



Universidad Politécnica
de Madrid

**Escuela Técnica Superior de
Ingenieros Informáticos**



Grado en Ingeniería Informática

Trabajo Fin de Grado

**Análisis topológico de trayectorias en
baja y alta dimensión**

Autor: Carlo Alessandro Cabrera
Tutor: Juan Antonio Fernández del Pozo

Madrid, JUNIO - 2025

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Grado
Grado en Ingeniería Informática

Título: Análisis topológico de trayectorias en baja y alta dimensión

JUNIO - 2025

Autor: Carlo Alessandro Cabrera

Tutor: Juan Antonio Fernández del Pozo

Departamento de Inteligencia Artificial (DIA)

Escuela Técnica Superior de Ingenieros Informáticos

Universidad Politécnica de Madrid

Índice general

1. Introducción	3
1.1. Motivación del proyecto	3
1.2. Contexto del proyecto	3
1.3. Objetivos	4
1.4. Estructura del Documento	4
2. Trabajo relacionado y Estado del Arte	6
2.1. Introducción al Análisis de Trayectorias	6
2.1.1. Definición y relevancia del estudio de trayectorias	6
2.1.2. Aplicaciones y Principales Retos	7
2.2. Fundamentos Matemáticos y Teóricos	7
2.2.1. Representación de Trayectorias	8
2.2.2. Introducción al TDA	8
2.3. Estado del Arte	8
2.3.1. Métodos de Análisis de Trayectorias	9
2.3.2. Herramientas y Frameworks	9
2.4. Trabajos Previos	10
2.4.1. Avances recientes en representación topológica de trayectorias	10
2.4.2. Integración de TDA con modelos de aprendizaje profundo	11
2.4.3. Aplicaciones emergentes del TDA en trayectorias biológicas y medioambientales	11
3. Análisis topológico de trayectorias en alta y baja dimensión	12
3.1. Modelización de los datos	12
3.1.1. Selección de los datos	12
3.1.2. Preparación de los datos	13
3.2. Metodología y herramientas para el análisis	13
3.2.1. Metodología de desarrollo	13
3.2.2. Herramientas utilizadas	14
3.3. Diseño e implementación de soluciones con TDA	14
3.3.1. Carga de Trayectorias	14
3.3.2. Visualización de Nube de Puntos	15
3.3.3. Cálculo de Homologías Persistentes	16
3.3.4. Extracción de Descriptores con Mapper	17
3.3.5. Análisis y comparación con PCA	20
3.4. Técnicas estadísticas para el análisis de trayectorias GPS	21
4. Resultados de TDA sobre trayectorias	23

4.1. Resultados	23
4.1.1. Análisis de clustering y comparación con métodos de TDA	28
5. Conclusiones	32
5.1. Evaluación de objetivos	32
5.2. Trabajo futuro	33
5.3. Conclusiones personales	34
5.4. Análisis del Impacto	35
5.4.1. Impacto general	35
5.4.2. Impacto en los Objetivos de Desarrollo Sostenible (ODS)	36
Bibliografía	38
A. Anexo	42
A.1. Materiales Desarrollados	42
A.2. Uso de Inteligencia Artificial Generativa	42
A.3. Turnitin	42

Índice de Figuras

3.1. Visualización de la Nube de Puntos de las trayectorias de Geolife	15
3.2. Diagrama de Persistencia de un Conjunto de trayectorias	16
3.3. Definición del objeto VietorisRipsPersistence	17
3.4. Grafo de Mapper generado a partir de una trayectoria del Dataset	19
4.1. Nube de puntos de una Trayectoria	24
4.2. Diagrama de Persistencia de una trayectoria	25
4.3. Matriz de Distancia de los puntos de la trayectoria	25
4.4. Análisis con PCA	26
4.5. Grafo Mapper de múltiples trayectorias	27
4.6. Diagrama de Persistencia para el conjunto de trayectorias	27
4.7. Análisis PCA de características topológicas	29
4.8. Resultados de clustering sobre reducción t-SNE: DBSCAN (izquierda) y KMeans (derecha)	30
4.9. PCA tradicional sobre trayectorias	31
5.1. Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030 [3]	36

Resumen

Este Trabajo de Fin de Grado explora el uso de herramientas del Análisis Topológico de Datos (TDA) para el estudio de trayectorias espaciales en contextos de baja y alta dimensión. A través de técnicas como la homología persistente y el algoritmo Mapper, se extraen características estructurales de trayectorias GPS, permitiendo identificar patrones de conectividad, ciclos y componentes relevantes que escapan a los métodos clásicos de análisis. El trabajo incluye una implementación práctica utilizando la biblioteca Giotto-TDA en Python, aplicada sobre el conjunto de datos Geolife. Además, se comparan los resultados topológicos con enfoques tradicionales como PCA, t-SNE y algoritmos de agrupamiento como DBSCAN y KMeans.

Los resultados demuestran que el enfoque topológico permite detectar estructuras globales robustas, como bucles o trayectorias anómalas, con mayor sensibilidad que las técnicas estadísticas convencionales. Asimismo, se destacan las ventajas del uso de TDA en entornos ruidosos o con alta complejidad geométrica. Se discuten también las implicaciones de este análisis para aplicaciones en movilidad urbana, transporte aéreo y sistemas dinámicos, así como su contribución a objetivos de desarrollo sostenible. Finalmente, se proponen líneas futuras de investigación orientadas a integrar TDA con aprendizaje automático profundo y explorar nuevas representaciones vectoriales topológicas.

Palabras Clave: Análisis Topológico de Datos, Homología Persistente, Mapper, Trayectorias GPS, Clustering, PCA, t-SNE, Geolife.

Abstract

This Bachelor Thesis investigates the application of Topological Data Analysis (TDA) techniques to the study of spatial trajectories in both low and high-dimensional contexts. Using tools such as persistent homology and the Mapper algorithm, the project extracts structural features from GPS trajectory data, enabling the identification of connectivity patterns, loops, and relevant components that are often missed by classical methods. The practical implementation is based on the Giotto-TDA Python library and is applied to the Geolife dataset. Results from TDA are compared to traditional approaches such as PCA, t-SNE, and clustering algorithms like DBSCAN and KMeans.

The findings show that topological methods offer a more sensitive detection of robust global structures—such as loops or outlier trajectories—than conventional statistical techniques. TDA proves particularly advantageous in noisy or geometrically complex settings. The work discusses its relevance for applications in urban mobility, air traffic, and dynamic systems, as well as its alignment with Sustainable Development Goals. Future research directions are proposed, including integration with deep learning models and the exploration of novel topological vector representations.

Keywords: Topological Data Analysis, Persistent Homology, Mapper, GPS Trajectories, Clustering, PCA, t-SNE, Geolife.

Capítulo 1

Introducción

1.1. Motivación del proyecto

El desarrollo de este Trabajo Fin de Grado surge de la necesidad actual de contar con metodologías robustas que permitan extraer información cualitativa y cuantitativa a partir de datos complejos. El análisis topológico y estadístico se presenta como una herramienta innovadora para la interpretación de fenómenos en ámbitos tan diversos como la movilidad urbana (por ejemplo, seguimiento de vehículos y dispositivos móviles) o procesos abstractos de optimización y sistemas dinámicos (como el estudio de atractores en modelos caóticos). La capacidad de capturar la "forma" subyacente de los datos mediante técnicas de análisis topológico de datos (TDA) y combinarlas con métodos de clustering y análisis estadístico, justifica el objetivo de este proyecto, ofreciendo potenciales aplicaciones en inteligencia artificial, robótica y análisis de sistemas complejos.

1.2. Contexto del proyecto

Este Trabajo de Fin de Grado se enmarca en el emergente campo del análisis topológico de datos, que ha ganado relevancia en los últimos años gracias a su eficacia para detectar estructuras y patrones en datos de alta complejidad. Diversos estudios y avances hechos en los últimos años han demostrado el valor de aplicar herramientas topológicas para el análisis de trayectorias, tanto en contextos físicos como en aplicaciones más abstractas [1]. La revisión del estado del arte evidencia la utilidad de integrar enfoques teóricos –como la construcción de complejos simpliciales y el uso de homología persistente – con técnicas estadísticas y de clustering para lograr una interpretación integral de los datos. Asimismo, cabe destacar la relevancia de trabajos previos que han establecido las bases para la modelización y análisis de series temporales y datos funcionales [2], proporcionando el contexto necesario para este proyecto.

1.3. Objetivos

El objetivo principal de este Trabajo Fin de Grado es desarrollar una metodología integral para el análisis topológico y estadístico de trayectorias en espacios de baja y alta dimensión, series temporales multivariantes, que permita extraer información relevante para la interpretación y clasificación de dichos datos.

- **Definir y documentar el problema:** Delimitar las variables y estructuras de datos que representan las trayectorias, tanto en escenarios físicos (movilidad, dispositivos, etc.) como en contextos abstractos (modelos dinámicos y optimización).
- **Diseñar la metodología de análisis:** Integrar técnicas de TDA para extraer características topológicas con métodos estadísticos y algoritmos de clustering que faciliten la identificación de patrones y agrupaciones en los datos.
- **Implementar y evaluar la propuesta:** Realizar un estudio experimental que valide la eficacia de la metodología, evaluando aspectos como la robustez frente al ruido, la sensibilidad de los algoritmos y la complejidad computacional.
- **Interpretar y sintetizar resultados:** Analizar los resultados obtenidos para ofrecer conclusiones que aporten al conocimiento en el análisis de trayectorias y que puedan abrir nuevas líneas de investigación en el campo.

1.4. Estructura del Documento

Este Trabajo Fin de Grado se organiza en diversos capítulos, cada uno de ellos enfocado en aspectos específicos del desarrollo del proyecto. A continuación, se describe brevemente el contenido de cada capítulo

- **Capítulo 1: Introducción.** Se expone la motivación para realizar el proyecto, el contexto en el que se encuentra, los objetivos del mismo y una breve descripción de la estructura del documento.
- **Capítulo 2: Trabajo relacionado y Estado del Arte.** Se realiza una revisión de la literatura y se presentan los conceptos y herramientas fundamentales relacionados con TDA, modelización de trayectorias y técnicas de clustering, contextualizando el trabajo en el marco de estudios previos.
- **Capítulo 3: Análisis Topológico de trayectorias en alta y baja dimensión** Se detalla el proceso de elección, preprocesamiento y modelización de datos, y se describe la metodología propuesta para la extracción y análisis de características topológicas de las trayectorias, así como la implementación de la propuesta, el desarrollo de algoritmos específicos y la realización de estudios experimentales para evaluar la eficacia del enfoque.
- **Capítulo 4: Resultados de TDA sobre trayectorias** Se presentan y analizan los resultados obtenidos durante la realización del trabajo, realizando comparativas con métodos existentes y discutiendo las implicaciones prácticas y teóricas del trabajo.

Introducción

- **Capítulo 5: Conclusiones** Se evalúan los objetivos declarados en el trabajo, se proponen futuras líneas de investigación que podrían extender los hallazgos obtenidos, se exponen conclusiones personales del autor sobre el trabajo realizado, se identifican las limitaciones y se realiza un análisis del impacto del trabajo, y las decisiones tomadas a lo largo del mismo, que tienen como base el impacto y el potencial impacto respecto a los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030 que sean relevantes para el trabajo realizado [3] .
- **Anexos y Bibliografía.** Se incluye material complementario, detalles técnicos y la documentación de las referencias utilizadas a lo largo del proyecto.

Esta estructura permite una presentación clara y ordenada de todo el proceso de análisis topológico de trayectorias, facilitando la comprensión de la metodología, los resultados y la relevancia del trabajo dentro del campo de la inteligencia computacional y el análisis de datos.

Capítulo 2

Trabajo relacionado y Estado del Arte

El presente capítulo tiene como objetivo contextualizar el trabajo realizado dentro del marco teórico y tecnológico actual. Para ello, se analiza en primer lugar el concepto de trayectoria, su relevancia en diferentes ámbitos de aplicación y los principales retos asociados a su estudio. A continuación, se presentan los fundamentos del Análisis Topológico de Datos (TDA), una técnica emergente para el estudio estructural de datos complejos como las trayectorias. Finalmente, se recogen las principales metodologías y herramientas utilizadas en el estado del arte, así como los trabajos previos que han explorado enfoques similares en distintos dominios.

2.1. Introducción al Análisis de Trayectorias

El análisis de trayectorias es una disciplina interdisciplinar que estudia la evolución espacial y temporal de objetos móviles. Estos datos, representados generalmente como secuencias de coordenadas en el espacio-tiempo, permiten modelar y comprender comportamientos dinámicos en una amplia variedad de contextos: desde la movilidad urbana y el tráfico aéreo, hasta aplicaciones biomédicas o industriales. Dada su naturaleza compleja, el análisis de trayectorias requiere herramientas matemáticas y computacionales que permitan extraer información relevante, detectar patrones y representar de manera eficiente la geometría del movimiento.

2.1.1. Definición y relevancia del estudio de trayectorias

El análisis de trayectorias se refiere al estudio de las rutas o caminos que siguen objetos o agentes a lo largo del tiempo en un espacio determinado. Una trayectoria puede definirse como la secuencia de posiciones (o puntos) en un espacio métrico, lo cual permite modelar dinámicas y comportamientos en una amplia variedad de contextos. La relevancia del estudio radica en que estas secuencias no solo capturan información sobre la posición, sino que también nos permiten observar patrones temporales y espaciales críticos para entender fenómenos complejos.

