

## **Aplicaciones del Análisis Topológico de Datos en Economía**

Heriberto Espino Montelongo

Universidad de las Américas Puebla

P24-LDS1041-1: Análisis Topológico de Datos

Dr. Hugo Villanueva Méndez

5 de mayo de 2024

## **Aplicaciones del Análisis Topológico de Datos en Economía**

El análisis topológico de datos (ATD) ha surgido como una herramienta en el estudio de los mercados financieros, ofreciendo una perspectiva única para comprender los complejos patrones de comportamiento que caracterizan a estos sistemas dinámicos. Al representar los datos de precios de activos como series temporales y aplicar ATD, se puede identificar y analizar cambios estructurales en los mercados, lo que tiene importantes implicaciones para la toma de decisiones en inversiones y políticas económicas.

En este ensayo, exploraremos las aplicaciones del ATD en economía, más específicamente en mercados bursátiles y cómo esta disciplina puede beneficiarse de esta perspectiva innovadora. Recientes estudios han demostrado que el ADT puede tener un impacto significativo aplicándose al estudio de mercados financieros para identificar y analizar patrones de comportamiento.

En el ámbito de la investigación financiera, el ADT ha demostrado un considerable potencial en una amplia variedad de dominios, particularmente en aquellos que manejan datos de alta dimensionalidad (Bubenik, 2015; Carlsson, 2009; Wasserman, 2017). El éxito del ADT abarca desde la investigación médica (McInemey y Terzopoulos, 1999; Nicolau et al., 2011) hasta la investigación financiera (Baitinger y Flegel, 2021; Gidea y Katz, 2018; Ismail et al., 2022). Aunque numerosos campos como la medicina y la biología, su incorporación en han investigado extensamente las aplicaciones del ADT, la literatura financiera no ocurrió hasta tuvo su primer artículo de investigación en ADT solo en 2017, lo cual refleja tanto su novedad como su potencial emergente en el análisis económico.

De acuerdo con García Colin N, el ATD se ha practicado con éxito en los últimos 15 años para estudiar cómo se puede inferir información de un sistema de datos a partir de muestras

representadas como un espacio topológico. Proporciona un conjunto de herramientas para analizar información multidimensional, lo que lo convierte en un método prometedor para economía. Carlsson and de Silva (2008) destacan que el ATD ha permitido el desarrollo de nuevas herramientas para el análisis y la visualización de conjuntos en datos complejos, lo que ha llevado a avances significativos en estadística, el aprendizaje automático e inteligencia artificial. En lugar de depender únicamente de medidas métricas, como la distancia euclidiana, el ATD se centra en las propiedades topológicas de los datos, lo que permite revelar patrones subyacentes que pueden pasar desapercibidos con enfoques tradicionales.

Una de las bases fundamentales del ATD son los simplejos, que son generalizaciones de los conceptos de puntos, líneas y polígonos en geometría euclidiana. En el contexto del ATD, los simplejos pueden representar cualquier conjunto finito de puntos que están conectados entre sí. Por ejemplo, un simplejo de dimensión cero es un punto, un simplejo de dimensión uno es una línea, y así sucesivamente.

Uno de los algoritmos clave en ATD es el algoritmo Mapper, que se utiliza para visualizar la estructura de datos de alta dimensionalidad en términos de simplejos. El algoritmo Mapper divide el espacio de datos en regiones solapadas y construye un complejo simplicial a partir de estas regiones, lo que permite visualizar las relaciones entre los datos en términos de simplejos y sus intersecciones.

Entre las herramientas más mencionadas sobre la materia, la Homología persistente que es una técnica para identificar características topológicas importantes en un conjunto de datos, con representaciones visuales como códigos de barras o diagramas de persistencia, que nos permiten visualizar las características topológicas de nuestros datos. Al emplear la homología persistente, los investigadores pueden descubrir patrones ocultos pero fundamentales en los

datos, lo que conduce a descubrimientos novedosos en los datos (Bubenik, 2015; Carlsson, 2009; Wasserman, 2017).

