden a evitarse.
- Si  $J(r) < 1$ , indica que los puntos están agrupados o que existe atracción espacial. Este comportamiento es típico en procesos cluster, donde los puntos tienden a formar conglomerados.
- Si  $J(r) = 1$ , el patrón es consistente con un proceso completamente aleatorio, como un proceso de Poisson homogéneo.

En conjunto, estas estadísticas funcionales resumen permiten analizar diferentes aspectos de los patrones puntuales. Mientras que  $F(r)$  se centra en la relación del patrón con un punto arbitrario en el espacio,  $G(r)$  se enfoca en las distancias entre los puntos del patrón, y  $J(r)$  integra ambas perspectivas para proporcionar una visión más completa de las interacciones espaciales. Este conjunto de herramientas es esencial para la caracterización y comparación de patrones espaciales tanto en estudios exploratorios como en la validación de modelos.

### 2.1.4. Modelos de Procesos Puntuales

Los modelos de procesos puntuales son herramientas fundamentales para describir y analizar patrones puntuales espaciales. Estos modelos se clasifican según las características de las interacciones espaciales que representan, desde patrones completamente aleatorios hasta distribuciones agrupadas o inhibidas.

El proceso de Poisson es el modelo más simple y ampliamente utilizado para patrones espaciales. Se caracteriza por la independencia entre puntos y por el hecho de que el número de puntos en cualquier subconjunto  $A \subseteq S$  sigue una distribución de Poisson con media  $\lambda\vartheta(A)$ , donde  $\lambda$  es la intensidad del proceso y  $\vartheta$  es la medida de referencia en  $S$ . En el caso homogéneo,  $\lambda$  es constante en todo el espacio, mientras que en el caso inhomogéneo,  $\lambda(u)$  varía en función de la posición  $u$ . Este modelo no considera interacciones espaciales, lo que lo hace apropiado para describir patrones completamente aleatorios. Formalmente, un proceso puntual  $Y$  en  $S$  es un proceso de Poisson si cumple las siguientes propiedades fundamentales:

- Propiedad de independencia: El número de puntos en subconjuntos disjuntos de  $S$  son variables aleatorias independientes.
- Propiedad de conteo de Poisson: Para cualquier subconjunto  $A \subseteq S$ , el número de puntos en  $A$ , denotado  $N(A)$ , sigue una distribución de Poisson con media  $\lambda\nu(A)$ .

Los procesos cluster son modelos diseñados para describir patrones donde los puntos tienden a agruparse en torno a centros específicos, reflejando interacciones espaciales de atracción. En estos modelos, los centros de los clústeres suelen distribuirse siguiendo un proceso de Poisson, mientras que los puntos secundarios, que representan la estructura de agrupamiento, se generan en torno a estos centros según una distribución dada.

Existen diferentes modelos de procesos cluster, cada uno caracterizado por la forma en que los puntos secundarios se distribuyen alrededor de los centros. En el modelo de Thomas, los puntos secundarios se disponen siguiendo una distribución normal isotrópica en torno a cada centro, lo que implica que la densidad de puntos decrece

suavemente conforme nos alejamos del centro del clúster. En el modelo de Neyman-Scott, en cambio, se permite que las distribuciones alrededor de los centros sigan otras densidades, como exponenciales o uniformes, proporcionando mayor flexibilidad en la estructura del agrupamiento. Un aspecto clave de estos modelos es la función de correlación pareada  $g(u, v)$ , la cual toma valores mayores a 1 para distancias pequeñas, lo que indica una mayor probabilidad de encontrar puntos cercanos entre sí debido al agrupamiento.

En contraste, los procesos inhibidos representan patrones donde los puntos tienden a evitarse entre sí, lo que da lugar a distribuciones espaciales más regulares. Un ejemplo destacado es el modelo de Strauss, que introduce un parámetro  $\gamma$  para controlar la repulsión entre puntos: cuando  $\gamma < 1$ , la probabilidad de observar un patrón dado disminuye si los puntos están demasiado cerca unos de otros. Los procesos de Gibbs, más generales, incluyen términos de interacción basados en funciones potenciales, permitiendo describir una amplia gama de patrones, desde agrupados hasta inhibidos. Otro modelo importante dentro de los procesos inhibidos es el Proceso Puntual Determinante (DPP), que surge en física cuántica y se ha utilizado en diversas aplicaciones estadísticas y computacionales. A diferencia del modelo de Strauss, que introduce interacción entre puntos a través de un parámetro explícito, los DPPs modelan la repulsión de manera implícita mediante la estructura espectral de una matriz núcleo  $K(x, y)$ . Este núcleo define la probabilidad conjunta de observar puntos en ciertas ubicaciones, favoreciendo configuraciones donde los puntos estén más separados entre sí. Matemáticamente, la probabilidad de observar un conjunto de puntos  $Y = \{x_1, x_2, \dots, x_n\}$  en un dominio  $S$  está dada por:

$$P(Y) \propto \det(K(x_i, x_j))_{1 \leq i, j \leq n}.$$

Los DPPs presentan ventajas computacionales en relación con otros procesos inhibidos, ya que permiten muestreo eficiente y tienen propiedades analíticas útiles, como la capacidad de calcular momentos y funciones de correlación de manera más directa. Estos modelos han sido aplicados en diversas áreas, incluyendo la distribución de árboles en bosques, análisis de distribución de átomos en mecánica estadística y

selección de subconjuntos en aprendizaje automático.

Cada modelo tiene sus propias ventajas y limitaciones. Mientras que los procesos de Poisson son simples de interpretar y calcular, no capturan interacciones espaciales. Los procesos cluster permiten modelar patrones agrupados, pero a menudo requieren estimaciones complejas de parámetros. Los procesos inhibidos, aunque útiles para describir patrones regulares, suelen implicar una implementación computacional más exigente. Estas características determinan la elección del modelo según el tipo de patrón puntual que se desea analizar o modelar.

### **Propiedades en Modelos Clásicos**

En un proceso de Poisson homogéneo,  $\rho^{(2)}(u, v) = \rho^2$ , lo que implica que los puntos son independientes entre sí y no existe interacción espacial. En contraste, en procesos cluster, como el modelo de Thomas,  $\rho^{(2)}(u, v) > \rho^2$  para distancias pequeñas, indicando agrupamiento. Por otro lado, en procesos inhibidos, como el modelo de Strauss,  $\rho^{(2)}(u, v) < \rho^2$ , reflejando repulsión entre puntos cercanos.

## 2.2. Topología

El análisis topológico de datos se basa en herramientas de la topología para estudiar la estructura subyacente en conjuntos de datos. Su principal objetivo es identificar características invariantes bajo transformaciones continuas, permitiendo extraer información que los métodos tradicionales pueden pasar por alto. En esta sección se presentan los conceptos fundamentales que servirán de base para su aplicación en el análisis de datos espaciales [Joharinad and Jost \(2023\)](#); [Vakil \(2017\)](#).

### 2.2.1. Homología simplicial

Un  $n$ -símplice es la envolvente convexa de  $n + 1$  puntos en  $\mathbb{R}^n$  que no están en un hiperplano. Se denota un  $n$ -símplice por el conjunto ordenado de sus vértices  $[v_0, v_1, \dots, v_n]$ . Esto equivale a darle una orientación.

$$[v_0, v_1] = \begin{array}{c} \bullet \longrightarrow \bullet \\ v_0 \qquad v_1 \end{array} \neq \begin{array}{c} \bullet \longleftarrow \bullet \\ v_0 \qquad v_1 \end{array} = [v_1, v_0]$$

Por ejemplo, un 1-símplice es un segmento de línea que conecta dos puntos ordenados, mientras que un 2-símplice es un triángulo formado por tres vértices con sus respectivos lados, y un 3-símplice es un tetraedro. La orientación asignada a los vértices es importante porque permite definir las operaciones algebraicas que se explicarán posteriormente.

El  $n$ -símplice estándar es un caso especial de los símplices, denotado como  $\Delta^n := [e_0, \dots, e_n]$ , definido por los vértices de la base canónica de  $\mathbb{R}^{n+1}$ , y se representa como

$$\Delta^n = \left\{ (t_0, \dots, t_n) \in \mathbb{R}^{n+1}, \sum_{i=0}^n t_i = 1, t_i \geq 0, \forall i \right\} \subseteq \mathbb{R}^{n+1}.$$

Cada  $n$ -símplice en un espacio puede describirse mediante un homeomorfismo lineal

canónico que lo relaciona con  $\Delta^n$ . Este mapeo es definido como

$$\begin{aligned} \beta : \Delta^n &\xrightarrow{\sim} [v_0, v_1, \dots, v_n] \\ (t_0, \dots, t_n) &\mapsto \sum_{i=0}^n t_i v_i \end{aligned}$$

es el único homeomorfismo lineal que envía  $e_i \mapsto v_i$ . Los  $(t_0, \dots, t_n)$  son las coordenadas baricéntricas de  $[v_0, \dots, v_n]$ , es decir, permite usar coordenadas baricéntricas para describir cualquier punto dentro del símplice.

Los símplices pueden combinarse para formar estructuras más complejas conocidas como  $\Delta$ -complejos. Un  $\Delta$ -complejo es el espacio cociente de una colección de símplices disjuntos  $\{\Delta_\alpha\} \alpha \in I$  que se obtiene identificando algunas de sus caras vía homeomorfismos lineales canónicos preservando el orden de vértices, y se representa como un espacio cociente.

Sea  $X = \left( \bigsqcup_{\alpha \in I} \Delta_\alpha \right) / \sim$  un  $\Delta$ -complejo. El símplice abierto  $e_\alpha^n$  asociado al  $n$ -símplice  $\Delta_\alpha$  es el interior de  $\Delta_\alpha$  (es decir,  $\Delta_\alpha$  sin sus caras). Así, cada  $e_\alpha^n$  viene con un mapeo canónico

$$\sigma_\alpha : \Delta^n \xrightarrow{\sim} \Delta_\alpha \rightarrow X$$

llamado mapeo característico de  $e_\alpha^n$  y con

$$\sigma_\alpha|_{\text{int}(\Delta^n)} : \text{int}(\Delta^n) \xrightarrow{\sim} e_\alpha^n$$

homeomorfos.

Para estudiar las propiedades topológicas de los  $\Delta$ -complejos, es necesario asociarles grupos abelianos mediante una estructura algebraica llamada complejo de cadenas.

Un complejo de cadenas  $C_\bullet$  es una sucesión de morfismos de grupos abelianos

$$C_\bullet : \cdots \rightarrow C_{n+1} \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} \cdots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

que satisfacen la propiedad  $\partial_n \circ \partial_{n+1} = 0$  para todo  $n \geq 1$ , es decir,  $\text{Im}(\partial_{n+1}) \subseteq \ker(\partial_n)$ .

Los elementos de  $C_n$  se llaman cadenas, los de  $\ker(\partial_n)$  ciclos y los de  $\text{Im}(\partial_{n+1})$  bordes.

El grupo abeliano cociente  $H_n(C_\bullet) := \ker(\partial_n) / \text{Im}(\partial_{n+1})$  es el  $n$ -ésimo grupo de homología de  $C_\bullet$ . Decimos que dos ciclos  $\alpha, \alpha' \in \ker(\partial_n)$  son homólogos si tienen la misma clase de homología, es decir, existe un  $\beta \in C_{n+1}$  tal que  $\alpha = \alpha' + \partial_{n+1}(\beta)$ .

Un concepto clave en homología simplicial es la función borde, que describe cómo los  $n$ -símplices contribuyen a las caras de menor dimensión.

La función borde se define como,

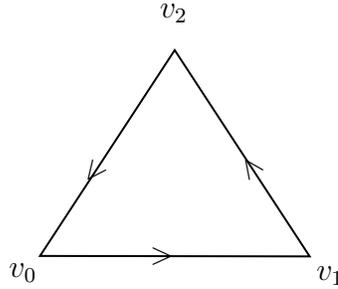
$$\begin{aligned} \partial_n : \Delta_n(X) &\rightarrow \Delta_{n-1}(X) \\ [v_0, \dots, v_n] &\mapsto \sum_{i=0}^n (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_n] \end{aligned}$$

o equivalentemente

$$\partial_n(\sigma_\alpha) := \sum_{i=0}^n (-1)^i \sigma_\alpha|_{[v_0, \dots, \hat{v}_i, \dots, v_n]}$$

donde  $[v_0, \dots, \hat{v}_i, \dots, v_n]$  representa el  $n - 1$ -símplice obtenido al omitir el vértice  $v_i$ .

El siguiente ejemplo ilustra cómo opera la función borde para un 2-símplice. Sea  $X = \Delta = [v_0, v_1, v_2]$ , un triángulo orientado cuyos vértices están ordenados. Entonces, la frontera de  $\Delta$  es



es decir,

$$\partial_2(\Delta) = (-1)^0 [\hat{v}_0, v_1, v_2] + (-1)^1 [v_0, \hat{v}_1, v_2] + (-1)^2 [v_0, v_1, \hat{v}_2]$$

es decir,  $\partial_2(\Delta) = [v_1, v_2] - [v_0, v_2] + [v_0, v_1]$  por tanto "  $\partial_n$  recorre el borde de forma orientada".

La función borde permite construir una estructura llamada complejo de cadenas, que conecta símplices de diferentes dimensiones mediante sus fronteras. En este contexto, un complejo simplicial  $\Delta_\bullet(x)$  está compuesto por grupos abelianos  $\Delta_n(x)$  y funciones borde  $\partial_n$ , formando la siguiente secuencia  $\Delta_\bullet(x) : \dots \rightarrow \Delta_n(x) \xrightarrow{\partial_n} \Delta_{n-1}(x) \xrightarrow{\partial_{n-1}} \Delta_{n-2}(x) \rightarrow \dots \rightarrow \Delta_0(x) \rightarrow 0$ .

### 2.2.2. Homología Singular

Cuando los  $\Delta$ -complejos no son suficientes para modelar un espacio topológico, se utiliza la homología singular, que extiende el concepto de homología a espacios arbitrarios. En este caso, un  $n$ -símplice singular en  $X$  es un mapeo  $\sigma : \Delta^n \rightarrow X$ , donde  $X$  es un espacio topológico en estudio. Tal como antes,

$$\Delta^n = \left\{ (t_0, \dots, t_n) \in \mathbb{R}^{n+1}, \sum_{i=0}^n t_i = 1 \text{ y } t_i \geq 0 \forall i \right\}$$

La colección de todos estos mapeos genera el grupo de cadenas singulares  $C_n(x)$ , con

la misma función borde que en la homología simplicial

$$\begin{aligned} \partial_n : C_n(X) &\rightarrow C_{n-1}(X) \\ \sigma &\mapsto \sum_{i=0}^n (-1)^i \sigma|_{[v_0, \dots, \hat{v}_i, \dots, v_n]} \end{aligned}$$

donde identificamos con  $\Delta^{n-1}$  vía el homeomorfismo lineal canónico.

La diferencia clave aquí es que los símlices singulares no necesitan ser homeomorfismos, lo que permite modelar espacios mucho más generales. Al igual que antes, los grupos de homología  $H_n(X)$  se definen como el cociente entre ciclos y bordes.

Sea  $X$  un espacio topológico y sea  $n \in \mathbb{N}$ . El  $n$ -ésimo grupo de homología singular de  $X$  es el  $n$ -ésimo grupo de homología de  $C_\bullet(X)$ , es decir,  $H_n(X) := H_n(X, \mathbb{Z}) := H_n(C_\bullet(X)) = \ker(\partial_n) / \text{Im}(\partial_{n+1})$

Además, se destaca lo siguiente

- (1) El adjetivo singular se refiere a que  $\sigma$  no es necesariamente un homeomorfismo de  $\text{int}(\Delta^n)$  a su imagen
- (2) El grupo  $C_n(X)$  es mucho más grande que los  $\Delta_n(X)$  (típicamente no es finitamente generado).
- (3) Tal como antes,

$$C_\bullet(X) : \dots \rightarrow C_n(x) \xrightarrow{\partial_n} C_{n-1}(x) \rightarrow \dots \rightarrow C_0(x) \rightarrow 0$$

es un complejo de cadenas.

### 2.2.3. Números de Betti

Los números de Betti son invariantes topológicos fundamentales que capturan información clave sobre la estructura de un espacio topológico. Formalmente, se definen a través de los grupos de homología, los cuales miden ciclos independientes que no son bordes en diferentes dimensiones. Por ejemplo:

- $b_0$  indica el número de componentes conexas del espacio.
- $b_1$  representa el número de agujeros en una superficie bidimensional.
- $b_2$  cuenta las cavidades tridimensionales.

Para entender la relación entre los números de Betti y los complejos simpliciales, considere el  $n$ -símplice estándar, definido como

$$\Delta^n = \left\{ (t_0, \dots, t_n) \in \mathbb{R}^{n+1}, \sum_{i=1}^n t_i = 1, \quad t_i \geq 0, \forall i \right\}$$

donde  $\Delta^n \cong \mathbb{D}^n$  y  $\partial\Delta^n \cong \partial\mathbb{D}^n = \mathbb{S}^{n-1}$ , es decir, el borde de  $\Delta^n$  es homeomorfo a una esfera de dimensión  $n - 1$ . Dado que  $(\mathbb{D}^n, \partial\mathbb{D}^n)$  es un buen par y  $\mathbb{D}^n / \langle \partial\mathbb{D}^n \rangle \cong \mathbb{S}^n$ , se deduce que los grupos de homología relativos satisfacen lo siguiente

$$H_i(\Delta^n, \partial\Delta^n) \cong H_i(\mathbb{D}^n, \partial\mathbb{D}^n) \cong \widetilde{H}_i(\mathbb{S}^n) \cong \begin{cases} \mathbb{Z}, & \text{si } i = n \\ 0, & \text{si } i \neq n \end{cases}$$

Esta relación establece la conexión entre los símlices y las propiedades topológicas del espacio. En el caso de un  $\Delta$ -complejo  $X$ , que es el espacio cociente de una unión disjunta de símlices  $\Delta_\alpha^n$ , el mapeo cociente

$$p : \bigsqcup_{\substack{\alpha \in I \\ n \in \mathbb{N}}} \Delta_\alpha^n \rightarrow X$$

permite definir el  $k$ -esqueleto de  $X$  (con  $k \in \mathbb{N}$ ) denotado como

$$X^k := p \left( \bigsqcup_{\substack{n \leq k \\ \alpha \in I}} \Delta_\alpha^n \right)$$

Dado que cada  $n$ -símplice  $\Delta_\alpha^n$  induce un mapa característico

$$\sigma_{\alpha,n} := p|_{\Delta_\alpha^n} : \Delta_\alpha^n \cong \Delta^n \rightarrow X,$$

lo que permite construir los grupos de cadenas  $\Delta_n(X) \rightarrow C_n(X)$  y como  $\partial_n$  tiene la misma expresión en ambos casos, definen morfismos en homología  $H_n^\Delta(X) \rightarrow H_n(X)$ .

Sea  $A$  un sub- $\Delta$ -complejo de  $X$ , es decir, la imagen por  $p$  de un subconjunto de  $\{\Delta_\alpha^n, \alpha \in I, n \in \mathbb{N}\}$ . Tal como para homología singular, definimos  $\Delta_n(X, A) := \Delta_n(X)/\Delta_n(A)$  y  $H_n^\Delta(X, A)$  su homología asociada, donde la sucesión

$$0 \rightarrow \Delta_\bullet(A) \xrightarrow{i_\bullet} \Delta_\bullet(X) \xrightarrow{\pi_\bullet} \Delta_\bullet(X, A) \rightarrow 0$$

exacta corta, induce una sucesión

$$\dots \rightarrow H_n^\Delta(A) \rightarrow H_n^\Delta(X) \rightarrow H_n^\Delta(X, A) \xrightarrow{\delta_n} H_{n-1}^\Delta(A) \rightarrow \dots$$

exacta larga en homología, y  $\Delta_n(X, A) \rightarrow C_n(X, A)$  induce morfismos  $H_n^\Delta(X, A) \rightarrow H_n(X, A)$  de homología relativa.

Esto establece una correspondencia entre la homología de los *Delta*-complejos y la homología singular, garantizada por el siguiente teorema.

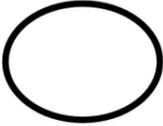
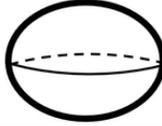
**Teorema 2.2.1.** *Para todo par de  $\Delta$ -complejos  $(X, A)$  se tienen isomorfismos  $H_n^\Delta(X, A) \xrightarrow{\cong} H_n(X, A)$  para todo  $n \in \mathbb{N}$ .*

Una consecuencia importante de esta construcción es que si  $X$  es un espacio topológico homotópicamente equivalente a un  $\Delta$ -complejo con un número finito de  $n$ -símplices para cierto  $n \in \mathbb{N}$ , entonces los grupos de homología  $H_n(X)$  son finitamente generados. Según el teorema de estructura de grupos abelianos finitamente generados, se tiene que

$$H_n(X) \cong \mathbb{Z}^{b_n(X)} \oplus \text{Tors}(H_n(X)).$$

donde  $b_n(X) := \text{rg } H_n(X)$  es el  $n$ -ésimo número de Betti, y  $\text{Tors}(H_n(X))$  denota la parte torsional del grupo. Esto significa que  $b_n(X)$  mide el número de ciclos independientes de dimensión  $n$  en el espacio, mientras que la parte torsional describe

características adicionales que no se pueden capturar mediante los números de Betti. A continuación, en la Figura 2.1 se muestran ejemplos concretos de los números de Betti en dimensiones bajas.

				
$\beta_0$	1	1	1	1
$\beta_1$	0	1	0	2
$\beta_2$	0	0	1	1
$\beta_3$	0	0	0	0
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮

**Figura 2.1:** Números de Betti para un punto, un círculo, una esfera y un toro. Cada fila indica los valores de  $\beta_i$  asociados a las características topológicas de cada espacio.

#### 2.2.4. Diagramas

Un aspecto importante del TDA es la necesidad de cuantificar similitudes entre estructuras topológicas, lo que llevó a [Frosini \(1990\)](#) a definir una distancia para comparar clases de similitud de subvariedades en un espacio euclidiano.

A lo largo del desarrollo del Análisis de Datos Topológicos (TDA), las representaciones visuales de la homología persistente han evolucionado significativamente. Una de las primeras herramientas introducidas fueron los códigos de barras, los cuales ofrecen una representación compacta de los invariantes topológicos. Cada barra representa una característica topológica (como componentes conexas, agujeros o cavidades) y su longitud refleja el intervalo en el cual dicha característica persiste en una filtración. Formalmente, para cada par de persistencia  $(p, q)$ , donde  $p$  es el tiempo de nacimiento y  $q$  el de muerte, se define un intervalo  $[p, q)$ . Este enfoque, popularizado por [Ghrist \(2008\)](#) y previamente desarrollado en trabajos como [Carlsson et al. \(2004\)](#),

es ampliamente utilizado debido a su simplicidad e intuición visual. Sin embargo, su estructura discreta dificulta la integración en contextos estadísticos y de aprendizaje automático.

Posteriormente, se introdujeron los diagramas de persistencia, que extienden los códigos de barras al representar cada par  $(p, q)$  como un punto en el plano cartesiano. En este caso, el eje  $x$  indica el tiempo de nacimiento  $p$  y el eje  $y$  el tiempo de muerte  $q$ . Este enfoque permite un análisis geométrico más detallado, pero presenta limitaciones similares a los códigos de barras en términos de aplicabilidad estadística y computacional (Fugacci et al., 2016).

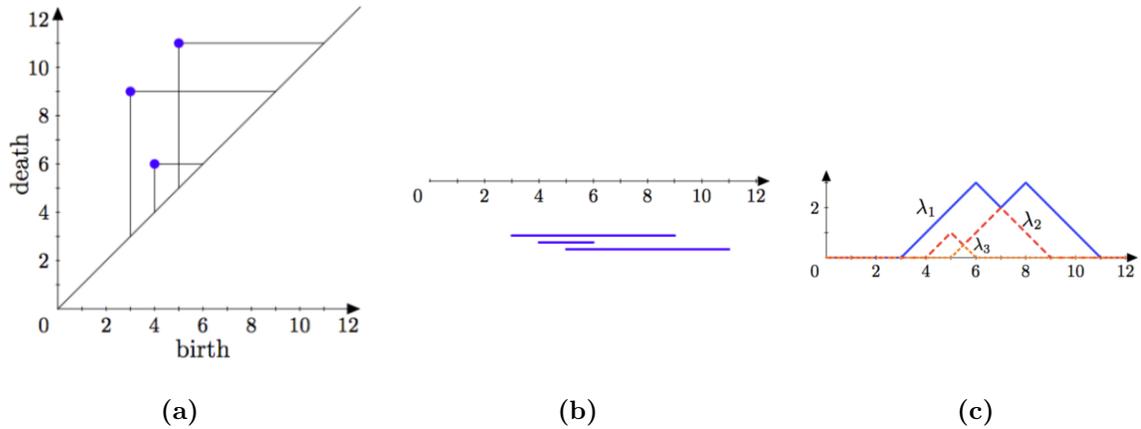
Para abordar estas limitaciones, Bubenik et al. (2015) introdujo los paisajes de persistencia, que ofrecen una representación funcional más adecuada para el análisis estadístico. Formalmente, dado un par  $(p, q)$ , se define una función triangular

$$\lambda(p, q, x) = \max\{0, \min(x - p, q - x)\}, \quad \text{para } x \in \mathbb{R}.$$

A partir de estas funciones triangulares, se construye una familia de funciones  $\lambda_k(x)$ , denominada paisaje de persistencia, donde  $k$  representa el nivel del paisaje. Esta representación permite transformar los datos topológicos en funciones definidas sobre  $\mathbb{R}$ , que son manipulables mediante técnicas estadísticas. Entre sus propiedades destacan:

- Robustez: Al igual que los diagramas de persistencia, los paisajes son estables frente a pequeñas perturbaciones en los datos.
- Representación numérica: Los paisajes permiten representar datos topológicos como vectores o funciones, facilitando su integración en modelos estadísticos.
- Cálculo de métricas: Las distancias entre paisajes, como las normas  $L_p$ , permiten comparar datos de manera eficiente.

A continuación, en la Figura 2.2 se presenta a modo de ejemplo cada uno de los diagramas descritos.



**Figura 2.2:** Visualizaciones de pares de persistencia en grado 1: (a) Diagrama de persistencia, (b) Diagrama de código de barras, y (c) Diagrama de persistencia lanscape.