2.1.2. Aplicaciones y Principales Retos

La versatilidad del análisis de trayectoria permite que estas técnicas pueden ser usadas para una gran variedad de campos sin relación aparente, con una alta eficacia, entre ellos destacan:

- **Movilidad urbana:** Optimización de rutas y gestión del tráfico de vehículos y dispositivos móviles. [4]
- **Transporte aéreo:** Clasificación y detección de patrones en datos de vuelo, identificación de retrasos y desviaciones.[5] [2]
- **Medicina personalizada:** Análisis de trayectorias clínicas para predecir progresión de enfermedades. [6] [7]
- **Sistemas físicos y dinámicos:** Modelado de trayectorias en espacios de fase o espacios de búsqueda en problemas complejos. [8]

No obstante, el análisis de trayectorias también presenta retos importantes:

- **Alta dimensionalidad:** Las trayectorias suelen estar definidas por múltiples variables (coordenadas espaciales, velocidad, tiempo, etc.), lo que dificulta su visualización y análisis sin aplicar técnicas de reducción de la dimensionalidad.
- **Ruido e incertidumbre:** Los datos reales pueden estar contaminados por ruido o inconsistencias, lo que exige métodos robustos para extraer patrones significativos.
- **Complejidad en la modelización:** La representación adecuada de trayectorias requiere elegir modelos matemáticos que capturen tanto la parte local (geometría) como la global (topología) de los datos.

2.2. Fundamentos Matemáticos y Teóricos

El análisis riguroso de trayectorias y su estructura subyacente requiere una base matemática sólida. En esta sección se presentan los conceptos fundamentales que sustentan el tratamiento teórico de las trayectorias desde una perspectiva geométrica y topológica. Por un lado, se describen las distintas formas de representar trayectorias como objetos matemáticos, ya sea como series temporales, curvas en espacios métricos o conjuntos de puntos en espacios de dimensión arbitraria. Por otro, se introduce el Análisis Topológico de Datos (TDA), un marco emergente basado en la topología algebraica, que permite capturar propiedades globales de los datos como la conectividad, la presencia de ciclos y la persistencia de estructuras en distintas escalas.

Estos fundamentos proporcionan las herramientas necesarias para abordar el análisis de trayectorias de forma robusta y explicativa, especialmente en contextos con ruido, alta dimensionalidad o complejidad estructural.

2.2.1. Representación de Trayectorias

La representación de trayectorias se fundamenta en distintos enfoques matemáticos. Tradicionalmente, las trayectorias se modelan como series temporales o datos funcionales, es decir, como funciones que asignan una posición en el espacio a cada instante de tiempo. Esta representación permite aplicar técnicas de análisis estadístico y geométrico para extraer características relevantes (por ejemplo, tendencias, periodicidad y anomalías).

Además, la representación geométrica de trayectorias como conjuntos de puntos en un espacio de dimensión \mathcal{D} posibilita el uso de distancias y métricas (como la distancia Euclidiana, la Distancia de Dynamic Time Warping o la Frechét) para evaluar similitudes entre trayectorias. Estos métodos tradicionales se han complementado, en los últimos años, con enfoques que integran herramientas de análisis topológico, ofreciendo una visión más global e intrínseca de la forma subyacente de los datos [9].

2.2.2. Introducción al TDA

El Análisis Topológico de Datos (TDA) es un marco matemático que utiliza herramientas de la topología algebraica para extraer características robustas e invariantes de conjuntos de datos complejos. Su principal ventaja reside en la capacidad de capturar la “forma” o estructura global de los datos, incluso en presencia de ruido y en contextos de alta dimensionalidad.

Una de las técnicas fundamentales de TDA es la homología persistente, que consiste en analizar cómo cambian las estructuras topológicas (como componentes conexas, ciclos y vacíos) a medida que se varía un parámetro de escala en la construcción de complejos simpliciales. Esta metodología permite generar diagramas de persistencia y barcodes, que sintetizan la vida de las características topológicas a lo largo de diferentes escalas [10].

El uso de TDA se ha extendido en el campo de la ciencia de datos y el aprendizaje automático, ya que ofrece una representación compacta y estable de las características estructurales de los datos, complementando modelos basados en geometría o estadísticas convencionales [11].

2.3. Estado del Arte

El análisis de trayectorias ha experimentado un notable desarrollo en los últimos años, impulsado por la disponibilidad creciente de datos espaciales y temporales procedentes de sensores, sistemas de navegación y registros clínicos, entre otros. Esta sección presenta una revisión de los enfoques más relevantes en la literatura, agrupándolos en metodologías tradicionales y técnicas basadas en el Análisis Topológico de Datos (TDA).

Los métodos tradicionales, como los basados en geometría, estadística o clustering, han sido ampliamente utilizados para modelar, comparar y clasificar trayectorias. Sin embargo, estos métodos a menudo presentan limitaciones frente a datos ruidosos o de alta complejidad. En respuesta a ello, el TDA ha emergido como una alternativa robusta, capaz de extraer información sobre la estructura global de los datos mediante herramientas como la homología persistente y el algoritmo Mapper.

En esta sección se describen los métodos actuales más representativos, las bibliotecas y frameworks más utilizados, así como trabajos previos destacados en distintos dominios de aplicación, con especial énfasis en movilidad, transporte y sistemas dinámicos.

2.3.1. Métodos de Análisis de Trayectorias

En los últimos años se han desarrollado diferentes métodos para el análisis de trayectorias, que pueden agruparse en dos grandes categorías:

- **Métodos tradicionales:** Incluyen técnicas estadísticas y geométricas como el clustering con k-means, DBSCAN, o técnicas de reducción de dimensionalidad como PCA y t-SNE. Estos métodos se centran en capturar información local o global a partir de distancias y similitudes entre trayectorias.
- **Métodos topológicos:** Basados en TDA, permiten extraer características invariantes ante transformaciones suaves. Entre ellos:
 - **Homología Persistente:** Detecta y cuantifica la aparición y desaparición de ciclos, componentes conexas y cavidades a múltiples escalas [10].
 - **Algoritmo Mapper:** Construye una representación en forma de grafo de la topología de los datos, útil para exploración visual y descubrimiento de clases latentes [11].

Técnicas recientes combinan estos métodos con aprendizaje automático, dando lugar al campo emergente del *topological machine learning*, que aprovecha estructuras topológicas para mejorar la generalización y robustez de los modelos. [9]

2.3.2. Herramientas y Frameworks

El avance del TDA ha sido posible gracias al desarrollo de bibliotecas especializadas, entre las que destacan en el entorno de Python:

- **GUDHI:** Biblioteca en C++/Python para la construcción de complejos simpliciales y cálculo de diagramas de persistencia. [12]
- **Dionysus y DIPHA:** Herramientas eficientes para la computación de homología persistente. [13]
- **Giotto-TDA:** Librería en Python que integra TDA con pipelines de machine learning, permitiendo la vectorización de diagramas y su uso directo en clasificadores o modelos de clustering. [10] [14]

En el entorno R, destacan paquetes como **tdaverse** [15], un conjunto de herramientas diseñadas para análisis topológico en R, incluyendo el paquete **ripserr** [16] que proporciona acceso rápido a cálculo de homología persistente mediante envoltorios a bibliotecas C++; **phutil** [17] para manipulación avanzada de datos topológicos; y **tdarec** [18], orientado al estudio de curvas de entropía y complejidad topológica.

2.4. Trabajos Previos

Diversos estudios han abordado el análisis de trayectorias en distintos dominios, combinando herramientas estadísticas, geométricas y topológicas.

En el ámbito de la aviación, se ha utilizado clustering sobre trayectorias de vuelo para gestionar el tráfico aéreo y anticipar desviaciones significativas [5]. En medicina, el análisis de trayectorias temporales de exposición de pacientes ha permitido predecir recaídas en enfermedades crónicas como la esclerosis múltiple, mediante modelos que integran secuencias temporales en pipelines de aprendizaje automático [6].

Por ejemplo, en contextos de movilidad urbana, el análisis de trayectorias posibilita el seguimiento y optimización del tráfico de vehículos. En la gestión del tráfico aéreo, este ha sido utilizado para detectar y predecir desviaciones o anomalías en trayectorias de vuelo, mediante el uso de técnicas de clustering y predicción, véase: [5]. Asimismo, en el ámbito biomédico, se ha aplicado para la predicción de recaídas en esclerosis múltiple, modelando trayectorias de exposición del paciente, lo cual resalta la flexibilidad del enfoque para representar dinámicas longitudinales en medicina personalizada [6]. Adicionalmente, se han desarrollado estudios que aplican herramientas topológicas al diagnóstico temprano de enfermedades como el cáncer [19] [20], la diabetes [21] [22] o trastornos neurodegenerativos como el Alzheimer y el Parkinson [23] [24] [25] [26].

En un contexto más puramente matemático, el análisis de trayectorias permite modelar la evolución de sistemas dinámicos, como ocurre con los atractores caóticos, los cuales representan patrones emergentes en el espacio de fases. Estos patrones pueden capturarse de forma más robusta mediante herramientas topológicas, como el Análisis Topológico de Datos (TDA), que ha mostrado su eficacia en múltiples dominios científicos y de ingeniería.

Desde un punto de vista metodológico, los trabajos de revisión como [9] [11] presentan un panorama exhaustivo del uso del TDA en machine learning, destacando su papel para la robustez en entornos de datos ruidosos o no estructurados. Estas contribuciones consolidan al TDA como una herramienta clave para la representación y el aprendizaje a partir de trayectorias en contextos diversos.

2.4.1. Avances recientes en representación topológica de trayectorias

La representación topológica de trayectorias ha incorporado innovaciones centradas en resumir las propiedades globales de las curvas de movimiento. Herramientas como los *diagramas de persistencia* se traducen ahora en características vectoriales mediante paisajes de persistencia, imágenes de persistencia u otras superficies topológicas, preservando la información global del conjunto de datos a múltiples escalas. Por ejemplo, Cuerno et al. construyen «huellas» topológicas de aeropuertos a partir de los retrasos y desviaciones de vuelo usando paisajes de persistencia, capturando así patrones globales de comportamiento en el tráfico aéreo [27]. De forma similar, la biblioteca *tramoTDA* representa trayectorias a través de diagramas e imágenes de persistencia para su clasificación en contextos de planificación urbana y navegación marítima, identificando patrones sutiles que escapan a los métodos clásicos [1]. Estas herramientas integran rigor científico con diseños accesibles, ampliando el análisis visual de datos de movilidad en sistemas complejos [1].

Además, la importancia de estas representaciones se refleja en trabajos recientes de clasificación de trayectorias: por ejemplo, Esteve y Falcó ofrecen una perspectiva completa de métodos topológicos para clasificar trayectorias, destacando cómo los invariantes de persistencia enriquecen los enfoques tradicionales [28].

2.4.2. Integración de TDA con modelos de aprendizaje profundo

La convergencia de TDA y aprendizaje profundo ha dado lugar al emergente campo del «aprendizaje profundo topológico». En este paradigma, las redes neuronales incorporan componentes topológicos (por ejemplo, diagramas de persistencia o capas basadas en homología persistente) para capturar propiedades globales de los datos [29]. Zia et al. revisan este área emergente, describiendo cómo las técnicas topológicas se han integrado progresivamente en distintas etapas de los modelos de aprendizaje profundo [29].

De manera complementaria, TDA también se emplea para analizar modelos existentes: por ejemplo, Liu et al. aplicaron homología persistente a las salidas de una red profunda compleja, transformando las predicciones en un «paisaje topológico» que facilita la interpretación del modelo. Este análisis topológico de las salidas revela detalles sutiles de la estructura de predicción que pasan desapercibidos con métodos de reducción de dimensión convencionales (como t-SNE o UMAP) [30].

2.4.3. Aplicaciones emergentes del TDA en trayectorias biológicas y medioambientales

Más allá de la movilidad urbana o aérea, el TDA se ha explorado en datos de trayectorias biológicas. Por ejemplo, Bhaskar et al. mostraron que configuraciones espaciales de poblaciones celulares de dos tipos pueden representarse eficazmente mediante imágenes de persistencia, logrando clasificar con alta precisión distintas arquitecturas celulares asociadas a variaciones en adhesión [31]. Este trabajo ilustra cómo los invariantes persistentes pueden capturar dinámicas de organización en tejidos multicelulares, con potenciales aplicaciones en el estudio de procesos de desarrollo y enfermedad.

En el ámbito medioambiental, Evans-Lee y Lamb proponen un método novedoso para detectar anomalías en trayectorias geoespaciales marinas. Al incrustar las coordenadas espaciales y temporales de los datos de posicionamiento de embarcaciones (AIS) en \mathbb{R}^3 , calculan la homología persistente para identificar bucles topológicos en las rutas de navegación. La presencia de dichos bucles, que no debería ocurrir en condiciones normales, señala trayectorias atípicas de tipo «crop circle» (movimientos circulares repetidos). Este enfoque robusto frente a perturbaciones locales ofrece nuevas herramientas para la navegación marítima y el monitoreo ambiental, permitiendo detectar patrones inusuales en la movilidad de buques con alta sensibilidad [32].

Capítulo 3

Análisis topológico de trayectorias en alta y baja dimensión

El desarrollo de este proyecto se puede dividir principalmente en 4 secciones, selección y preparación de los datos a utilizar durante el desarrollo del proyecto, en segundo lugar la elección de la metodología de análisis a seguir durante el desarrollo del mismo, y las herramientas computacionales y de software que se emplearon para realizar dicho análisis. En tercer lugar, el diseño e implementación de la solución, explicando las decisiones tomadas durante el diseño de la misma, y finalmente, estudio de la solución y pruebas realizadas para comprobar la eficacia de la misma.