En conjunto, el análisis topológico de datos, con su enfoque en los simplejos, el algoritmo Mapper, y, principalmente, homología persistente, ofrecen una poderosa herramienta para explorar y comprender la estructura y las relaciones en conjuntos de datos financieros revelando patrones y características que pueden no ser evidentes con enfoques tradicionales de análisis de datos.

En cuanto a las aplicaciones económicas el ATD puede partir de temas simples como aplicarse al análisis de grandes conjuntos de datos comerciales, como transacciones de compra y venta, para identificar patrones de comportamiento del consumidor, segmentar mercados y mejorar la eficacia de las estrategias de marketing, hasta cosas más complejas como representar los datos de precios de activos como una serie temporal, se puede aplicar el ATD para identificar cambios estructurales en los mercados, como burbujas especulativas o cambios en la volatilidad. También se puede usar en análisis de flujos de datos económicos, como el comercio internacional y la migración laboral, identificando patrones emergentes, como comunidades de países con patrones comerciales similares o corredores migratorios importantes, comprendiendo mejor las interconexiones económicas a nivel global y formulando políticas que fomenten un mejor crecimiento. Además, puede ser aplicado al análisis de datos macroeconómicos, como el producto interno bruto y la inflación, identificando regiones de estabilidad y cambios abruptos en la economía, previniendo crisis financieras.

A través de una revisión diversas fuentes relevantes en este campo, exploraremos las diferentes aplicaciones y problemas estudiados.

Guo et al, se centran en el análisis de crisis financieras investigando la dinámica de las crisis financieras a través del análisis de series temporales financieras de varios países y regiones, examinando las relaciones entre diferentes mercados financieros e identificando que indican la ocurrencia de crisis económicas a través de las covarianzas.

Goel et al, exploran la toma de decisiones de inversión con ATD, observando la volatilidad que puede ayudar a identificar patrones complejos en los datos financieros y mejorando la toma de decisiones de inversión.

Otra vez, Guo et al. se centran en analizar las conexiones entre los mercados bursátiles globales, con especial énfasis en el impacto de la pandemia de COVID-19, investigando cómo se intensificaron las relaciones entre los mercados y llegando a proporcionar información sobre la interconexión de los mercados financieros internacionales durante períodos de crisis.

Baitinger y Flegel proponen un índice de turbulencia basado en homología persistente para detectar adversidades en los mercados financieros, investigando la capacidad de previsión de diferentes índices de turbulencia en condiciones de mercado difíciles, concluyendo que el enfoque basado en la homología persistente supera a otros índices tradicionales en la detección de condiciones adversas del mercado.

Gobato Souto investiga la dependencia de la cola topológica en la predicción de la volatilidad realizada en los mercados financieros. Proponiendo un enfoque basado en prever la volatilidad en períodos turbulentos del mercado. Demostrando la eficacia y la identificación temprana de condiciones de mercado adversas a través del ATD.

Guo et al. (2020) realizaron un análisis de las series temporales financieras de Estados Unidos, específicamente los índices DIJA y NASDAQ, durante el período comprendido entre el 2 de enero de 2003 y el 31 de diciembre de 2013. Observaron la volatilidad realizada y vieron un

crecimiento brusco en los diagramas de persistencia homológica durante la segunda mitad de 2008, coincidiendo con una rápida disminución de los índices bursátiles. Por otro lado, la persistencia se mantuvo en niveles bajos durante períodos de rápido aumento de los índices, como en 2009. Concluyendo que el ADT puede ser una herramienta eficaz para identificar señales tempranas de crisis y entender la dinámica de los mercados durante períodos turbulentos. Las conclusiones obtenidas son de tipo inferencial, ya que se deduce que los cambios bruscos en los diagramas de persistencia se interpretan como indicadores de crisis económicas. Los resultados se interpretaron en términos de la relación entre los cambios en las el diagrama de persistencia y los eventos económicos significativos.