Para comparar diagramas de persistencia, se han desarrollado métricas específicas como la distancia de Bottleneck y la distancia Wasserstein (Fugacci et al., 2016). La distancia de Bottleneck se define como:

$$d_B(D_1, D_2) = \inf_{\phi} \sup_{x \in D_1} \|x - \phi(x)\|_{\infty},$$

donde  $\phi$  es una correspondencia entre los puntos de los diagramas  $D_1$  y  $D_2$ . Por otro lado, la distancia Wasserstein pondera las diferencias entre los puntos de los diagramas:

$$d_W(D_1, D_2) = \inf_{\phi} \left( \sum_{x \in D_1} \|x - \phi(x)\|^p \right)^{1/p}.$$

Ambas métricas son fundamentales para evaluar la estabilidad de los invariantes topológicos y comparar características persistentes entre conjuntos de datos.

## Capítulo 3

---

# ALGORITMOS

En esta sección se estudiarán los algoritmos implementados para el análisis topológico de datos, describiendo en detalle las librerías utilizadas, las figuras simuladas y las funciones aplicadas. Se presentarán las herramientas computacionales empleadas, justificando la elección de cada una y su relevancia en el contexto del estudio. Además, se explicará el procedimiento seguido para la simulación de las figuras y la extracción de características topológicas, destacando las funciones clave que permitieron el procesamiento y análisis de los datos. Este capítulo proporciona una base metodológica fundamental para la correcta interpretación de los resultados obtenidos en secciones posteriores.

### 3.1. Aplicación del TDA a patrones puntuales

De acuerdo con lo presentado en el Capítulo 2, surge la pregunta de cómo integrar el enfoque del Análisis de Datos Topológicos (TDA) al estudio de patrones puntuales. Para entender esta conexión, primero es necesario reconocer que un patrón puntual se puede modelar como un conjunto de puntos  $\mathcal{P} = \{x_1, x_2, \dots, x_n\}$  distribuidos en un espacio métrico  $(\mathbb{R}^d, d)$ , donde  $d$  denota la dimensión del espacio. Cada punto

en este conjunto representa la localización de un evento, y el objetivo del análisis de patrones puntuales es identificar propiedades espaciales relevantes, como agrupamientos, vacíos o estructuras globales. Aquí es donde el TDA llega a complementar este análisis al proporcionar herramientas para estudiar las propiedades topológicas de estos datos, especialmente aquellas que no son evidentes mediante enfoques estadísticos convencionales, es decir, ayuda a comprender y estudiar la estructura subyacente de los datos.

La clave de la conexión entre patrones puntuales y el TDA radica en que un patrón puntual puede ser visto como una nube de puntos en un espacio métrico, sobre la cual se construyen objetos topológicos mediante complejos simpliciales. Estos complejos permiten modelar cómo los puntos interactúan a diferentes escalas, introduciendo un parámetro de control  $r$  (radio) que regula dichas interacciones.

Por esto el TDA es particularmente adecuado para estudiar patrones puntuales porque se enfoca en características geométricas y topológicas que describen la estructura de un conjunto de puntos. Las características topológicas, como componentes conexas, agujeros y cavidades, están directamente relacionadas con la forma y distribución de los puntos en el espacio. Por ejemplo al trabajar con datos espaciales, los números de Betti dan a entender lo siguiente:

- **Componentes conexas** ( $b_0$ ): Reflejan cómo los puntos forman grupos o están separados en el espacio.
- **Agujeros** ( $b_1$ ): Representan regiones vacías rodeadas por puntos, como vacíos entre agrupamientos.
- **Cavidades** ( $b_2$ ): Aparecen en datos tridimensionales y reflejan volúmenes cerrados sin puntos en su interior.

La combinación de ambas disciplinas comienza transformando el conjunto de puntos  $\mathcal{P}$  en un objeto topológico mediante la construcción de un complejo simplicial. Este enfoque permite analizar las relaciones espaciales entre los puntos, identificando características topológicas significativas a través de diferentes escalas mediante un

parámetro de escala  $r$  (generalmente un radio). A continuación, se detalla el proceso que se ilustra en la Figura 3.1.

1. Representar la interacción entre puntos mediante un complejo simplicial: Dado el conjunto de puntos  $\mathcal{P}$ , se construye un complejo simplicial que modela las conexiones entre puntos.

Los dos complejos más utilizados en este contexto son:

- **Complejo de Čech:**

$$\mathcal{C}_r(\mathcal{P}) = \left\{ \sigma \subseteq \mathcal{P} : \bigcap_{x_i \in \sigma} B(x_i, r) \neq \emptyset \right\},$$

donde  $B(x_i, r)$  es una bola de radio  $r$  centrada en el punto  $x_i$ . Un  $n$ -símplice pertenece al complejo si las bolas asociadas a sus vértices tienen intersección no vacía.

- **Complejo de Vietoris-Rips:**

$$\mathcal{R}_r(\mathcal{P}) = \left\{ \sigma \subseteq \mathcal{P} : \|x_i - x_j\| \leq 2r, \forall x_i, x_j \in \sigma \right\}.$$

En este caso, un  $n$ -símplice se incluye si todos los pares de puntos en sus vértices están conectados por una arista de longitud menor o igual a  $2r$ . Aunque es más eficiente computacionalmente, el complejo de Vietoris-Rips puede sobreestimar ciertas características topológicas.

Ambos complejos forman representaciones discretas del espacio métrico subyacente, permitiendo capturar cómo los puntos interactúan a diferentes escalas.

2. Crear una filtración: A medida que el radio  $r$  varía, el complejo simplicial crece, comenzando con puntos individuales ( $r \rightarrow 0$ ) y aumentando hasta formar una estructura completamente conectada ( $r \rightarrow \infty$ ). Esto genera una filtración:

$$\mathcal{C}_{r_1}(\mathcal{P}) \subseteq \mathcal{C}_{r_2}(\mathcal{P}) \subseteq \dots, \quad \text{con } r_1 < r_2 < \dots.$$

La filtración permite observar cómo las características topológicas (componentes conexas, agujeros, cavidades, etc.) aparecen, persisten y desaparecen a medida que  $r$  aumenta.

3. Calcular los números de Betti: En cada etapa de la filtración, se calculan los números de Betti  $b_n$ . Estos números se obtienen a partir de los grupos de homología:

$$H_n(\mathcal{C}_r(\mathcal{P})) = \ker(\partial_n) / \text{im}(\partial_{n+1}),$$

donde  $\partial_n$  es la función borde definida como:

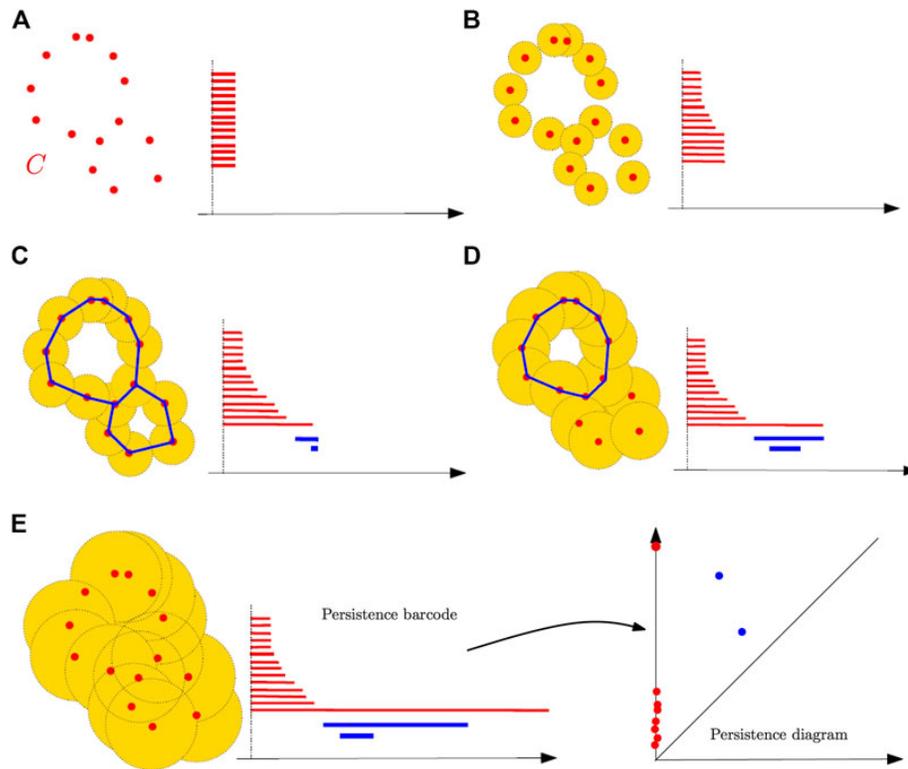
$$\partial_n([v_0, \dots, v_n]) = \sum_{i=0}^n (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_n].$$

4. Visualizar los resultados:

Una vez calculadas las características topológicas a lo largo de la filtración, estas pueden ser representadas de la siguiente forma.

- Código de barras: Cada barra representa una característica topológica, y su longitud indica cuánto persiste a través de las escalas  $r$ .
- Diagrama de persistencia: Cada característica se representa como un punto  $(b, d)$  en el plano, donde  $b$  es el valor de  $r$  en el que la característica aparece (nacimiento) y  $d$  el valor en el que desaparece (muerte).
- Diagrama landscape: Un persistence landscape es una representación funcional de los diagramas de persistencia, construida mediante funciones triangulares asociadas a cada par nacimiento-muerte  $(b, d)$ . Estas funciones se combinan para producir curvas suavizadas organizadas por niveles. Las ventajas de esta representación incluyen

$$\lambda_k(t) = \max \{ \min(t - b_i, d_i - t, 0) : (b_i, d_i) \text{ pertenece al } k\text{-ésimo nivel} \}$$



**Figura 3.1:** Idea intuitiva de la construcción de códigos de barras y diagramas de persistencia a partir de una nube de puntos, ilustrando el concepto de filtración (Imagen obtenida de [Chazal and Michel, 2021](#)).

En términos prácticos, el proceso puede entenderse como un análisis de cómo la nube de puntos  $\mathcal{P}$  interactúa en diferentes escalas. Para  $r$  pequeño, solo se observan puntos individuales y grupos pequeños ( $b_0$ ), mientras que para  $r$  grande, las estructuras más grandes como agujeros ( $b_1$ ) y cavidades ( $b_2$ ) emergen. Las características que persisten en un rango amplio de  $r$  son consideradas las más significativas, ya que reflejan patrones robustos en los datos.

Esta combinación de TDA y patrones puntuales proporciona una perspectiva única, capturando propiedades espaciales que los métodos estadísticos tradicionales no detectan.

### 3.2. Herramientas computacionales

Para implementar esta metodología, se utilizaron diversas librerías del lenguaje de programación **R**, fundamentales para la simulación, el análisis y la visualización de patrones puntuales e invariantes topológicos. A continuación, se describen las principales librerías empleadas y sus funciones clave.

El análisis de patrones espaciales se realizó con la librería **spatstat** (Baddeley and Turner, 2005), que permitió estructurar los datos como patrones puntuales en el plano, definiendo los límites de observación con `owin` y representando los puntos mediante `ppp`. Las funciones `Fest`, `Gest`, `Jest` y `Kest` se usaron para calcular las funciones espaciales  $F$ ,  $G$ ,  $J$  y  $K$ , esenciales para caracterizar las propiedades espaciales de los patrones.

Para las simulaciones, se emplearon las funciones `rpoispp` y `rpois` para generar puntos siguiendo un proceso puntual de Poisson homogéneo, modelando el número de puntos con un parámetro de intensidad  $\lambda$ . Esto introdujo variabilidad en las simulaciones al modelar el número de puntos como una variable estocástica.

El análisis topológico se realizó con la librería **TDA** (Fasy et al., 2024), utilizando `ripsDiag` para calcular diagramas de persistencia mediante complejos de Vietoris-Rips. La librería **TDAstats** (Wadhwa et al., 2018) complementó este análisis con la función `calculate_homology`, que calculó diagramas de persistencia a partir de matrices de distancias, y `plot_barcode` para representar los resultados gráficamente.

Para comparar y clasificar invariantes topológicos, se usó la librería **TDAkit** (You and Yu, 2021), con herramientas como `fsdist` para calcular distancias funcionales entre paisajes y `fskmedoids` para realizar agrupamientos  $k$ -medoids. La librería **tdaunif** (Brunson et al., 2024) fue clave para generar puntos en superficies tridimensionales usando `rejection_sampler` y funciones predefinidas como `sphere` y `tori`. Estas herramientas permitieron integrar análisis espaciales y topológicos de forma

eficiente, optimizando los cálculos y facilitando las visualizaciones.

### 3.3. Metodología

El proceso de simulación se diseñó para generar y analizar patrones puntuales distribuidos sobre diversas figuras geométricas, con el objetivo de evaluar cómo las herramientas del TDA capturan las características geométricas y topológicas de estas estructuras.

La primera etapa consistió en la generación de puntos distribuidos uniformemente sobre figuras tridimensionales como el toro, la esfera, el cilindro, el paraboloides y otras configuraciones geométricas. Para capturar la variabilidad en la cantidad de puntos generados, se utilizó un proceso de Poisson con un parámetro de intensidad definido en función del área de la figura. La función  $rpois$  se empleó para generar valores de  $n$ , lo que permitió incorporar un componente estocástico en las simulaciones.

En una segunda etapa, se introdujeron modificaciones estructurales en las figuras para estudiar casos borde. Se generaron agujeros en superficies como el plano, la esfera, el cilindro y el paraboloides, eliminando puntos dentro de un radio definido para crear regiones vacías. Estas modificaciones permitieron evaluar cómo las características topológicas responden a discontinuidades estructurales.

La tercera etapa consistió en el cálculo de la homología persistente directamente sobre los datos tridimensionales generados. A partir de estos cálculos, se obtuvieron diagramas de persistencia, diagramas de código de barras y paisajes de persistencia. Los diagramas de persistencia permiten visualizar la aparición y desaparición de estructuras topológicas a lo largo de la filtración. Los diagramas de código de barras representan esta información de manera equivalente, pero con un formato que facilita la comparación entre distintas figuras.

Los paisajes de persistencia se obtienen por dimensión y proporcionan una representación funcional de los ciclos persistentes. En figuras tridimensionales como el toro o la esfera, es común que los paisajes en dimensión 2 sean significativos, ya que capturan la presencia de cavidades tridimensionales. Sin embargo, para figuras

bidimensionales como el plano, los paisajes de dimensión 2 carecen de significado, ya que estas figuras no presentan cavidades, por lo que la información relevante se encuentra solo en las dimensiones 0 y 1.

Posteriormente, los puntos generados en tres dimensiones fueron proyectados al plano bidimensional con el objetivo de facilitar comparaciones y analizar diferencias en la estructura de los datos. Para ello, se utilizó Análisis de Componentes Principales (PCA), lo que permitió reducir la dimensionalidad preservando la mayor cantidad de información posible. Esta reducción facilitó la comparación entre TDA y funciones estadísticas tradicionales utilizadas en el análisis de patrones puntuales. Como alternativa, se consideró la selección directa de dos coordenadas, aunque este enfoque podía ocasionar pérdida de información. Los puntos proyectados se limitaron a una ventana de observación uniforme definida mediante la función *owin*, asegurando consistencia en las comparaciones entre figuras y facilitando el análisis en dos dimensiones.

A partir de los patrones puntuales proyectados, se calcularon las funciones espaciales  $F$ ,  $G$ ,  $J$  y  $K$ . Estas funciones caracterizan la estructura espacial de los puntos, describiendo la distribución de distancias entre vecinos más cercanos y la dependencia espacial en distintas escalas. La función  $F$  analiza la distribución de distancias desde un punto aleatorio hasta el punto más cercano, mientras que  $J$  combina  $F$  y  $G$  para proporcionar información adicional sobre las interacciones espaciales. El análisis de estas funciones permitió detectar desviaciones respecto a patrones aleatorios y evaluar la influencia de la estructura geométrica de las figuras en las distribuciones de puntos.

Este proceso metodológico permitió evaluar la capacidad del TDA para diferenciar estructuras geométricas, identificar características topológicas persistentes y comprender mejor la organización espacial de los datos analizados.

# EXPERIMENTOS Y DISCUSIÓN

El objetivo de este capítulo es evaluar la eficacia del Análisis Topológico de Datos (TDA) en la caracterización de patrones puntuales a partir de simulaciones. En particular, buscamos analizar la capacidad del TDA para diferenciar entre distintas figuras geométricas y extraer información estructural relevante a partir de sus características topológicas persistentes. Para ello, generaremos diversas figuras simuladas y aplicaremos herramientas computacionales que nos permitirán explorar su estructura mediante homología persistente y funciones resumen asociadas a patrones puntuales. Las simulaciones se dividen en dos grupos principales: **figuras simples** y **figuras modificadas**.

- En la primera categoría, consideramos figuras geométricas elementales que presentan estructuras bien definidas, como esferas, toros y cilindros.
- En la segunda categoría, analizamos figuras con modificaciones estructurales que introducen nuevas características topológicas, como perforaciones o entrelazamientos.

Esta clasificación nos permitirá estudiar cómo el TDA responde a cambios en la estructura de las figuras y evaluar su capacidad para identificar diferencias topológicas entre ellas.

Las rutinas de simulación fueron desarrolladas en R versión 4.4.1 (Race for Your Life, 2024-06-14) y ejecutadas en dos entornos distintos: un entorno virtual con 2 vCPUs Intel Xeon a 2.2 GHz y 13 GiB de RAM, y un MacBook Pro con chip M1 (arquitectura aarch64-apple-darwin20), equipado con una CPU de 8 núcleos y 16 GiB de RAM.

#### 4.1. Generación de puntos mediante procesos de Poisson

Para ambas categorías de figuras, la generación de puntos se realizó utilizando un **proceso puntual de Poisson homogéneo**. Este proceso permite distribuir puntos aleatoriamente sobre la superficie de cada figura sin interacciones espaciales. La cantidad total de puntos,  $N$ , se determinó mediante la función `rpois(1,  $\lambda$ )`, donde  $\lambda$  representa la intensidad del proceso y se calcula como:

$$\lambda = \text{Área} \times \text{Densidad esperada de puntos.}$$

Al variar  $\lambda$ , se controla la densidad de puntos en la figura y, en consecuencia, la precisión en la representación de su estructura topológica.

#### 4.2. Simulaciones de figuras simples

El propósito de esta sección es evaluar el desempeño del TDA en la identificación de estructuras básicas en figuras geométricas bien definidas. Estas figuras servirán como referencia para comparar posteriormente con versiones modificadas.

Las figuras simuladas en esta categoría incluyen:

- **Toro**
- **Esfera**
- **Cono**

- **Cilindro**
- **Paraboloide Positivo**
- **Paraboloide Negativo**
- **Plano**

Cada una de estas figuras se parametrizó matemáticamente para describir su estructura en el espacio tridimensional. La generación de puntos sobre sus superficies se realizó utilizando el jacobiano de la transformación, lo que garantiza una distribución uniforme. Para este proceso, empleamos la función `make_rejection_sampler` de la librería **TDAunif**, basada en el método de aceptación y rechazo.

#### 4.2.1. Toro

En el caso del toro, su parametrización en coordenadas tridimensionales está dada por:

$$\begin{aligned}x &= (R + r \cos \theta) \cos \phi, \\y &= (R + r \cos \theta) \sin \phi, \\z &= r \sin \theta,\end{aligned}$$

donde  $R$  es el radio mayor del toro,  $r$  es el radio menor,  $\theta \in [0, 2\pi]$ , y  $\phi \in [0, 2\pi]$ . Esta parametrización describe cómo se mapea el toro en el espacio tridimensional. El jacobiano asociado a esta parametrización, que asegura una distribución uniforme de puntos en la superficie, está dado por:

$$J(\theta, \phi) = r(R + r \cos \theta).$$

La herramienta `make_rejection_sampler` utiliza estas ecuaciones para generar puntos distribuidos uniformemente sobre la superficie del toro.

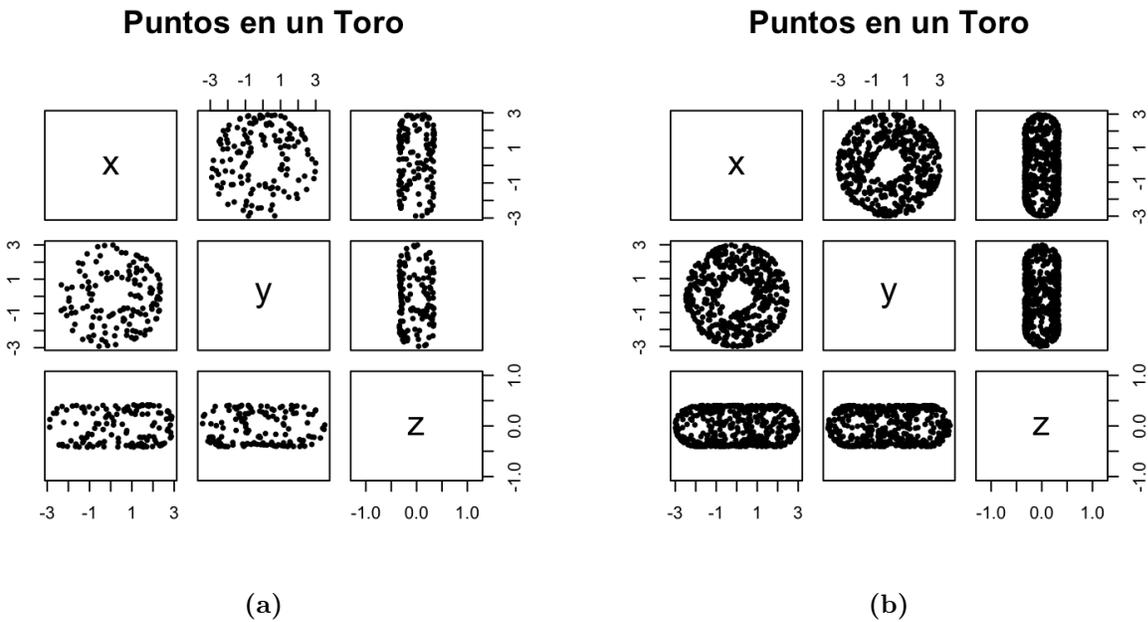
El área teórica de un toro es:

$$\text{Área} = 4\pi^2 Rr,$$

donde para esta simulación se consideraron  $R = 2$  y  $r = 1$ , resultando en un área aproximada de 238.87.

Para garantizar una representación adecuada de la figura, se exploraron diferentes valores del parámetro  $\lambda$ , que controla la densidad de puntos generados en la simulación. Un valor bajo de  $\lambda$  implica una menor cantidad de puntos, lo que puede dificultar la identificación de las características topológicas fundamentales de la figura, mientras que valores más altos permiten una mayor precisión en la reconstrucción geométrica. En particular, se probó con  $\lambda = 150$ , lo que resultó en un número de puntos suficiente para representar la estructura general del toro y detectar sus principales características topológicas, como los dos ciclos persistentes. Sin embargo, debido a la menor densidad de puntos, algunas regiones de la figura pueden no estar tan bien definidas, lo que podría afectar la interpretación visual y la estabilidad de los invariantes topológicos. En contraste, al aumentar  $\lambda$  a 500, la distribución de puntos se volvió más densa, proporcionando una representación más detallada de la figura y facilitando la identificación precisa de sus ciclos persistentes en los diagramas de persistencia.

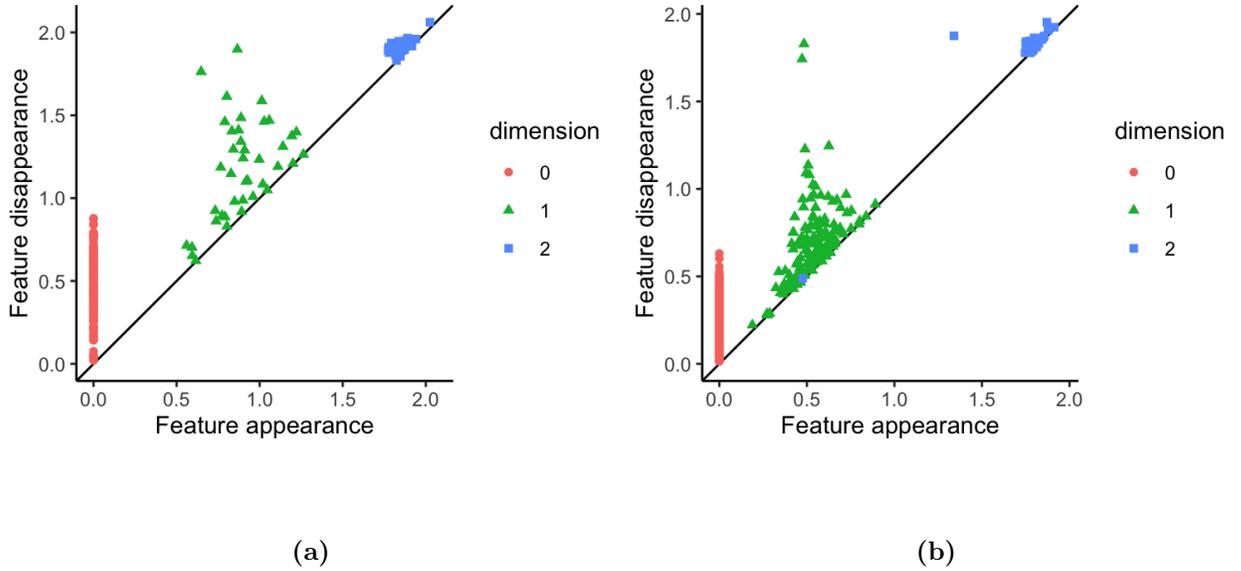
A pesar de que un valor mayor de  $\lambda$  mejora la calidad de la representación y la estabilidad de los resultados,  $\lambda = 150$  sigue siendo una elección adecuada, ya que permite capturar las propiedades topológicas esenciales del toro sin incurrir en un costo computacional excesivo. Por lo tanto, para aplicaciones donde la precisión geométrica sea crucial, un mayor número de puntos puede ser recomendable, pero en términos de análisis topológico, la cantidad de puntos generada con  $\lambda = 150$  es suficiente para extraer la información clave de la estructura.