3.1. Modelización de los datos

3.1.1. Selección de los datos

A lo hora de seleccionar los datasets de trayectorias utilizados para el desarrollo delo trabajo, se tuvieron en cuenta varios factores. El primero de ellos fue las licencias de uso de los mismos, reduciendo la búsqueda a datasets con licencia CC0, es decir, aquellos de datos que fueran denominados de dominio público. En segundo lugar, se tuvieron en cuenta los criterios estándar para la calidad del dato, es decir, se tuvo en la precisión (falta de errores), la integridad (ausencia de columnas vacías o datos inexistentes), consistencia y unicidad de los mismos. Asimismo, se tuvo en cuenta el formato de los datasets, evaluando la compatibilidad de estos con las librerías que se usarían durante el desarrollo del proyecto, y las posibilidad de deformaciones en caso de haber la necesidad de oficiar los mismos para adaptarlos a las herramientas. Finalmente, se evaluó el contenido de los mismos, analizando minuciosamente las posibilidades que la trayectorias contenidas en los mismos aportaban de cara a su manipulación y análisis durante el desarrollo del proceso.

Atendido a los criterios anteriormente mencionados, finalmente se tomó la decisión de usar el dataset GeoLife GPS Trajectories [33] . Esto se debe a la gran variedad de datos que nos presenta el dataset, teniendo trayectorias pequeños, fácilmente manejables sin gastar en exceso en recursos computacionales, como trayectorias más extensas, con más posibilidad de obtener resultados más vistosos de cara a analizar su topología, así como la posibilidad de realizar análisis con un conjunto de estas trayectorias simultáneamente.

Estas trayectorias se encuentran en formato `.plt`, ofreciéndonos la oportunidad de utilizar las herramientas de análisis sin realizar ningún tipo de modificación sobre los mismos. En lo que al contenido se refiere, estos datos contienen datos de trayectorias GPS de 178 usuarios durante un período de cuatro años, siendo cada trayectoria representada por una secuencia de puntos marcados en el tiempo, conteniendo aparte de su información temporal, información sobre la latitud, longitud y altitud en cada uno de estos instantes temporales, pudieron realizar en primer instante análisis sobre las 3 dimensiones del espacio, pudiendo escalarlo añadiéndole lag o desfases usando funciones internas de la librería en caso de desear realizar análisis más complejos sobre el conjunto de datos, siendo este obligado en el ámbito de las dependencias temporales y espaciales de las trayectorias. [34] [35] [36]

3.1.2. Preparación de los datos

De cara a la preparación de los datos a utilizar, se realizó un estudio del contenido de los mismos durante la selección, como se mencionó en la sección anterior, para garantizar la integridad y completitud de los mismos. Posteriormente, se revisó más en detalles el formato de los campos y los valores de los mismos para preparar código para la lectura y la composición de un dataframe de *Pandas* (librería de Python, una de las herramientas que se usará durante el desarrollo) de cara a generar el array de puntos necesaria para el uso de las funciones de la librería *giotto-tda* [14]. Este array de la nube puntos se trata de un *ndarray*, de la librería *NumPy*. Estos array se caracterizan por ser estructuras n-dimensionales, homogéneas y de tamaño fijo, construyéndose a partir del dataframe generado, seleccionando los valores de latitud, longitud y altitud, haciendo uso de la función `.values()` propia de los dataframe de *Pandas*.

3.2. Metodología y herramientas para el análisis

3.2.1. Metodología de desarrollo

En lo que a la metodología se refiere, se ha seguido un enfoque iterativo-exploratorio. En primer lugar, se realizó una etapa de preprocesamiento de los datos, que consistía en la lectura de trayectorias desde diferentes fuentes (`.plt`, `LINESTRING`, etc.), su transformación a nubes de puntos uniformes en dimensión y formato, y su normalización o reducción si fuera necesario. A continuación, se aplican las técnicas de análisis topológico, comenzando con el cálculo de diagramas de persistencia mediante filtraciones de Vietoris–Rips y la computación de medidas asociadas como la entropía de persistencia.

Posteriormente, se integró la herramienta Mapper, configurando un pipeline con funciones de filtrado (como proyecciones PCA), coberturas cúbicas e implementaciones de clustering como DBSCAN. Finalmente, se generaron grafos para facilitar la interpretación de la estructura topológica de las trayectorias a analizar.

Este enfoque ha permitido estudiar una variedad de configuraciones relativas al análisis, adaptándose al tamaño y características de la trayectoria o conjunto de trayectorias analizado en el momento, y facilitando la extracción de patrones generales y particularidades topológicas del movimiento en diferentes contextos.

3.2.2. Herramientas utilizadas

Como se mencionó anteriormente en el capítulo de Estado del Arte, existen una gran variedad de herramientas computacionales para la realización de análisis topológico de datos, permitiendo al usuario desarrollar con libertad en una variedad de entornos y lenguajes de programación las tareas de análisis que él desee. Nos hemos decidido por usar el lenguaje **Python**, debido a su amplia adopción en la comunidad científica, su sintaxis sencilla y su gran oferta de bibliotecas para análisis de datos, visualización y aprendizaje automático. También ha ayudado a tomar esta decisión la familiaridad y experiencia con el uso de este lenguaje, y las facilidades que este da a la hora de ejecutar desde varias máquinas o entornos distintos, habiendo ejecutado las distintas implementaciones realizadas tanto en local como en entornos como **Google Colab**, para aprovechar los recursos extra que pudieran aportar de cara a agilizar los tiempos de ejecución de los procesos propios del análisis.

La elección de Python como lenguaje para basar nuestro desarrollo nos da acceso a una amplia variedad de librerías que nos ayudarán en nuestro estudio, entre las cuáles cabe destacar las siguientes:

- **NumPy** y **Pandas**: utilizadas para el tratamiento de datos y la conversión de trayectorias en arrays o dataframes.
- **Matplotlib** y **Seaborn**: utilizadas para la visualización de trayectorias, resultados intermedios, representaciones PCA y diagramas de persistencia.
- **Scikit-learn**: empleada para tareas de preprocesamiento (como la reducción de trayectorias mediante $KMeans$) y para clustering dentro del algoritmo Mapper.
- **Giotto-TDA** [14]: una biblioteca especializada en análisis topológico de datos, desarrollada en Python sobre interfaces en C++, que ofrece implementaciones optimizadas para el cálculo de homología persistente y la construcción del grafo Mapper, entre otras funcionalidades.

3.3. Diseño e implementación de soluciones con TDA

Durante el desarrollo del trabajo se hizo uso de una serie de scripts de **Python** para realizar análisis sobre una sola trayectoria del dataset Geolife y un conjunto de trayectorias del mismo, con herramientas TDA, y un script aplicando métodos no TDA como PCA, clustering tradicional y análisis t-SNE. Adicionalmente se añadió desplazamiento a los datos de las trayectorias para generar más diversidad al análisis, al darnos una visión adicional de las características topológicas de los datos.

3.3.1. Carga de Trayectorias

Para poder realizar cualquiera de estos análisis es fundamental realizar una correcta lectura y transformación de los datos de las trayectorias. Haciendo uso de la función `read_csv` de **Pandas**, y la librería `os`, se realiza una lectura del archivo o carpeta dependiendo del análisis que se desea realizar, generando un dataframe, y usando funciones de **Pandas** se eliminan los valores nulos en caso de haberlos, y se genera un dataframe de **Pandas** que contiene los valores de latitud, longitud y altitud de las trayectorias. Se establecieron también criterios para la toma de trayectorias, estableciendo como 50 el número mínimo de puntos que debe poseer el dataset para tenerlo

Análisis topológico de trayectorias en alta y baja dimensión

en cuenta. Finalmente no hizo falta realizar ninguna tarea de limpieza ni transformación adicional, al poseer los datos escogidos una muy buena calidad en términos de lectura.

3.3.2. Visualización de Nube de Puntos

Después de la correcta lectura de los datos, se utiliza la función *plot_point_cloud* para realizar una representación y visualización en tres dimensiones de la nube de puntos. Esta visualización nos permite obtener una idea preliminar de la distribución que siguen las trayectorias objeto del análisis. Cada trayectoria es representada como una secuencia de coordenadas geográficas, transformadas en puntos en tres dimensiones. En la figura 3.1 se adjunta la nube de puntos generada por un conjunto de 110 trayectorias consecutivas, permitiéndonos observar la evolución de los recorridos y las estructuras topológicas que se van a analizar posteriormente.

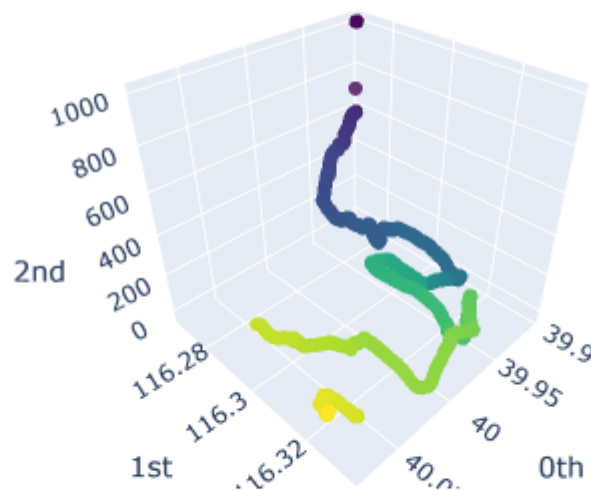


Figura 3.1: Visualización de la Nube de Puntos de las trayectorias de Geolife

3.3.3. Cálculo de Homologías Persistentes

Una vez visualizados los datos, se procede a realizar un análisis utilizando la técnica de homologías persistentes, uno de los pilares fundamentales del análisis topológico de datos. Para la presente implementación, se decidió hacer uso de este método para el cálculo de homologías persistentes de complejos de *Vietoris-Rips*, permitiéndonos extraer información clave sobre las características topológicas de los datos a lo largo de diferentes escalas. Esta implementación se realizó utilizando el módulo *VietorisRipsPersistence* de la librería *giotto-tda*, especificando como argumentos las dimensionales de homología que son de interés en el análisis, siendo en el ejemplo que se muestra en la Figura 4.2 tres:

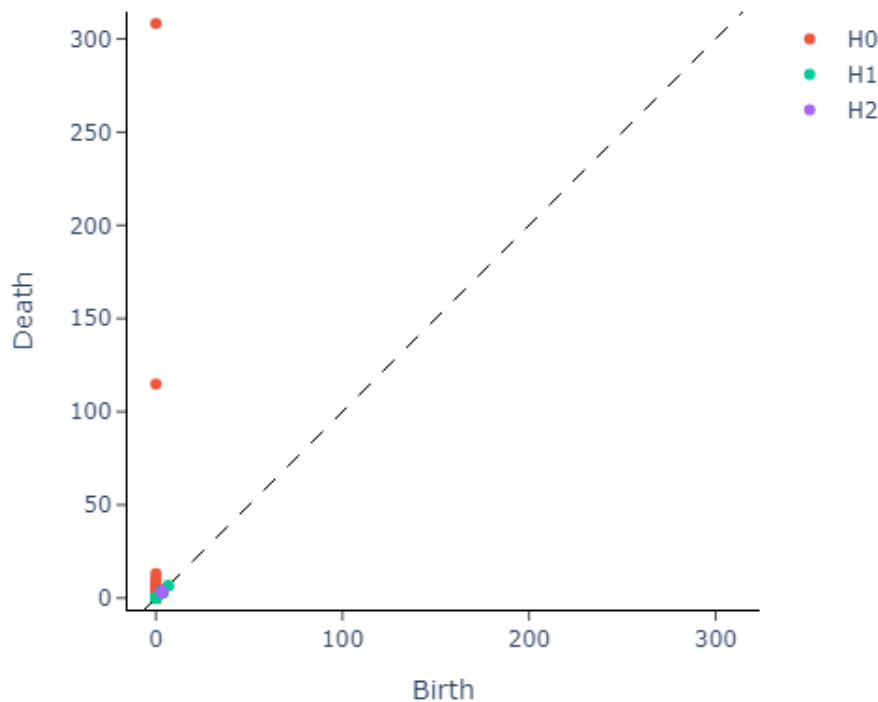


Figura 3.2: Diagrama de Persistencia de un Conjunto de trayectorias

Estas tres dimensiones nos permiten observar diferentes estructuras topológicas:

1. H0: Nos permite identificar las componentes conexas de la nube de puntos
2. H1: Nos indica los ciclos presentes en el conjunto de los datos
3. H2: Nos permite observar las cavidades, o huecos tridimensionales presentes en la estructura

Análisis topológico de trayectorias en alta y baja dimensión

A la hora de realizar las ejecuciones de esta función, al usar volúmenes de datos de gran dimensiones, los tiempos de ejecución en la misma eran bastantes elevados, con un alto coste computacional, por tanto, se hizo uso de la integración de la librería *giotto-tda* con paralelismo *multi-thread* para agilizar estos procesos de cálculo y generación de gráficas, mejorando los tiempos de ejecución en alrededor de un 200% usando 15 hilos ejecutando en una sola máquina.

```
homology_dimensions = [0, 1, 2]
persistence = VietorisRipsPersistence(
    metric="euclidean",
    homology_dimensions=homology_dimensions,
    n_jobs=15,
    collapse_edges=True,
)
```

Figura 3.3: Definición del objeto VietorisRipsPersistence

Una vez hemos definido el objeto *persistence* y realizados los cálculos, se procede a la representación gráfica del diagrama de persistencia, veáse Figura 3.2 . Esta representación se genera haciendo las transformaciones necesarias al objetivo persistencia generando mediante el uso de la función *fit_transform*, generando así un objeto para mandar de argumento a la función *plot_diagram* para generar el diagrama mencionado anteriormente.