El artículo de Goel, Pasricha y Mehra (2020) explora las aplicaciones del ADT en el campo financiero, abordando especialmente el problema de asignación de activos. En primer lugar, explican por qué el ADT puede ser una mejor alternativa que los indicadores de riesgo tradicionales, como la desviación estándar, utilizando conjuntos de datos reales. Aplican el teorema de incrustación de Takens para reconstruir las series temporales de rendimientos en un espacio de alta dimensionalidad. Adoptaron el enfoque de deslizante para obtener conjuntos de datos de nubes de puntos dependientes del tiempo y les asociaron un espacio topológico. Luego aplicaron homología persistente para descubrir los patrones que aparecen en las series temporales. Los cambios temporales en los diagramas de persistencia se capturaron mediante la norma  $L_p$ . La serie temporal de las normas  $L_p$  muestra que es mejor para medir la dinámica de los rendimientos que la desviación estándar. Aplicaron el ADT para la Indexación Mejorada (EI) que tiene como objetivo construir una cartera con menos activos que la del índice para superarlo. Proponen un procedimiento de dos pasos para lograr esta tarea. En el primer paso, utilizan las normas  $L_p$  de los activos para proponer una técnica de filtrado para seleccionar una cartera de

activos de la clase de activos filtrados para la EI. Para probar la eficiencia de este algoritmo mejorado, se realizaron experimentos en diez conjuntos de datos de mercados financieros de todo el mundo. Su análisis muestra que la estrategia propuesta ofrece un rendimiento superior en varias medidas, incluidos los retornos medios respecto al índice de referencia y las relaciones de recompensa-riesgo de cola en comparación con algunos de los modelos existentes. Los resultados empíricos demuestran que tanto los portafolios lograron el mayor Exceso de Retorno Medio (EMR) para 16 casos, de entre 40 casos, superando al índice en ambos casos. Además, las ratios de recompensa-riesgo muestran un desempeño superior para los portafolios en la mayoría de los casos. Demostrando que el ADT es una buena herramienta para la toma de decisiones de inversión, ofreciendo resultados prometedores y una mejora significativa en el rendimiento de las carteras, especialmente en períodos de alta volatilidad del mercado financiero.

El estudio de Guo, Zhao, Yu y Zhang (2021) se centra en explorar las conexiones de los mercados bursátiles globales durante las crisis financieras o riesgos desde 1995, con especial énfasis en la situación bajo COVID-19. Los autores eligen 40 países/regiones y toman un índice de cada uno, luego calculan los coeficientes de correlación y las distancias entre cada par de índices con un deslizante. Construyeron complejos y realizaron análisis a través de diagramas de persistencia. Establecieron un sistema de detección de fechas críticas basado en los diagramas de persistencia. Las características topológicas de las redes se muestran en las fechas críticas y en las fechas anteriores a ellas. Todos los resultados muestran que las conexiones entre los mercados se hicieron aún más estrechas cuando se propagó COVID-19 en todo el mundo que en cualquier otro riesgo. La eficacia de estos métodos proporciona orientación para el análisis de crisis financieras en el futuro. Los autores examinan específicamente tres crisis: la crisis financiera global de 2007-2008, la crisis de deuda europea de 2010-2011 y la crisis bajo COVID-

19 en 2020. Utilizan el Análisis Topológico de Datos (ADT) para analizar la estructura topológica de las series temporales financieras y detectar puntos críticos con cambios abruptos. Construyeron un sistema de detección de fechas críticas basado en las normas  $L_p$  de los diagramas de persistencia y utilizaron árboles de expansión mínima (MST) para estudiar el sistema financiero y el mecanismo de transmisión del riesgo. En cuanto a las conclusiones obtenidas, el estudio proporciona información valiosa para entender la interconexión de los mercados bursátiles globales durante las crisis financieras, con la detección de fechas críticas y la identificación de cambios significativos en la estructura topológica de los mercados. Estas conclusiones pueden utilizarse para hacer inferencias sobre la propagación del riesgo financiero y la interdependencia de los mercados en tiempos de crisis. El estudio sugiere que el ADT es útil para clasificar y analizar las crisis financieras basadas en la topología de los datos financieros, lo que llevaría a entender mejor la dinámica del mercado y la toma de decisiones de inversión.

