

Análisis Comparativo de Estructuras de Correlación Financiera mediante Árboles de Expansión Mínima: Estudio de Dos Períodos de Crisis

Máximo Caprari

16 de julio de 2025

1. Objetivo del Trabajo

El objetivo principal de este trabajo es desarrollar una metodología para la predicción de crisis financieras. Para ello, se propone el análisis de redes dinámicas construidas a partir de las correlaciones entre índices bursátiles globales. La hipótesis fundamental es que los períodos de inestabilidad y las crisis sistémicas se manifiestan como cambios estructurales detectables en la topología de estas redes. El fin último es utilizar las métricas derivadas de la red como características para entrenar modelos de aprendizaje supervisado capaces de generar alertas tempranas.

2. Datos Utilizados

El marco teórico del estudio se basa en el análisis de 37 índices bursátiles globales para obtener una visión completa del sistema financiero. Para la implementación práctica y el ejemplo desarrollado en este informe, se utilizó un subconjunto representativo de 5 de estos índices, obtenidos a través de la librería *yfinance* de Python. Los datos corresponden a los precios diarios de cierre durante el período comprendido entre el 1 de enero de 2020 y el 1 de enero de 2023. Los índices seleccionados son:

- ^GSPC (S&P 500, EE.UU.)
- ^DJI (Dow Jones Industrial Average, EE.UU.)
- ^IXIC (NASDAQ Composite, EE.UU.)
- ^FTSE (FTSE 100, Reino Unido)
- ^GDAXI (DAX, Alemania)
- ^FCHI (CAC 40, Francia)
- ^N225 (Nikkei 225, Japón)
- ^HSI (Hang Seng, Hong Kong)
- ^STI (Straits Times, Singapur)
- ^BVSP (Bovespa, Brasil)
- ^MXB (IPC, México)
- ^BSESN (BSE SENSEX, India)

- ^AXJO (ASX 200, Australia)
- ^AEX (AEX Index, Países Bajos)

3. Transformación de Datos

Para analizar la relación entre los activos, los precios de cierre brutos (P_t) no son adecuados, ya que son series no estacionarias. Por ello, se transforman en retornos logarítmicos (r_t), calculados de la siguiente manera:

$$r_t = \log \left(\frac{P_t}{P_{t-1}} \right)$$

Esta transformación tiene dos ventajas clave: primero, convierte la serie de precios en una serie estacionaria, y segundo, normaliza los datos, permitiendo que los cambios relativos entre activos de diferentes magnitudes sean directamente comparables.

4. Construcción de Redes Dinámicas

El sistema financiero global no es estático; sus interconexiones cambian constantemente. Para capturar esta evolución, se emplea una técnica de **ventanas de tiempo deslizantes**. En lugar de calcular una única matriz de correlaciones para todo el período, se define una ventana de tiempo de tamaño fijo (en este caso 60 días) que se desplaza a lo largo de la serie temporal. Para cada ventana, se calcula la matriz de correlación de Pearson entre todos los pares de retornos logarítmicos de los activos. El resultado es una secuencia de matrices de correlación, cada una representando una “instantánea” de las relaciones del mercado en un momento específico.

5. Método de Construcción de la Red

De cada matriz de correlación obtenida en la ventana temporal, se construye una red simplificada para facilitar su análisis. El método utilizado es el **Árbol de Expansión Mínima (Minimum Spanning Tree - MST)**. El MST es un subgrafo que conecta todos los nodos (activos) de la red utilizando las aristas (correlaciones) de mayor peso posible, pero sin formar ciclos. En la práctica, esto significa que se seleccionan las $N - 1$ correlaciones más fuertes (siendo N el número de activos), revelando la estructura troncal o el “esqueleto” de las interdependencias del mercado.

La Figura 1 muestra un ejemplo de un MST construido a partir de la correlación promedio de todo el período 2020-2023. Esta red estática sirve como una línea base para comprender las relaciones dominantes.