**Figura 4.1:** (a) Toro simulado usando  $\lambda = 150$ . (b) Toro simulado usando  $\lambda = 500$ .

La elección del valor de  $\lambda$  no solo impacta la representación geométrica de la figura, sino también la precisión del análisis topológico. Un  $\lambda$  más bajo puede introducir ruido en los diagramas de persistencia, mientras que valores más altos permiten una identificación clara de las características topológicas, como las componentes conexas y los ciclos.

En la siguiente figura se muestra los diagramas de persistencia (Figura 4.2), que permite ilustrar cómo estas diferencias en  $\lambda$  afectan la interpretación topológica de los datos simulados.



**Figura 4.2:** (a) Diagrama de persistencia para el toro simulado usando  $\lambda = 150$ . (b) Diagrama de persistencia para el toro simulado usando  $\lambda = 500$ .

La Figura 4.2 muestra los diagramas de persistencia calculados para dos simulaciones de un toro tridimensional con diferentes valores de  $\lambda$ : (a)  $\lambda = 150$  y (b)  $\lambda = 500$ . Estos diagramas representan las características topológicas identificadas en las simulaciones, clasificadas por su dimensión.

- **Dimensión 0:** Representan componentes conexas. En ambos diagramas, se observa una componente conexa dominante que persiste durante toda la filtración, mientras que las demás componentes desaparecen rápidamente.
- **Dimensión 1:** Corresponden a ciclos. En el caso de  $\lambda = 150$ , los ciclos principales del toro son visibles, aunque aparecen más dispersos debido a la menor densidad de puntos. Para  $\lambda = 500$ , los dos ciclos principales son más evidentes, reflejando la estructura topológica característica del toro.
- **Dimensión 2:** Reflejan cavidades tridimensionales. En ambos diagramas, los

puntos cercanos a la diagonal representan cavidades que aparecen y desaparecen rápidamente, como se esperaría en un toro tridimensional.

En general, ambos diagramas son consistentes con las características topológicas esperadas de un toro, mostrando dos ciclos persistentes y una componente conexa dominante. Sin embargo, el diagrama asociado a  $\lambda = 500$  refleja con mayor claridad la estructura del toro, debido a la mayor densidad de puntos, lo que permite una mejor resolución de las características topológicas. No obstante, debido a limitaciones de capacidad computacional, en las simulaciones de las siguientes figuras se trabajará con un máximo de  $\lambda = 150$ . Esta decisión busca optimizar los tiempos de procesamiento y evitar problemas de rendimiento, manteniendo un equilibrio entre la representación adecuada de la estructura topológica y la viabilidad computacional del análisis.

#### 4.2.2. Esfera

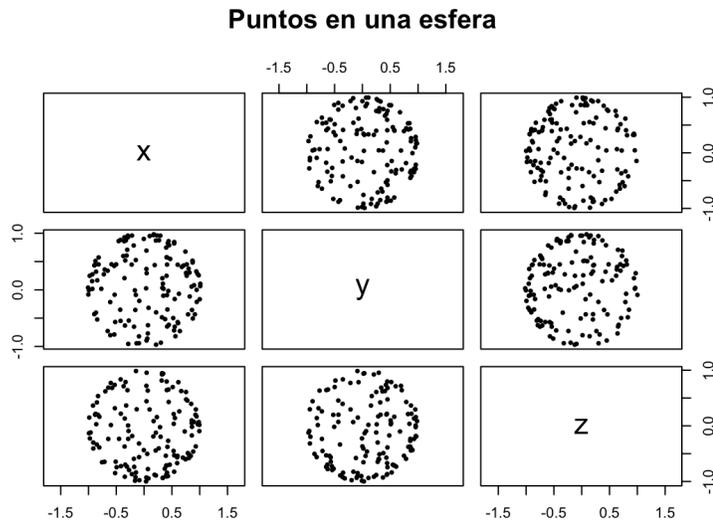
La simulación de la esfera fue realizada usando una parametrización tridimensional estándar que permite distribuir puntos de manera uniforme en su superficie, lo que garantiza una representación precisa de la geometría de la esfera, con un número total de puntos determinado mediante un proceso puntual de Poisson homogéneo. La parametrización usada es la siguiente:

$$\begin{aligned}x &= r \sin \theta \cos \phi, \\y &= r \sin \theta \sin \phi, \\z &= r \cos \theta,\end{aligned}$$

donde  $r$  es el radio de la esfera, y el jacobiano de la esfera está dado por:

$$J(\theta, \phi) = r^2 \sin \theta.$$

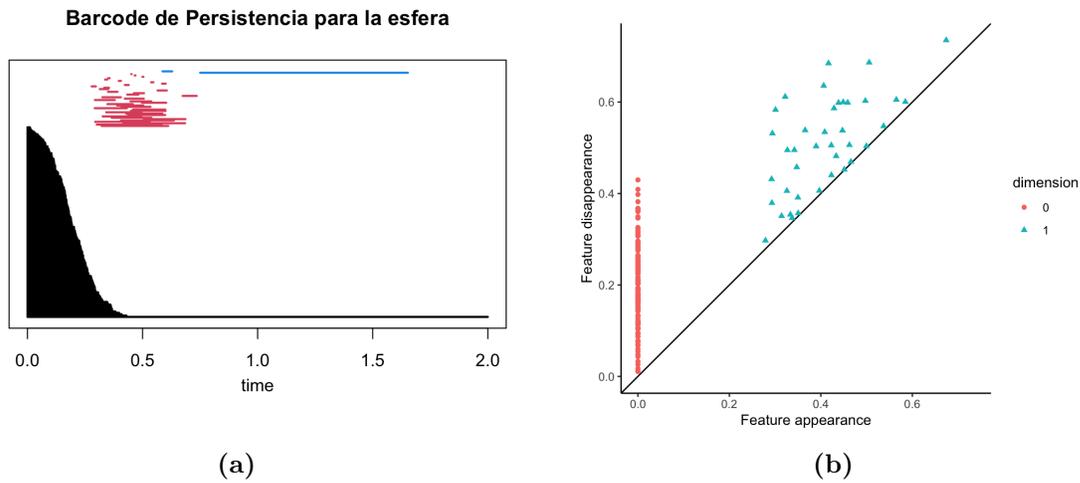
A continuación, se presenta en la Figura 4.3 la esfera simulada.



**Figura 4.3:** Proyecciones de los puntos generados en la esfera simulada con  $\lambda = 150$ .

Luego, en los diagramas de persistencia (Figura 4.4 (a)) y el diagrama de código de barras (Figura 4.4 (b)), se pueden observar las características topológicas identificadas en la esfera.

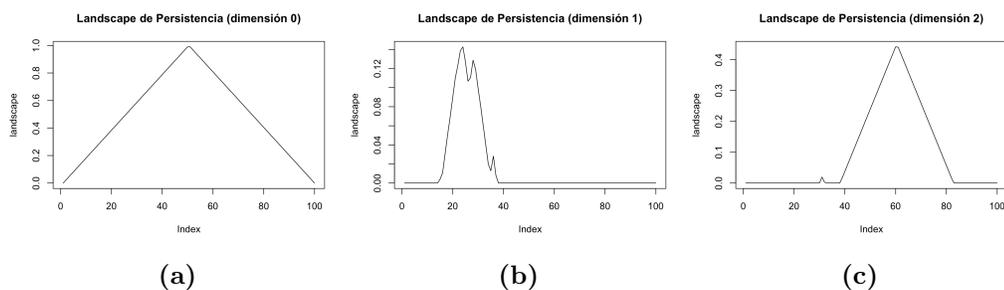
- **Dimensión 0:** Se aprecia que hay una única componente conexa que persiste durante toda la filtración.
- **Dimensión 1:** Los ciclos desaparecen rápidamente, lo cual es consistente con la topología de una esfera, ya que no presenta agujeros persistentes en dimensión 1.
- **Dimensión 2:** En este caso, la cavidad principal de la esfera es altamente persistente.



**Figura 4.4:** (a) Diagrama de código de barras. (b) Diagrama de persistencia.

En los paisajes de persistencia (Figuras 4.5), se representan de forma funcional las características topológicas observadas en cada dimensión:

- **Dimensión 0:** Refleja la única componente conexa persistente.
- **Dimensión 1:** Muestra picos bajos asociados a ciclos que desaparecen rápidamente.
- **Dimensión 2:** Presenta un pico predominante, correspondiente a la cavidad tridimensional persistente de la esfera.



**Figura 4.5:** Paisajes de persistencia para la esfera simulada. Se presentan los resultados para distintas dimensiones: (a) Dimensión 0, (b) Dimensión 1 y (c) Dimensión 2.

El análisis de la esfera confirma la coherencia entre su estructura geométrica y las características topológicas identificadas mediante TDA. La persistencia en dimensión 0 refleja la conectividad global de la figura, mientras que la rápida desaparición de los ciclos en dimensión 1 es consistente con la ausencia de agujeros en la esfera. En dimensión 2, la cavidad principal persiste a lo largo de la filtración, lo que concuerda con la naturaleza tridimensional cerrada de la esfera. Estos resultados validan la capacidad del TDA para capturar la estructura esperada en figuras sin agujeros en dimensión 1, demostrando su efectividad en la identificación de características topológicas clave en superficies cerradas.

### 4.2.3. Cono

En esta simulación se analiza el comportamiento topológico de un cono, generado mediante la parametrización en coordenadas polares. La parametrización utilizada se define como:

$$\begin{aligned}x &= r \cos(\theta), \\y &= r \sin(\theta), \\z &= h - r,\end{aligned}$$

donde  $r$  representa el radio de la base del cono,  $\theta \in [0, 2\pi]$  es el ángulo polar, y  $h = 2$  corresponde a la altura del cono. En esta formulación,  $z$  disminuye a medida que  $r$  aumenta, generando la forma cónica deseada.

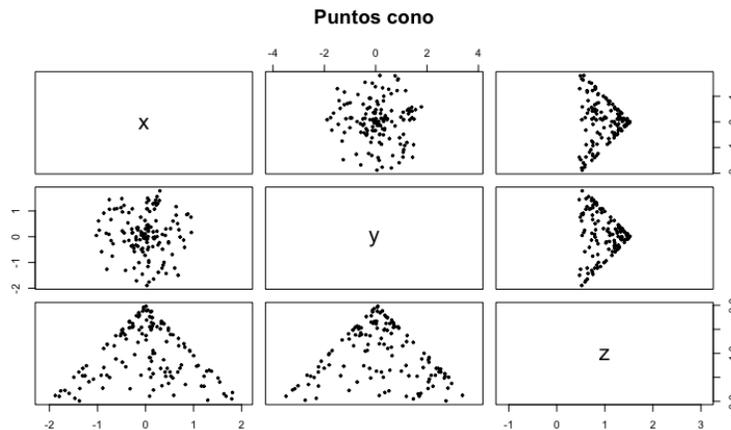
El jacobiano correspondiente a esta parametrización es:

$$J(r, \theta) = r,$$

lo cual refleja la dependencia del área diferencial con respecto al radio en coordenadas polares.

Para la simulación de puntos sobre la superficie del cono, se utilizó el método de aceptación y rechazo mediante la función `make_rejection_sampler`, estableciendo los parámetros de  $r$  en el intervalo  $[0, 2]$  y de  $\theta$  en  $[0, 2\pi]$ .

A continuación, se presenta la visualización de los puntos generados en el cono (Figura 4.6).



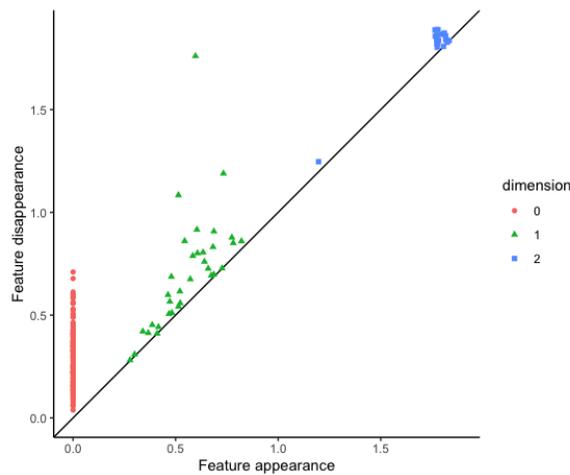
**Figura 4.6:** Distribución de puntos simulados en la superficie de un cono.

Posteriormente, se calculó la homología persistente hasta dimensión 2. En el diagrama de persistencia (Figura 4.7), se identifican las siguientes características topológicas:

- **Dimensión 0** : Representadas por puntos rojos, estas indican la presencia

de varias componentes conexas que eventualmente se fusionan en una única componente principal a medida que avanza la filtración.

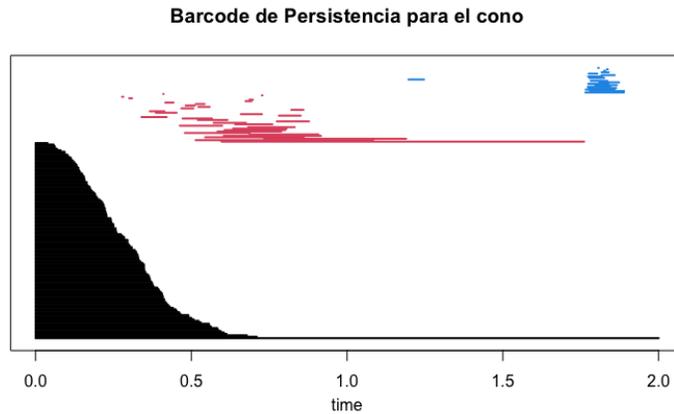
- **Dimensión 1** : Representados por triángulos verdes, estos ciclos reflejan la presencia de agujeros temporales formados por la disposición de los puntos en la superficie del cono. La intensidad de estos ciclos es baja debido a que la estructura del cono no presenta ciclos persistentes de forma natural, aunque algunos puedan aparecer por la distribución de los puntos.
- **Dimensión 2** : Representadas por cuadrados azules, estas características corresponden a cavidades tridimensionales que surgen brevemente durante la filtración. Como el cono es una figura tridimensional, la presencia de estas cavidades es esperable, aunque su persistencia varía dependiendo de la densidad de los puntos simulados.



**Figura 4.7:** Diagrama de persistencia para el cono simulado.

Además, se presenta el código de barras de persistencia (Figura 4.8), que ofrece una representación visual de la duración de las características topológicas a lo largo de la filtración. Las barras más largas indican estructuras que persisten por más tiempo y, por lo tanto, son más relevantes topológicamente. En el caso del cono, se observa

que las barras de dimensión 2 son las más persistentes, lo cual refleja la presencia de una cavidad tridimensional asociada a la estructura cónica.



**Figura 4.8:** Código de barras de persistencia para el cono simulado.

Finalmente, los paisajes de persistencia (Figura 4.9) permiten una interpretación funcional de las características topológicas en cada dimensión:

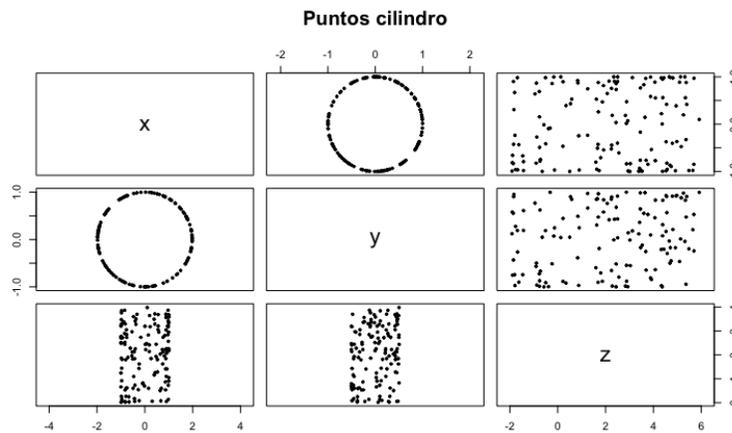
- **Dimensión 0:** Refleja la aparición y fusión de componentes conexas. Se observa un único pico dominante, lo que indica que la mayoría de las componentes se fusionan rápidamente.
- **Dimensión 1:** Muestra la presencia de ciclos temporales que surgen debido a la disposición de los puntos. La baja intensidad de estos ciclos es consistente con la topología del cono, que no presenta agujeros persistentes de forma natural.
- **Dimensión 2:** Destaca cavidades tridimensionales de corta duración, típicas de estructuras cónicas. En este caso, la presencia de un pico indica la existencia de una cavidad significativa que persiste durante un intervalo de la filtración.



donde  $r = 1$  corresponde al radio del cilindro,  $\theta \in [0, 2\pi]$  es el ángulo polar y  $z$  varía entre 0 y 4 para definir la altura del cilindro. El jacobiano de la parametrización es constante, dado por:

$$J(z, \theta) = r = 1.$$

Para la generación de puntos, se utilizó la función `make_rejection_sampler`, la cual emplea el método de aceptación y rechazo para obtener una distribución uniforme de puntos en la superficie del cilindro.

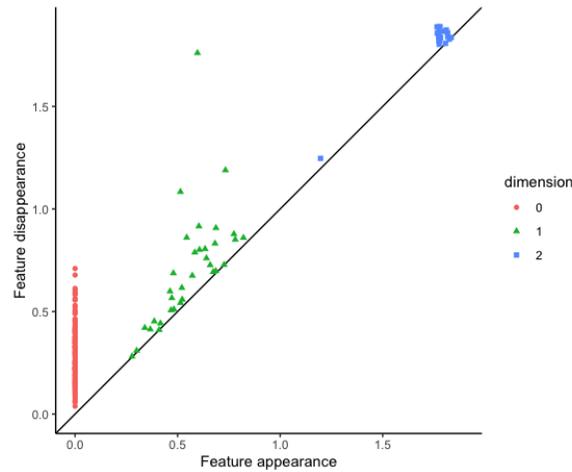


**Figura 4.10:** Proyecciones de los puntos generados en el cilindro simulado con  $\lambda = 150$ .

Posteriormente, se calculó la homología persistente utilizando la función `calculate_homology`, con una dimensión máxima de 2. En el diagrama de persistencia (Figura 4.11), se identifican características topológicas relevantes:

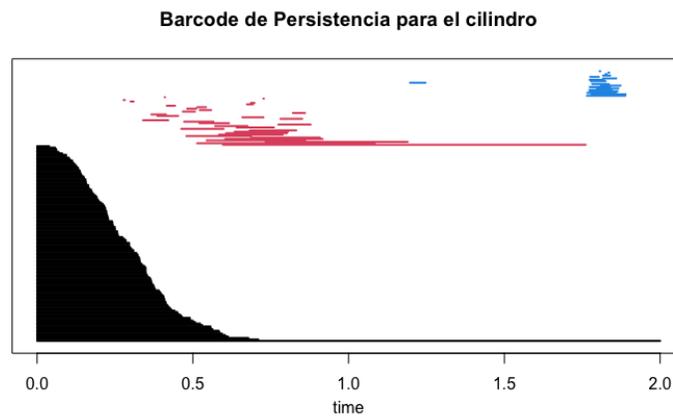
- **Dimensión 0:** Se observa una única componente conexa que persiste durante toda la filtración.
- **Dimensión 1:** Se identifican ciclos persistentes que reflejan la estructura tubular del cilindro.

- **Dimensión 2:** Se observan cavidades tridimensionales, aunque menos predominantes.



**Figura 4.11:** Diagrama de persistencia para el cilindro simulado.

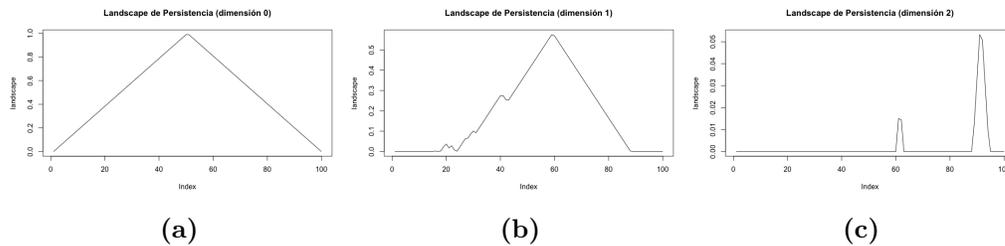
El análisis de persistencia se complementa con el diagrama de código de barras (Figura 4.12), que ilustra la duración de las características topológicas a lo largo de la filtración.



**Figura 4.12:** Diagrama de código de barras para el cilindro simulado.

Finalmente, los paisajes de persistencia (Figura 4.13) permiten visualizar de forma funcional las características topológicas en cada dimensión:

- **Dimensión 0:** Refleja la componente conexa persistente.
- **Dimensión 1:** Muestra picos asociados a ciclos que representan la estructura del cilindro.
- **Dimensión 2:** Indica cavidades tridimensionales de menor persistencia.



**Figura 4.13:** Paisajes de persistencia para el cilindro simulado. Se presentan los resultados para distintas dimensiones: (a) Dimensión 0, (b) Dimensión 1 y (c) Dimensión 2.

El análisis topológico del cilindro revela que su estructura presenta características distintivas en dimensión 1, con ciclos persistentes que reflejan su forma tubular. A diferencia del cono, donde los ciclos en dimensión 1 son de corta duración, el cilindro muestra una persistencia más estable en esta dimensión, lo que es coherente con su estructura cilíndrica sin puntos de convergencia. En dimensión 2, las cavidades tridimensionales tienen menor relevancia en comparación con figuras cerradas como la esfera, lo que sugiere que el TDA identifica correctamente la falta de una estructura completamente cerrada en esta geometría. Este resultado confirma que el TDA logra capturar la forma característica del cilindro y diferenciarlo de otras superficies con propiedades topológicas similares pero con diferencias estructurales clave.

#### 4.2.5. Paraboloide Positivo (sin agujero)

En esta simulación se generaron puntos sobre un paraboloide positivo, utilizando un proceso de Poisson con parámetro  $\lambda = 150$  para definir la cantidad de puntos:

$$n \sim \text{Poisson}(\lambda), \quad \lambda = 150.$$

La parametrización utilizada para el paraboloide positivo está definida por:

$$\phi(x, y) = (x, y, z), \quad \text{donde } z = x^2 + y^2.$$

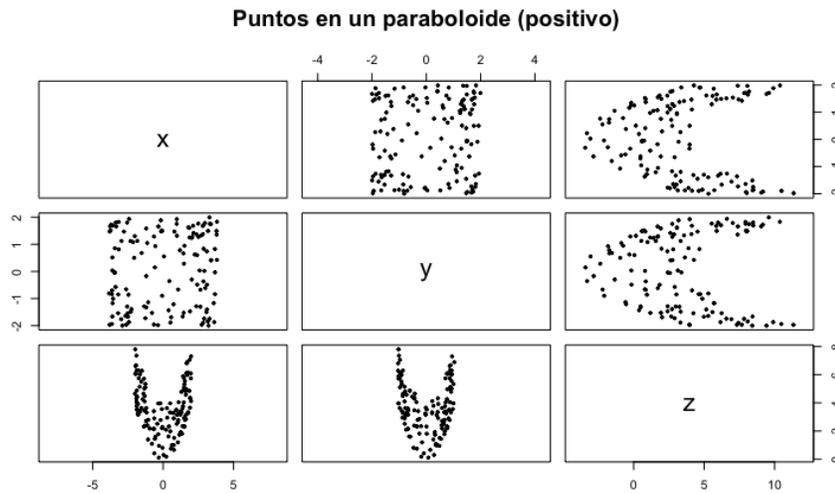
Esta relación describe la superficie de un paraboloide orientado en dirección positiva del eje  $z$ , donde el vértice se encuentra en el origen.

El jacobiano de la transformación es:

$$J(x, y) = 1 + 4x^2 + 4y^2,$$

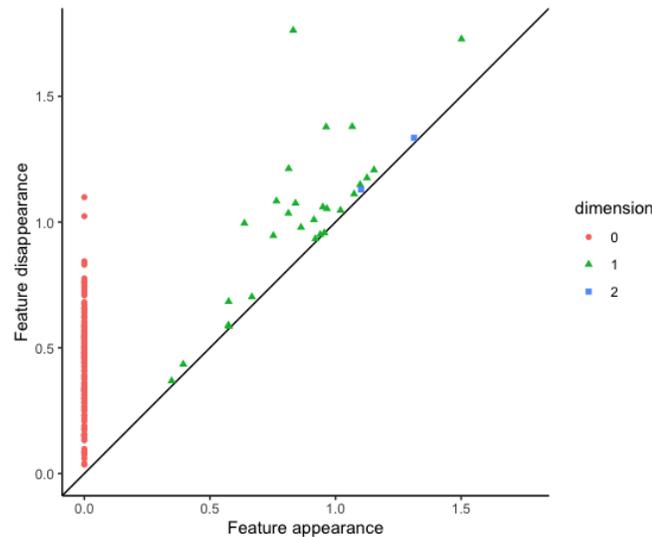
lo que refleja la variación en la densidad de los puntos sobre la superficie. Se utilizó un valor máximo para el jacobiano de 33, correspondiente a los límites definidos para los parámetros.

La generación de los puntos se realizó utilizando el método de aceptación y rechazo, implementado a través de la función `make_rejection_sampler` del paquete `tdaunif`. La distribución final de los puntos se muestra en la Figura [4.14](#).



**Figura 4.14:** Puntos simulados en un paraboloides positivo.

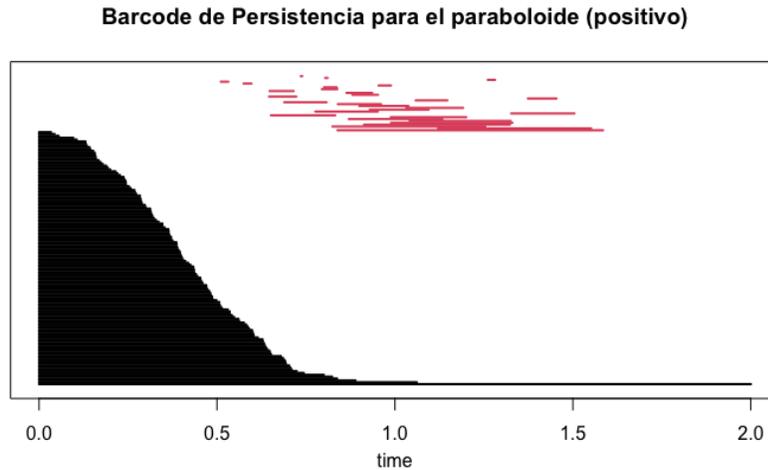
En la Figura 4.14, se observa la disposición característica de los puntos que forman la superficie curva del paraboloides. La densidad de puntos aumenta a medida que nos acercamos al vértice, debido a la influencia del jacobiano en la distribución. Posteriormente, se aplicó el análisis de homología persistente utilizando la función `ripsDiag` del paquete **TDA**, junto con la función `calculate_homology` del paquete **TDAstats**, para identificar las características topológicas relevantes.