3.3.4. Extracción de Descriptores con Mapper

Llegados a este punto, habiendo realizado ya la visualización de los datos y el cálculo de las homología persistentes, precedemos a la preparación de un pipeline de Mapper, y la representación en forma de grafo de este haciendo uso de las funciones de la librería *giotto-tda*. En primera instancia, es importante dar una breve explicación de en que consiste el algoritmo: " Un conjunto de datos S , con los siguientes hiperparámetros, tiene asociado el complejo simplicial de Mapper:

- Un algoritmo de clusterización y un conjunto de n clústers.
- Una función filtro que capture variables latentes a un espacio de dimensión K .
- Una cubierta abierta de nuestro espacio de variables latentes.
- Para cada abierto U del espacio latente y cada $i \leq K$, construimos un conjunto de puntos en S determinado por el i -ésimo clúster de la imagen inversa de U , es decir, todos los elementos de la base de datos que bajo f van a dar a U .

La familia de todos los subconjuntos $S(U, i)$ de la base de datos determinada por todas las parejas (U, i) serán los puntos de nuestro complejo simplicial de Mapper. Las caras de este complejo simplicial están determinadas por los subconjuntos que contengan registros de nuestro dataset en común." [37]

3.3. Diseño e implementación de soluciones con TDA

Teniendo esto en cuenta, procedemos al diseño y configuración del pipeline. Este proceso se compone de los siguientes cuatro pasos:

- **Cartografía** En primer lugar, cartografiamos D a un espacio de menor dimensión usando una función de filtrado de tipo $f : R^n \rightarrow R^m$. Se puede elegir de una variedad de funciones para realizar esta tarea, en nuestro análisis decimos hacer uso de una proyección, al darnos la opción la opción de realizar distintos análisis con cambios mínimos en los parámetros del análisis. Todo esto se consigue haciendo uso de la función `FilterFunctionName` de la librería `giotto-tda`, que contiene una gran variedad de funciones de filtrado, dando libertad al usuario para ajustar el análisis que desea realizar a las características del problema o datos a tratar.
- **Filtrado** Una vez realizado la proyección, se procede al filtrado de valores de tipo $\mathcal{U} = (U_i)_{i \in I}$, para generar un recubrimiento de un conjunto de intervalos que se solapan, con una longitud constante. En esta implementación, se hace uso de un recubrimiento de tipo `CubicalCover` con `n_intervals = 10` y `overlap_frac = 0.4`, lo que significa que el espacio filtrado se divide en 10 intervalos por dimensión, cada uno con un solapamiento del 40%. Esta superposición es esencial para capturar estructuras conectadas en regiones vecinas y evitar una fragmentación excesiva del espacio.
- **Clusterización** Para cada conjunto preimagen $f^{-1}(U_i)$ asociado a cada intervalo del recubrimiento, se aplica un algoritmo de agrupamiento con el objetivo de identificar componentes conexas. En nuestro caso, se emplea el algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise), que tiene la ventaja de no requerir especificar el número de clusters y es robusto frente al ruido y las formas arbitrarias de los datos. Esto lo convierte en una opción adecuada para trayectorias geográficas, donde pueden existir regiones de alta densidad intercaladas con trayectorias dispersas o inusuales.
- **Construcción del grafo** Finalmente, se construye un grafo Mapper, en el cual cada nodo representa un cluster obtenido en algún intervalo, y se crea una arista entre dos nodos si comparten al menos un punto de datos (es decir, si sus conjuntos de puntos tienen intersección no vacía). Este grafo sintetiza la estructura topológica del conjunto de datos original, revelando la conectividad global, la presencia de ciclos o componentes separados, y sirve como herramienta de exploración visual y estructural.

Para la visualización del grafo Mapper resultante del proceso descrito anteriormente, se utiliza la función `plot_static_mapper_graph` de `giotto-tda`, que genera una representación estática del grafo que puede ser interpretada fácilmente en el contexto de las trayectorias geoespaciales. Adicionalmente, se hace uso nuevamente de la integración de la librería con paralelismo *multi-thread*, para agilizar la ejecución de la creación de los pipelines de Mapper.

A continuación, se puede ver el resultado de ejecutar este proceso, con un pipeline definido de la siguiente manera: una función de proyección sobre los 3 ejes cartesianos, con un recubrimiento cúbico, y un solapamiento de un 30%:

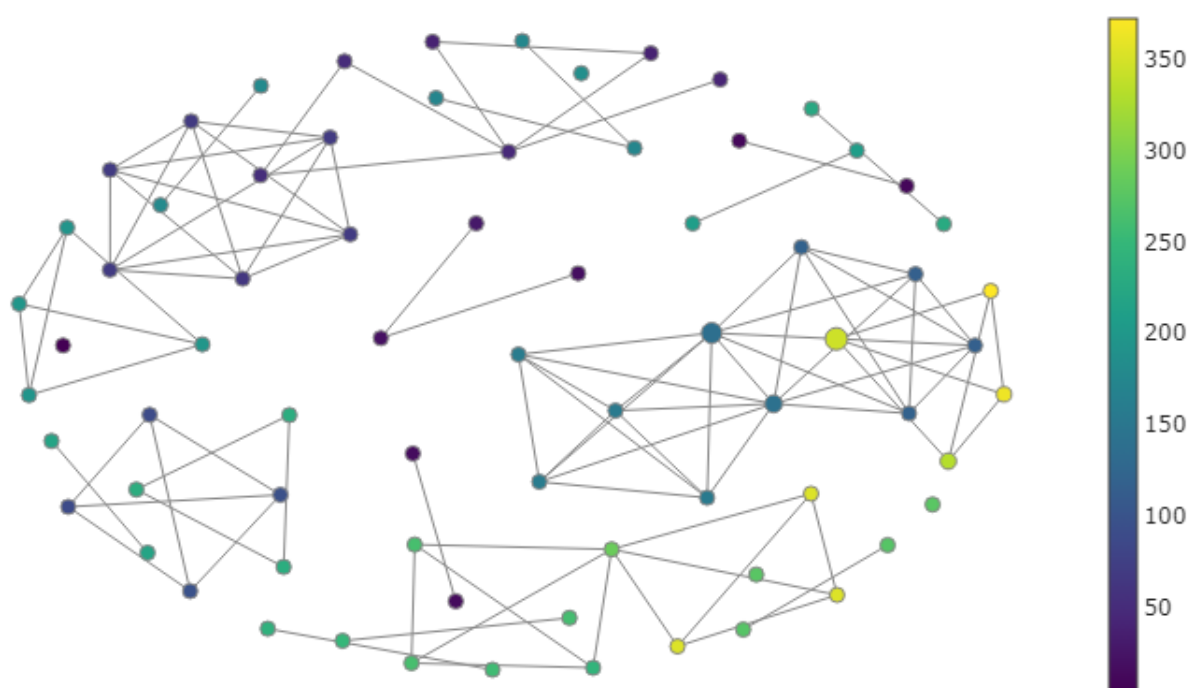


Figura 3.4: Grafo de Mapper generado a partir de una trayectoria del Dataset

Adicionalmente, para algunas trayectorias en concreto se generó la matriz de distancias haciendo uso de la librería dedicada para Machine Learning de Python, *sklearn*, más concretamente la función *pairwise_distances*, utilizando la métrica euclidiana, y representando esta gráfica en formato mapa de calor, para visualizar de mejor manera las distancias entre las componentes de la trayectoria.

3.3.5. Análisis y comparación con PCA

Para complementar el análisis topológico y obtener una representación reducida de los datos de trayectoria, se ha empleado la técnica de Análisis de Componentes Principales (PCA). Este método estadístico nos permite proyectar un conjunto de datos multidimensional en un espacio de menor dimensión, conservando la mayor parte de la varianza de los datos originales.

En primer lugar, se ha aplicado PCA sobre la nube de puntos generada anteriormente, cada una de estas trayectorias, inicialmente expresada en coordenadas cartesianas tridimensionales, ha sido transformada mediante PCA a un espacio de dos dimensiones, facilitando de esta manera la visualización de la evolución temporal de los puntos. El resultado se representa mediante un gráfico de dispersión en el que cada punto está coloreado según su componente temporal, permitiendo observar cómo varía la trayectoria a lo largo del tiempo en el nuevo espacio de representación.

A continuación, se ha aplicado también PCA sobre las características extraídas tras aplicar técnicas topológicas, como los diagramas de persistencia y sus correspondientes representaciones vectorizadas. En este caso, la proyección a dos dimensiones permite explorar visualmente si existen agrupaciones o patrones diferenciables en las características topológicas. La visualización se realiza nuevamente como un gráfico de dispersión, donde se espera que trayectorias similares en cuanto a su estructura topológica se agrupen en regiones cercanas del espacio proyectado.

En ambas representaciones, se ha utilizado *matplotlib* y *seaborn* para la visualización de los resultados. Estas herramientas permiten no solo visualizar los componentes principales, sino también identificar posibles agrupaciones, transiciones o tendencias que puedan estar ocultas en la representación original de los datos. La comparación entre el espacio proyectado de los datos originales y el espacio proyectado de las características topológicas constituye un recurso valioso para validar y entender las transformaciones realizadas mediante TDA.

Este enfoque no topológico se ha utilizado como una técnica complementaria, tanto para interpretar los datos de entrada como para evaluar la riqueza estructural de los descriptores generados mediante homología persistente, estableciendo así una comparación visual y cuantitativa entre representaciones convencionales y topológicas.

3.4. Técnicas estadísticas para el análisis de trayectorias GPS

A diferencia de los métodos basados en Análisis Topológico de Datos (TDA) [10] [11], el enfoque clásico emplea técnicas convencionales de análisis estadístico y de aprendizaje automático. Estas técnicas se utilizan como línea base comparativa, aprovechando la facilidad de interpretación y la robustez de su metodología [38]. A continuación se enumeran de manera breve los pasos principales del pipeline clásico implementado, las herramientas utilizadas y cómo estas metodologías permiten capturar la forma y similitud de las trayectorias usando métricas tradicionales:

1. **Preprocesamiento de datos:** se limpian las trayectorias GPS eliminando puntos erróneos o valores atípicos, se corrigen saltos de señal y se muestrean las trayectorias a longitudes uniformes [39]. Este tratamiento inicial facilita el análisis posterior al homogenizar la representación de cada ruta.
2. **Extracción de características:** cada trayectoria se convierte en un vector numérico que describe su forma o recorrido. Esto puede lograrse mediante el muestreo regular de coordenadas espaciales (latitud/longitud) o la obtención de estadísticas agregadas (distancia total recorrida, velocidad media, ángulos de cambio de dirección, etc.) [40]. Estas características resumen las propiedades geométricas y dinámicas de cada trayectoria.
3. **Reducción de dimensionalidad (PCA):** se aplica Análisis de Componentes Principales (PCA) para proyectar los datos en un espacio de menor dimensión [41]. El PCA identifica las direcciones de máxima varianza en el conjunto de trayectorias y permite conservar las componentes principales (generalmente las primeras 2 o 3), capturando la forma global de las rutas con pocos parámetros [41].
4. **Agrupamiento (clustering):** sobre los datos reducidos se ejecutan algoritmos de clustering clásicos. Por ejemplo, K-means agrupa las trayectorias en un número fijo de clústeres, minimizando la varianza interna de cada grupo [42]. En paralelo, DBSCAN (Density-Based Spatial Clustering) identifica grupos basados en la densidad de puntos en el espacio de características, permitiendo descubrir clústeres de forma arbitraria sin conocer de antemano el número de grupos y diferenciando ruido de datos dispersos [43]. Estas técnicas agrupan trayectorias con formas o patrones similares.
5. **Cálculo de métricas de similitud:** para cuantificar la similitud entre trayectorias se emplean métricas convencionales. Esto incluye distancias euclídeas o Manhattan entre vectores de características, así como distancias propias de trayectorias (por ejemplo, la distancia de Fréchet o el Dynamic Time Warping (DTW)) [44] [45]. Estas medidas numéricas permiten evaluar cuán semejantes son las rutas en cuanto a su forma y recorrido, complementando el agrupamiento con una evaluación explícita de la similitud.

3.4. Técnicas estadísticas para el análisis de trayectorias GPS

6. **Visualización de resultados:** finalmente, se utilizan herramientas de visualización para interpretar los resultados. Se emplean bibliotecas de Python como Matplotlib y Seaborn para generar gráficos bidimensionales de las componentes principales y mapas de calor [46] [47]. Asimismo, es posible trazar directamente las rutas GPS en un plano geográfico, coloreando cada trayectoria según el clúster asignado. Estas visualizaciones facilitan la inspección visual de patrones de movilidad y la comparación con los patrones obtenidos mediante TDA.

Los métodos descritos capturan la forma de las trayectorias mediante su representación numérica y evalúan la similitud a través de métricas clásicas de distancia. Aunque este enfoque clásico no explora explícitamente la estructura topológica de los datos, proporciona una base sólida y comprensible que puede compararse con los resultados del análisis topológico [48]. En conjunto, las técnicas de PCA y clustering tradicionales permiten obtener patrones globales en los datos de movilidad y sirven como complemento para evaluar las ventajas del enfoque topológico.