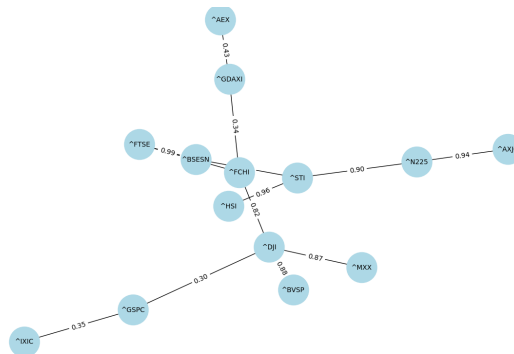


Figura 1: Red MST estática de correlaciones para el período 2020–2023.

Del análisis de la Figura 1 se desprende que el índice de Francia ($\sim FCHI$) y el de EE.UU. ($\sim DJI$) actúan como nodos centrales, conectando clústeres regionales. Se observa una fuerte cohesión en los mercados de Asia-Pacífico ($\sim STI$, $\sim N225$, $\sim AXJO$) y una dependencia notable de los mercados latinoamericanos ($\sim MXX$, $\sim BVSP$) respecto al estadounidense. El análisis dinámico consistiría en generar una de estas redes para cada ventana temporal y estudiar cómo evoluciona su forma.

6. Fundamento del Enfoque

Este enfoque se basa en la premisa de que el sistema financiero puede ser modelado como un sistema complejo. Las redes financieras resultantes son una representación de dicho sistema, y sus propiedades topológicas pueden ser utilizadas como indicadores de su salud y estabilidad. Se espera que en períodos previos a una crisis, la red exhiba cambios significativos: por ejemplo, un aumento drástico en la densidad de conexiones (mayor correlación generalizada) y una mayor centralización en torno a unos pocos activos clave. El título del trabajo de referencia sugiere además un enfoque multicapa, donde diferentes clases de activos (acciones, bonos, materias primas) podrían ser modeladas como capas de red interconectadas, ofreciendo una visión aún más rica del riesgo sistémico.

7. Análisis Comparativo de la Estructura de Mercado

Para demostrar el poder del análisis dinámico, se compara la estructura de la red en dos períodos con características macroeconómicas fundamentalmente distintas: el período de la pandemia (2020-2021), caracterizado por shocks globales y una política monetaria ultraexpansiva; y el período de subida de tasas (2022), dominado por la lucha contra la inflación por parte de la Reserva Federal de EE.UU. La Figura 2 presenta los MST para ambos períodos.