**Figura 4.15:** Diagrama de persistencia para el paraboloido positivo.

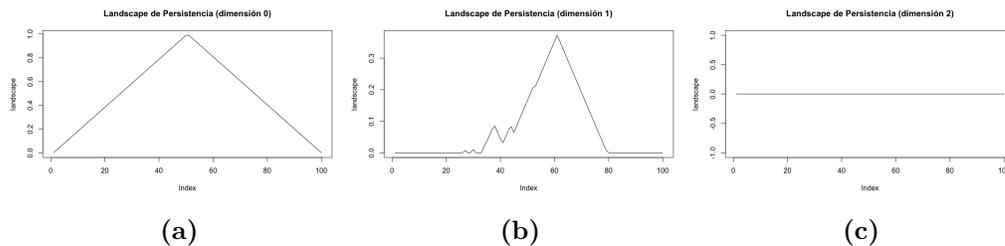
En el diagrama de persistencia de la Figura 4.15, se identifican las siguientes características topológicas:

- **Dimensión 0:** Representa las componentes conexas. Se observa una única componente conexa dominante que persiste a lo largo de toda la filtración, lo cual es consistente con la estructura continua del paraboloido.
- **Dimensión 1:** Corresponde a ciclos o bucles. Aunque el paraboloido no tiene agujeros evidentes, se observan ciclos de corta duración que representan estructuras locales formadas por la curvatura de la superficie. Estos ciclos emergen debido a la disposición irregular de los puntos en algunas regiones, generando conexiones temporales durante la filtración.
- **Dimensión 2:** Refleja la presencia de cavidades tridimensionales de corta duración. Estas cavidades son artefactos del proceso de filtración, causados por la forma curva del paraboloido y la distribución de puntos, y no representan cavidades significativas en la topología global de la superficie.



**Figura 4.16:** Código de barras de persistencia para el paraboloido positivo.

El código de barras de persistencia (Figura 4.16) muestra la duración de las características topológicas. Se observan barras largas en dimensión 0, lo que indica la persistencia de la conectividad global. En dimensión 1, las barras son más cortas, reflejando la presencia de ciclos temporales.



**Figura 4.17:** Paisajes de persistencia para el paraboloido positivo. Se presentan los resultados para distintas dimensiones: (a) Dimensión 0, (b) Dimensión 1 y (c) Dimensión 2.

En los paisajes de persistencia (Figura 4.17):

- **Dimensión 0:** Muestra un pico alto, reflejando la persistencia de la componente conexa principal.

- **Dimensión 1:** Presenta picos pequeños, indicando la presencia de ciclos de corta duración asociados a la curvatura de la superficie.
- **Dimensión 2:** Apenas se observan valores significativos, lo que coincide con la falta de cavidades tridimensionales persistentes en la topología del paraboloido.

El análisis topológico del paraboloido positivo revela que su principal característica es la fuerte conectividad global, evidenciada por la persistencia en dimensión 0. A diferencia de figuras como el toro o el cilindro, donde existen ciclos significativos en dimensión 1, en el paraboloido los ciclos detectados son efímeros y producto de la distribución local de los puntos. La ausencia de cavidades tridimensionales persistentes en dimensión 2 es consistente con su estructura abierta. En general, estos resultados muestran que el TDA logra capturar la curvatura y la continuidad del paraboloido sin identificar falsos invariantes topológicos, lo que confirma la robustez del método en el análisis de este tipo de superficies.

#### 4.2.6. Paraboloido Negativo (sin agujero)

Para la simulación del paraboloido negativo, se empleó una parametrización que describe la curvatura invertida de la superficie. La parametrización utilizada es la siguiente:

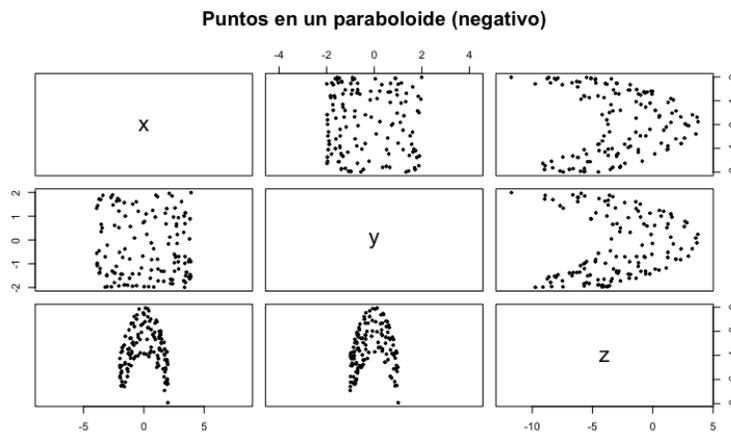
$$\begin{aligned}x &= x, \\y &= y, \\z &= -(x^2 + y^2),\end{aligned}$$

donde  $x, y \in [-2, 2]$  y  $z$  representa la curvatura negativa del paraboloido. El jacobiano correspondiente a esta parametrización es:

$$J(x, y) = 1 + 4x^2 + 4y^2.$$

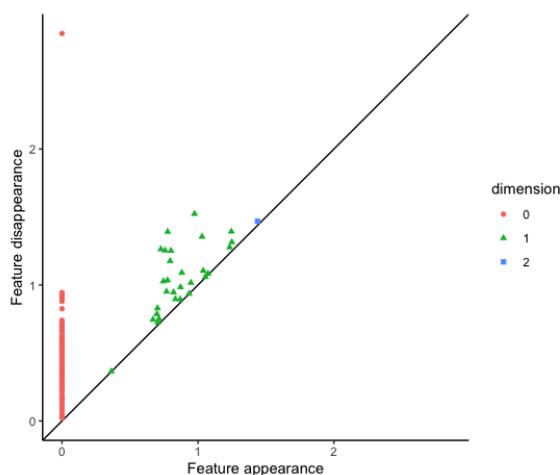
Para la generación de los puntos se utilizó la función `make_rejection_sampler`, estableciendo la parametrización `paraboloide_neg_parametrizacion`, que define la superficie del paraboloide negativo, y el jacobiano `paraboloide_neg_jacobiano`, que garantiza una distribución uniforme de los puntos. Los parámetros de simulación incluyen el intervalo de  $x, y \in [-2, 2]$ , estableciendo los límites de la simulación, y un jacobiano máximo de 33 para normalizar la densidad de puntos.

La Figura 4.18 muestra la distribución de los puntos generados sobre el paraboloide negativo.



**Figura 4.18:** Distribución de puntos en un paraboloide negativo.

Posteriormente, se calculó la homología persistente utilizando la función `ripsDiag`, obteniendo el diagrama de persistencia mostrado en la Figura 4.19.

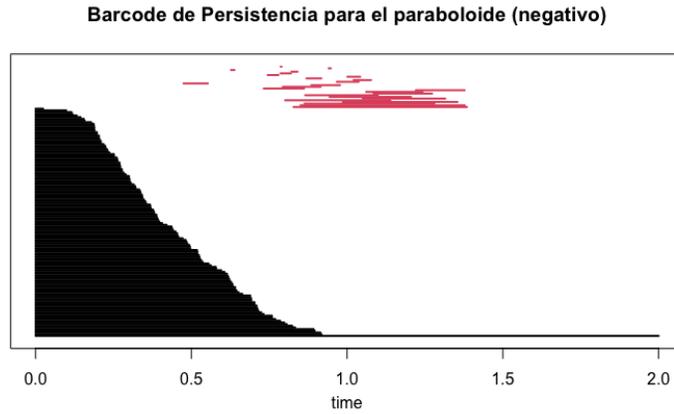


**Figura 4.19:** Diagrama de persistencia para el paraboloides negativo.

En el diagrama de persistencia se observan las siguientes características topológicas:

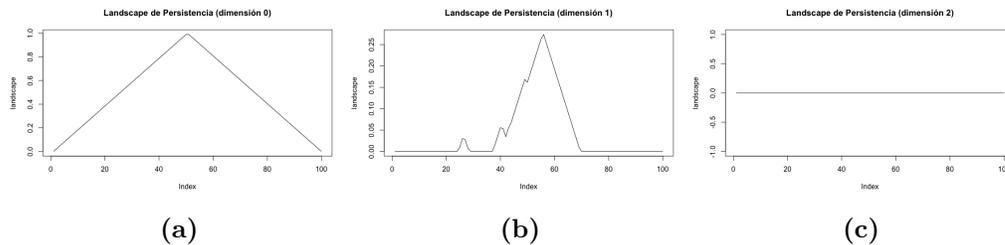
- **Dimensión 0 :** Representa las componentes conexas. Se observa una única componente conexa dominante que persiste a lo largo de toda la filtración, lo cual indica que el conjunto de puntos está conectado globalmente.
- **Dimensión 1 :** Corresponde a ciclos o bucles. Aunque el paraboloides negativo no presenta ciclos evidentes, se observan algunos de corta duración que emergen debido a la disposición de los puntos y la curvatura de la superficie.
- **Dimensión 2 :** Refleja la presencia de cavidades tridimensionales transitorias, aunque estas son de corta duración, lo cual es esperable dado que el paraboloides negativo no presenta cavidades internas significativas en su topología.

Estos resultados muestran que el paraboloides negativo presenta una estructura topológica simple, dominada por una única componente conexa y ciclos de corta duración. Las cavidades observadas son artefactos del proceso de filtración y no reflejan características topológicas significativas de la superficie.



**Figura 4.20:** Código de barras de persistencia para el paraboloido negativo.

En el código de barras de persistencia (Figura 4.20), se observa una barra larga correspondiente a la componente conexa persistente, mientras que las barras más cortas representan ciclos de corta duración. La falta de barras largas en dimensiones superiores refuerza la simplicidad topológica del paraboloido negativo.



**Figura 4.21:** Paisajes de persistencia para el paraboloido negativo. Se presentan los resultados para distintas dimensiones: (a) Dimensión 0, (b) Dimensión 1 y (c) Dimensión 2.

En los paisajes de persistencia (Figura 4.21), se representan de forma funcional las características topológicas observadas en cada dimensión:

- **Dimensión 0:** Refleja la única componente conexa persistente.
- **Dimensión 1:** Muestra picos bajos asociados a ciclos de corta duración.

- **Dimensión 2:** Presenta picos esporádicos que corresponden a cavidades tridimensionales transitorias.

Al igual que en el caso del paraboloides positivo, los resultados obtenidos para el paraboloides negativo reflejan una estructura topológica simple, dominada por su conectividad global y la ausencia de cavidades tridimensionales persistentes. La presencia de ciclos de corta duración en dimensión 1 es consecuencia de la curvatura de la superficie y la disposición de los puntos, pero no representan características topológicas significativas. En dimensión 2, los paisajes muestran la falta de cavidades persistentes, lo que refuerza la idea de que el paraboloides negativo no contiene estructuras internas cerradas. En general, estos resultados confirman que el TDA captura correctamente la geometría del paraboloides negativo y permite diferenciarlo de otras figuras con características topológicas más complejas.

#### 4.2.7. Plano

Para la simulación de un plano, se generaron puntos distribuidos uniformemente en una superficie bidimensional. El número de puntos se obtuvo a partir de un proceso de Poisson con parámetro  $\lambda = 150$ :

$$n \sim \text{Poisson}(\lambda), \quad \lambda = 150.$$

Este enfoque introduce variabilidad en la cantidad de puntos simulados, lo que facilita el análisis de cómo dicha variabilidad afecta la identificación de características topológicas mediante homología persistente.

La parametrización del plano se definió de manera simple para reflejar su naturaleza bidimensional en un espacio tridimensional. La función utilizada para la parametrización fue:

$$\phi(x, y) = (x, y, 0),$$

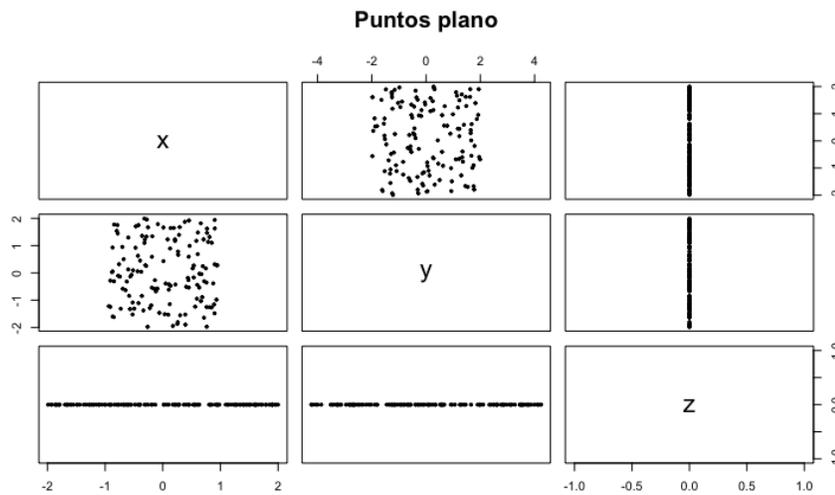
donde  $z = 0$  para todos los puntos, ya que la figura es plana. Esta parametrización mapea directamente los puntos en el plano  $xy$ .

El jacobiano asociado es constante, dado que no existe una transformación que altere

la densidad de los puntos sobre la superficie. Por lo tanto:

$$J(x, y) = 1.$$

Para la generación de los puntos, se utilizó la función `make_rejection_sampler` del paquete `tdaunif`. Los parámetros utilizados incluyeron la parametrización definida previamente y el jacobiano constante. Se establecieron límites para los parámetros  $x$  e  $y$  en el rango de  $-2$  a  $2$ , lo que definió el área de observación del plano. Además, se fijó el jacobiano máximo en  $1$ , dado que la densidad de los puntos es uniforme en toda la superficie.

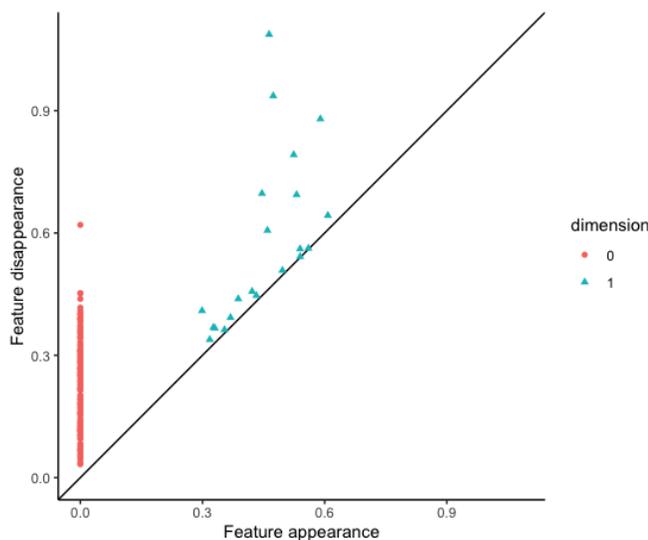


**Figura 4.22:** Puntos simulados en un plano.

En la Figura 4.22, se muestra la distribución de los puntos simulados. Dado que se trata de un plano, los puntos están confinados en una superficie bidimensional, lo cual se refleja en la proyección de las coordenadas  $z = 0$ , indicando que la variación ocurre únicamente en los ejes  $x$  e  $y$ . Las proyecciones permiten observar la distribución espacial de los puntos, destacando la uniformidad en el plano.

Posteriormente, se aplicó el análisis de homología persistente utilizando la función `ripsDiag` del paquete `TDA`, configurada para detectar características topológicas

hasta dimensión 1, dado que en un plano no se esperan cavidades tridimensionales. Además, se utilizó la función `calculate_homology` del paquete **TDAstats** para obtener la homología persistente a partir de la matriz de coordenadas de los puntos.



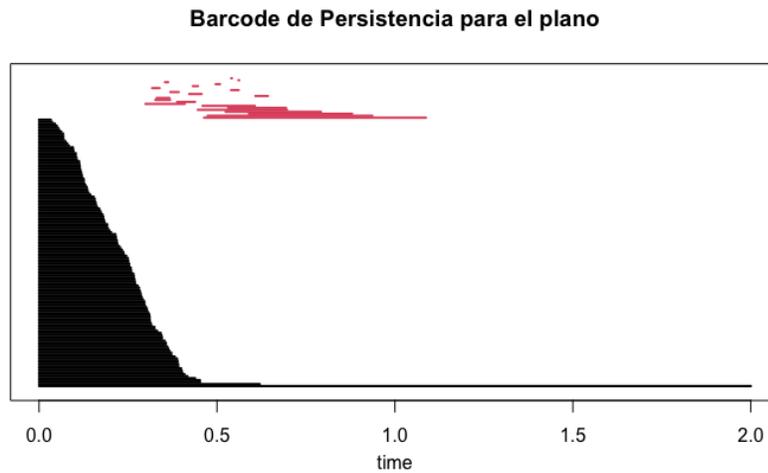
**Figura 4.23:** Diagrama de persistencia para el plano.

La Figura 4.23 presenta el diagrama de persistencia obtenido, donde cada punto representa una característica topológica. El eje horizontal indica el momento de aparición (nacimiento) de la característica, mientras que el eje vertical indica su desaparición (muerte).

- **Dimensión 0:** Representa las componentes conexas. Se observa una única componente conexa dominante que persiste durante toda la filtración, lo cual indica que el conjunto de puntos está conectado globalmente.
- **Dimensión 1:** Corresponde a ciclos o bucles. Se identifican algunos ciclos de corta duración, lo cual es esperado en un plano sin estructuras topológicas complejas. Estos ciclos emergen debido a la distribución aleatoria de los puntos, más que por la presencia de estructuras geométricas significativas.

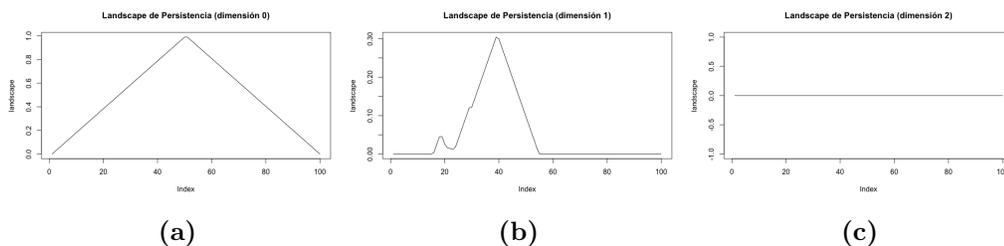
En comparación con figuras más complejas, como el toro entrelazado o la esfera, el análisis topológico del plano muestra una estructura más simple, con menos ciclos

persistentes y sin cavidades tridimensionales. Esto es consistente con la topología trivial del plano, donde la conectividad se mantiene sin la presencia de agujeros o vacíos significativos.



**Figura 4.24:** Código de barras de persistencia para el plano.

El código de barras de persistencia (Figura 4.24) muestra la duración de las características topológicas en el tiempo de filtración. Se observa una barra negra dominante correspondiente a la componente conexa principal en dimensión 0, que persiste durante toda la filtración. Las barras rojas cortas reflejan la aparición de ciclos efímeros en dimensión 1, que desaparecen rápidamente, lo cual es coherente con la ausencia de estructuras cíclicas significativas en el plano.



**Figura 4.25:** Paisajes de persistencia para el plano simulado. Se presentan los resultados para distintas dimensiones: (a) Dimensión 0, (b) Dimensión 1 y (c) Dimensión 2.

En los paisajes de persistencia (Figura 4.25), se destacan las siguientes observaciones:

- **Dimensión 0:** Muestra un único pico pronunciado que corresponde a la componente conexa persistente en el conjunto de puntos.
- **Dimensión 1:** Presenta picos de baja intensidad, indicando la existencia de ciclos de corta duración creados por la disposición aleatoria de los puntos. Estos ciclos no representan estructuras topológicas significativas.
- **Dimensión 2:** No se observan picos, lo que confirma la ausencia de cavidades tridimensionales en la estructura del plano.

Al igual que en el caso de otras figuras bidimensionales, los resultados obtenidos para el plano reflejan una estructura topológica simple, dominada por su conectividad global y la ausencia de cavidades tridimensionales persistentes. La presencia de ciclos de corta duración en dimensión 1 se debe a la disposición aleatoria de los puntos, y no a características intrínsecas de la topología del plano. En dimensión 2, los paisajes confirman la ausencia de cavidades, lo que concuerda con la estructura geométrica esperada. Estos resultados reafirman que el TDA permite capturar correctamente la simplicidad topológica del plano y diferenciarlo de figuras más complejas que presentan estructuras cíclicas o volumétricas significativas.

### 4.3. Simulaciones de figuras modificadas

En esta sección, extendemos el análisis a figuras geométricas que han sido modificadas con la intención de evaluar cómo el TDA detecta cambios estructurales. Estas modificaciones incluyen la introducción de agujeros y entrelazamientos, lo que permite observar cómo la homología persistente captura diferencias en la conectividad y en la presencia de cavidades.

Las figuras consideradas en esta categoría incluyen:

- **Toro Entrelazado**
- **Toro Plano**
- **Paraboloide Positivo con agujero**
- **Paraboloide Negativo con agujero**
- **Plano con agujero**

Estas figuras fueron seleccionadas porque introducen características topológicas distintas que permiten evaluar la capacidad del TDA para distinguir entre estructuras con la misma forma general pero con diferencias significativas en su conectividad.

El mismo procedimiento de generación de puntos utilizado en las figuras simples se aplicó a estas figuras modificadas, asegurando que las comparaciones sean consistentes. Al analizar los diagramas de persistencia obtenidos, podremos determinar cómo las modificaciones estructurales afectan las características topológicas detectadas por el TDA.

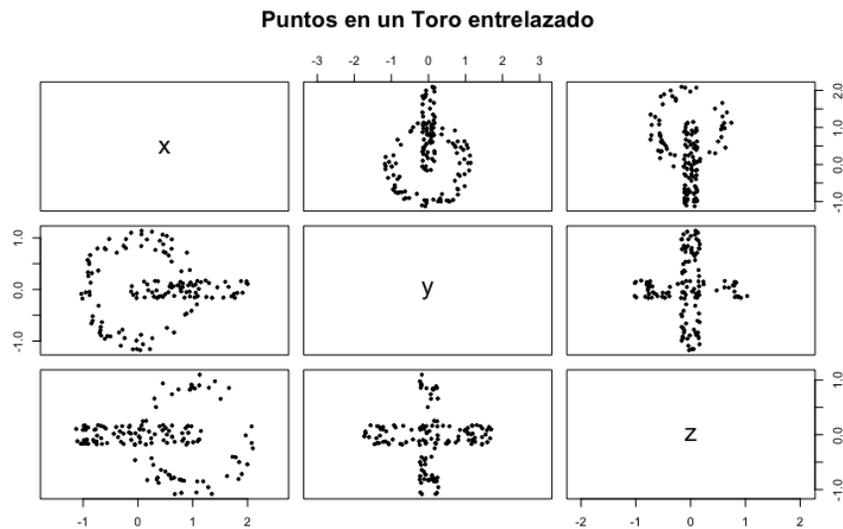
Se espera que la introducción de agujeros y entrelazamientos genere cambios en la homología persistente, como la aparición de ciclos más persistentes en dimensión 1 o cavidades adicionales en dimensión 2, dependiendo de la naturaleza de la modificación. Esto permitirá evaluar la sensibilidad del TDA a alteraciones en la estructura geométrica y su capacidad para distinguir entre figuras similares en apariencia pero con propiedades topológicas diferentes.

### 4.3.1. Toro Entrelazado

Dentro de las figuras simuladas modificadas, se encuentra el **toro entrelazado**, también conocido como *interlocked torus*, que presenta una estructura más compleja debido a la presencia de bucles entrelazados que actúan como agujeros adicionales en la figura. Para su simulación, se utilizó la función `sample_tori_interlocked` del paquete `tdaunif`, la cual permite generar dos toros entrelazados en el espacio tridimensional, simulando así una estructura con interacciones topológicas más ricas. El número de puntos se determinó siguiendo una distribución de Poisson con parámetro  $\lambda = 150$ :

$$n \sim \text{Poisson}(\lambda), \quad \lambda = 150.$$

Este enfoque introduce variabilidad en la cantidad de puntos, permitiendo observar cómo esta afecta la identificación de características topológicas en estructuras complejas.

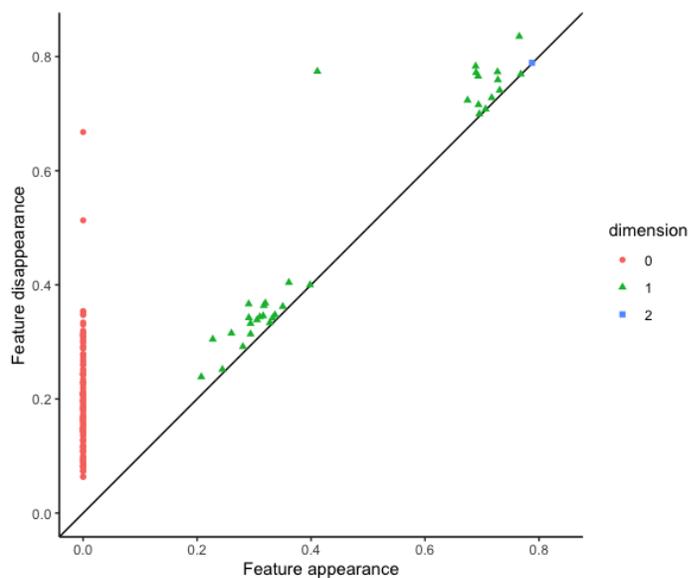


**Figura 4.26:** Puntos simulados en un Toro entrelazado.

En la Figura 4.26, se muestra la distribución de los puntos simulados para el toro entrelazado. Las proyecciones bidimensionales permiten observar la complejidad de

la figura, donde la intersección de los toros genera patrones de puntos que reflejan la naturaleza entrelazada de la estructura. Esta representación destaca las conexiones entre los bucles, que actúan como .<sup>a</sup>agujeros.<sup>a</sup>adicionales, aumentando la complejidad topológica.

