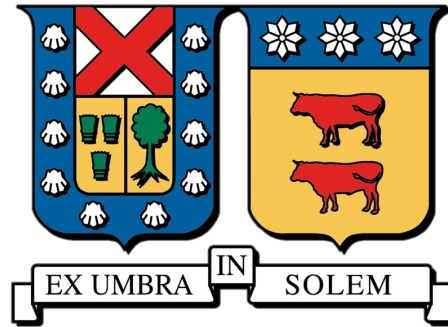




UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS

**PREDICCIÓN DE LA IMPORTANCIA SISTÉMICA FUTURA
RELATIVA DE INSTRUMENTOS ACCIONARIOS MEDIANTE REDES
FINANCIERAS DINÁMICAS Y GRAPH NEURAL NETWORKS:
EVIDENCIA GLOBAL Y APLICACIÓN EN PORTAFOLIOS**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

AUTOR

Joaquín Mauricio Espejo Núñez

PROFESOR GUÍA

Dr. Werner Kristjanpoller Rodríguez

PROFESOR CO-REFERENTE

Dr.(c), M.Sc. Felipe Mora Rojas

Valparaíso, agosto de 2026



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Predicción de la importancia sistémica futura relativa de instrumentos accionarios mediante redes financieras dinámicas y Graph Neural Networks: evidencia global y aplicación en portafolios

Nombre del candidato(a): Joaquín Mauricio Espejo Núñez

Carrera / Grado: Ingeniería Civil Industrial / Ingeniero Civil Industrial

Campus: Casa Central **Departamento:** Industrias

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Werner Kristjanpoller Rodríguez, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 22-09-2026

Firma:

Estudiante o Candidato(a):

Fecha: 21-09-2026

Firma:

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Tabla de Contenidos

1	Resumen ejecutivo	9
2	Introducción	10
3	Problema de Investigación	12
3.1	Preguntas de investigación	13
4	Objetivos	15
4.1	Objetivo general	15
4.2	Objetivos específicos	15
5	Marco teórico	17
5.1	Riesgo sistémico financiero	17
5.2	Redes financieras y modelación de interconexiones	18
5.3	Redes de correlación en mercados accionarios	19
5.4	Redes financieras dinámicas y ventanas temporales	20
5.5	Formalización mínima: retornos, aristas, <i>strength</i> y clases	21
5.6	Métricas de centralidad e importancia sistémica futura relativa	23
5.7	Machine Learning supervisado para clasificación financiera	24
5.8	Modelos clásicos de referencia	24
5.9	Graph Neural Networks	25
5.9.1	Message passing y matriz de adyacencia	25
5.9.2	Graph Convolutional Network (GCN)	26

5.9.3	GraphSAGE	27
5.9.4	Graph Attention Networks (GAT)	28
5.10	Métricas de evaluación	28
5.10.1	Métrica primaria: quadratic weighted kappa (QWK)	29
5.10.2	Métricas complementarias	29
5.11	Significancia estadística y bootstrap por ventanas	31
5.11.1	Remuestreo temporal y moving-block bootstrap	31
5.11.2	Multiplicidad y alcance de la inferencia	32
5.12	Trabajos relacionados y contraste con el problema de investigación	33
5.12.1	Elementos compartidos en la lógica de investigación	34
5.12.2	Elementos que no se trasladan	34
5.12.3	Brechas que aborda el presente estudio	34
5.13	Síntesis del marco teórico	35
6	Metodología	36
6.1	Universo y período	36
6.2	Ventanas, grafo y variable objetivo	37
6.3	Uso del grafo en GCN, GraphSAGE y GAT	39
6.4	Variables por bloque	39
6.5	Ablación geográfica (filtrado de instrumentos)	40
6.6	Modelos y entrenamiento	41
6.7	Partición temporal y <i>embargo</i>	42
6.8	Métricas e inferencia estadística	43

6.9	Horizonte calendario de evaluación	44
6.10	Diagnósticos de valor incremental y parsimonia	45
6.11	Sensibilidad del horizonte del target	45
6.12	Score ordinal y exposición sistémica relativa	46
6.13	Portafolio Variante A (mínima varianza penalizada, fuera de muestra)	47
6.14	Portafolio Variante B (mapa retrospectivo sobre la frontera media–varianza)	48
7	Alcance	49
8	Resultados	50
8.1	Bloque 1: seis variables individuales	50
8.1.1	Diseño	50
8.1.2	Resultados en el conjunto de prueba	50
8.1.3	Inferencia	52
8.2	Bloque 2: región-v2, sector (24 variables) y ablación geográfica	53
8.2.1	Diseño	53
8.2.2	Resultados en el conjunto de prueba	54
8.2.3	Inferencia	54
8.2.4	Ablación geográfica a–f	55
8.3	Bloque 3: conjunto ampliado (47 variables) y parsimonia	57
8.3.1	Diseño	57
8.3.2	Resultados en el conjunto de prueba	58
8.3.3	Inferencia	59
8.3.4	Parsimonia 24 vs 47	60