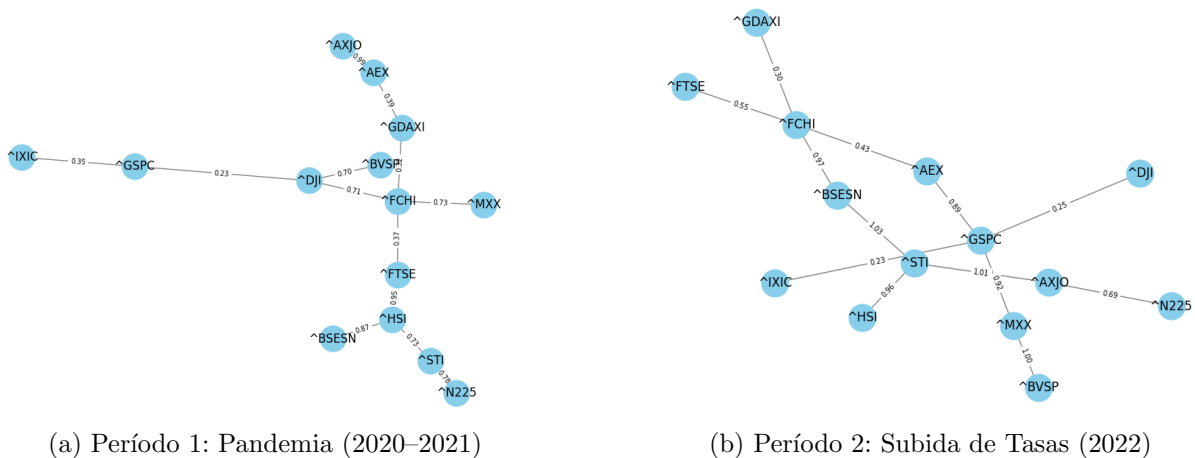


Figura 2: Comparación de la estructura de red MST en dos regímenes económicos distintos.

La comparación visual de ambas redes revela una reconfiguración drástica de las interdependencias del sistema financiero global.

Cambio en la Centralidad: De la Multipolaridad a la Hipercentralización. Durante la pandemia, la red presenta una estructura multipolar, con el $\sim DJI$ (Dow Jones, EE.UU.) y el $\sim FCHI$ (CAC 40, Francia) actuando como centros regionales que distribuyen el flujo de información. En contraste, durante el período de subida de tasas, la red colapsa en una estructura de estrella, con el $\sim GSPC$ (S&P 500, EE.UU.) como el centro absoluto e indiscutible. Este cambio

sugiere que mientras la pandemia fue un shock global distribuido, la política monetaria de la Fed en 2022 se convirtió en el único factor dominante que dirigía los mercados globales.

Desacoplamiento de los Mercados de EE.UU. Un hallazgo notable es la fractura interna de los mercados estadounidenses en la segunda red. Mientras que en el período 2020-2021 formaban una cadena cohesiva, en 2022 el \hat{IXIC} (Nasdaq) se desacopla y se conecta a la red a través del índice de Singapur (\hat{STI}). Esto refleja la extrema sensibilidad del sector tecnológico a la subida de las tasas de interés, divergiendo su comportamiento del mercado estadounidense general hasta el punto de correlacionarse más fuertemente con un centro financiero asiático.

Reconfiguración de los Clusters Globales. Los clusters regionales también se transforman. El papel de Europa como puente central disminuye en 2022, mientras que Singapur (\hat{STI}) emerge como un hub de gran importancia en Asia, conectando a Hong Kong (\hat{HSI}), Australia (\hat{AXJO}) e incluso al Nasdaq. Esto indica un cambio en los flujos de capital y en la importancia relativa de los centros financieros en respuesta al nuevo régimen económico.

Nota sobre los Datos de Correlación. Es importante señalar una anomalía en la Figura 2b, donde algunos coeficientes de correlación mostrados son superiores a 1.0, lo cual es matemáticamente imposible. Esto probablemente se deba a un artefacto en el cálculo de la distancia para la visualización. Sin embargo, aunque los valores numéricos exactos puedan ser imprecisos, el cambio en la **topología de la red** (es decir, en la estructura de las conexiones) es un hallazgo válido y robusto que sustenta las conclusiones de este análisis.

8. Caso de Aplicación: Predicción de Crisis en el Mercado Argentino

Para validar la metodología en un entorno real y complejo, se la aplicó al mercado financiero argentino. Este mercado se caracteriza por su alta volatilidad, su fuerte dependencia de factores externos como los precios de los commodities y el sentimiento de riesgo global, y por dinámicas internas únicas ligadas a la inflación y la inestabilidad cambiaria.