Posteriormente, se aplicó el análisis de homología persistente utilizando la función `ripsDiag` del paquete **TDA**, configurada para detectar características topológicas hasta dimensión 2. Además, se utilizó la función `calculate_homology` del paquete **TDASTATS** para calcular la homología persistente a partir de la matriz de coordenadas de los puntos.



**Figura 4.27:** Diagrama de persistencia para el Toro entrelazado.

La Figura 4.27 presenta el diagrama de persistencia obtenido, donde cada punto representa una característica topológica. El eje horizontal indica el momento de aparición (nacimiento) de la característica, mientras que el eje vertical indica su desaparición (muerte).

- **Dimensión 0:** Representa las componentes conexas. Se observa una única componente conexa persistente, lo que indica la conectividad global de la figura.

- **Dimensión 1:** Corresponde a ciclos o bucles. En este caso, se identifican múltiples ciclos persistentes que reflejan la compleja estructura entrelazada de los toros. La presencia de varios ciclos con alta persistencia indica la existencia de bucles significativos que caracterizan la topología del toro entrelazado.
- **Dimensión 2:** Representa cavidades tridimensionales. Se observan algunas cavidades, aunque con menor persistencia, lo cual es consistente con la estructura del toro entrelazado, donde los vacíos tridimensionales son menos prominentes en comparación con los ciclos de dimensión 1.

El análisis muestra que, a pesar de la complejidad añadida por la estructura entrelazada, el TDA es capaz de identificar con claridad las características topológicas más relevantes. En particular, el mayor número de ciclos persistentes en dimensión 1 refleja la presencia de múltiples bucles interconectados, lo que distingue al toro entrelazado del toro simple. Esto demuestra la capacidad del TDA para diferenciar estructuras geométricas similares pero con conectividades distintas, capturando interacciones topológicas no evidentes mediante métodos geométricos tradicionales. Así, la homología persistente se confirma como una herramienta eficaz para el análisis de estructuras con entrelazamientos y conexiones complejas.

### 4.3.2. Toro Plano

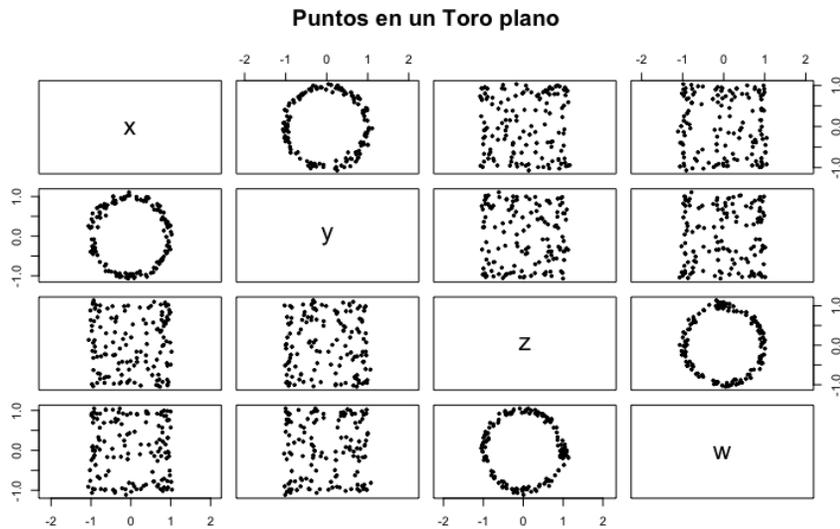
Para la simulación del toro plano, se utilizó la función `sample_torus_flat` del paquete `tdaunif`, que permite generar puntos distribuidos uniformemente sobre la superficie de un toro representado en un espacio plano. Esta función facilita la generación de un toro parametrizado en cuatro dimensiones, proyectado en coordenadas  $(x, y, z, w)$  para su visualización.

El número de puntos generados se modeló siguiendo una distribución de Poisson con parámetro  $\lambda = 150$ :

$$n \sim \text{Poisson}(\lambda), \quad \lambda = 150.$$

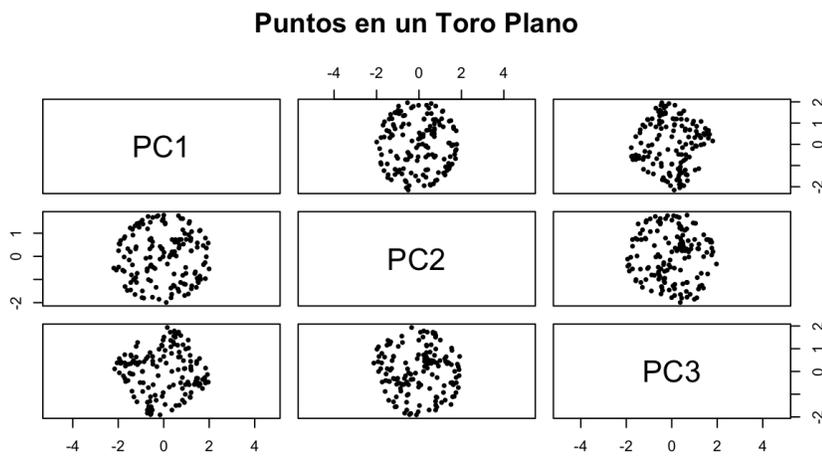
Este enfoque introduce una variabilidad controlada en la cantidad de puntos, permitiendo analizar cómo dicha variabilidad afecta la detección de características to-

pológicas en el análisis de homología persistente.



**Figura 4.28:** Puntos simulados en un Toro plano.

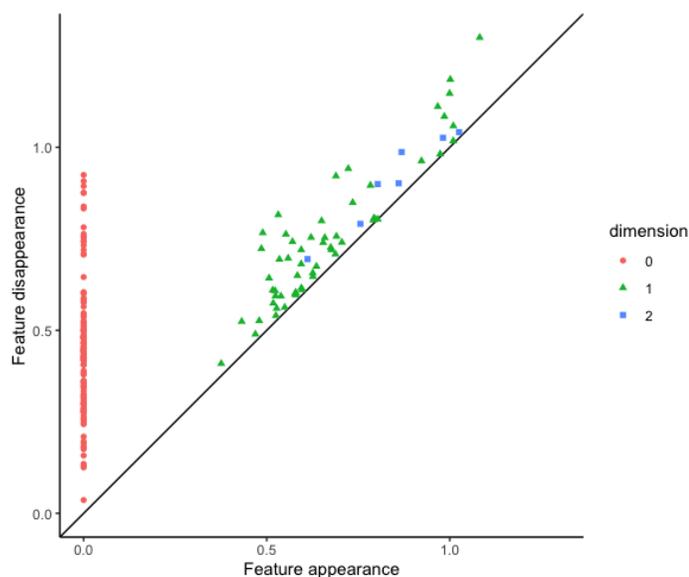
En la Figura 4.28 se muestra la distribución de los puntos simulados.



**Figura 4.29:** Puntos simulados en un Toro plano proyectados a 3 dimensiones.

En la Figura 4.29 se muestra la distribución de los puntos simulados después de la proyección a tres dimensiones. Las proyecciones bidimensionales entre las combinaciones de coordenadas permiten observar patrones circulares característicos del toro en algunas vistas  $(xy, xz)$  y distribuciones más uniformes en otras. Esta estructura refleja la topología subyacente del toro, donde la conexión de bordes opuestos genera la forma toroidal en el espacio tridimensional reducido. Cabe destacar que esta proyección a 3D se realizó porque, en los análisis topológicos, los indicadores obtenidos incluían números de Betti de dimensión 0, 1 y 2. Dado que las demás figuras utilizadas en la comparación están en 3D, la proyección asegura una representación coherente y evita la pérdida de información estructural al comparar directamente con las otras figuras.

Posteriormente, se aplicó el análisis de homología persistente utilizando la función `ripsDiag` del paquete **TDA**, configurada para calcular características topológicas hasta dimensión 2. Además, se utilizó la función `calculate_homology` del paquete **TDAstats** para obtener la homología persistente directamente a partir de la matriz de coordenadas.



**Figura 4.30:** Diagrama de persistencia para el Toro plano.

La Figura 4.30 presenta el diagrama de persistencia obtenido, donde cada punto representa una característica topológica. El eje horizontal indica el momento de aparición (nacimiento) de la característica, mientras que el eje vertical indica el momento de su desaparición (muerte).

- **Dimensión 0** : Representa las componentes conexas. Se observa una componente conexa dominante que persiste durante toda la filtración, lo que refleja la conectividad global de la figura.
- **Dimensión 1** : Corresponde a ciclos o bucles. Se identifican dos ciclos principales, característicos de la topología del toro, que representan la existencia de dos agujeros: uno alrededor del eje principal del toro y otro a lo largo de su superficie.
- **Dimensión 2** : Representa cavidades tridimensionales. Aunque el toro plano es una representación en un espacio de menor dimensión, se identifican algunas cavidades de menor persistencia, posiblemente debido a la proyección de la estructura en el espacio plano.

El análisis del toro plano destaca la preservación de sus características topológicas fundamentales a pesar de su representación en un espacio de menor dimensión. En particular, la identificación clara de dos ciclos persistentes en dimensión 1 confirma la estructura toroidal esperada, diferenciándola del toro entrelazado, donde la mayor complejidad estructural generaba más ciclos. La relativa simplicidad de la proyección en el espacio plano facilita la visualización de estos invariantes topológicos, lo que se refleja en una distribución más uniforme de los puntos y una mayor estabilidad en la detección de las características clave del toro. Aunque se observan cavidades en dimensión 2, su baja persistencia sugiere que pueden ser artefactos derivados de la simulación y no propiedades intrínsecas de la estructura. En general, estos resultados confirman que el TDA es una herramienta eficaz para diferenciar variaciones geométricas y topológicas dentro de estructuras toroidales.

### 4.3.3. Paraboloide Positivo con Agujero

La simulación del paraboloide positivo fue realizada introduciendo un agujero en su superficie. Para la generación de los puntos, se utilizó una parametrización sencilla que describe la forma del paraboloide:

$$\begin{aligned}x &= x, \\y &= y, \\z &= x^2 + y^2,\end{aligned}$$

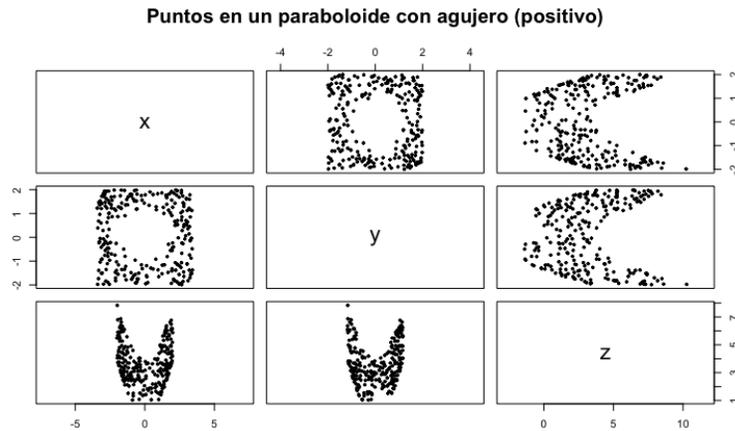
donde  $(x, y)$  son coordenadas en el plano y  $z$  representa la altura en función de estos valores. El jacobiano correspondiente a esta parametrización se define como:

$$J(x, y) = 1 + 4x^2 + 4y^2.$$

Se utilizó la función `make_rejection_sampler` para generar puntos sobre la superficie del paraboloide, con los siguientes parámetros:

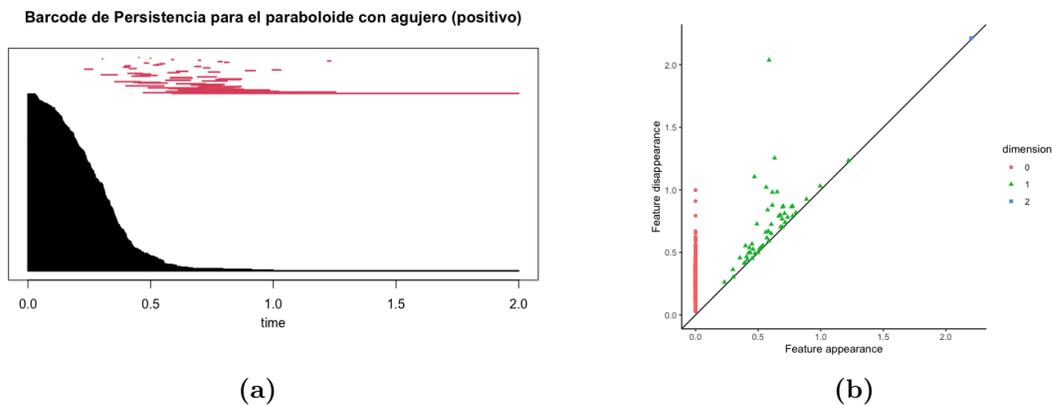
- **Parámetros mínimos:**  $(x, y) = (-2, -2)$
- **Parámetros máximos:**  $(x, y) = (2, 2)$
- **Máximo del jacobiano:** 33

Posteriormente, se filtraron los puntos que caían dentro de un círculo de radio 1 centrado en el origen, creando así un agujero en la superficie del paraboloide. Esta eliminación de puntos genera una estructura topológica diferente, la cual será analizada a continuación.



**Figura 4.31:** Proyecciones de los puntos generados en el paraboloides positivo con agujero.

En la Figura 4.31, se observa claramente la presencia del agujero en el centro del paraboloides, resultado del filtrado de puntos. Esta característica topológica se refleja en los diagramas de persistencia presentados a continuación.



**Figura 4.32:** (a) Diagrama de código de barras. (b) Diagrama de persistencia para el paraboloides positivo con agujero.

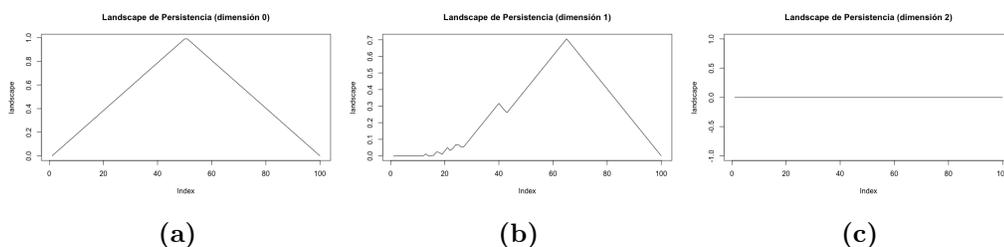
En la Figura 4.32, se observan las siguientes características topológicas:

- **Dimensión 0:** Se aprecia la presencia de múltiples componentes conexas que

desaparecen rápidamente, lo que indica que los puntos inicialmente no están completamente conectados pero se unen con el aumento del parámetro de filtración.

- **Dimensión 1:** Destacan varios ciclos persistentes, representados por los triángulos verdes. El ciclo más persistente se asocia directamente con el agujero introducido en la superficie del paraboloides, confirmando la alteración topológica esperada.
- **Dimensión 2:** Se observan algunos puntos azules, aunque en menor cantidad, lo que indica la presencia de cavidades tridimensionales formadas por la estructura global del paraboloides.

A continuación, se presentan los paisajes de persistencia para cada dimensión, los cuales permiten una visualización funcional de las características topológicas detectadas:



**Figura 4.33:** Paisajes de persistencia para el paraboloides positivo con agujero. Se presentan los resultados para distintas dimensiones: (a) Dimensión 0, (b) Dimensión 1 y (c) Dimensión 2.

En los paisajes de la Figura 4.33, se puede apreciar lo siguiente:

- **Dimensión 0:** Refleja la rápida fusión de las componentes conexas en una sola, lo cual es esperado en una estructura conectada globalmente como el paraboloides.

- **Dimensión 1:** Se observa un pico prominente que corresponde al ciclo generado por el agujero introducido. Este pico indica una persistencia significativa, destacando el impacto de la modificación en la superficie.
- **Dimensión 2:** Muestra una menor actividad, con picos menos pronunciados, lo cual es coherente con la ausencia de cavidades tridimensionales dominantes en la estructura.

La introducción de un agujero en el paraboloides positivo genera un cambio topológico significativo en comparación con su versión sin modificaciones. El análisis de homología persistente muestra un ciclo en dimensión 1 con alta persistencia, reflejando la estructura del agujero introducido. En contraste, en dimensión 2 no se observan cavidades altamente persistentes, lo que confirma que la alteración afecta principalmente la conectividad en el plano. Estos resultados evidencian la capacidad del TDA para detectar y cuantificar modificaciones estructurales en superficies geométricas y resaltan la importancia de los ciclos persistentes como indicadores de alteraciones topológicas.

#### 4.3.4. Paraboloides Negativo con Agujero

En este experimento, se generaron puntos sobre un paraboloides negativo al cual se le introdujo un agujero en su estructura. Para la simulación, se utilizó la siguiente parametrización:

$$\begin{aligned}x &= x, \\y &= y, \\z &= -(x^2 + y^2),\end{aligned}$$

donde  $(x, y)$  son los parámetros que definen la superficie, y la coordenada  $z$  se obtiene como la suma negativa de los cuadrados de  $x$  y  $y$ , característica típica de un paraboloides cóncavo hacia abajo.

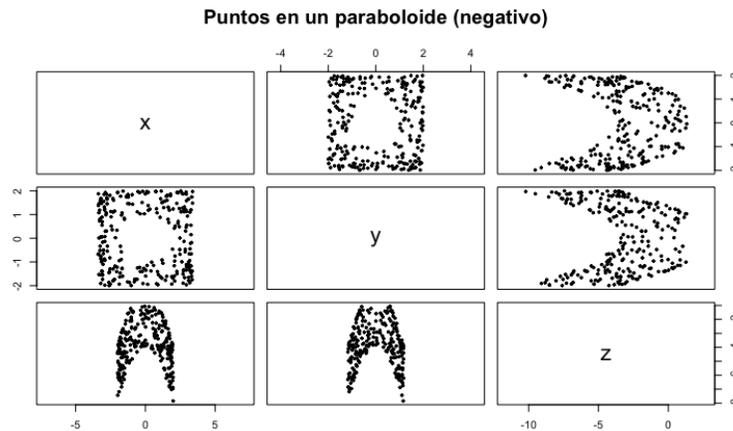
El jacobiano asociado a esta parametrización está dado por:

$$J(x, y) = 1 + 4x^2 + 4y^2,$$

el cual se utilizó en el muestreador de rechazo `make_rejection_sampler` para asegurar una distribución uniforme de los puntos sobre la superficie.

Posteriormente, se introdujo un agujero eliminando los puntos que cayeran dentro de un radio predefinido  $r = 1$  alrededor del origen. Esto se realizó mediante una función que filtra los puntos generados, manteniendo solo aquellos que están fuera del área circular definida:

$$\sqrt{x^2 + y^2} > r.$$

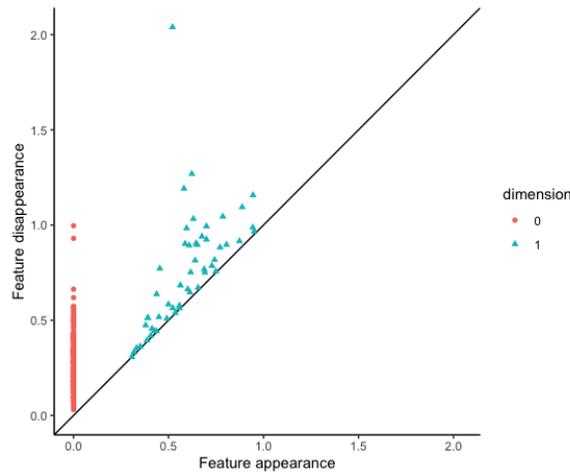


**Figura 4.34:** Distribución de puntos en un paraboloides negativo con agujero.

La Figura 4.34 muestra la distribución de los puntos simulados, donde se observa claramente la ausencia de puntos en la región central correspondiente al agujero.

Se calculó la homología persistente de los puntos utilizando la función `calculate_homology`, considerando hasta la segunda dimensión ( $\text{dim} = 2$ ). El diagrama de persistencia re-

sultante se presenta en la Figura 4.35.



**Figura 4.35:** Diagrama de persistencia para el paraboloides negativo con agujero.

En el diagrama de persistencia se pueden identificar las siguientes características topológicas:

- **Dimensión 0** : Se observa una componente dominante y varias componentes que desaparecen rápidamente.
- **Dimensión 1** : La presencia del agujero contribuye a la formación de estos ciclos persistentes.
- **Dimensión 2** : Aunque son menos frecuentes, reflejan la existencia de cavidades tridimensionales formadas por la distribución de los puntos.

El paraboloides negativo con agujero presenta una topología distinta a la de su versión original, con un ciclo persistente en dimensión 1 que captura la presencia del agujero introducido. A diferencia del paraboloides positivo, en este caso la estructura cóncava favorece la aparición de cavidades transitorias en dimensión 2, aunque estas no son dominantes. La comparación con la versión sin agujero confirma que la eliminación de puntos en la superficie altera de manera clara los invariantes topológicos, lo que demuestra la efectividad del TDA para identificar discontinuidades y diferencias estructurales en modelos geométricos.

### 4.3.5. Plano con Agujero

En esta simulación se utilizó la misma parametrización que para el plano simple, pero se introdujo una restricción para generar un agujero circular en el centro del plano. El número de puntos se obtuvo a partir de un proceso de Poisson con parámetro  $\lambda = 150$ :

$$n \sim \text{Poisson}(\lambda), \quad \lambda = 150.$$

La parametrización del plano se define de la siguiente forma:

$$\phi(x, y) = (x, y, 0),$$

donde  $z = 0$  para todos los puntos, lo que refleja que los puntos están confinados en el plano  $xy$ . El jacobiano de la transformación es constante:

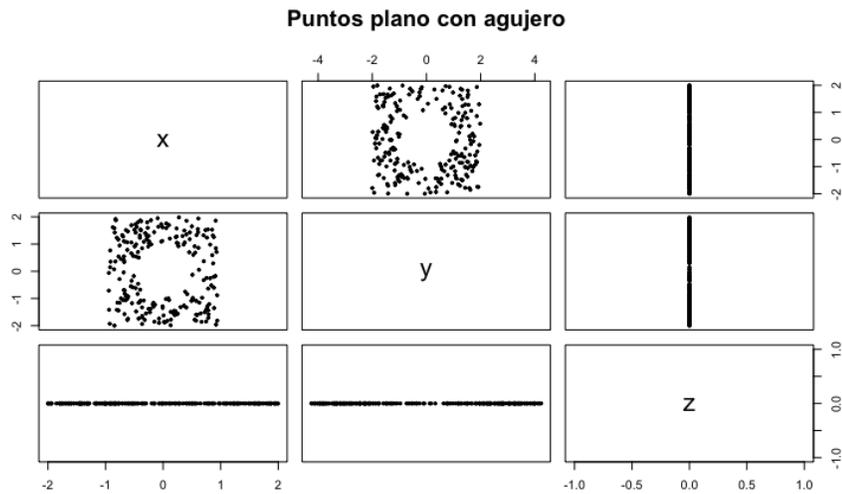
$$J(x, y) = 1,$$

ya que no existe una transformación que altere la densidad de los puntos.

Para generar el agujero en el plano, se utilizó una función de filtrado que elimina los puntos dentro de un círculo de radio  $r$  centrado en el origen. El proceso de filtrado se basa en la siguiente condición:

$$\sqrt{x^2 + y^2} > r,$$

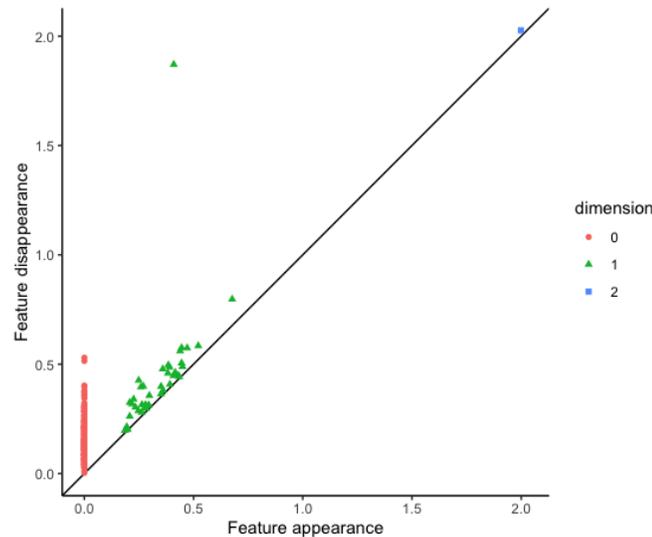
lo que asegura que solo se mantengan los puntos fuera del círculo definido por el radio  $r$ . Este enfoque permite modelar un patrón puntual con una región vacía en el centro, creando un agujero que afecta la estructura topológica del conjunto de puntos.



**Figura 4.36:** Puntos simulados en un plano con un agujero central.

En la Figura 4.36, se observa claramente la ausencia de puntos en la región central, lo que indica la presencia del agujero. Las proyecciones muestran cómo la distribución de puntos se mantiene uniforme fuera del área restringida.

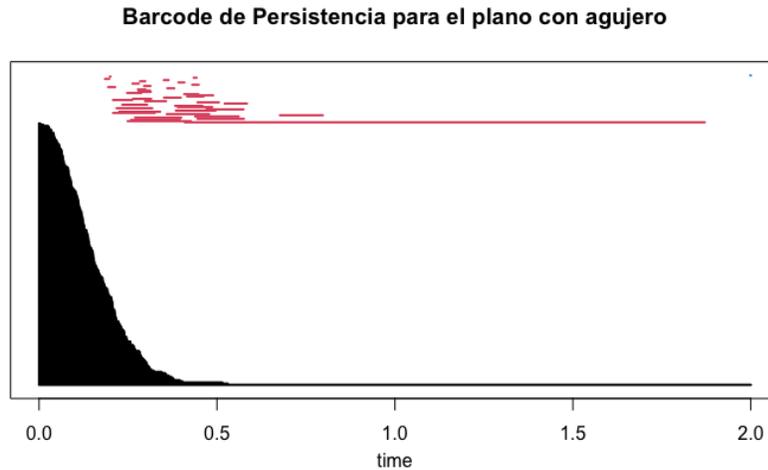
Posteriormente, se aplicó el análisis de homología persistente utilizando la función `ripsDiag` del paquete **TDA**, configurada para detectar características topológicas hasta dimensión 2. También se utilizó la función `calculate_homology` del paquete **TDAstats** para obtener los invariantes topológicos.