Baitinger, E. y Flege, S. (2021) propusieron un índice de turbulencia basado en homología persistente para la detección de mercado adversos. Demostraron que las estrategias de inversión que se basan en la detección de turbulencia mediante homología persistente superan a aquellas basadas en otros índices de turbulencia populares. Proporcionando una herramienta efectiva para los gestores de carteras en la identificación de períodos de alta volatilidad y baja rentabilidad en los mercados financieros. Al realizar un análisis de estabilidad, los autores confirmaron la superioridad de las estrategias de inversión basadas en la homología persistente, incluso en diversas configuraciones. Mencionan “Desde una perspectiva a largo plazo, los gestores de carteras de acciones se enfrentan regularmente a períodos turbulentos, caracterizados por bajos rendimientos promedio, alta volatilidad y correlaciones aumentadas. Durante estos períodos, el efecto de diversificación desaparece, lo que plantea un desafío fundamental para la



industria de gestión de activos (Chua et al., 2009; Leibowitz y Bova, 2009; Asness et al., 2011; Yang et al., 2017)”. El artículo también destaca dos enfoques básicos propuestos por la academia financiera para enfrentar las consecuencias no deseadas de períodos turbulentos. El primero implica la construcción de carteras robustas, que son las que esperan el peor retorno, mediante la integración de estimaciones negativas o de peor caso en la función objetivo de la optimización de cartera (Bauer et al., 2004; Tütüncü y Koenig, 2004; Scherer, 2007). Sin embargo, estas carteras robustas tienden a infravalorar el retorno en tiempos tranquilos. El segundo enfoque se basa en la detección dinámica de regímenes, permitiendo estrategias de inversión con menor riesgo y mayor rendimiento total en comparación con la estrategia de comprar y mantener (Ang y Bekaert, 2004, 2015; Chen, 2009). La propuesta de Baitinger y Flege representa una metodología para identificar condiciones de alta volatilidad en los mercados financieros basándose en homología persistente, facilitando la identificación de cambios. Demostrando el rendimiento superior del índice de turbulencia basado en la homología persistente para el mercado de valores de EE. UU. Además, realizaron un análisis estadístico de las señales comerciales respectivas, confirmando sus análisis.

El estudio realizado por Gobato Souto (2023) propone la utilización de la homología persistente en la predicción de la volatilidad, proponiendo una nueva teoría que vincula la homología persistente y la teoría financiera, llamada la Dependencia de Cola Topológica, probando las consecuencias de la teoría propuesta a través de una experimentación. Las principales contribuciones del estudio incluyen la propuesta de un nuevo algoritmo para medir los cambios topológicos de series temporales de acciones, adaptando el índice de turbulencia basado en persistencia homológica propuesto originalmente por Baitinger y Flegel (2021), para ser empleado efectivamente en la predicción de volatilidad. Además, el estudio incorpora la

métrica de cambio topológico propuesta en modelos lineales, no lineales y de redes neuronales para pronosticar la volatilidad realizada. Se demuestra que esta incorporación propuesta logra pronósticos estadísticamente significativamente más precisos que los modelos existentes en períodos financieramente turbulentos para todos los modelos. Las conclusiones obtenidas son de tipo inferencial, ya que se demuestra que la incorporación de información de persistencia homológica permite a los modelos no lineales y de redes neuronales predecir mejor la volatilidad realizada durante períodos turbulentos. Los resultados se interpretan en términos de mejora en la precisión de la predicción durante períodos turbulentos, respaldando la teoría propuesta.