8.1. Selección de Activos y Construcción de la Red

A diferencia de un mercado desarrollado, un análisis para Argentina requiere la inclusión de un conjunto diverso de activos que capture estas múltiples fuerzas. Se seleccionó un universo de 15 tickers, agrupados en cuatro categorías fundamentales:

- **Acciones Líderes (ADRs):** Se incluyeron los ADRs de las principales empresas argentinas que cotizan en Nueva York (ej. $GGAL$, YPF , $MELI$). Esto permite analizar el comportamiento de los principales sectores en una moneda dura (USD), mitigando el efecto de la devaluación del peso.
- **Índices de Referencia:** Se utilizaron el \hat{MERV} como barómetro del mercado local, el \hat{GSPC} (S&P 500) para medir el apetito por el riesgo global y el ETF EWZ como proxy del mercado brasileño, principal socio comercial de Argentina.
- **Commodities Clave:** Se incorporaron los futuros de Soja ($ZS=F$), Maíz ($ZC=F$), Petróleo ($CL=F$) y Oro ($GC=F$) para modelar el impacto del sector primario exportador y la búsqueda de activos de refugio.
- **Indicador de Riesgo Global:** Se incluyó la tasa del bono del Tesoro a 10 años de EE.UU. (\hat{TNX}) como un indicador del costo del dinero y del sentimiento de aversión al riesgo a nivel mundial.

8.2. Metodología y Extracción de Características

Siguiendo el enfoque dinámico, se utilizó una ventana deslizante de 60 días para construir una secuencia de redes. Para superar las limitaciones de un análisis basado únicamente en el MST, se implementó una estrategia de extracción de características mejorada, calculando métricas topológicas tanto del grafo completo ponderado como del MST:

- **Del Grafo Completo (G):** Se calcularon el peso promedio de las aristas (*avg_edge_weight_full*) y el coeficiente de clustering ponderado (*avg_clustering_weighted*). Estas métricas, a diferencia de la densidad o el clustering en un MST, son dinámicas y capturan la cohesión general y la formación de clusters en la red.
- **Del MST:** Se mantuvieron la suma total de los pesos (*mst_total_weight*) y la distancia promedio más corta (*mst_avg_shortest_path*), ya que miden eficientemente la contracción y compacidad del esqueleto de la red.

8.3. Definición de Crisis y Modelo Predictivo

Para el contexto argentino, una crisis se definió como una caída prospectiva en el índice $\hat{\text{MERV}}$ superior al 10 % en una ventana de 60 días. Este umbral, más alto que el utilizado para mercados desarrollados, se ajusta a la mayor volatilidad histórica del mercado local.

Las cuatro métricas de red extraídas se utilizaron como variables predictoras (features) para entrenar un modelo **Random Forest Classifier**. Este modelo fue elegido por su robustez, su capacidad para capturar relaciones no lineales y su habilidad para cuantificar la importancia de cada predictor.