**Figura 4.37:** Diagrama de persistencia para el plano con agujero.

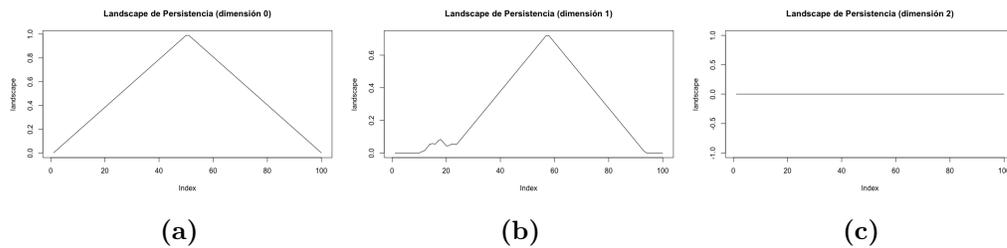
La Figura 4.37 muestra el diagrama de persistencia resultante. En este se pueden identificar las siguientes características topológicas:

- **Dimensión 0 :** Representa las componentes conexas. Se observa una única componente conexa dominante que persiste a lo largo de toda la filtración, lo cual indica que, a pesar del agujero, el conjunto de puntos se mantiene conectado globalmente.
- **Dimensión 1 :** Corresponde a ciclos o bucles. Destaca la presencia de un ciclo altamente persistente, que representa el agujero central en el plano. Este ciclo es una característica topológica significativa, ya que persiste durante un amplio rango de la filtración, reflejando la estructura vacía generada en la simulación.
- **Dimensión 2 :** No se observan cavidades tridimensionales significativas, lo cual es esperado dado que la figura es esencialmente bidimensional. La ausencia de estos invariantes es consistente con la estructura del plano con agujero.



**Figura 4.38:** Código de barras de persistencia para el plano con agujero.

El código de barras de persistencia (Figura 4.38) permite observar la duración de las características topológicas durante el proceso de filtración. Se destaca una barra larga en dimensión 1, que corresponde al ciclo persistente asociado al agujero, reflejando su relevancia topológica.



**Figura 4.39:** Paisajes de persistencia para el plano con agujero. Se presentan los resultados para distintas dimensiones: (a) Dimensión 0, (b) Dimensión 1 y (c) Dimensión 2.

En los paisajes de persistencia (Figura 4.39), se representa de forma funcional la prominencia de las características topológicas:

- **Dimensión 0:** Refleja la fusión de las componentes conexas hacia una única estructura conectada.

- **Dimensión 1:** Destaca un pico pronunciado correspondiente al ciclo persistente generado por el agujero.
- **Dimensión 2:** No se observan picos significativos, en coherencia con la ausencia de cavidades tridimensionales.

La simulación del plano con agujero confirma que la eliminación de puntos en una región central introduce un ciclo altamente persistente en dimensión 1, lo cual no estaba presente en la versión original del plano. La homología persistente captura esta alteración con claridad, mostrando una estructura simple pero con un cambio topológico evidente. La ausencia de características relevantes en dimensión 2 es coherente con la naturaleza bidimensional del plano. En general, estos resultados demuestran que el TDA es una herramienta eficaz para detectar cambios en la conectividad de un patrón puntual y diferenciar estructuras con y sin discontinuidades espaciales.

#### 4.4. Cluster de superficies

El análisis topológico de datos permite extraer descriptores que caracterizan la estructura geométrica y topológica de las figuras simuladas. En esta sección, se describe el procedimiento de clusterización aplicado a estos descriptores con el fin de agrupar automáticamente las distintas superficies en función de sus propiedades topológicas. La clusterización se realizó utilizando el algoritmo k-medoids sobre una matriz de distancias construida a partir de los diagramas de paisajes. Para cada figura, se generó su diagrama de persistencia y se calcularon los diagramas de paisajes en dimensiones 0, 1 y 2. Posteriormente, se definió la distancia  $L_2$  entre estos diagramas como métrica de comparación, permitiendo evaluar la similitud topológica entre figuras.

##### 4.4.1. Construcción de Descriptores Topológicos

Para cada figura generada, se calcularon diagramas de persistencia y se extrajeron sus correspondientes diagramas de paisajes. Estos descriptores resumen la evolución

de componentes conexas y ciclos a medida que varía la escala  $\varepsilon$ , proporcionando información clave sobre la estructura topológica de cada figura.

Dado que la clasificación se basa en similitud topológica, la distancia  $L_2$  entre los diagramas de paisajes fue utilizada como métrica de comparación. Esto permitió representar las diferencias topológicas de manera robusta y reducir la dimensionalidad de los datos para facilitar la agrupación.

#### 4.4.2. Algoritmo k-medoids

Para la clusterización, se utilizó el algoritmo k-medoids, implementado en la función `pam()` de la librería `cluster` en R. Este algoritmo es una variante más robusta de k-means, donde en lugar de calcular la media de los puntos en un cluster, se seleccionan representantes reales (medoids), minimizando la distancia total dentro de cada cluster.

Dado un conjunto de  $n$  observaciones y una matriz de distancias  $D$ , el algoritmo k-medoids busca un conjunto de  $k$  medoids  $\{m_1, m_2, \dots, m_k\}$  que minimicen la función de costo:

$$\sum_{i=1}^n \min_{j \in \{1, \dots, k\}} d(x_i, m_j)$$

donde  $d(x_i, m_j)$  representa la distancia entre la observación  $x_i$  y el medoid más cercano. A diferencia de k-means, que es sensible a valores extremos debido al cálculo de medias, k-medoids es más robusto frente a la presencia de valores atípicos.

El procedimiento del algoritmo es el siguiente:

1. Se inicializan  $k$  medoids seleccionando aleatoriamente  $k$  observaciones del conjunto de datos.
2. Cada punto se asigna al medoid más cercano según la matriz de distancias  $D$ .
3. Se actualizan los medoids iterativamente intercambiando puntos dentro y fuera de los clusters, buscando minimizar la distancia intra-cluster.

4. Se repite el proceso hasta que los medoids converjan y la asignación de clusters se estabilice.

El uso de k-medoids en el contexto del Análisis Topológico de Datos ha sido explorado en diversas aplicaciones. Por ejemplo, en [Shin \(2019\)](#), el algoritmo fue utilizado para la clasificación de géneros literarios en un problema de Procesamiento de Lenguaje Natural (NLP), donde los datos fueron representados mediante descriptores topológicos y luego agrupados utilizando k-medoids. Este enfoque permitió identificar patrones significativos en los textos mediante técnicas de TDA.

Adicionalmente, en [Ohanuba et al. \(2021\)](#), se aplicaron técnicas de aprendizaje no supervisado para la detección de patrones en fenómenos atmosféricos, demostrando que la combinación de TDA con algoritmos de clustering es efectiva para la segmentación de datos espaciales complejos. Aunque este estudio no menciona explícitamente el uso de k-medoids, proporciona un marco conceptual en el que el clustering de patrones topológicos es una estrategia viable para la clasificación de datos geoespaciales.

#### 4.4.3. Procedimiento de Clusterización

El proceso de clusterización siguió los siguientes pasos:

1. Se generaron puntos en cada figura simulada utilizando un proceso puntual de Poisson y se calcularon los diagramas de persistencia.
2. Se construyeron los diagramas de paisajes en dimensiones 0, 1 y 2, representando la evolución de los ciclos topológicos.
3. Se calculó la matriz de distancias entre los diagramas de paisajes utilizando la distancia  $L_2$ .
4. Se aplicó el algoritmo k-medoids para identificar grupos de figuras con características topológicas similares.
5. Se determinó el número óptimo de clusters mediante el método del codo.

Este procedimiento permite analizar la capacidad del TDA para clasificar distintas figuras geométricas basándose en su estructura persistente. En las siguientes secciones, se evaluará la efectividad del método considerando distintas configuraciones: la clusterización de un subconjunto de figuras, la agrupación por categorías geométricas y la clasificación de todas las figuras en conjunto.

#### 4.5. Clasificación y comparación de resultados

El objetivo de esta sección es analizar los resultados obtenidos tras la aplicación del algoritmo de clusterización sobre las figuras simuladas. Mientras que la sección anterior detalló el procedimiento y las herramientas utilizadas para la clusterización, en esta parte se examina en qué medida la clasificación obtenida refleja la estructura topológica esperada de las figuras analizadas.

El análisis se enfoca en evaluar la capacidad del método para distinguir figuras geométricas con distintas estructuras topológicas, agrupándolas de manera coherente en función de sus características. Además, se explora cómo la cantidad de figuras y su complejidad afectan la clasificación y qué patrones emergen en la segmentación automática de estas estructuras.

Para ello, se realizan pruebas considerando diferentes conjuntos de figuras:

- **Clusterización de subconjuntos específicos:** Se analizan separadamente grupos de figuras con propiedades similares, como los toros, los planos y las superficies parabólicas, con el fin de evaluar la capacidad del algoritmo para distinguir estructuras dentro de una misma categoría.
- **Clusterización con todas las figuras:** Se agrupan todas las figuras en un mismo análisis para evaluar la robustez del método al enfrentar una mayor diversidad de estructuras topológicas.
- **Impacto de las modificaciones estructurales:** Se estudia cómo la presencia de agujeros u otras alteraciones en las figuras afecta la clasificación y si el algoritmo es capaz de diferenciarlas de sus versiones originales.

Además, se comparan los resultados obtenidos con distintas configuraciones de parámetros en  $k$ -medoids y se evalúa el desempeño del método mediante la coherencia de los clusters formados. Finalmente, se analizan las implicancias de estos hallazgos para la aplicación del TDA en problemas de clasificación de patrones espaciales.

Las siguientes subsecciones presentan en detalle los experimentos realizados, los agrupamientos obtenidos y la interpretación de los resultados.

#### 4.5.1. Clusterización con todas las figuras

El propósito de esta sección es analizar cómo el algoritmo de clusterización agrupa todas las figuras simuladas en un mismo experimento. La idea es evaluar si el método logra diferenciar las estructuras topológicas presentes y si la clasificación obtenida tiene sentido desde el punto de vista de la persistencia de homología.

Para ello, se aplicará  $k$ -medoids sobre la matriz de distancias entre los diagramas de paisajes, utilizando como métrica la norma  $L_2$ . Antes de ejecutar la clusterización, se determinará el número óptimo de clusters  $k$  mediante el método del codo y el coeficiente de silueta. Luego, se interpretará la agrupación resultante a partir de un mapa de calor de la matriz de distancias, con el objetivo de visualizar las relaciones entre las figuras según sus características topológicas.

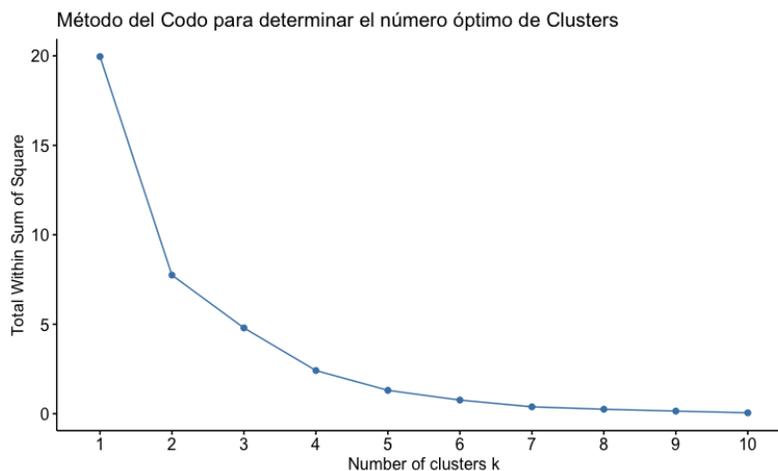
Dado que el análisis se basa en información topológica y no en formas geométricas exactas, se espera que las figuras con agujeros sean separadas de aquellas sin perforaciones, y que estructuras con similitudes en términos de componentes conexas, ciclos y cavidades terminen en clusters similares. En particular:

- Las figuras con agujeros deberían formar un grupo separado de las que no tienen perforaciones.
- El toro, por su estructura con un ciclo de dimensión 1 persistente, debería diferenciarse de otras figuras sin esa característica.
- Figuras con estructuras muy similares, como distintos tipos de paraboloides, podrían ser agrupadas en un mismo cluster si la diferencia en su topología no es significativa.

El siguiente análisis revisará si estas expectativas se cumplen y qué patrones emergen al considerar todas las figuras en conjunto.

Para evaluar la capacidad del algoritmo de clusterización en la diferenciación de las figuras simuladas, se aplicó el método k-medoids sobre la matriz de distancias entre los diagramas de paisajes. Antes de proceder con la segmentación, fue necesario determinar un número óptimo de clusters  $k$ , para lo cual se emplearon dos enfoques estándar: el método del codo y el coeficiente de silueta.

El método del codo se basa en la evaluación de la *suma de las distancias intra-cluster* para distintos valores de  $k$ . La idea principal es identificar el punto donde la reducción de la varianza dentro de los clusters deja de ser significativa, es decir, donde la curva de la inercia forma un "codo". La Figura 4.40 muestra que este punto ocurre en  $k = 3$ , lo que sugiere que este valor es una elección razonable para la segmentación de las figuras.

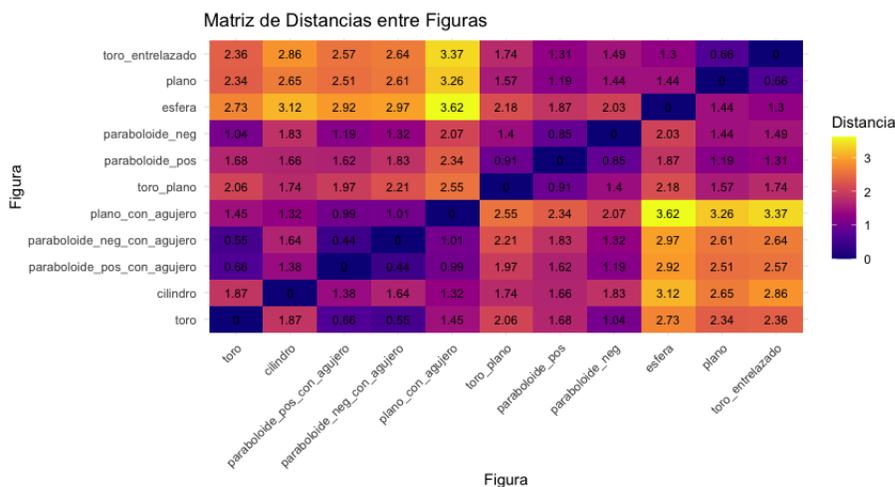


**Figura 4.40:** Método del Codo para determinar el número óptimo de clusters.

Para complementar esta selección, se calculó el coeficiente de silueta para  $k = 3$ , el cual mide la coherencia de los grupos obtenidos comparando la proximidad de cada punto a su propio cluster frente a los clusters vecinos.

Con  $k = 3$  confirmado como el número óptimo de clusters, se aplicó k-medoids a la matriz de distancias obtenida. La Figura 4.41 muestra la distribución de las figuras

en los tres grupos resultantes. Se observa que las distintas superficies han sido agrupadas en categorías diferenciadas, lo que sugiere que las características topológicas capturadas por los diagramas de paisajes permiten distinguir adecuadamente entre las figuras.



**Figura 4.41:** Distribución de Figuras por Cluster.

Los resultados obtenidos muestran que, en términos generales, la clusterización refleja algunas de las expectativas iniciales. En particular:

- Las figuras con agujeros han sido agrupadas juntas, lo que indica que la métrica utilizada logra capturar correctamente esta característica topológica.
- La separación del toro respecto a otras figuras no es completamente clara, lo que sugiere que su persistencia en dimensión 1 podría no ser lo suficientemente distintiva en comparación con otras estructuras.
- Algunos paraboloides y planos han sido agrupados en clusters similares, lo que puede deberse a la similitud en su estructura de componentes conexas y cavidades en términos de persistencia.

Si bien el agrupamiento obtenido sigue en parte la estructura esperada, existen ciertos casos donde las distancias entre figuras no reflejan intuitivamente sus diferencias

geométricas. Esto sugiere que, aunque la métrica  $L_2$  sobre los diagramas de paisajes es efectiva en algunos aspectos, podrían explorarse otras métricas o representaciones topológicas para mejorar la segmentación.

Para facilitar la consulta de los resultados, la Tabla 6.1 presenta la lista de figuras asignadas a cada cluster.

#### 4.5.2. Clusterización de subconjuntos específicos

Se realizaron pruebas de clusterización sobre subconjuntos de figuras con el objetivo de evaluar la capacidad del algoritmo para diferenciar estructuras similares con y sin modificaciones topológicas. En este apartado se presentan los resultados obtenidos para un subconjunto compuesto por planos y paraboloides, dejando espacio para un segundo experimento con distintas variantes de toros.

##### Prueba con planos y paraboloides

En este experimento se analiza cómo el algoritmo de clusterización agrupa figuras con estructuras similares, pero con diferencias clave en su topología. Se incluyeron un plano y un plano con agujero, junto con paraboloides positivo y negativo, tanto en sus versiones originales como con perforaciones. El objetivo es determinar si la presencia de agujeros afecta significativamente la clasificación y cómo se organizan las figuras en función de sus características persistentes.

Para evaluar la segmentación, se realizaron pruebas con  $k = 2$  y  $k = 3$ . La idea es observar si con  $k = 2$  la clusterización separa correctamente las figuras con y sin agujero, y si con  $k = 3$  aparecen agrupaciones más específicas según la forma de los paraboloides. Se espera que:

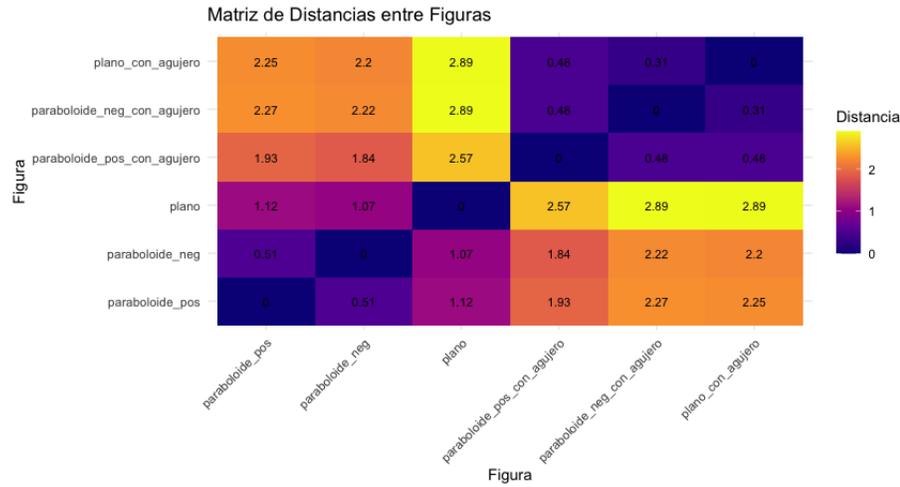
- Con  $k = 2$ , el algoritmo distinga las figuras con agujero de aquellas sin perforaciones.
- Con  $k = 3$ , se formen clusters más detallados, diferenciando los planos de los paraboloides.

- La distancia  $L_2$  utilizada para comparar los diagramas de paisajes refleje correctamente la presencia o ausencia de agujeros como criterio de agrupación.

Para entender mejor la agrupación obtenida en los experimentos de clusterización, se presentan las matrices de distancia entre las figuras utilizadas. Estas matrices reflejan la similitud entre los diagramas de paisajes asociados a cada figura, utilizando la distancia  $L_2$  como métrica de comparación.

En la Figura 4.42, correspondiente a la prueba con  $k = 2$ , se observa que las figuras con agujeros presentan mayores distancias respecto a aquellas sin agujeros. En particular:

- El *plano con agujero* tiene una distancia elevada ( $> 3.3$ ) respecto al *plano*, lo que sugiere que la presencia del agujero introduce una diferencia topológica significativa.
- De manera similar, los paraboloides con agujero tienen mayores distancias respecto a sus versiones sin perforaciones, lo que indica que el algoritmo es capaz de capturar esta diferencia.
- Las distancias entre el *paraboloide positivo* y el *paraboloide negativo* son pequeñas ( $< 1$ ), lo que sugiere que el método los percibe como estructuras similares desde una perspectiva topológica.

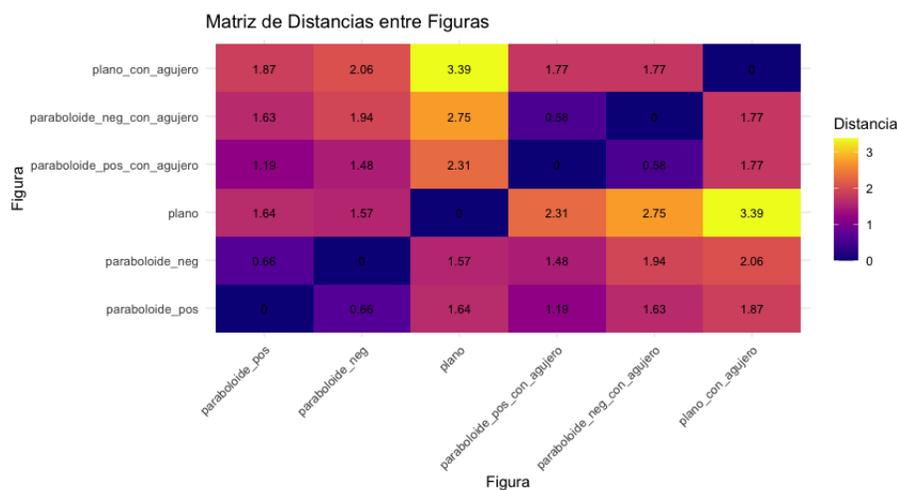


**Figura 4.42:** Distribución de figuras por cluster con  $k = 2$ .

La clusterización para  $k = 2$  reflejó esta segmentación esperada, separando las figuras en dos grupos: aquellas con agujeros y aquellas sin agujeros.

Por otro lado, al aumentar el número de clusters a  $k = 3$ , la segmentación obtenida no siguió la estructura esperada, como se observa en la Figura 4.43. En este caso:

- La distancia entre el *paraboloide negativo* y el *paraboloide positivo* sigue siendo pequeña, indicando que el algoritmo los sigue agrupando en una misma categoría.
- Sin embargo, las distancias entre el *plano* y el *plano con agujero* siguen siendo elevadas ( $> 2.8$ ), lo que sugiere que el método sigue detectando la diferencia estructural pero no la refleja tan claramente en los clusters generados.
- En este caso, la segmentación resultante parece más dispersa, indicando que agregar más clusters no necesariamente mejora la clasificación.



**Figura 4.43:** Distribución de figuras por cluster con  $k = 3$ .

La Figura 6.2 muestra los resultados de ambas clasificaciones.

#### Prueba con distintas variantes de toros

En este experimento se evaluó cómo el algoritmo de clusterización diferencia entre distintas configuraciones del toro. Se consideraron tres variantes:

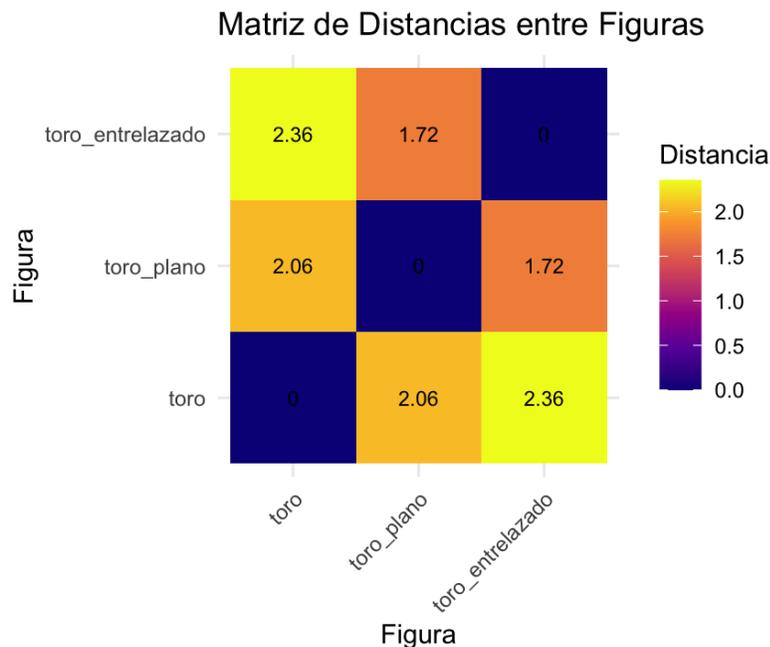
- Toro estándar.
- Toro entrelazado.
- Toro plano.

El propósito de esta prueba es determinar si k-medoids puede capturar las diferencias topológicas entre estas figuras, a pesar de que todas comparten un ciclo en dimensión 1. En particular, se quiere analizar si la estructura entrelazada o la forma más achatada del toro plano afectan la clasificación, o si el método los agrupa en función de su número de agujeros.

Para ello, se aplicó el algoritmo sobre la matriz de distancias construida a partir de los diagramas de paisajes, utilizando la norma  $L_2$  como métrica de comparación. Se espera que:

- El toro estándar sea diferenciado de las otras dos configuraciones debido a su estructura más uniforme.
- El toro entrelazado y el toro plano puedan agruparse juntos si presentan patrones de homología persistente similares.
- La clasificación refleje las características topológicas más que la geometría visual de las figuras.

Para este conjunto de figuras, se utilizó  $k = 2$ . La Figura 4.44 muestra la segmentación obtenida, donde el toro estándar se encuentra en un cluster separado, mientras que el toro entrelazado y el toro plano fueron agrupados en el mismo cluster debido a que la distancia entre ellos es menor. Esto sugiere que, en términos de homología persistente, el toro estándar posee una estructura más distintiva, mientras que el toro entrelazado y el toro plano comparten más características en sus diagramas de persistencia.