Capítulo 4

Resultados de TDA sobre trayectorias

En este capítulo se presentan los resultados obtenidos a lo largo del análisis topológico de trayectorias utilizando técnicas de TDA, junto con algunos enfoques comparativos mediante métodos clásicos de análisis de datos. El objetivo principal de esta sección es mostrar de forma clara y estructurada cómo las herramientas aplicadas permiten extraer información significativa sobre la estructura de los datos, tanto a nivel local como global.

Se incluyen visualizaciones de nubes de puntos generadas a partir de trayectorias GPS del conjunto Geolife, así como sus correspondientes diagramas de persistencia, representaciones PCA y grafos Mapper. También se analizan las propiedades topológicas identificadas —como ciclos persistentes o conectividad— y se discuten sus implicaciones. Adicionalmente, se presentan resultados de métodos alternativos no topológicos (como clustering y reducción de dimensionalidad) con el fin de establecer una comparación cualitativa entre enfoques.

La interpretación de los resultados se centra tanto en trayectorias individuales como en el comportamiento agregado del conjunto de datos, permitiendo evaluar la capacidad del TDA para detectar patrones globales, anomalías y propiedades invariantes que no son fácilmente identificables mediante técnicas convencionales.

4.1. Resultados

Durante este capítulo, se van a plantear los resultados del análisis topológico de datos explicada a lo largo del capítulo 3 sobre una sola trayectoria del dataset *Geolife*, sobre un conjunto de trayectorias consecutiva del mismo, y sobre otro conjunto de trayectorias no relacionadas entre sí, así como la aplicación de estos mismos datos en su contraparte implementada con métodos tradicionales. Posteriormente, se presentarán las características topológicas observadas a lo largo del desarrollo, y se hará una breve comparativa entre las 2 soluciones, y se intentará dar una visión a las diferencias entre los resultados en la ejecución de las mismas.

En primer lugar, vamos a observar el resultado de aplicar estos análisis a una sola trayectoria del conjunto *Geolife*. Como se ha visto anteriormente en el capítulo 3, en primer lugar se realizó una visualización de mapa de puntos en 3 dimensiones de la trayectoria en sí.

Se observa en este caso que al ser una trayectoria de puntos consecutivos, se ve una línea bien definida en su movimiento, pudiendo deducirse que produjo un desplazamiento mayormente vertical en primera, seguido de una serie de desplazamientos horizontales que podrían asimilar la toma de una curva.

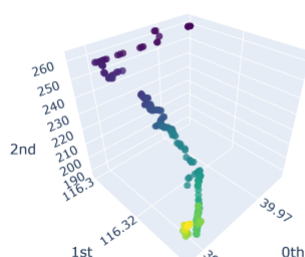


Figura 4.1: Nube de puntos de una Trayectoria

Esta visualización nos ofrece una primera idea de las características topológicas que podría tener esta trayectoria. En general, la visualización 3D no es eficaz, al ser compleja, y no es compatible con trayectorias de mayor dimensión. Para poder obtener resultados, debemos aplicar las herramientas mencionadas anteriormente, como sería la generación del diagrama de persistencia 4.2 y el grafo Mapper 3.4, así como la matriz de distancia entre los puntos de la trayectoria 4.3.

A partir de estos tres componentes del análisis, podemos afirmar una serie de hechos sobre la naturaleza de esta trayectoria. En primer lugar, la visualización en tres dimensiones (Figura 4.1) nos indica que la trayectoria es continua en el espacio, sin cambios abruptos en altitud ni desplazamientos bruscos en el plano horizontal. Este comportamiento sugiere que el movimiento del agente fue progresivo y suave, lo que es coherente con trayectorias humanas o de vehículos en contextos urbanos o semiurbanos.

En segundo lugar, el análisis topológico de la trayectoria mediante homología persistente (Figura 4.2) confirma que la trayectoria se comporta como una única componente conexa (grupo de homología H_0), ya que todos los puntos rojos aparecen cerca del origen y mueren rápidamente. Esto indica que no hay trayectos desconectados ni segmentos aislados dentro del recorrido.

Por otro lado, la presencia de un ciclo detectado en la dimensión uno (H_1), representado por un punto verde alejado de la diagonal, revela una característica topológica significativa: la trayectoria contiene una curva cerrada o un bucle. Esta información sería difícil de detectar a simple vista, y gracias al enfoque topológico, puede

Resultados de TDA sobre trayectorias

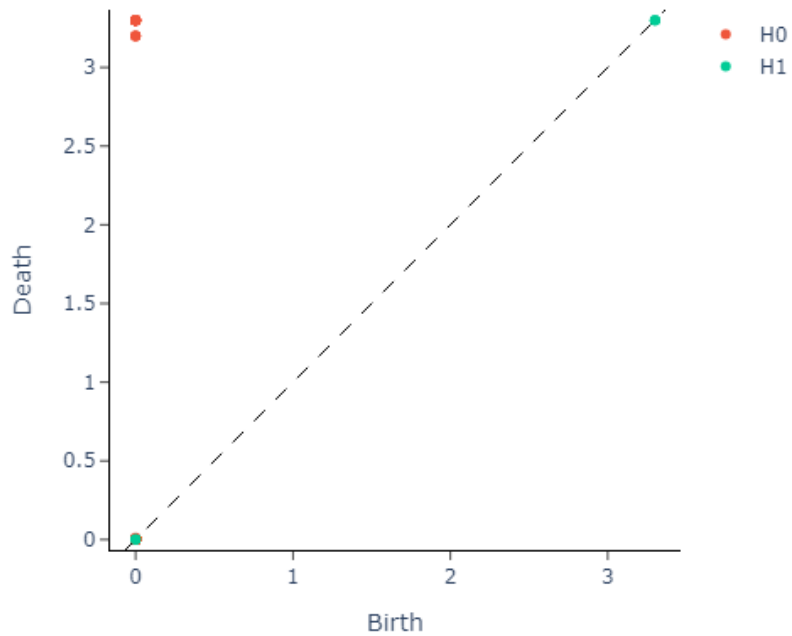


Figura 4.2: Diagrama de Persistencia de una trayectoria

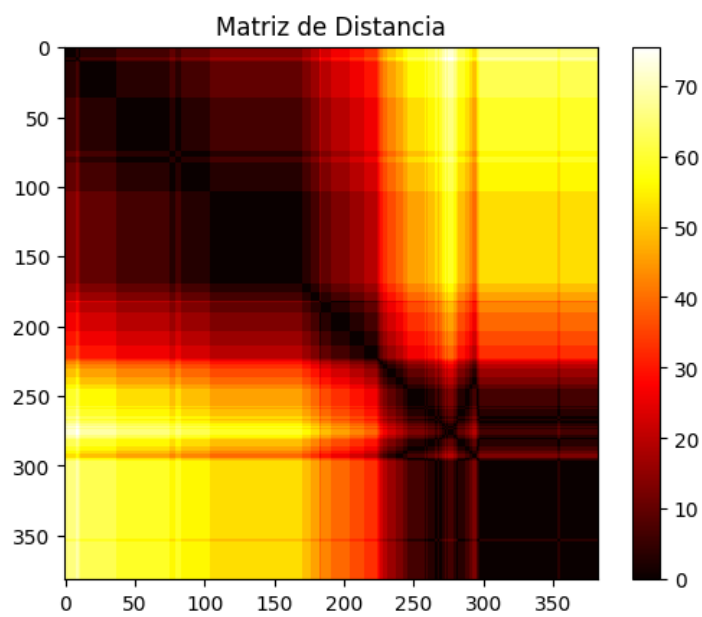


Figura 4.3: Matriz de Distancia de los puntos de la trayectoria

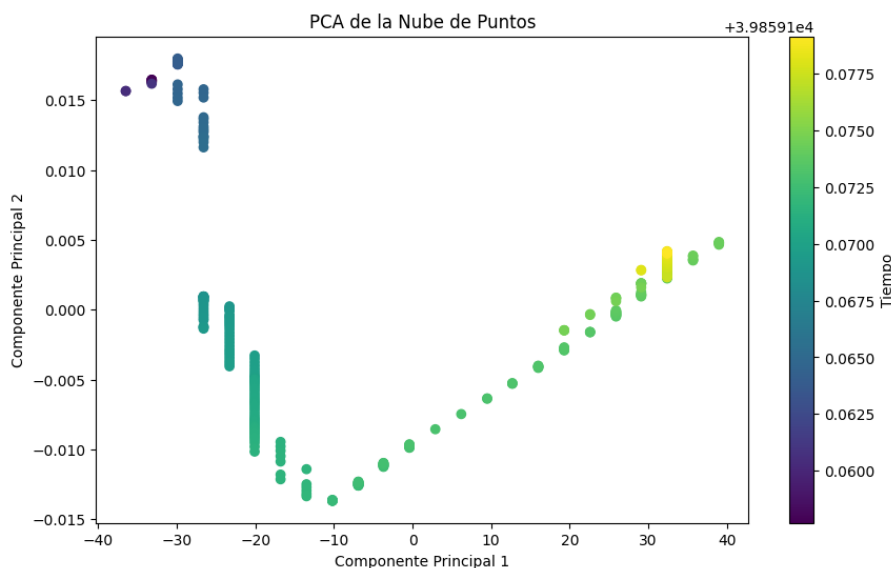


Figura 4.4: Análisis con PCA

interpretarse como una posible zona de retorno, giro o patrón de navegación circular realizado por el agente.

Finalmente, la ausencia de características en dimensión dos (H_2) corrobora que no existen cavidades tridimensionales cerradas dentro del conjunto de puntos, lo cual es esperable dado que el movimiento ocurre sobre una superficie terrestre y no en un volumen tridimensional cerrado.

Además del análisis topológico, se aplicó una reducción de dimensionalidad utilizando Análisis de Componentes Principales (PCA) con el objetivo de visualizar la trayectoria en un espacio bidimensional conservando la mayor varianza posible de los datos originales. En este análisis, se observó cómo la trayectoria proyectada mantiene una estructura coherente, lo que valida que la geometría global del recorrido puede conservarse parcialmente incluso al reducir la dimensionalidad. No obstante, a diferencia del análisis topológico, el PCA no es capaz de identificar la presencia de bucles o estructuras conexas, ya que estas propiedades no se manifiestan directamente en la varianza lineal de los datos.

En este apartado, vamos a observar el resultado de aplicar herramientas topológicas a un **conjunto de trayectorias**, utilizando el framework desarrollado y descrito en el capítulo 3. Para ello, se aplicaron técnicas de Análisis Topológico de Datos (TDA), comenzando con la construcción del grafo Mapper, el cálculo de la homología persistente y la proyección mediante Análisis de Componentes Principales (PCA).

En la Figura 4.5, se presenta el resultado del algoritmo Mapper aplicado al conjunto completo. Esta visualización revela una estructura compuesta por varias componentes conectadas y otras aisladas. Se puede observar una componente densa (a la izquierda del grafo), que agrupa trayectorias con características similares, mientras que otros nodos aislados indican trayectorias con comportamientos significativamente distintos. La coloración de los nodos representa algún valor escalar asignado, como una función de filtro (por ejemplo, centralidad o densidad), lo que permite identificar posibles agrupamientos o trayectorias atípicas.

Resultados de TDA sobre trayectorias

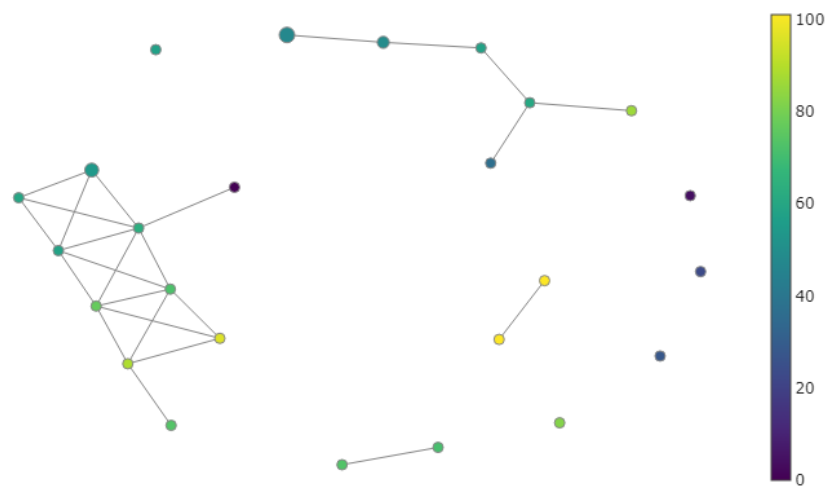


Figura 4.5: Grafo Mapper de múltiplas trayectorias

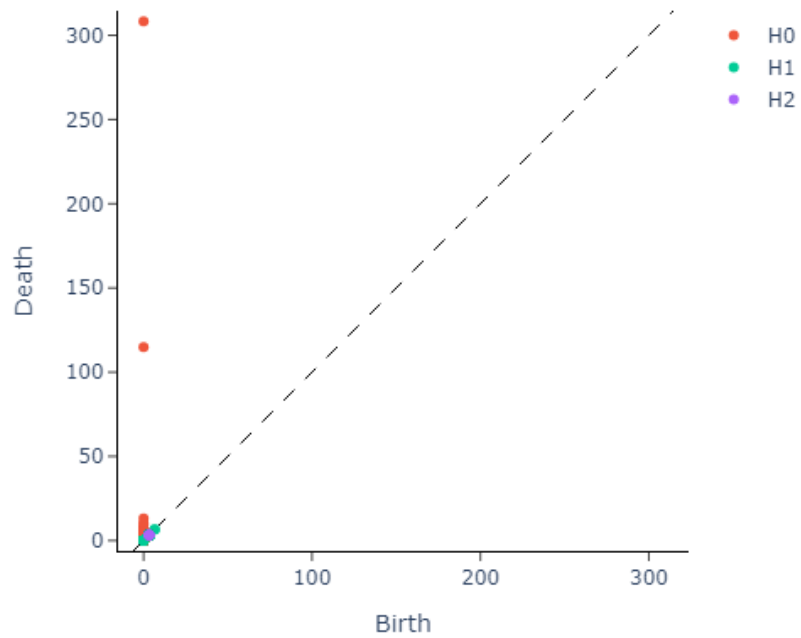


Figura 4.6: Diagrama de Persistência para o conjunto de trayectorias

La Figura 4.6 muestra el diagrama de persistencia del conjunto, el cual permite analizar las características topológicas de las trayectorias en distintas dimensiones. En el módulo de homología cero (H_0), correspondiente a componentes conexas, observamos un gran número de puntos que nacen y mueren rápidamente, lo cual sugiere que muchas trayectorias comienzan como componentes aisladas que se conectan gradualmente. Algunos puntos rojos más lejanos de la diagonal indican componentes que persisten más tiempo, es decir, trayectorias bien definidas y separadas del resto.