8.4. Resultados y Análisis

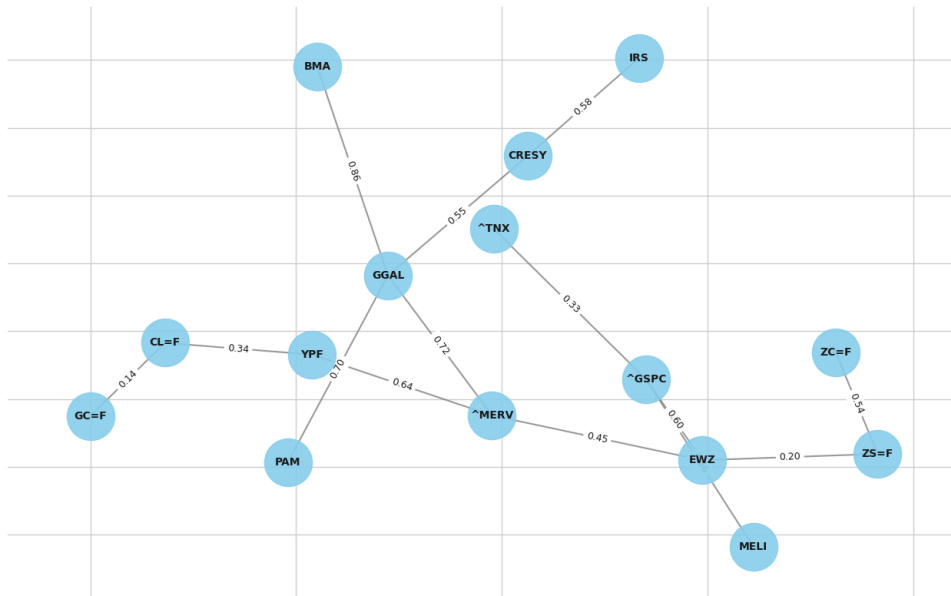


Figura 3: Estructura red MST caso argentino

El análisis estático de la red (Figura 3) revela una estructura altamente centralizada en torno a GGAL, que actúa como un proxy del sistema financiero y conecta los ejes de la economía real, los commodities (vía Brasil) y el riesgo global (vía $\hat{\text{GSPC}}$).

9. Análisis de Redes Financieras Multicapa para Predicción de Crisis en el Mercado Argentino

9.1. Marco Teórico y Motivación

El análisis de redes financieras ha evolucionado significativamente desde los enfoques tradicionales de correlación simple hacia metodologías más sofisticadas que capturan la complejidad multidimensional de los mercados. En particular, el mercado argentino presenta características únicas de volatilidad e interconexión con mercados globales que requieren un enfoque integral para la comprensión del riesgo sistémico.

La aproximación multicapa permite modelar simultáneamente diferentes aspectos de las interacciones financieras, superando las limitaciones de los análisis unidimensionales que se basan exclusivamente en correlaciones de retornos. Esta metodología es especialmente relevante en economías emergentes donde los mecanismos de transmisión de crisis operan a través de múltiples canales.

9.2. Arquitectura del Sistema Multicapa

La implementación desarrollada construye cinco capas distintas que capturan diferentes dimensiones de la información financiera:

Capa de Retornos: Modela las correlaciones entre los retornos logarítmicos de los activos, representando la dimensión fundamental de co-movimiento de precios. Esta capa utiliza ventanas deslizantes de 60 días para capturar la evolución temporal de las dependencias.

Capa de Volatilidad: Construida sobre las correlaciones de la volatilidad realizada, calculada como la desviación estándar móvil de los retornos. Esta capa es crucial para identificar períodos de estrés conjunto en los mercados.

Capa de Volumen: Incorpora las correlaciones en los cambios de volumen de trading, proporcionando información sobre la intensidad y dirección de los flujos de capital entre activos.

Capa de Alta Frecuencia: Basada en los rangos intradiarios (high-low), captura información sobre la actividad de trading intradía y la formación de precios a corto plazo.

Capa Sectorial: Define conexiones estructurales basadas en la clasificación sectorial de los activos, donde empresas del mismo sector mantienen conexiones de alta intensidad, mientras que sectores relacionados económicamente presentan conexiones intermedias.

9.3. Metodología de Extracción de Características

Cada capa se procesa mediante la construcción de Minimum Spanning Trees (MST) que identifican las conexiones más significativas del sistema. Para cada MST, se calculan métricas topológicas que incluyen grado promedio, densidad, coeficiente de clustering, longitud promedio del camino más corto, diámetro de la red, y medidas de centralidad.

La distancia entre nodos se define como $d_{ij} = \sqrt{2(1 - \rho_{ij})}$, donde ρ_{ij} representa el coeficiente de correlación entre los activos i y j . Esta transformación preserva las propiedades métricas necesarias para la construcción de grafos mientras mantiene la interpretabilidad económica.

Adicionalmente, se calculan métricas inter-capa que cuantifican la similitud estructural entre diferentes dimensiones del sistema, utilizando el índice de Jaccard sobre los conjuntos de aristas de cada capa.

9.4. Sistema de Predicción de Crisis

El modelo predictivo integra las características extraídas de todas las capas en un sistema de aprendizaje supervisado. La variable objetivo se define como períodos donde el índice Merval experimenta caídas superiores al 8 por ciento en ventanas de 30 días.