**Figura 4.44:** Distribución de toros por cluster con  $k = 2$ .

En particular, la distancia entre el toro estándar y el toro entrelazado es de aproximadamente 2.36, mientras que la distancia entre el toro estándar y el toro plano es de 2.06. Por otro lado, el toro entrelazado y el toro plano tienen una menor distancia mutua de 1.72, lo que indica que presentan estructuras topológicas más similares entre sí que respecto al toro estándar. Estos resultados reflejan que el algoritmo de clustering capta diferencias topológicas más allá de la geometría visual de las figuras. En consecuencia, el toro estándar se diferencia claramente de las otras configuraciones, mientras que el toro entrelazado y el toro plano tienden a agruparse juntos debido a sus similitudes en términos de homología persistente.

La Tabla 6.3 presenta la asignación de cada figura dentro de los clusters obtenidos.

#### **Prueba con esfera, cono y cilindro**

En este experimento se evaluó cómo el algoritmo k-medoids diferencia entre figuras geométricas con distintas estructuras, considerando una esfera, un cono y un cilindro. El objetivo es determinar si la homología persistente logra capturar sus diferencias y si la cantidad de figuras afecta la agrupación final.

Para ello, se realizaron experimentos con  $k = 2$  y  $k = 3$ :

- Con  $k = 2$ , se busca evaluar si la esfera se diferencia del cono y el cilindro.
- Con  $k = 3$ , se realizaron dos configuraciones distintas para analizar cómo la cantidad y tipo de figuras afecta la clasificación:
  1. Se incluyeron dos esferas, dos conos y dos cilindros para observar si el algoritmo logra identificar correctamente grupos equivalentes.
  2. Se utilizó una esfera, un cono y dos cilindros, para evaluar si el aumento de una única figura de un tipo modifica la estructura de los clusters.

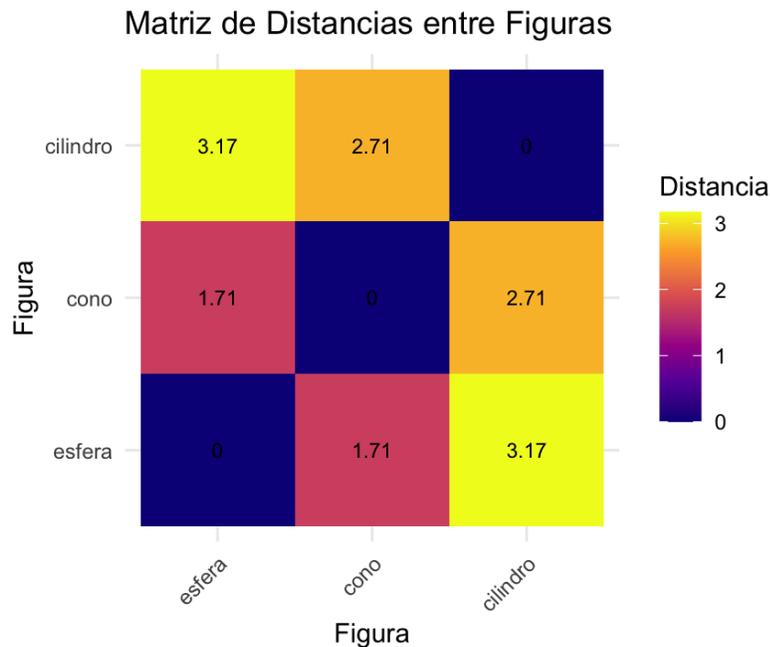
Se espera que:

- La esfera se diferencie del cono y el cilindro, ya que su estructura topológica carece de bordes o vértices.

- En  $k = 3$ , el algoritmo agrupe correctamente las figuras equivalentes en la primera configuración.
- En la segunda prueba con  $k = 3$ , la presencia de dos cilindros pueda influir en la segmentación, posiblemente agrupándolos juntos y dejando la esfera y el cono en clusters separados.

A continuación, se presentan los resultados obtenidos y se analiza cómo la configuración de las figuras afecta la estructura de los clusters formados.

Para la prueba con  $k = 2$ , se utilizaron únicamente una esfera, un cono y un cilindro. La Figura 4.45 muestra la segmentación obtenida, donde el cilindro fue asignado a un cluster separado, mientras que la esfera y el cono fueron agrupados juntos.

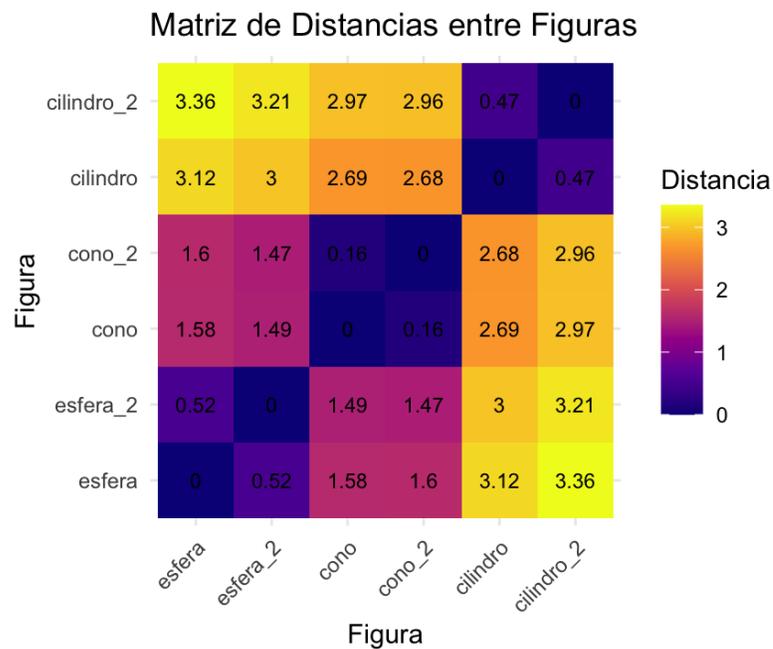


**Figura 4.45:** Distribución de esfera, cono y cilindro por cluster con  $k = 2$ .

En la primera matriz de distancias, se observa cómo la esfera, el cono y el cilindro son comparados en términos de sus características topológicas. La distancia entre la esfera y el cono es de aproximadamente 1.71, mientras que la distancia entre el

cilindro y el cono es de 2.71. La mayor distancia en la matriz se da entre la esfera y el cilindro, con un valor de 3.17. Se destaca que la esfera y el cilindro, a pesar de que ambos son superficies sin borde, tienen una diferencia más marcada debido a la topología del cilindro, que presenta una estructura más extendida.

Para la prueba con  $k = 3$ , se incluyeron dos figuras adicionales del mismo tipo (una segunda esfera, un segundo cono y un segundo cilindro) para evaluar si el método mantenía la coherencia agrupando correctamente figuras del mismo tipo. La Figura 4.46 muestra la segmentación obtenida en este escenario.

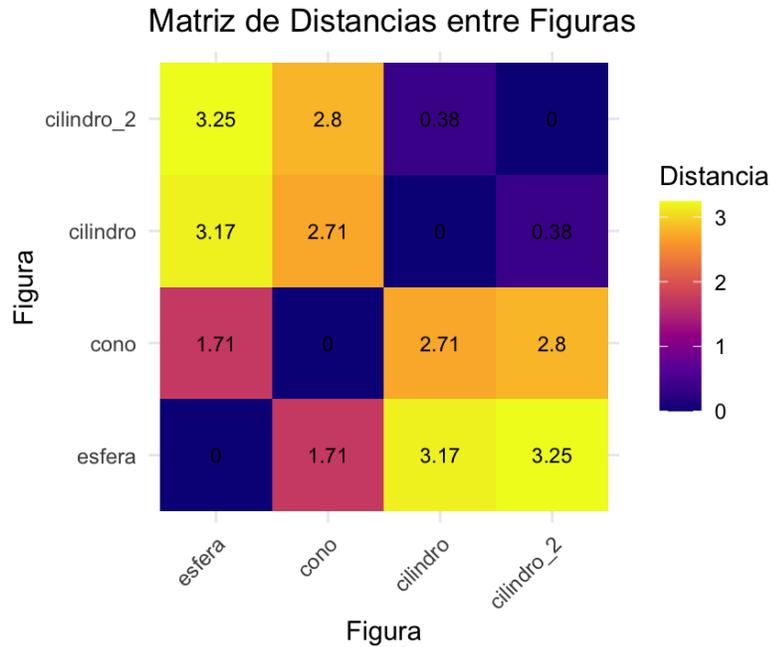


**Figura 4.46:** Distribución de dos esferas, dos conos y dos cilindros por cluster con  $k = 3$ .

La similitud entre pares de figuras equivalentes indica que la homología persistente captura correctamente las diferencias topológicas dentro del conjunto de datos. Se observa que los conos (cono y cono\_2) tienen distancias menores entre sí (1.47 a 1.6), lo que significa que su estructura topológica es más homogénea. La clasificación obtenida en los experimentos con  $k = 3$  es coherente con lo esperado, ya que el algoritmo tiende a agrupar figuras geoméricamente similares basándose en su estructura

topológica.

Adicionalmente, se probó otra configuración con  $k = 3$ , donde se incluyeron dos cilindros en lugar de dos figuras de cada tipo, con el objetivo de evaluar si los cilindros se mantenían en el mismo cluster. La Figura 4.47 muestra los resultados de esta variante.



**Figura 4.47:** Distribución de una esfera, un cono y dos cilindros por cluster con  $k = 3$ .

En esta configuración, se agrega una segunda instancia del cilindro (llamada cilindro\_2). Los valores de distancia muestran que la diferencia entre cilindro y cilindro\_2 es muy pequeña (0.38), lo que indica que el método reconoce su similitud. Por otro lado, la distancia entre el cono y los cilindros sigue siendo mayor (2.71 - 2.80), mientras que la distancia entre la esfera y los cilindros es aún más significativa (3.25).

La distancia del cono con respecto a los cilindros sigue siendo significativa, lo que indica que la estructura topológica del cono es lo suficientemente diferente como para separarlo en otro cluster.

En este caso, se esperaba que los cilindros se agruparan juntos, lo cual se cumple

según la matriz de distancias.

La Tabla 6.4 presenta la asignación de cada figura en los clusters obtenidos en las distintas configuraciones.

En la siguiente sección, se analizarán los resultados obtenidos y se discutirá la coherencia de las agrupaciones respecto a la estructura geométrica y topológica de las figuras simuladas.

### 4.5.3. Impacto de las modificaciones estructurales

Los experimentos realizados muestran que el método de clusterización basado en k-medoids logra capturar diferencias significativas entre las figuras simuladas, reflejando tanto su estructura geométrica como sus características topológicas. A continuación, se discuten los principales hallazgos en función de cómo el algoritmo agrupó las figuras y qué propiedades pueden haber influido en esta segmentación.

Uno de los patrones más consistentes fue la capacidad del método para identificar figuras del mismo tipo cuando se introdujeron duplicados. Por ejemplo, en la prueba con esferas, conos y cilindros, al incluir dos figuras de cada tipo, el algoritmo logró asignarlas al mismo cluster, indicando que los descriptores topológicos utilizados reflejan correctamente la similitud entre objetos equivalentes. Este resultado es coherente con la matriz de distancias obtenida (Figura 4.46), donde se observa que las distancias entre figuras del mismo tipo son significativamente menores en comparación con figuras de diferente categoría. En particular, la distancia entre dos cilindros fue de solo 0.38, mientras que la distancia entre una esfera y un cilindro alcanzó valores de hasta 3.17, lo que evidencia la separación topológica entre estos objetos.

Otro resultado relevante fue la segmentación de las variantes del toro. Se observó que el algoritmo separó el toro estándar del toro entrelazado y del toro plano, agrupando estos últimos en un mismo cluster. Este resultado se puede analizar a través de la matriz de distancia correspondiente (Figura 4.44), donde se aprecia que las distancias entre el toro entrelazado y el toro plano son considerablemente menores en comparación con su distancia al toro estándar. Esto refuerza la idea de que, en

términos de persistencia de homología, estos dos toros comparten una estructura más similar que con el toro estándar.

En cuanto a las modificaciones estructurales relacionadas con la presencia de agujeros, el algoritmo también mostró una clara diferenciación entre figuras con y sin perforaciones. La matriz de distancias correspondiente muestra que las distancias entre figuras con y sin agujeros son significativamente mayores, lo que indica que la introducción de perforaciones genera alteraciones topológicas que afectan la clasificación. En particular, la presencia de agujeros altera los ciclos de dimensión 1 en los diagramas de persistencia, modificando la forma en que el algoritmo percibe la similitud entre estructuras.

Finalmente, en la clusterización con todas las figuras utilizando  $k = 3$ , se observó que el algoritmo generó tres grandes grupos que, en principio, no correspondían a categorías geométricas evidentes. Sin embargo, al analizar la matriz de distancias (Figura 4.41), se pudo notar que los clusters reflejaban patrones de conectividad y persistencia topológica más que la geometría convencional de las figuras. Esto sugiere que el método responde fuertemente a características intrínsecas del espacio de configuración de los puntos, lo que refuerza la idea de que los descriptores topológicos utilizados capturan información estructural que va más allá de la simple forma tridimensional de los objetos.

#### 4.5.4. Análisis de Clustering con el Toro, Esfera y Plano

En esta sección se presentan los resultados obtenidos al aplicar Análisis de Datos Topológicos (TDA) sobre simulaciones generadas en tres figuras geométricas fundamentales: toro, esfera y plano. Para cada una de estas figuras, se generaron 50, 100 y 500 muestras siguiendo un proceso de Poisson con parámetro  $\lambda = 150$ . Posteriormente, se calcularon los landscapes de persistencia para cada simulación y se aplicó un algoritmo de clustering  $k$ -medoids con  $k = 3$  para identificar posibles agrupaciones basadas en sus características topológicas.

En la Tabla 4.1, se muestra la distribución de simulaciones entre los tres clusters detectados, organizados por figura y número de muestras. Se observa que la distribución

es relativamente estable a medida que aumenta el número de simulaciones.

**Tabla 4.1:** Distribución de simulaciones en cada cluster

Número de Simulaciones	Cluster 1 (Toro)	Cluster 2 (Esfera)	Cluster 3 (Plano)
50	50	50	50
100	100	99	101
500	500	499	501

Para evaluar la calidad del clustering obtenido, se utilizó el índice de silueta, el cual mide qué tan bien están separadas las agrupaciones. En la Tabla 4.2, se presentan los valores promedio del índice de silueta obtenidos para cada cantidad de muestras.

**Tabla 4.2:** Índice de Silueta promedio para  $k = 3$

Número de Simulaciones	Cluster 1 (Toro)	Cluster 2 (Esfera)	Cluster 3 (Plano)
50	0.75	0.71	0.76
100	0.75	0.68	0.75
500	0.75	0.74	0.78

Los valores del índice de silueta indican que los clusters están bien definidos, con valores relativamente altos en todas las configuraciones analizadas. Esto sugiere que cada figura se agrupa de manera coherente con su estructura topológica y que la separación entre clusters es clara.

Los resultados obtenidos confirman que el TDA permite diferenciar de manera efectiva estructuras geométricas distintas a través del análisis de homología persistente y landscapes de persistencia. La estabilidad de los clusters a medida que se incrementa el número de muestras sugiere que los invariantes topológicos utilizados son adecuados para la caracterización de estas figuras. Adicionalmente, los valores del índice de

silueta indican una clara separación entre los clusters, lo que refuerza la validez del enfoque aplicado.

#### 4.5.5. Análisis de Clustering con 100 Repeticiones y 11 Figuras Simuladas

En este experimento, se aplicó el algoritmo **k-medoids** utilizando **paisajes de persistencia** como insumo para la clasificación de 11 figuras geométricas simuladas. Cada figura fue generada 100 veces, obteniendo un total de 1100 simulaciones. Se empleó la métrica de distancia  $L_2$  para comparar los paisajes de persistencia y posteriormente agrupar las estructuras topológicas similares.

El proceso de clustering permitió identificar 11 grupos distintos, cada uno caracterizado por una estructura topológica predominante. La Tabla 6.5 muestra la cantidad de instancias de cada figura dentro de cada cluster.

A partir de la clasificación obtenida, se observa que figuras con estructuras similares tienden a agruparse en los mismos clusters. En particular, los **paraboloides con agujeros** se concentran en los clusters 1 y 3, mientras que los **toros** aparecen en varios grupos, reflejando diferencias dentro de sus paisajes de persistencia.

Para evaluar la coherencia del clustering, se calculó el **índice de silueta**, el cual mide qué tan bien separadas están las observaciones dentro de sus clusters. Los valores obtenidos para cada cluster se presentan en la Tabla 4.3.

---

Cluster	Tamaño	Índice de Silueta
1	86	0.16
2	67	0.24
3	124	0.14
4	74	0.37
5	92	0.18
6	75	0.34
7	110	0.21
8	69	0.41
9	99	0.75
10	201	0.68
11	103	0.67

---

**Tabla 4.3:** Índice de silueta promedio por cluster para  $k = 11$ .

A partir de los resultados obtenidos, se pueden destacar las siguientes observaciones:

- Se identificaron 11 clusters con una distribución de figuras coherente con sus características topológicas.
- Los toros y los paraboloides con agujeros presentan mayor dispersión, apareciendo en múltiples clusters debido a sus similitudes en términos de persistencia.
- El índice de silueta mostró valores variados, sugiriendo que algunas figuras tienen estructuras intermedias que las acercan a más de un cluster.

- 
- Este análisis confirma que la homología persistente es una herramienta útil para diferenciar estructuras geométricas complejas basadas en sus características topológicas.

# APLICACIÓN DEL ANÁLISIS A LOS DATOS DE BCI

En este capítulo se presenta la aplicación del análisis a los datos de *Barro Colorado Island* (BCI), un conjunto de datos ecológicos ampliamente utilizado en el estudio de patrones espaciales en comunidades forestales. Se detallará la estructura de los datos, las transformaciones necesarias para su procesamiento y las metodologías aplicadas para extraer información relevante sobre la distribución de las especies.

Inicialmente, se describirá la naturaleza de los datos, incluyendo las variables disponibles y su relevancia en el contexto del análisis espacial. Posteriormente, se abordarán los procedimientos de preprocesamiento y normalización, asegurando la coherencia y calidad de los datos antes de aplicar herramientas de análisis topológico. A partir de estos datos, se calculará la homología persistente y se generarán diagramas para caracterizar la estructura espacial de las especies en diferentes censos.

Además, se aplicará un análisis de clusterización basado en k-medoids, utilizando diagramas de paisaje como insumo, con el objetivo de identificar patrones de distribución espacial entre especies. Se explorará cómo estos patrones varían a lo largo de los censos y qué información ecológica se puede extraer de las agrupaciones obtenidas. Finalmente, se analizarán los resultados obtenidos, evaluando la efectividad del Análi-

sis de Datos Topológicos (TDA) en la identificación de estructuras espaciales en comunidades forestales y su capacidad para capturar relaciones entre especies en términos de persistencia topológica.

Este capítulo proporciona la base empírica del estudio, permitiendo evaluar la aplicabilidad del TDA en el reconocimiento de estructuras espaciales en comunidades forestales y su capacidad para capturar relaciones entre especies en términos de persistencia topológica.

### 5.1. Datos usados en el análisis

En el artículo presentado por [Condit et al. \(2019\)](#) se presenta la base de datos que se describe a continuación, el objetivo de este artículo es fomentar la investigación en ecología, la conservación de las especies y el análisis de patrones espaciales, ofreciendo un conjunto de datos que documenta el crecimiento, mortalidad, entre otras características de las especies a lo largo de varias décadas.

La base de datos utilizada en este análisis corresponde a la parcela de 50 hectáreas de la Isla de Barro Colorado, Panamá. Esta parcela se encuentra delimitada dentro de un rectángulo de bosque de 1000 metros por 500 metros, destinado exclusivamente a investigaciones científicas. Desde el año 1982, todos los árboles leñosos y arbustos con tallos de al menos 1 cm de diámetro han sido censados de manera exhaustiva en esta área.

El primer censo completo de la parcela se realizó en 1982, durante el cual cada árbol fue identificado y etiquetado permanentemente con una placa de aluminio. Posteriormente, se han realizado un total de ocho censos, llevados a cabo aproximadamente cada cinco años, abarcando el período desde 1985 hasta 2015. Estos censos incluyen información detallada de cada individuo, como sus coordenadas espaciales precisas, lo que permite analizar la distribución espacial y la dinámica ecológica del bosque a lo largo del tiempo.

En esta base de datos se han registrado más de 500 especies de árboles leñosos y

arbustos, distribuidas en los 50 hectáreas de la parcela. Con las coordenadas espaciales precisas de cada árbol en cada censo, se cuenta con alrededor de 424.000 puntos observados en este espacio, lo que constituye una de las colecciones de datos más completas para estudios de biodiversidad en bosques tropicales. Esta riqueza de datos proporciona una base sólida para analizar patrones espaciales, dinámicas de poblaciones y otros fenómenos ecológicos.

Los datos utilizados en este análisis corresponden a las tablas analíticas en formato R, que contienen la información necesaria para realizar estudios espaciales detallados. Estas tablas incluyen los datos de los censos y una tabla adicional que describe las especies presentes en la parcela. Estos datos han sido procesados cuidadosamente para garantizar su calidad y consistencia, permitiendo obtener resultados confiables y replicables.

La profundidad y amplitud de esta base de datos la convierten en un recurso invaluable para el estudio de ecosistemas tropicales, ofreciendo una oportunidad única para investigar patrones espaciales y temporales de biodiversidad. Este trabajo busca contribuir al conocimiento generado a partir de estas observaciones, explorando nuevas perspectivas en el análisis de datos espaciales mediante herramientas avanzadas como el Análisis de Datos Topológicos (TDA).

## 5.2. Metodología

El número de puntos usado por especie fue de un mínimo de 100 puntos para obtener una buena representación y de un máximo de 500, ya que debido a limitaciones computacionales, calcular la homología de especies con un mayor número de puntos resultaba demasiado costoso. Con la ventana de observación predefinida y este límite de puntos, el análisis se redujo a un total de 85 especies. Los datos fueron almacenados en estructuras tabulares (`data frames`) para facilitar su posterior procesamiento.

Para cada conjunto de puntos observados, se realizó una proyección al plano bidimensional. Este paso permitió la comparación directa entre las distribuciones tridimensionales generadas en simulaciones y las coordenadas bidimensionales observadas

en las especies. La proyección se realizó definiendo una ventana de observación mediante la función `owin`, asegurando que todos los puntos quedaran limitados a un espacio consistente. Posteriormente, estos puntos fueron convertidos en un formato de patrón puntual `ppp`, el cual sirvió como base para el cálculo de funciones espaciales y la aplicación de herramientas topológicas.

Una vez obtenidos los patrones puntuales, se calculó su homología persistente. A partir de la matriz de distancias entre los puntos, se identificaron características topológicas clave: componentes conexas (dimensión 0) y ciclos (dimensión 1). Estas estructuras fueron analizadas en función de un parámetro de escala  $\varepsilon$ , describiendo cómo emergen y desaparecen a medida que aumenta  $\varepsilon$ . Los resultados se representaron mediante diagramas de persistencia y códigos de barras, visualizando el nacimiento y la muerte de cada característica. Además, se generaron paisajes, una transformación funcional de los diagramas de persistencia que permite comparar cuantitativamente diferentes patrones.

Para resumir la información topológica de cada patrón puntual, se extrajeron características numéricas relevantes, tales como el número total de componentes conexas, la vida media de los ciclos persistentes y el ciclo más persistente. Estas métricas proporcionaron un resumen compacto de las propiedades topológicas, sirviendo como insumo para análisis posteriores. Es importante destacar que, al trabajar con datos bidimensionales, las características topológicas se limitaron a dimensiones 0 (componentes conexas) y 1 (ciclos), ya que las cavidades (dimensión 2) solo aparecen en estructuras tridimensionales.

Para analizar la relación entre especies, se aplicó el método de clusterización **k-medoids** utilizando las representaciones de paisajes como insumo. Este enfoque permitió segmentar las especies en grupos según su similitud topológica, utilizando una métrica de distancia basada en las diferencias entre sus paisajes. La selección del número óptimo de clusters se realizó a partir del método del codo y del coeficiente de silueta, como en los experimentos previos con figuras geométricas. De esta manera, se identificaron patrones estructurales comunes entre especies y se evaluó cómo sus distribuciones espaciales reflejan similitudes en términos de persistencia topológica.

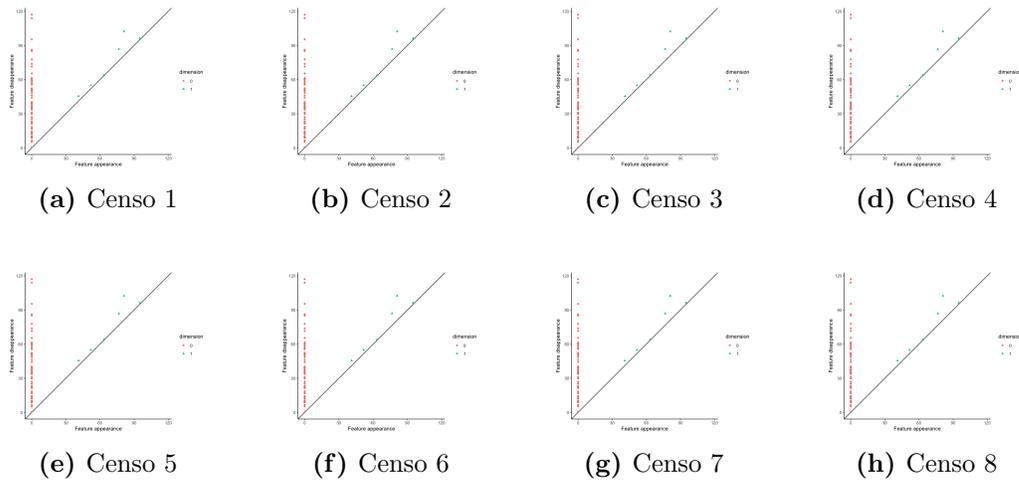
Finalmente, se calcularon las funciones espaciales  $F$ ,  $G$ ,  $J$  y  $K$ , las cuales resumen propiedades de primer y segundo orden de los patrones proyectados en dos dimensiones. Estas funciones permitieron analizar la distribución de las distancias entre puntos y la interacción espacial en cada patrón. En combinación con los invariantes topológicos calculados, estas funciones proporcionaron una descripción robusta y detallada de la estructura de los datos, tanto observados como simulados.