En el módulo de homología (H_1), representada por los puntos verdes, se observa la presencia de varios ciclos, algunos de los cuales son persistentes. Esto indica que dentro del conjunto existen trayectorias que forman bucles o retornos. Estas estructuras topológicas son relevantes, ya que pueden estar relacionadas con patrones de desplazamiento cíclico, como recorridos circulares, vueltas o trayectorias repetitivas en ciertos espacios.

Se detecta un solitario punto en módulo de homología (H_2), siendo este punto una anomalía entre los valores de las trayectorias, posiblemente un error a la hora de medir los datos en alguno de los instantes tomados, ya que al representar las trayectorias movimientos sobre superficies planas o cuasi-planas (como mapas urbanos) y no en espacios tridimensionales cerrados, no debería haber cavidades dentro de este conjunto consecutivo de trayectorias. Al revisar los datos, se observa que efectivamente un error en los valores de una de las 110 trayectorias leídas es el causante de este punto. Este valor habría sido casi imposible de detectar sin hacer uso de este tipo de métodos.

En la Figura 4.7, se presenta la proyección de las trayectorias en un espacio de dos dimensiones mediante PCA. Esta técnica se aplicó a los vectores de características topológicas derivados (como la persistencia en H_0 , H_1 , entre otros). La mayoría de los puntos se agrupan en una región cercana al eje vertical, lo cual indica que comparten una estructura topológica similar. Sin embargo, destaca un punto que se aleja notablemente en el eje $PC1$, lo que sugiere la presencia de una trayectoria con una topología significativamente diferente al resto. Esta trayectoria podría corresponder a un *outlier*, detectado gracias al poder de discriminación del análisis topológico.

En conjunto, estos tres componentes del análisis nos permiten concluir que el conjunto de trayectorias analizado presenta una variedad rica de estructuras topológicas. Existen trayectorias que forman patrones recurrentes (loops), algunas con comportamiento similar agrupadas en clústeres, y otras más atípicas o anómalas. Herramientas como Mapper, homología persistente y PCA permiten explorar estos datos más allá de su geometría directa, capturando propiedades de conectividad, ciclos y diversidad estructural.

4.1.1. Análisis de clustering y comparación con métodos de TDA

Además del análisis topológico previamente presentado, se exploraron otros métodos más tradicionales para el análisis de las trayectorias, con el fin de contrastar sus resultados y comprender mejor la naturaleza del conjunto de datos. Para ello, se aplicaron técnicas de agrupamiento como DBSCAN y KMeans, así como un Análisis de Componentes Principales (PCA) clásico sobre todas las trayectorias reducidas.

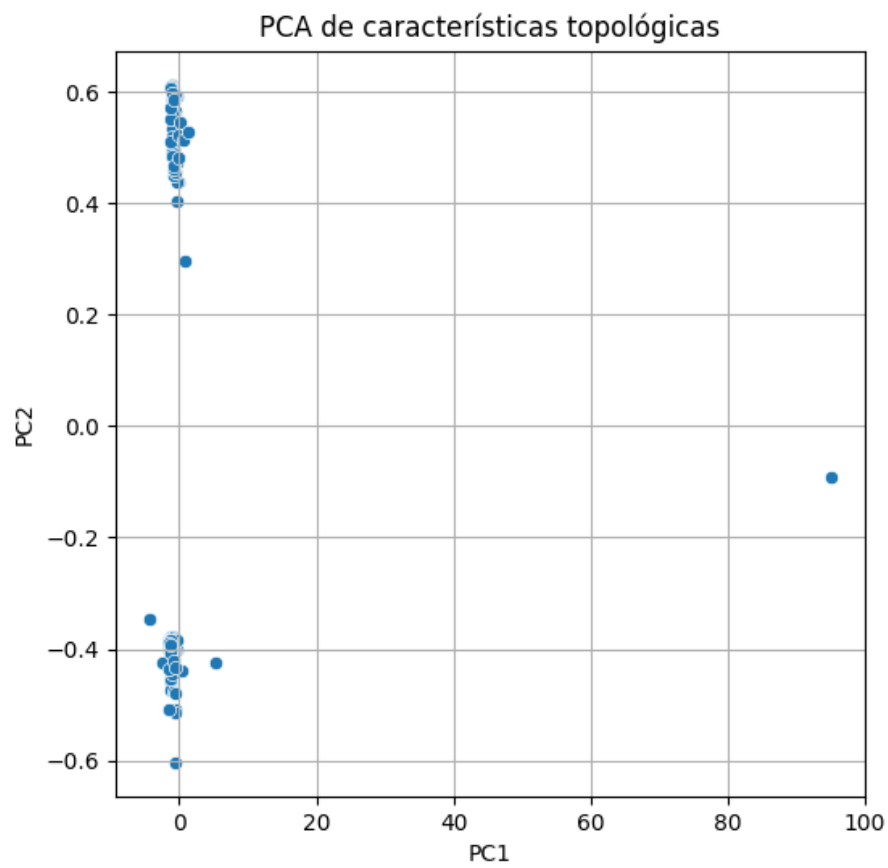


Figura 4.7: Análisis PCA de características topológicas

En primer lugar, en la Figura 4.8 se presentan los resultados de la reducción de dimensionalidad mediante t-SNE, sobre la cual se aplicaron los algoritmos de agrupamiento DBSCAN y KMeans.

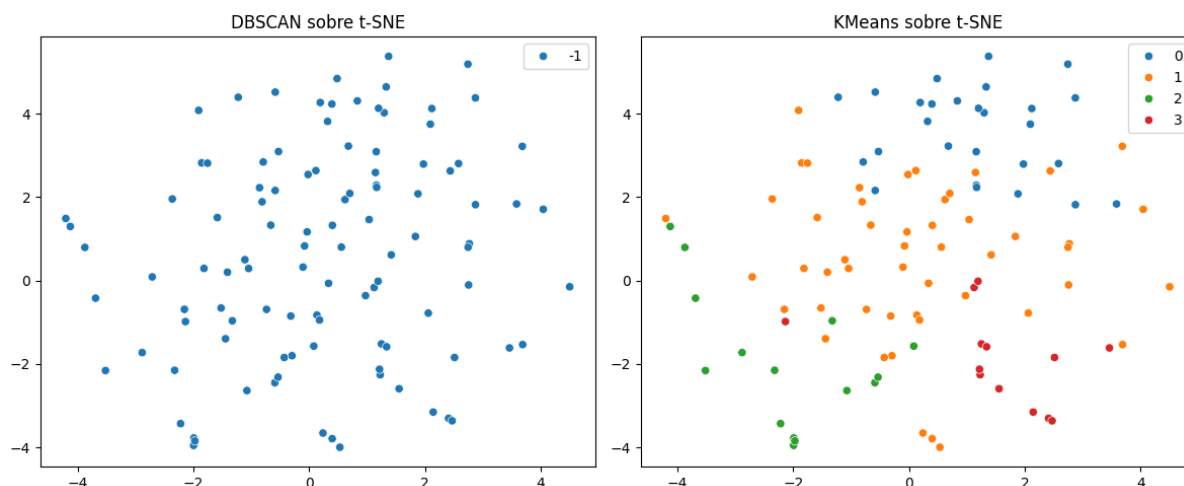


Figura 4.8: Resultados de clustering sobre reducción t-SNE: DBSCAN (izquierda) y KMeans (derecha)

En el caso de DBSCAN, se observa que el algoritmo no logra identificar estructuras claras de agrupamiento, asignando a todos los puntos la misma etiqueta (o marcándolos como ruido). Esto sugiere que las trayectorias no presentan densidades locales suficientemente diferenciadas para que DBSCAN pueda formar clústeres robustos, o que los parámetros utilizados no eran adecuados para la distribución de los datos en este espacio reducido. Este comportamiento contrasta con los resultados del análisis topológico, donde se identificaban componentes conectadas y ciclos, lo cual indica que sí existe una estructura no trivial en los datos que DBSCAN no ha podido capturar.

Por otro lado, el método KMeans sí fue capaz de dividir el conjunto en cuatro clústeres distintos. Estos grupos parecen estar distribuidos con cierta coherencia espacial, aunque el método se basa únicamente en distancias euclidianas en el espacio reducido por t-SNE. Sin embargo, KMeans no puede detectar bucles o estructuras de conectividad, lo cual representa una limitación frente al análisis de homología persistente. A pesar de ello, los resultados de KMeans pueden ser útiles para la clasificación general de patrones de trayectorias en función de su forma proyectada.

También se realizó un Análisis de Componentes Principales (PCA) tradicional sobre el conjunto de trayectorias para visualizar la distribución global de las mismas en un espacio linealmente reducido (Figura 4.9).

En esta representación, se aprecia una nube de puntos concentrada cerca del origen, con algunas trayectorias dispersas hacia los extremos. Esto indica una cierta homogeneidad en el comportamiento general de las trayectorias, con algunas excepciones que podrían representar trayectos atípicos. No obstante, al igual que en el PCA aplicado a características topológicas, esta visualización no permite identificar características como ciclos o componentes desconectadas, las cuales son evidenciadas únicamente a través del análisis topológico.

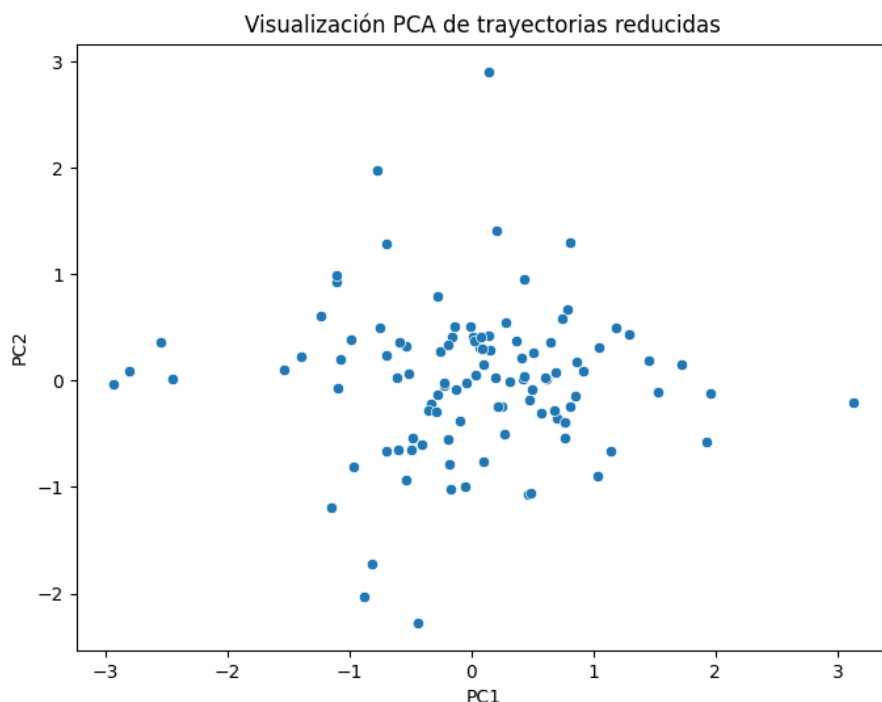


Figura 4.9: PCA tradicional sobre trayectorias

Comparación con métodos topológicos:

- **PCA y t-SNE:** Ambos métodos de reducción de dimensionalidad son útiles para representar visualmente la distribución de trayectorias, pero no capturan propiedades topológicas como la conectividad o la existencia de ciclos. Esto limita su capacidad para detectar patrones más complejos de movimiento.
- **DBSCAN y KMeans:** Estos algoritmos permiten segmentar los datos en grupos, pero dependen fuertemente de la métrica y del espacio de representación. En especial, DBSCAN no resultó efectivo en este contexto. KMeans, aunque más útil, tampoco proporciona información sobre la estructura topológica de las trayectorias.
- **Homología Persistente y Mapper:** En comparación, el análisis topológico ofrece una visión más rica del comportamiento de las trayectorias. Permite detectar bucles, componentes conexas y variaciones estructurales que los métodos convencionales no logran identificar.

En conclusión, los métodos clásicos como PCA, t-SNE y clustering son complementarios al análisis topológico. Mientras los primeros proporcionan una visión global basada en distancias y varianza, el segundo aporta información clave sobre la forma y conectividad de las trayectorias, siendo particularmente útil para identificar patrones complejos de movimiento.