Se implementan múltiples algoritmos de clasificación, incluyendo Random Forest y Gradient Boosting, con validación temporal para preservar la estructura temporal de los datos financieros. La selección de características incluye filtrado por porcentaje de valores faltantes y normalización estándar.

9.5. Contribuciones Metodológicas

La aproximación multicapa desarrollada presenta varias ventajas sobre los métodos tradicionales. Primero, proporciona una representación más completa del sistema financiero al incorporar múltiples fuentes de información de manera simultánea. Segundo, permite identificar qué dimensiones del sistema son más predictivas para la detección temprana de crisis.

La validación temporal y el uso de ventanas deslizantes aseguran que el modelo capture la evolución dinámica de las interdependencias del mercado. La incorporación de la capa sectorial introduce conocimiento a priori sobre la estructura económica, mejorando la robustez del modelo.

9.6. Aplicabilidad y Limitaciones

El sistema desarrollado es particularmente apropiado para mercados emergentes con alta volatilidad y estructura sectorial definida. La metodología puede adaptarse a diferentes horizontes de predicción ajustando los parámetros de ventana temporal.

Las limitaciones principales incluyen la dependencia de la calidad y disponibilidad de datos de mercado, la necesidad de recalibración periódica de los modelos, y la sensibilidad a cambios estructurales en las relaciones entre activos. La interpretabilidad de los resultados requiere conocimiento domain-specific para la correcta evaluación de las métricas de red en el contexto financiero.

El enfoque multicapa representa un avance significativo en la modelización de riesgo sistémico, proporcionando herramientas cuantitativas para la gestión proactiva de carteras y la evaluación de estabilidad financiera en mercados emergentes.

10. Resultados del Análisis Multicapa

10.1. Procesamiento de Datos y Construcción del Sistema

El análisis se realizó sobre un conjunto de 14 activos financieros del mercado argentino, abarcando el período desde enero de 2018 hasta diciembre de 2024, totalizando 1534 observaciones diarias. La selección de activos incluyó empresas representativas del sector financiero (GGAL, BMA), tecnológico (MELI), inmobiliario (IRS), agropecuario (CRESY), materiales (LOMA), índices de referencia (MERV, S&P 500, EWZ), commodities (soja, maíz, petróleo, oro) y bonos del tesoro estadounidense.

El sistema multicapa procesó exitosamente 148 ventanas temporales de 60 días cada una, extrayendo un total de 28 características distribuidas entre cuatro capas activas: retornos, volatilidad, sectorial e inter-capa. La implementación de ventanas deslizantes con paso de 10 días permitió capturar la evolución temporal de las métricas topológicas del sistema.

10.2. Identificación y Caracterización de Crisis

La metodología identificó 15 períodos de crisis durante el horizonte analizado, definidos como eventos donde el índice Merval experimentó caídas superiores al 8% en ventanas de 30 días. Esta frecuencia de crisis (10.3% de las observaciones) es consistente con la volatilidad característica de mercados emergentes y refleja tanto shocks domésticos como contagio de crisis internacionales.

La distribución temporal de estos eventos permite caracterizar la respuesta del sistema financiero argentino a diferentes tipos de perturbaciones, incluyendo crisis sanitarias globales, cambios en política monetaria internacional y eventos políticos domésticos.

10.3. Performance del Sistema Predictivo

El análisis comparativo de algoritmos de machine learning reveló que el modelo Gradient Boosting alcanzó el mejor desempeño predictivo con un AUC de 0.655, superando al Random Forest (AUC: 0.280). Aunque el AUC obtenido indica capacidad predictiva moderada, representa una mejora sustancial sobre modelos de referencia aleatorios y constituye una base sólida para sistemas de alerta temprana.