8.4	Score ordinal (24 variables)	61
8.5	Portafolio Variante A (mínima varianza penalizada, fuera de muestra, 24 variables)	62
8.5.1	Diseño	62
8.5.2	Resultados en el conjunto de prueba	62
8.5.3	Inferencia	63
8.6	Portafolio Variante B (mapa retrospectivo sobre la frontera media–varianza, 24 variables)	64
8.7	Síntesis del capítulo	65
9	Discusión	66
9.1	Respuesta a la pregunta principal	66
9.2	Paridad tabular y procesamiento de la estructura relacional	67
9.3	Parsimonia: 24 frente a 47 variables	68
9.4	Sensibilidad a la composición geográfica	69
9.5	Del clasificador al score ordinal	70
9.6	Utilidad económica: Variante A (fuera de muestra)	70
9.7	Variante B como mapa retrospectivo	71
9.8	Relación con la evidencia previa y aporte académico	72
10	Limitaciones	74
10.1	Qué mide el proxy	74
10.2	Qué no permite afirmar	74
10.3	Hasta dónde se puede generalizar	76

11 Conclusiones	78
11.1 Trabajo futuro	80
Referencias	81
12 Anexos	87
12.1 Puente descriptivo: especificación preliminar de 26 variables a 24 variables	87
12.2 Protocolo y artefactos	87
12.3 Elegibilidad del Universo Global 2015	88
12.4 Ilustración de la red Global 2015 por región	88
12.5 Diccionario de variables	91
12.5.1 Bloque 1 (seis individuales).	91
12.5.2 Bloque 2 (región-v2 y sector; 24 columnas).	91
12.5.3 Bloque 3 (contexto y red tabular; 47 columnas).	92
12.6 Hiperparámetros seleccionados por validación	92
12.6.1 Selecciones históricas (no organizan el eje del estudio).	93
12.7 Glosario de artefactos GNN	93
12.8 Corrida archivada sin red tabular (histórico; no comparable con 6/24/47)	94
12.9 Figuras complementarias de portafolio	94
12.10 Inferencia extendida	95
12.11 Notas sobre material histórico	95
12.12 Diagnóstico lineal y ablación históricos (26 y 49 variables; no comparable con 24/47)	95

Lista de Tablas

1	Convención de clases ordinales a partir del <i>strength</i> futuro.	23
---	---	----

2	Conjuntos de variables del diseño experimental.	39
3	Casos de ablación geográfica sobre el Bloque 2 (24 variables).	41
4	Criterio de selección, semillas y agregación según modelo.	42
5	Partición temporal del Universo Global 2015 (fechas de ventana).	43
6	Bloque 1 (seis variables). Panel en el conjunto de prueba.	51
7	Bloque 1. Inferencia ΔQWK (GNN <i>hard vote</i> vs MLP).	52
8	Bloque 2 (24 variables). Panel en el conjunto de prueba.	54
9	Bloque 2. Inferencia ΔQWK (GNN <i>hard vote</i> vs MLP).	54
10	Bloque 2. Ablación geográfica: nivel de QWK en el conjunto de prueba según la composición del universo.	56
11	Bloque 3 (47 variables). Panel en el conjunto de prueba.	58
12	Bloque 3. Inferencia ΔQWK (GNN <i>hard vote</i> vs XGBoost).	59
13	Variante A. Métricas del conjunto de prueba (costo 10 pb).	62
14	Variante A. Inferencia ΔESR (P4 vs referencias). La última columna indica si el contraste es significativo tras el ajuste de Holm al 5 %.	63
15	Criterios operativos de elegibilidad del Universo Global 2015.	88
16	Variables individuales (Bloque 1).	91
17	Variables de contexto (18) y de red tabular (5) añadidas sobre la base de 24 (total 47).	92
18	Hiperparámetros seleccionados por QWK de validación. Bloque 1 (6 variables). . .	92
19	Hiperparámetros seleccionados por QWK de validación. Bloque 2 (24 variables). .	92
20	Hiperparámetros seleccionados por QWK de validación. Bloque 3 (47 variables). .	93

21	Diagnóstico lineal histórico (logit elastic-net). QWK en validation y test (26/49 variables).	96
22	Ranking histórico de importancia lineal (elastic-net), 49 variables (filas 1–25). No comparable con 24/47.	98
23	Ranking histórico de importancia lineal (elastic-net), 49 variables (filas 26–37). No comparable con 24/47.	98
24	Ranking histórico de importancia lineal (elastic-net), 49 variables (filas 38–49). No comparable con 24/47.	99