Capítulo 5

Conclusiones

5.1. Evaluación de objetivos

La presente sección tiene como finalidad valorar el grado de cumplimiento de los objetivos definidos al inicio de este Trabajo Fin de Grado. El objetivo general consistía en desarrollar una metodología integral para el análisis topológico y estadístico de trayectorias en espacios de baja y alta dimensión, permitiendo extraer información relevante para su interpretación y clasificación. Para lograrlo, se establecieron una serie de objetivos específicos que han guiado el desarrollo del proyecto y que, en su conjunto, han sido abordados satisfactoriamente.

En primer lugar, se ha llevado a cabo una correcta definición y documentación del problema, delimitando las variables relevantes para el estudio de trayectorias en contextos tanto físicos como abstractos. Esta tarea incluyó la selección del conjunto de datos (GeoLife), la identificación de los atributos espaciales y temporales clave y la representación de las trayectorias en forma de nubes de puntos tridimensionales, preparadas para su análisis mediante herramientas topológicas.

A continuación, se diseñó una metodología de análisis que integró con éxito técnicas de Análisis Topológico de Datos (TDA) como la homología persistente y el algoritmo Mapper, junto con métodos estadísticos tradicionales y algoritmos de agrupamiento. Este diseño metodológico permitió identificar patrones estructurales en los datos, tales como bucles, componentes conexas y trayectorias atípicas, aportando una perspectiva más rica que la proporcionada por enfoques puramente estadísticos.

La implementación de la propuesta se realizó utilizando un conjunto de herramientas en el lenguaje Python, incluyendo librerías especializadas como Giotto-TDA, Scikit-learn y Pandas. Esta implementación fue puesta a prueba mediante diversos experimentos que validaron la eficacia de la metodología propuesta. Se evaluaron aspectos como la robustez frente al ruido, la sensibilidad ante anomalías y la eficiencia computacional, especialmente al aplicar técnicas de paralelización para mejorar los tiempos de ejecución.

Finalmente, los resultados fueron interpretados y sintetizados de forma rigurosa. Se analizaron las estructuras topológicas detectadas en los datos, se compararon con los resultados obtenidos por métodos tradicionales como PCA y clustering, y se discutieron las implicaciones prácticas del enfoque propuesto.

Conclusiones

Esta evaluación permitió extraer conclusiones relevantes para el campo del análisis de trayectorias, destacando las ventajas del uso de TDA frente a técnicas convencionales y proponiendo líneas futuras de investigación orientadas a su integración con modelos de aprendizaje profundo.

En resumen, los objetivos planteados en este trabajo han sido cumplidos de manera sólida, evidenciando la utilidad del enfoque topológico-estadístico en el análisis de trayectorias y demostrando su potencial para aplicaciones en movilidad, sistemas dinámicos y análisis de datos complejos.

5.2. Trabajo futuro

Este proyecto nos ofrece una visión relativamente limitada de la potencia que poseen las herramientas de análisis específicas para TDA, al no tener a disposición ni el tiempo ni los recursos computacionales que harían falta para una investigación más en detalle. Se podrían abrir varias líneas de investigaciones relativas a una variedad de factores, ya sea los factores relativos al uso de recursos computacionales, aquellos relativos a los datos usados, etc.

En primer lugar, desde el punto de vista de la mayor explotación de la potencia de las herramientas de análisis, se plantearía el uso de otras herramientas o librerías específicas de TDA, ya sea haciendo uso de otras funciones disponibles dentro de la propia librería *giotto-tda*, o haciendo uso de alternativas como las mencionadas en la sección 2.3.2 del capítulo de estado del arte. Adicionalmente, se podría hacer uso de volúmenes de datos mayores, aprovechando la capacidad de integración de las librerías de análisis con multi-threading, ya sea haciendo uso de recursos computacionales para alto rendimiento en servicios en la nube, o solicitando el uso de súper computadores como podría ser el Magerit-3.

En segundo lugar, el proyecto se podría expandir modificando los datos que son objetos del análisis, ya que el uso de otro tipo de trayectorias físicas o del mundo real, como podría ser el uso de datos médicos, o plantear alternativas basadas en fenómenos más abstractos, como podrían ser los atractores caóticos o espacios fuera de métricas tradicionales como sería la euclidiana, podrían abrir líneas de desarrollo que podrían generar bastante interés.

Partiendo de la idea de modificar los datos, se podrían hacer ligeras transformaciones a las trayectorias usadas actualmente para ver la evolución de las características topológicas de las mismas, por ejemplo, aplicando desplazamientos, o lag, en ciertos intervalos de los datos, generando conjuntos de datos de mayor dimensionalidad, que revelan la dependencia espacial y temporal, y la oportunidad de analizar no solamente la naturaleza de los datos, sino también su robustez ante estos cambios.

Finalmente, se podría expandir el proyecto haciendo uso de soluciones híbridas, implementando las herramientas de clusterización y análisis habitualmente usadas en machine learning y análisis de datos, y combinándolas con las soluciones de TDA para poder un análisis más completo, que nos ofrece una visión más global de las características de un determinado conjunto de datos.

5.3. Conclusiones personales

La realización de este Trabajo de Fin de Grado ha supuesto un hito de gran valor, tanto en lo académico como en lo personal. Durante el proceso de realización del mismo, he tenido la oportunidad de adentrarme en el campo del Análisis Topológico de Datos (TDA), una disciplina que, si bien requiere de una base matemática sólida para entender algunos de los conceptos clave de la misma, con conocimientos básicos de programación, se nos ofrece un conjunto de herramientas muy potentes para el estudio de estructuras de datos no triviales, como lo son las trayectorias espaciales, que nos permiten realizar análisis de una complejidad considerable, ofreciéndonos una alternativa más optimizada respecto las que nos darían los métodos de análisis tradicional.

A lo hora de ejecutar con una sola trayectoria, se seleccionó una trayectoria de complejidad media, de alrededor de 90 puntos, todos en intervalos de tiempo consecutivos

En lo relativo a los conocimientos técnicos, este proyecto me ha permitido consolidar conocimientos en programación científica y herramientas de análisis de datos con Python, especialmente en el uso de bibliotecas especializadas como `giotto-tda`, `scikit-learn`, `matplotlib` y `pandas`, siendo algunas de estas herramientas de uso común el mercado profesional. También he mejorado mis conocimientos y he ganado experiencia a la hora de manipular conjuntos de datos reales de tamaño considerable (como el dataset de trayectorias *Geolife*), en la depuración de problemas relacionados con formatos de datasets y limpieza y transformación a conjunto de datos, en la evaluación de necesidad de cómputo y la optimización del rendimiento computacional con herramientas de software, y a gestionar entornos mixtos de desarrollo, tanto en local como en la nube, mediante Google Colab.

Desde el punto de vista metodológico, ha sido muy enriquecedor explorar la comparación entre métodos tradicionales como serían PCA, k-means, DBSC, entre otros, y métodos topológicos, como la homología persistente y el algoritmo Mapper. Esta comparación me ha permitido comprender no solo las ventajas del TDA en cuanto a la robustez frente al ruido o invariancia ante deformaciones suaves, sino también sus limitaciones en términos de complejidad computacional y la necesidad de interpretación de los resultados generados usando estos métodos.

Uno de los mayores retos fue la elección, limpieza y preparación de los datos, al haber una gran variedad de campos que podrían generar análisis de un interés general, y con referencias previas de otras investigaciones, así como la adaptación de las trayectorias a la tecnología usada en el diseño de las soluciones, dada su variabilidad en longitud y estructura. Esto me llevó a diseñar funciones de reducción dimensional adaptativas (como el uso de clustering K-means para homogeneizar longitudes), que posteriormente se integraron dentro de los pipelines de análisis. Asimismo, la visualización de los resultados y la generación de figuras comprensibles fue un aspecto crucial que me ayudó a reforzar mi capacidad de comunicación científica.

En cuanto al impacto personal, considero que este trabajo ha sido una introducción sólida al mundo de la investigación aplicada en ciencia de datos. Me ha motivado a seguir profundizando en campos como el aprendizaje automático basado en geometría y topología, y ha reforzado mi interés por el análisis de patrones espaciales y temporales en entornos reales, como el transporte o la salud.

Conclusiones

Quisiera destacar especialmente la ayuda y orientación que me ha aportado mi tutor, Juan Antonio Fernández del Pozo en todo momento, manteniendo una comunicación fluida y constante, y ofreciéndome información, en forma de documentación, ejemplos, y sugerencias muy valiosas que han ayudado enormemente al resultado de este trabajo, así como la fluidez y constancia en la comunicación, ayudando a mantener unas pautas y un tempo de trabajo constante, que ha ayudado a seguir los tiempos establecidos inicialmente en el plan de trabajo en la medida de lo posible.

Finalmente, valoro positivamente la capacidad que este trabajo me ha aportado a la hora de estructurar un proyecto de análisis desde cero, empezando con el planteamiento del problema a resolver, continuando con una planificación de las tareas a realizar en base, y seguido por las diferentes etapas de la realización del mismo, como la revisión extensa de literatura y trabajos previos, desarrollo y experimentación computacional, y síntesis de resultados. Ha sido una experiencia exigente, pero altamente satisfactoria, y considero que me ha preparado tanto para futuros estudios de posgrado como para desafíos profesionales en el ámbito de la analítica avanzada.

5.4. Análisis del Impacto

5.4.1. Impacto general

El presente trabajo, enfocado en el análisis topológico de trayectorias mediante homología persistente y algoritmos de Mapper con la librería de Python `Giotto-TDA`, tiene un amplio potencial de aplicación en una multitud de campos. Al extraer características topológicas como componentes conexas, ciclos y cavidades de conjuntos de trayectorias en espacios métricos —por ejemplo, \mathbb{R}^2 para movilidad terrestre o \mathbb{R}^3 para recorridos aéreos o de GPS, este método ofrece descripciones robustas de la estructura global del movimiento. Dichos resúmenes topológicos son invariantes ante deformaciones continuas y resistentes al ruido, lo cual los hace especialmente útiles en escenarios reales donde los datos no están completos, o no tienen la precisión deseada. Como ya se ha mencionado en otras secciones en este mismo documento, este tipo de prácticas tienen especial relevancia en las siguientes áreas:

- **Movilidad urbana:** El análisis topológico de trayectorias permite identificar rutas frecuentes, núcleos de congestión y patrones de movimiento recurrentes. Por ejemplo, ciclos persistentes en diagramas de persistencia pueden corresponder a bucles viales o rutas circulares predominantes. Esta información puede ayudar a optimizar semáforos, carriles exclusivos o zonas peatonales, favoreciendo un desarrollo urbano más eficiente y sostenible.
- **Transporte aéreo:** En este contexto, las trayectorias de vuelo se representan como curvas en \mathbb{R}^3 (latitud, longitud, altitud). La homología persistente permite detectar patrones recurrentes, intersecciones o corredores aéreos significativos, mejorando la gestión del tráfico aéreo y la planificación de rutas más eficientes desde el punto de vista energético.
- **Sistemas dinámicos:** En contextos matemáticos o físicos, las trayectorias de sistemas dinámicos (como atractores caóticos) muestran estructuras complejas en el espacio de fases. Aplicar homología persistente revela invariantes topológicos que caracterizan el comportamiento del sistema, mientras que el algoritmo Mapper facilita la visualización de estos patrones complejos.

- **Otros campos:** Incluyendo logística, análisis ecológico, planificación de robots móviles y salud pública. En todos estos casos, el análisis topológico permite extraer patrones globales de movimiento y detectar trayectorias anómalas con mayor eficacia que los métodos tradicionales.

Durante el desarrollo del trabajo se tomaron decisiones metodológicas con el objetivo de maximizar el impacto práctico. Por ejemplo, se empleó la métrica euclidiana en la representación de las trayectorias por su interpretación geométrica intuitiva, y se eligieron funciones de filtro semánticamente relevantes en Mapper (como la distancia recorrida o el tiempo transcurrido). El uso de bibliotecas optimizadas como Giotto-TDA también refuerza la aplicabilidad de los resultados a contextos reales, dada su escalabilidad y compatibilidad con flujos de trabajo de machine learning.

5.4.2. Impacto en los Objetivos de Desarrollo Sostenible (ODS)

El análisis topológico de trayectorias contribuye directa e indirectamente a diversos Objetivos de Desarrollo Sostenible (ODS) definidos en la Agenda 2030:



Figura 5.1: Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030 [3]

- **ODS 11: Ciudades y comunidades sostenibles.** Mejorar la movilidad urbana es fundamental para lograr ciudades más inclusivas, seguras y eficientes. Los métodos propuestos ayudan a reducir la congestión y optimizar las rutas, favoreciendo una planificación urbana basada en datos.
- **ODS 9: Industria, innovación e infraestructura.** El uso de herramientas avanzadas como TDA fomenta la innovación tecnológica en transporte e infraestructura, habilitando nuevos modelos de gestión urbana, control de tráfico o rutas logísticas inteligentes.

Conclusiones

- **ODS 3: Salud y bienestar.** Una red de movilidad más eficiente contribuye a la salud pública mediante la reducción de accidentes, tiempos de respuesta de emergencias y exposición a la contaminación. Además, el análisis de trayectorias se puede aplicar a estudios epidemiológicos.
- **ODS 13: Acción por el clima.** La optimización de trayectorias y flujos de transporte, ya sea terrestre o aéreo, contribuye a reducir las emisiones de gases de efecto invernadero. Las herramientas desarrolladas pueden emplearse para diseñar rutas de bajo consumo energético.