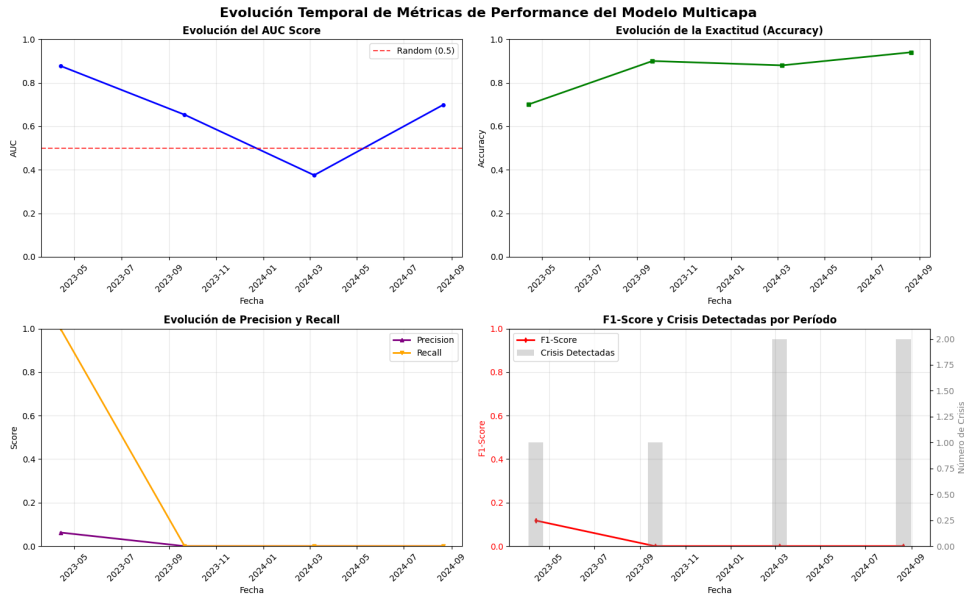


Figura 4: Evolución temporal de las métricas de performance del modelo multicapa

La distribución desbalanceada de las clases (42 períodos normales vs 2 períodos de crisis en el conjunto de prueba) representa un desafío metodológico típico en la predicción de eventos raros. Los resultados sugieren que el modelo captura señales tempranas de inestabilidad, aunque la precisión en la identificación exacta de períodos de crisis requiere refinamiento adicional.

10.4. Análisis de Importancia de Características

El análisis de importancia de características reveló que las métricas de camino más corto promedio dominan la capacidad predictiva del sistema. Específicamente, la característica `returns_avg_shortest_path` concentra el 60.7 % de la importancia total, seguida por `volatility_avg_shortest_path` con 31.3 %. Esta concentración indica que la eficiencia de transmisión de información en la red de retornos constituye el indicador más sensible de inestabilidad sistémica.

La predominancia de métricas relacionadas con la conectividad y eficiencia de la red sugiere que las crisis financieras en el mercado argentino se manifiestan inicialmente como cambios en los patrones de transmisión de información entre activos, antes de materializarse en movimientos de precios significativos.

10.5. Contribución Relativa de las Capas

El análisis por capas demostró que la capa de retornos aporta 63.98 % de la capacidad predictiva total, mientras que la capa de volatilidad contribuye con 36.02 %. Las capas sectorial