### 5.3. Análisis de la distribución espacial de especies a lo largo de los censos

En esta sección se analiza la distribución espacial de las especies seleccionadas a lo largo de los ocho censos de la Isla de Barro Colorado. Se presentan las visualizaciones de los patrones espaciales de tres especies representativas: *Posola*, *Guazul* y *Pourbi*. A través de estas visualizaciones, se busca identificar posibles cambios en la distribución a lo largo del tiempo.

#### 5.3.1. Distribución de Posola

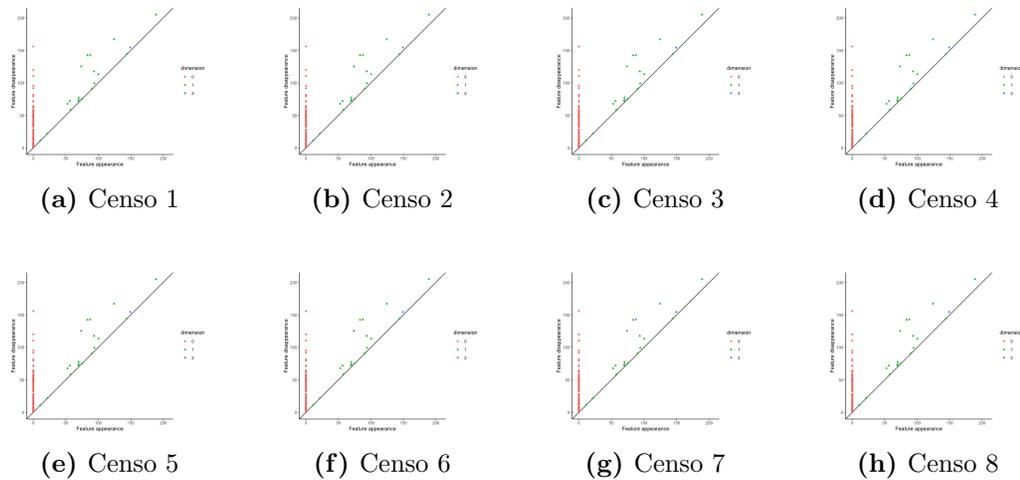
La Figura 5.1 muestra la distribución de Posola en los ocho censos. Se observa que la disposición de los puntos es relativamente estable, sin cambios significativos en su patrón espacial a lo largo del tiempo. La distribución de Posola mantiene una estructura dispersa sin indicios de agrupamiento fuerte en ninguna de las etapas analizadas. Esto sugiere que esta especie no ha experimentado grandes modificaciones en su patrón de crecimiento o en la interacción con su entorno.



**Figura 5.1:** Distribución espacial de *Posola* a lo largo de los censos.

### 5.3.2. Distribución de Guazul

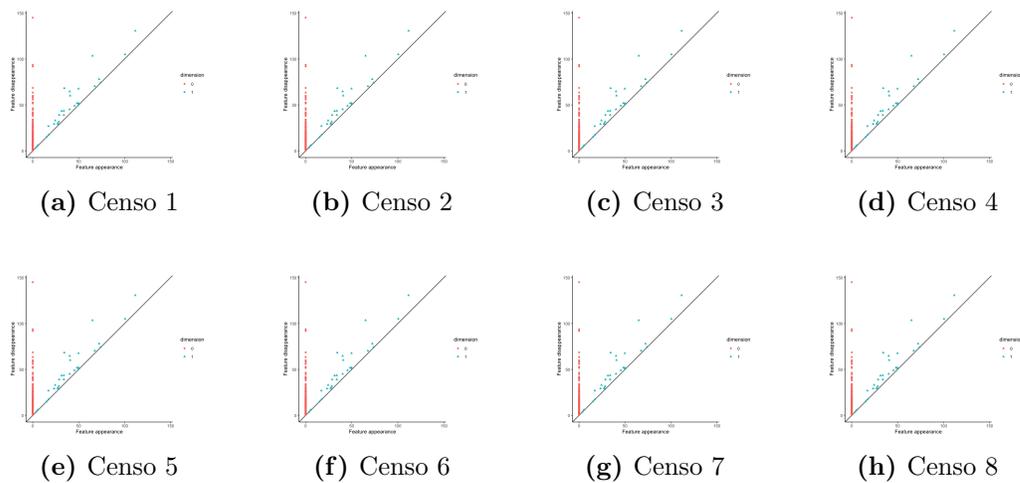
La Figura 5.2 presenta la distribución de Guazul en los ocho censos. Al igual que en *Posola*, no se observan cambios significativos en la estructura espacial de esta especie. Se mantiene un patrón disperso con ciertas regiones de mayor densidad, lo que podría indicar zonas preferenciales para su crecimiento. Sin embargo, no se identifican desplazamientos evidentes ni variaciones abruptas en la densidad de individuos en los diferentes censos.



**Figura 5.2:** Distribución espacial de *Guazul* a lo largo de los censos.

### 5.3.3. Distribución de Pourbi

Finalmente, la Figura 5.3 muestra la distribución de Pourbi en los censos. En este caso, tampoco se observan modificaciones significativas en la distribución espacial a lo largo del tiempo. La estructura general de los puntos permanece uniforme, sin una tendencia clara hacia la concentración o dispersión. Esto refuerza la idea de que muchas especies del bosque tropical mantienen una distribución estable a lo largo del tiempo, sin variaciones abruptas en su ubicación espacial.



**Figura 5.3:** Distribución espacial de *Pourbi* a lo largo de los censos.

Al finalizar el análisis de la distribución espacial de las especies a lo largo de los censos, se puede concluir que, aunque se revisaron las 85 especies seleccionadas, se eligieron tres de ellas de manera arbitraria (Posola, Guazul y Pourbi) para representar gráficamente los cambios observados en los ocho censos. A partir de estas visualizaciones, se observa que la distribución espacial de las especies se ha mantenido prácticamente inalterada en el tiempo, sin presentar patrones evidentes de expansión, contracción o desplazamiento significativo. Si bien en algunos censos se identificaron ligeras variaciones, como la aparición de unos pocos individuos adicionales en ciertas áreas, la estructura general de la distribución permanece estable.

Una posible explicación para esta estabilidad es que los datos corresponden a una parcela de una hectárea dentro de un área de 50 hectáreas destinada exclusivamente a la investigación, lo que significa que el ecosistema analizado ha estado libre de intervención humana directa, como deforestación o actividades agrícolas. Esto sugiere que la dinámica natural de las especies en este entorno no ha generado cambios sustanciales en la distribución espacial de los individuos a lo largo del período de estudio. Además, al tratarse de un bosque tropical maduro, es esperable que las especies analizadas mantengan una estructura poblacional relativamente constante, con patrones de regeneración y mortalidad que no alteran significativamente su

distribución general dentro de la parcela.

#### 5.4. Análisis de agrupaciones de especies mediante k-medoids

Dado que la distribución espacial de las especies analizadas en la Isla de Barro Colorado no presentó cambios significativos a lo largo de los censos, se decidió trabajar con los datos del primer censo disponible. A partir de estos datos, se aplicó el método de k-medoids para identificar grupos de especies con características topológicas similares en función de sus paisajes de persistencia.

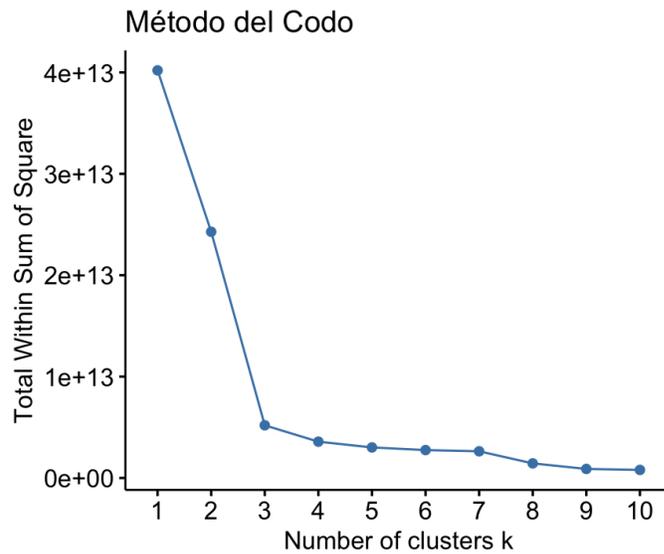
Para determinar el número óptimo de clusters, se utilizaron dos métodos complementarios: el método del codo y el método de la silueta, ambos aplicados a la distancia entre los paisajes obtenidos para cada especie.

El método del codo (Figura 5.4) evalúa la varianza explicada por los clusters en función del número de agrupaciones  $k$ . Se observa que la disminución en la suma total de cuadrados dentro de los clusters se estabiliza a partir de  $k = 3$ , lo que sugiere que tres agrupaciones explican de manera adecuada la variabilidad presente en los datos sin añadir complejidad innecesaria. Este comportamiento indica que, tras este punto, agregar más clusters no aporta una mejora significativa en la partición de los datos.

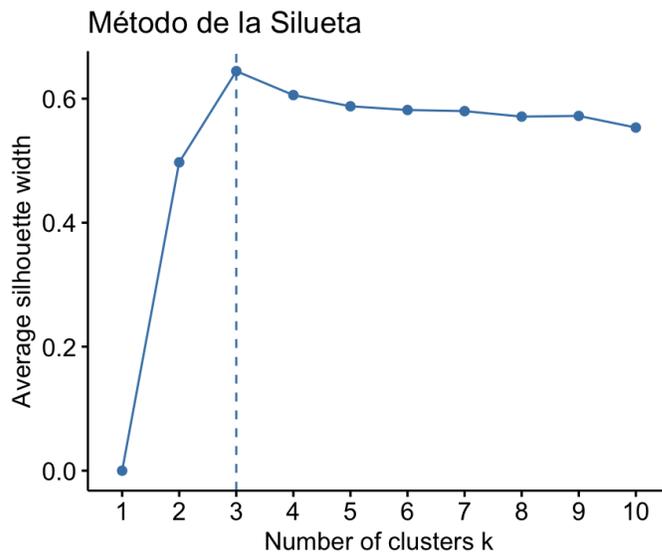
Por otro lado, el método de la silueta (Figura 5.5) mide la coherencia interna de los clusters mediante el coeficiente de silueta, que evalúa cuán similar es un punto con los elementos de su propio cluster en comparación con los de otros clusters. Se observa que el coeficiente de silueta alcanza su valor máximo en  $k = 3$ , lo que refuerza la elección de tres agrupaciones como la mejor configuración para la partición de los datos.

Ambos métodos convergen en la misma solución óptima, lo que indica la existencia de tres estructuras principales en la disposición espacial de las especies analizadas. Esta partición refleja patrones diferenciados en la distribución de los paisajes persistentes asociados a cada especie, proporcionando evidencia cuantitativa para la segmentación de los datos basada en la información topológica extraída.

---



**Figura 5.4:** Método del codo. Se observa un punto de inflexión en  $k = 3$ , lo que sugiere tres agrupaciones óptimas.

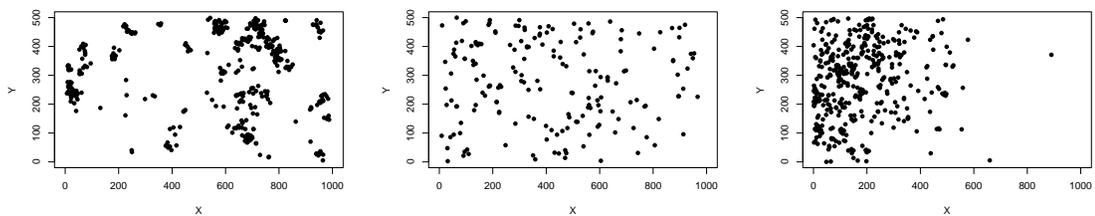


**Figura 5.5:** Método de la silueta. Se observa un valor máximo en  $k = 3$ , lo que indica la mejor configuración de clusters.

### 5.4.1. Resultados del clustering

Tras aplicar k-medoids, se identificaron tres clusters, cuyos elementos fueron determinados en base a la similitud de los paisajes de persistencia. En la Figura 5.6 se presentan los representantes de cada cluster, seleccionados para ilustrar la distribución espacial de las especies dentro de cada grupo. En la Figura 5.7 se muestran los paisajes de dimensión 0, y en la Figura 5.8, los paisajes de dimensión 1.

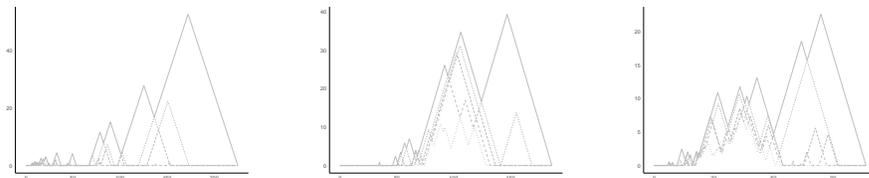
Los detalles de los clusters obtenidos se presentan en el Anexo 6.



**Figura 5.6:** Distribución espacial de las especies adeltr, aegipa e ingath en los clusters 1, 2 y 3.



**Figura 5.7:** Paisajes de persistencia de dimensión 0 para las especies adeltr, aegipa e ingath.



**Figura 5.8:** Paisajes de persistencia de dimensión 1 para las especies adeltr, aegipa e ingath.

Los resultados obtenidos muestran que el análisis topológico mediante k-medoids es capaz de agrupar especies con patrones espaciales similares. En particular, se observa que:

- **Cluster 1:** Incluye especies con una distribución más compacta y con mayor interconectividad entre los individuos.
- **Cluster 2:** Agrupa especies con una disposición más dispersa, con mayor separación entre los individuos.
- **Cluster 3:** Contiene especies con estructuras de agrupamiento intermedio, donde los individuos forman conglomerados dentro de una distribución más amplia.

Los paisajes de dimensión 0 permiten visualizar cómo los puntos permanecen conectados dentro de un intervalo de escala determinado, mientras que los paisajes de dimensión 1 reflejan la formación de ciclos en la distribución de las especies. La comparación entre los clusters indica que las diferencias en la organización espacial de las especies pueden ser capturadas de manera efectiva mediante la homología persistente y la agrupación con k-medoids.

Este análisis proporciona una nueva forma de segmentar especies en función de su distribución espacial, utilizando herramientas de TDA. La estabilidad en la distribución de las especies observada a lo largo de los censos sugiere que estos grupos reflejan estructuras espaciales intrínsecas más que dinámicas temporales. Esto abre la posibilidad de utilizar este enfoque para clasificar especies con base en patrones espaciales, lo que podría complementar estudios ecológicos tradicionales centrados en abundancia y diversidad.

# COMENTARIOS FINALES Y TRABAJOS A FUTURO

En esta tesis se exploró la aplicación del TDA para el estudio de patrones puntuales, tanto en figuras simuladas simples y modificadas como en datos reales del Bosque de Barro Colorado (BCI). Se logró identificar estructuras topológicas relevantes, como componentes conexas, ciclos y cavidades, además de evaluar la capacidad del TDA para diferenciar entre diferentes figuras y distribuciones espaciales de especies. La combinación del TDA con técnicas de clustering permitió observar patrones interesantes, especialmente en la diferenciación de figuras con características topológicas distintivas.

Durante el desarrollo del trabajo se identificaron algunas limitaciones y oportunidades de mejora que podrían ser abordadas en investigaciones futuras. En primer lugar, uno de los desafíos más relevantes fue el costo computacional asociado al cálculo de la homología persistente, especialmente al trabajar con un gran número de puntos o múltiples especies de manera simultánea. Optimizar el código y mejorar la eficiencia de los algoritmos permitiría abarcar un mayor número de especies y realizar análisis más complejos.

Además, el análisis actual se basó únicamente en las coordenadas espaciales  $x$  e  $y$ ,

asumiendo un terreno plano. Sin embargo, el Bosque de Barro Colorado presenta variaciones topográficas significativas, como cerros y pendientes, que podrían influir en la distribución de las especies, por lo tanto, incorporar otras métricas de distancia que consideren la geografía real del terreno podría mejorar la interpretación de los resultados, permitiendo un análisis más realista y robusto.

Finalmente, un enfoque interesante para trabajos futuros sería combinar el TDA con otras técnicas de análisis estadístico o de aprendizaje automático, lo que podría mejorar la capacidad de detección y clasificación de patrones complejos. El TDA ha demostrado su versatilidad al ser aplicado en disciplinas tan diversas como la biología, las redes sociales y la ciencia de materiales, conectando de manera efectiva la teoría matemática con herramientas computacionales para resolver problemas del mundo real ([Carlsson and Vejdemo-Johansson, 2021](#)). Además, el desarrollo de herramientas específicas como `giotto-tda`, diseñado para integrar el TDA con algoritmos de aprendizaje automático, ha abierto nuevas posibilidades para la exploración de datos y la mejora de modelos predictivos en diferentes contextos ([Tauzin et al., 2021](#)). La integración de estos métodos podría ofrecer nuevas perspectivas para el análisis de datos espaciales, ampliando el potencial del TDA en aplicaciones ecológicas y en otros campos científicos.

---

---

# ANEXO: TABLAS

## 6.1. Resultados de Clustering simulaciones

### 6.1.1. Clusterización con todas las figuras

Cluster	Figuras asignadas
1	Cilindro, Paraboloides positivo con agujero, Toro Paraboloides negativo con agujero, Plano con agujero
2	Toro plano, Paraboloides positivo, Paraboloides negativo
3	Esfera, Plano, Toro entrelazado

**Tabla 6.1:** Asignación de figuras a cada cluster según el algoritmo k-medoids.

### 6.1.2. Clusterización de subconjuntos específicos

#### Prueba con planos y paraboloides

Cluster	Figuras asignadas
$k = 2$	
1	Plano, Paraboloide positivo, Paraboloide negativo
2	Plano con agujero, Paraboloide positivo con agujero, Paraboloide negativo con agujero
$k = 3$	
1	Paraboloide negativo, Paraboloide positivo, Plano
2	Paraboloide negativo con agujero, Paraboloide positivo con agujero
3	Plano con agujero

**Tabla 6.2:** Distribución de figuras en los clusters para  $k = 2$  y  $k = 3$  en la prueba con planos y paraboloides.

#### Prueba con distintas variantes de toros

Cluster	Figuras asignadas
1	Toro
2	Toro entrelazado, Toro plano

**Tabla 6.3:** Asignación de figuras en los clusters para  $k = 2$  en la prueba con toros.

## Prueba con esfera, cono y cilindro

Cluster	Figuras asignadas
$k = 2$	
1	Esfera, Cono
2	Cilindro
$k = 3$ , incluyendo dos figuras de cada tipo	
1	Esfera, Esfera <sub>2</sub>
2	Cono, Cono <sub>2</sub>
3	Cilindro, Cilindro <sub>2</sub>
$k = 3$ , incluyendo dos cilindros	
1	Esfera
2	Cono
3	Cilindro, Cilindro <sub>2</sub>

**Tabla 6.4:** Asignación de figuras en los clusters para  $k = 2$  y distintas configuraciones de  $k = 3$  en la prueba con esfera, cono y cilindro.

**Tabla 6.5:** Frecuencia de cada figura dentro de los clusters obtenidos.

Cluster	Figura	Frecuencia
1	Toro	52
	Paraboloide_neg_con_agujero	9
	Paraboloide_pos_con_agujero	8

Cluster	Figura	Frecuencia
	Paraboloide_neg	7
	Paraboloide_pos	6
	Plano_con_agujero	4
2	Toro	26
	Paraboloide_pos_con_agujero	17
	Paraboloide_neg_con_agujero	16
	Plano_con_agujero	5
	Cilindro	1
	Paraboloide_pos	1
	Toro_plano	1
3	Paraboloide_neg_con_agujero	48
	Paraboloide_pos_con_agujero	45
	Toro	20
	Paraboloide_pos	4
	Plano_con_agujero	4
	Paraboloide_neg	3

Cluster	Figura	Frecuencia
4	Paraboloide_pos_con_agujero	29
	Paraboloide_neg_con_agujero	27
	Plano_con_agujero	16
	Toro	2
5	Paraboloide_neg	43
	Paraboloide_pos	36
	Toro_plano	13
6	Toro_plano	75
7	Paraboloide_pos	52
	Paraboloide_neg	46
	Toro_plano	10
	Plano	2
8	Plano_con_agujero	67
	Paraboloide_pos_con_agujero	1
	Toro_plano	1

---

Cluster	Figura	Frecuencia
9	Esfera	99
10	Toro_entrelazado	100
	Plano	98
	Esfera	1
	Paraboloide_neg	1
	Paraboloide_pos	1
11	Cilindro	99
	Plano_con_agujero	4

---

## 6.2. Resultados de Clustering para los datos BCI

### Dimensión 0

Cluster	Especies
1	acalma, adeltr, apeiti, ardife, ast2gr, bactba, bactc1, clidde, cupala, cuparu, geniam, guargr, guazul, heisac, ingaco, ingafa, lacmpa, licahy, licapl, lindla, malpro, maytsc, myrcga, nectpu, ormocr, ormoma, perexa, pipea1, pipecu, pipepe, psycde, psycg1, psycg3, quasam, sapiau, sponmo, sterap, tab1gu, termob, trataas, vismba
2	aegipa, allops, andiin, ast1st, casear, casesy, ceibpe, celtsc, cestme, chr2ca, coccco, cordal, dendar, dio2ar, ery1co, ery2pa, guapst, guetfo, hameax, hampap, huracr, hyeral, micoho, nectci, neeaam, phoeci, pipeae, pipere, pla1pi, pla2el, posola, solaha, sympgl, tropra, turpoc, virosu, xyl2ol, zantpr
3	ingape, ingath, macrgl, pourbi, psycli, thevah

### Dimensión 1

Cluster	Especies
1	acalma, adeltr, bactba, bactc1, cordal, guargr, ingafa, maytsc, nectci, posola, sapiau, sterap, sympgl, turpoc
2	aegipa, allops, ast1st, casear, casesy, cestme, chr2ca, coccco, ery1co, ery2pa, guapst, guetfo, hameax, hampap, lacmpa, ormoma, phoeci, pipere, pla1pi, pla2el, solaha, tropra, virosu, zantpr

Cluster	Especies
3	andiin, apeiti, ardife, ast2gr, ceibpe, celtsc, clidde, cupa- la, cuparu, dendar, dio2ar, geniam, guazul, heisac, hu- racr, hyeral, ingaco, ingape, ingath, licahy, licapl, lind- la, macrgl, malpro, micoho, myrega, nectpu, neeaam, ormocr, perexa, pipea1, pipeae, pipecu, pipepe, pourbi, psycde, psycg1, psycg3, psycli, quasam, sponmo, tab1gu, termob, thevah, trataas, vismba, xyl2ol

---

---

## Bibliografía

- Anscombe, F. J. (1973). Graphs in statistical analysis. *The american statistician*, 27(1):17–21.
- Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial point patterns: methodology and applications with R*. CRC press.
- Baddeley, A. and Turner, R. (2005). Spatstat: an r package for analyzing spatial point patterns. *Journal of statistical software*, 12:1–42.
- Biscio, C. A., Chenavier, N., Hirsch, C., and Svane, A. M. (2020). Testing goodness of fit for point processes via topological data analysis. *Electronic Journal of Statistics, Electron. J. Statist.*
- Brunson, J. C., Demkowicz, B., and Choudhary, S. (2024). *tdaunif: Uniform Manifold Samplers for Topological Data Analysis*. R package version 0.2.0.
- Bubenik, P. et al. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16(1):77–102.
- Carlsson, G. and Vejdemo-Johansson, M. (2021). *Topological data analysis with applications*. Cambridge University Press.
- Carlsson, G., Zomorodian, A., Collins, A., and Guibas, L. (2004). Persistence barcodes for shapes. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 124–135.

- Chazal, F. and Michel, B. (2021). An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, 4:667963.
- Condit, R., Pérez, R., Aguilar, S., Lao, S., Foster, R., and Hubbell, S. (2019). Complete data from the barro colorado 50-ha plot: 423617 trees, 35 years. URL <https://doi.org/10.15146/5xcp-0d46>, 2:13.
- Daley, D. J., Vere-Jones, D., et al. (2003). *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer.
- Delfinado, C. J. A. and Edelsbrunner, H. (1993). An incremental algorithm for betti numbers of simplicial complexes. In *Proceedings of the ninth annual symposium on Computational geometry*, pages 232–239.
- Edelsbrunner, Letscher, and Zomorodian (2002). Topological persistence and simplification. *Discrete & computational geometry*, 28:511–533.
- Fasy, B. T., Kim, J., Lecci, F., Maria, C., Millman, D. L., Rouvreau, V., and Kim, M. J. (2024). Package ‘tda’.
- Frosini, P. (1990). A distance for similarity classes of submanifolds of a euclidean space. *Bulletin of the Australian Mathematical Society*, 42:407 – 415.
- Fugacci, U., Scaramuccia, S., Iuricich, F., De Floriani, L., et al. (2016). Persistent homology: a step-by-step introduction for newcomers. In *STAG*, pages 1–10.
- Ghrist, R. (2008). Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75.
- Joharinad, P. and Jost, J. (2023). *Mathematical principles of topological and geometric data analysis*. Springer.
- Matejka, J. and Fitzmaurice, G. (2017). Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annea-

- ling. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 1290–1294.
- Moller, J. and Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. CRC press.
- Ohanuba, F. O., Ismail, M. T., Ali, M. K. M., Alih, E., and Ezra, P. N. (2021). On the cluster validity test (s) in unsupervised machine learning tda approach for atmospheric river patterns on flood detection in nigeria. *Scientific African*, 13.
- Shin, K. (2019). Genre classification: A topological data analysis approach. *Macalester College, MATH 471 Thesis*.
- Tauzin, G., Lupo, U., Tunstall, L., Pérez, J. B., Caorsi, M., Medina-Mardones, A. M., Dassatti, A., and Hess, K. (2021). giotto-tda: A topological data analysis toolkit for machine learning and data exploration. *Journal of Machine Learning Research*, 22(39):1–6.
- Vakil, R. (2017). *The Rising Sea: Foundations of Algebraic Geometry*. Stanford, <http://math.stanford.edu/~vakil/216blog/FOAGnov1817public.pdf>.
- Wadhwa, R. R., Williamson, D. F., Dhawan, A., and Scott, J. G. (2018). Tdastats: R pipeline for computing persistent homology in topological data analysis. *Journal of open source software*, 3(28):860.
- You, K. and Yu, B. (2021). *TDAkit: Toolkit for Topological Data Analysis*. R package version 0.1.2.