En resumen, aunque este trabajo es principalmente metodológica, sus implicaciones abarcan diversos dominios con relevancia social, científica y ambiental. La capacidad de extraer descripciones topológicas robustas de trayectorias contribuye a la toma de decisiones informadas y sostenibles en múltiples contextos de aplicación.

Bibliografía

- [1] Miriam Esteve y Antonio Falc'o. «tramoTDA: A trajectory monitoring system using Topological Data Analysis». En: *SoftwareX* 28 (2024), pág. 101953. DOI: 10.1016/j.softx.2024.101953. URL: <https://doi.org/10.1016/j.softx.2024.101953>.
- [2] Daniela Espinoza-Molina, Florian Breitenberger y Thomas A Henzinger. «Topological data analysis for anomaly detection in air traffic data». En: *Journal of Air Transport Management* 104 (2022), pág. 100136. DOI: 10.1016/j.jairtraman.2022.100136. URL: <https://www.sciencedirect.com/science/article/pii/S0969699722000332>.
- [3] URL: <https://www.un.org/sustainabledevelopment/es/2015/09/la-asamblea-general-adopta-la-agenda-2030-para-el-desarrollo-sostenible/>.
- [4] Jéssica D Lamosa et al. «Topological indexes and community structure for urban mobility networks: Variations in a business day». En: *PLoS ONE* 16.3 (2021), e0248126. DOI: 10.1371/journal.pone.0248126.
- [5] Manuel Cuerno et al. «Topological Data Analysis in Air Traffic Management: The Shape of Big Flight Data Sets». En: *PLOS ONE* (2025). DOI: 10.1371/journal.pone.0318108. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0318108>.
- [6] P. Bosoni et al. «Predicting Multiple Sclerosis Relapses Using Patient Exposure Trajectories». En: *Proceedings of the Intelligent Disease Progression Prediction (iDPP) CLEF 2024 Challenge*. 2024. URL: <https://aisberg.unibg.it/handle/10446/292385>.
- [7] T. Shaikhina et al. «Topological data analysis identifies distinct biomarker clusters in Alzheimer's disease». En: *Scientific Reports* 10.1 (2020), pág. 20357. DOI: 10.1038/s41598-020-77896-0. URL: <https://www.nature.com/articles/s41598-020-77896-0>.
- [8] Joshua D Garland y Elizabeth Bradley. «Topological data analysis of dynamical systems: Concepts, computation, and applications». En: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 31.6 (2021), pág. 063116. DOI: 10.1063/5.0039292. URL: <https://doi.org/10.1063/5.0039292>.
- [9] Daniel Leykam y Dimitris G. Angelakis. «Topological Data Analysis and Machine Learning». En: *Advances in Physics: X* 8.1 (2023). DOI: 10.1080/23746149.2023.2202331. URL: <https://www.tandfonline.com/doi/full/10.1080/23746149.2023.2202331>.
- [10] Frédéric Chazal y Bertrand Michel. «An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists». En: *Frontiers in*

- Artificial Intelligence* 4 (2021). DOI: 10.3389/frai.2021.667963. URL: <https://www.frontiersin.org/articles/10.3389/frai.2021.667963/full>.
- [11] Felix Hensel, Michael Moor y Bastian Rieck. «A Survey of Topological Machine Learning Methods». En: *Frontiers in Artificial Intelligence* 4 (2021). DOI: 10.3389/frai.2021.681108. URL: <https://www.frontiersin.org/articles/10.3389/frai.2021.681108/full>.
- [12] URL: <https://gudhi.inria.fr/python/latest/>.
- [13] URL: <https://github.com/DIPHA/dipha>.
- [14] A. Lumino et al. *Giotto-TDA: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration*. 2020. URL: <https://github.com/giotto-ai/giotto-tda>.
- [15] tdaverse contributors. *tdaverse: A family of R packages for Topological Data Analysis*. <https://github.com/tdaverse/tdaverse>. 2023.
- [16] tdaverse contributors. *ripserr: Fast persistent homology computation in R*. <https://github.com/tdaverse/ripserr>. 2023.
- [17] CRAN Team. *phutil: Utilities for working with persistent homology in R*. <https://cran.r-project.org/web/packages/phutil/index.html>. 2023.
- [18] CRAN Team. *tdarec: Tools for Topological Data Analysis based on Reconstruction Methods*. <https://cran.rediris.es/web/packages/tdarec/readme/README.html>. 2023.
- [19] Peiyuan Jiang et al. «A topological data analysis network model of genes for cancer outcome prediction». En: *Frontiers in Oncology* 12 (2022), pág. 1043411. DOI: 10.3389/fonc.2022.1043411. URL: <https://www.frontiersin.org/articles/10.3389/fonc.2022.1043411/full>.
- [20] V. Balasubramanian, K. Krishnasamy y S. Meenakshi. «Detection of Cancer in Gene Expression Data Using Topological Data Analysis». En: *International Journal of VLSI design & Communication Systems (VLSICS)* 6.1 (2015). URL: <https://www.airconline.com/vlsics/V6N1/6115vlsi01.pdf>.
- [21] Ramy Saleh y Amal Riad. «Machine Learning and Topological Data Analysis Approaches for Diagnosing Diabetes Mellitus». En: *Handbook of Research on Advanced Practical Approaches to Deepfake Detection and Applications*. IGI Global, 2023, págs. 170-187. DOI: 10.4018/979-8-3693-0570-3.ch008. URL: <https://www.igi-global.com/chapter/machine-learning-and-topological-data-analysis-approaches-for-diagnosing-diabetes-mellitus/379970>.
- [22] Leyi Chi et al. «DeepPatient: An Online Framework for Personalized Diabetes Risk Prediction Using Longitudinal Electronic Health Records». En: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2021, págs. 2723-2731. DOI: 10.1145/3462741.3466653. URL: <https://dl.acm.org/doi/10.1145/3462741.3466653>.
- [23] Wei Liu et al. «Topological feature extraction and analysis of Alzheimer's disease based on functional MRI data». En: *Frontiers in Neurology* 14 (2023), pág. 1150634. DOI: 10.3389/fneur.2023.1150634. URL: <https://www.frontiersin.org/articles/10.3389/fneur.2023.1150634/full>.
- [24] Christina Hofer y et al. «Topological features of EEG in Parkinson's disease». En: *arXiv preprint* (2020). arXiv: 2004.07384 [q-bio.NC]. URL: <https://arxiv.org/pdf/2004.07384>.
- [25] Jaewook Jeong, Kang Kim y Yong Jeong. «Functional brain connectivity from EEG data: An application of persistent brain network homology». En: *Journal*

- of Parkinson's Disease* 6.4 (2016), págs. 699-710. DOI: 10.3233/JPD-150662. URL: <https://journals.sagepub.com/doi/full/10.3233/JPD-150662>.
- [26] Tianyu Cheng et al. «Early diagnosis of neurodegenerative diseases using persistent homology and deep learning». En: *Heliyon* 10.4 (2024), e34858. DOI: 10.1016/j.heliyon.2024.e34858. URL: <https://doi.org/10.1016/j.heliyon.2024.e34858>.
- [27] Manuel Cuerno et al. «Topological data analysis in air traffic management: The shape of big flight data sets». En: *PLOS ONE* 20.2 (2025), e0318108. DOI: 10.1371/journal.pone.0318108. URL: <https://doi.org/10.1371/journal.pone.0318108>.
- [28] Miriam Esteve y Antonio Falc'o. «Trajectory Classification Through Topological Data Analysis: Perspectives». En: *IEEE Access* 13 (2025), págs. 32458-32469. DOI: 10.1109/ACCESS.2025.3543111. URL: <https://doi.org/10.1109/ACCESS.2025.3543111>.
- [29] Ali Zia et al. «Topological deep learning: a review of an emerging paradigm». En: *Artificial Intelligence Review* 57 (2024), pág. 77. DOI: 10.1007/s10462-024-10710-9. URL: <https://doi.org/10.1007/s10462-024-10710-9>.
- [30] Meng Liu, Tamal K. Dey y David F. Gleich. «Topological structure of complex predictions». En: *Nature Machine Intelligence* 5 (2023), págs. 1382-1389. DOI: 10.1038/s42256-023-00749-8. URL: <https://doi.org/10.1038/s42256-023-00749-8>.
- [31] Dhananjay Bhaskar et al. «Topological data analysis of spatial patterning in heterogeneous cell populations: clustering and sorting with varying cell-cell adhesion». En: *NPJ Systems Biology and Applications* 9 (2023), pág. 43. DOI: 10.1038/s41540-023-00302-8. URL: <https://doi.org/10.1038/s41540-023-00302-8>.
- [32] Kyle Evans-Lee y Kevin Lamb. «Identification of Anomalous Geospatial Trajectories via Persistent Homology». En: *arXiv preprint arXiv:2410.03889* (2024). DOI: 10.48550/arXiv.2410.03889. URL: <https://arxiv.org/abs/2410.03889>.
- [33] URL: <https://www.kaggle.com/datasets/arashnic/microsoft-geolife-gps-trajectory-dataset>.
- [34] Yu Zheng et al. «Mining interesting locations and travel sequences from GPS trajectories». En: *WWW '09: Proceedings of the 18th international conference on World wide web* (2009), págs. 791-800. DOI: <https://doi.org/10.1145/1526709.1526816>. URL: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/fp120-zheng.pdf>.
- [35] Yu Zheng et al. «Understanding Mobility Based on GPS Data. In Proceedings of ACM conference on Ubiquitous Computing». En: *ACM Press* (2008), págs. 312-321. DOI: <http://dx.doi.org/10.1145/1409635.1409677>. URL: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/Ubicomp270-yuzheng.pdf>.
- [36] Wei-Ying Ma Yu Zheng Xing Xie. «GeoLife: A Collaborative Social Networking Service among User, location and trajectory.» En: *IEEE Data Engineering Bulletin* 33.2 (2010), págs. 32-40. URL: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/GeoLife-A20Collaborative20Social20Netwo.pdf>.
- [37] Colegio Bourbaki. *El algoritmo de Mapper y el análisis topológico de datos*. Consultado el 27 de mayo de 2025. 2021. URL: <https://www.colegio-bourbaki>.

- com/blog/el-algoritmo-de-mapper-y-el-analisis-topologico-de-datos.
- [38] Jianbo Yuan, Xuefei Li y Anastasios Noulas. «A Survey on Trajectory Clustering Analysis Techniques for Trajectory Mining». En: *IEEE Transactions on Knowledge and Data Engineering* (2020). DOI: 10.1109/TKDE.2020.3034874.
 - [39] Yu Zheng, Xiaofang Zhou y Xing Xie. «Understanding Mobility Based on GPS Data». En: *Proceedings of the 10th International Conference on Ubiquitous Computing* (2008), págs. 312-321. DOI: 10.1145/1409635.1409677.
 - [40] C. Aggarwal y C. Reddy. *Data Clustering: Algorithms and Applications*. Chapman y Hall/CRC, 2013.
 - [41] I.T. Jolliffe. *Principal Component Analysis*. 2nd. Springer, 2002.
 - [42] J. MacQueen. «Some Methods for Classification and Analysis of Multivariate Observations». En: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1* (1967), págs. 281-297.
 - [43] Martin Ester et al. «A density-based algorithm for discovering clusters in large spatial databases with noise». En: (1996), págs. 226-231.
 - [44] Thomas Eiter y Heikki Mannila. «Computing Discrete Fréchet Distance». En: *Technical Report CD-TR 94/64* (1994).
 - [45] Donald J. Berndt y James Clifford. «Using Dynamic Time Warping to Find Patterns in Time Series». En: (1994), págs. 359-370.
 - [46] Matplotlib Development Team. *Matplotlib: Visualization with Python*. 2023. URL: <https://matplotlib.org>.
 - [47] Michael Waskom et al. *Seaborn: Statistical Data Visualization*. 2023. URL: <https://seaborn.pydata.org>.
 - [48] Mathieu Carrière, Bertrand Michel y Steve Oudot. «Statistical Analysis and Parameter Selection for Mapper». En: *Journal of Machine Learning Research* 19.1 (2018), págs. 478-516. URL: <http://jmlr.org/papers/v19/17-047.html>.
 - [49] OpenAI. *ChatGPT*. Disponible en: <https://chatgpt.com/>. 2022.

Apéndice A

Anexo

A.1. Materiales Desarrollados

El código, junto con los datos específicos utilizados para generar los resultados para este trabajo se encuentran en https://github.com/krlokbrera/TDA_CarloCabrera


A.2. Uso de Inteligencia Artificial Generativa

Para la realización de este trabajo, se ha hecho uso de herramientas de IA Generativa, como sería *ChatGPT* [49], para consultas relativas a la realización de la memoria (estructura de figura en overleaf, generación de citas).

A.3. Turnitin

Tanto el informe de originalidad como el recibo digital se encuentran disponibles en el repositorio de Github https://github.com/krlokbrera/TDA_CarloCabrera , más en concreto en la sección de documentos, junto con los archivos .tex utilizados para generar esta memoria.

Este documento esta firmado por



Firmante	CN=tfgm.fi.upm.es, OU=CCFI, O=ETS Ingenieros Informaticos - UPM, C=ES
Fecha/Hora	Wed Jun 04 10:20:02 CEST 2025
Emisor del Certificado	EMAILADDRESS=camanager@etsiinf.upm.es, CN=CA ETS Ingenieros Informaticos, O=ETS Ingenieros Informaticos - UPM, C=ES
Numero de Serie	561
Metodo	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)