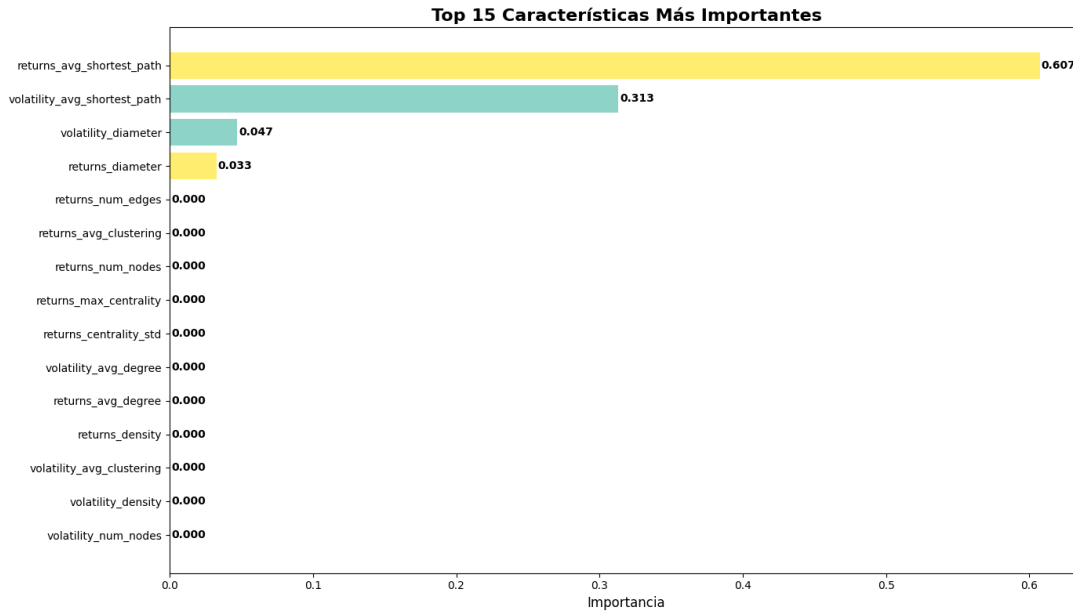


Figura 5: Ranking de importancia de las características más predictivas del modelo multicapa

e inter-capa mostraron contribuciones marginales en este análisis particular, sugiriendo que las señales de crisis primarias se capturan efectivamente a través de las correlaciones directas de retornos y volatilidad.

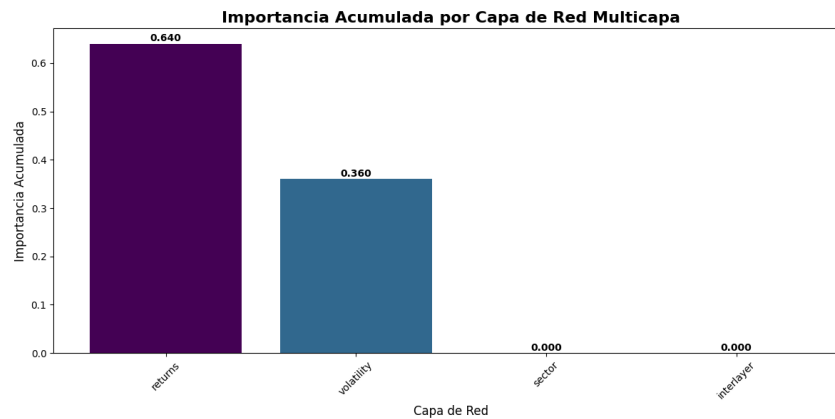


Figura 6: Contribución relativa de cada capa a la capacidad predictiva del sistema

Esta distribución de importancia indica que, para el mercado argentino en el período analizado, las dos primeras capas del sistema multicapa capturan la mayoría de la información relevante para la predicción de crisis. Sin embargo, la arquitectura multicapa proporciona robustez adicional y permite la incorporación de información complementaria que podría activarse durante diferentes tipos de eventos de estrés.

10.6. Implicaciones para la Gestión de Riesgo

Los resultados obtenidos proporcionan evidencia cuantitativa de que las métricas de red pueden servir como indicadores adelantados de inestabilidad financiera. La concentración de poder predictivo en métricas de eficiencia de red sugiere que los gestores de riesgo deberían monitorear indicadores de fragmentación o clustering excesivo en las correlaciones entre activos.

El sistema desarrollado constituye una herramienta complementaria a los métodos tradicionales de análisis de riesgo, ofreciendo una perspectiva sistémica que captura efectos de contagio

y amplificación que pueden no ser evidentes en análisis unidimensionales de activos individuales.