Guo et al. (2020) realizó un estudio empírico sobre las crisis financieras. Este estudio examina cómo el ADT puede ayudar en la detección temprana de crisis financieras mediante el análisis de la homología persistente en series temporales financieras. Identificando patrones que pueden indicar la proximidad de una crisis.

Por otro lado, Goel et al. (2020) exploraron el uso del ADT en la toma de decisiones de inversión. Este estudio sugiere que el ADT puede aplicarse no solo en el ámbito financiero, sino también en otros sectores económicos, como la industria manufacturera o el comercio internacional.

El estudio de Guo et al. (2021) se centró en analizar las conexiones entre los mercados bursátiles globales, con especial énfasis en el impacto de la pandemia de COVID-19. Utilizando el ATD, examinaron cómo se intensificaron las relaciones entre los mercados durante la crisis del COVID-19, observando información importante sobre la conexión de los mercados financieros internacionales durante períodos de crisis.

Baitinger y Flegel (2021) propusieron un nuevo índice de turbulencia basado en la homología persistente para la detección de adversidades en los mercados financieros. Este

estudio destaca cómo el ADT puede mejorar la comprensión y la gestión del riesgo en los mercados financieros, proporcionando una visión única de la dinámica del mercado.

Finalmente, Gobato Souto (2023) investigó la dependencia de la cola topológica en la predicción de la volatilidad realizada en los mercados financieros. Demostrando cómo el ADT puede utilizarse para pronosticar la volatilidad en los mercados financieros, identificando patrones significativos que indican cambios bruscos en las condiciones del mercado.

Los estudios revisados muestran que el análisis topológico de datos ofrece una perspectiva única y poderosa para abordar una variedad de problemas en economía y finanzas. Desde la detección temprana de crisis financieras hasta la predicción de la volatilidad en los mercados, el ADT ha demostrado ser una fuerte herramienta para comprender y gestionar la complejidad de los fenómenos económicos. Además, estos estudios sugieren que el ADT tiene el potencial de extenderse a nuevos campos económicos, proporcionando nuevas formas de abordar problemas económicos y comerciales en el futuro.

## References

- Baitinger, E., Flegel, S., 2021. *The better turbulence index? Forecasting adverse financial markets regimes with persistent homology*. *Financ. Mark. Portfolio Manag.* 35 (3), 277–308. <https://doi.org/10.1007/s11408-020-00377-x>.
- García-Colín, N. (s.f.). *Una introducción al análisis topológico de datos. En XXX Coloquio Víctor Neumann-Lara de Teoría de Gráficas, Combinatoria y sus Aplicaciones*. <https://infotec.repositorioinstitucional.mx/jspui/bitstream/1027/199/1/Una%20introducci%C2%B4on%20al%20an%C2%B4alisis%20topol%C2%B4ogico%20de%20datos.pdf>
- Gobato Souto, H. (2023). *Topological tail dependence: Evidence from forecasting realized volatility*. *The Journal of Finance and Data Science*, 9, 100107. <https://doi.org/10.1016/j.jfds.2023.100107>.
- Goel, A., Pasricha, P., Mehra, A. (2020). *Topological data analysis in investment decisions*. *Expert Syst. Appl.* 147, 113222. <https://doi.org/10.1016/j.eswa.2020.113222>.
- Guo, H., Xia, S., An, Q., Zhang, X., Sun, W., and Zhao, X. (2020). *Empirical study of financial crises based on topological data analysis*. *Physica A: Statistical Mechanics and its Applications*, 558:124956. <https://doi.org/10.1016/j.physa.2020.124956>.
- Guo, H., Zhao, X., Yu, H., & Zhang, X. (2021). *Analysis of global stock markets' connections with emphasis on the impact of COVID-19*. *Physica A: Statistical Mechanics and its Applications*, 569, 125774. <https://doi.org/10.1016/j.physa.2021.125774>
- Paredes Aguilera, C. (2023, 29 de mayo). *Utilización del análisis topológico de datos en el ámbito económico*. <https://www.researchgate.net/publication/371133996>