Lista de Figuras

1	Representación conceptual de una red financiera accionaria basada en correlaciones de retornos.	19
2	Contraste conceptual entre el dominio de Balmaseda et al. (2023) y el problema estudiado aquí.	33
3	Esquema temporal de una observación instrumento–ventana: atributos y grafo en la ventana histórica y clase futura a horizonte H	37
4	Universo Global 2015 en una ventana de prueba (2025-04-07). Color: región (ocho grupos del congelamiento); tamaño: <i>strength</i> . Esparsificación solo visual; el protocolo experimental usa la red densa. No identifica causalidad regional ni sustituye los contrastes de Resultados.	38
5	Bloque 1. QWK por ventana del conjunto de prueba (panel temporal; complementario al QWK <i>pooled</i>).	52
6	Bloque 1. Δ QWK vs MLP con IC95 % (MBB).	53
7	Bloque 2. Δ QWK vs MLP con IC95 % (MBB).	55

8	Bloque 2. QWK del conjunto de prueba por composición geográfica (GraphSAGE y clásico del conjunto de prueba).	57
9	Bloque 3. QWK por ventana del conjunto de prueba (panel temporal; complementario al QWK <i>pooled</i>).	59
10	Bloque 3. Δ QWK vs XGBoost con IC95 % (MBB).	60
11	Score de 24 variables: quintiles vs <i>strength</i> en el conjunto de prueba.	61
12	Variante A. Exposición sistémica relativa (ESR) en el tiempo (conjunto de prueba).	63
13	Variante B. Curvas de nivel de <i>ESR</i> sobre la frontera (estimación con la muestra completa, retrospectiva).	65
14	Red Global 2015 por región (I). Ventana de prueba 2025-04-07; layout idéntico al de Metodología. Región destacada en color; resto en gris.	89
15	Red Global 2015 por región (II). Misma ventana y layout que la Figura 14.	90
16	Variante A. Riqueza acumulada en el conjunto de prueba (complemento).	94
17	Variante B. Detalle (zoom) de curvas de nivel de <i>ESR</i> (complemento).	95
18	Quince variables con mayor participación de importancia lineal en el diagnóstico elastic-net histórico (49 variables).	97

1 Resumen ejecutivo

La importancia sistémica futura relativa de un instrumento accionario cambia con el estado de la red y no coincide con el riesgo sistémico regulatorio, la probabilidad de crisis ni el retorno. El problema que aborda esta memoria es anticipar esa posición estructural relativa (cuartiles del *strength*, la conectividad ponderada agregada, en redes dinámicas de correlación absoluta) fuera de muestra en el tiempo.

Se comparan GCN, GraphSAGE y GAT, cada una, con modelos clásicos de Machine Learning bajo el mismo vector tabular, con adyacencia adicional para esas arquitecturas. Sobre esa clasificación se construye un score ordinal y se examina su incorporación a portafolios que penalizan exposición sistémica relativa.

El universo es el Universo Global 2015: 598 instrumentos accionarios listados en Estados Unidos.

GraphSAGE es la arquitectura más consistente y se mantiene competitiva frente a los clásicos de referencia, sin evidencia suficiente de superioridad general; GCN y GAT no reproducen ese patrón. La especificación de 24 variables queda por encima de la de 47 para GraphSAGE y se adopta como base del score y de los portafolios. La señal reduce de forma detectable la exposición sistémica relativa realizada, con magnitud pequeña.

En conjunto, GraphSAGE es competitivo, no muestra superioridad general y su aporte al portafolio es pequeño.

2 Introducción

Anticipar la conectividad relativa futura de instrumentos accionarios importa porque la posición estructural de un activo en la red de mercado puede orientar monitoreo de riesgo, priorización de vigilancia y decisiones de exposición, sin exigir por ello un mapa contractual completo de contagio. En mercados donde las interdependencias se observan sobre todo a través de comovimientos de retornos, disponer de una señal temprana de importancia relativa permite complementar criterios financieros tradicionales con información estructural.

Esta memoria estudia esa predicción fuera de muestra en el **Universo Global 2015**. El objeto es la **importancia sistémica futura relativa**: cuartiles del *strength* en una red dinámica de correlaciones absolutas. El propósito es evaluar el desempeño comparativo de GCN, GraphSAGE y GAT, cada una frente a modelos clásicos de Machine Learning, cuando clásicos y modelos de grafo reciben la misma información tabular y esas arquitecturas procesan además la estructura del grafo. El arco empírico avanza desde amplitudes de 6, 24 (con ablación geográfica) y 47 variables hacia un score ordinal y dos variantes de portafolio (A fuera de muestra; B como mapa retrospectivo). El estudio no predice crisis ni retornos, no asume superioridad a priori de GCN, GraphSAGE o GAT y se limita a un único universo.

Tras el resumen ejecutivo, el documento se organiza del modo siguiente. El **problema de investigación** formula el objeto, las preguntas y su vínculo con los objetivos. Los **objetivos** fijan el compromiso empírico del estudio. El **marco teórico** sitúa redes financieras, aprendizaje relacional y criterios de evaluación. La **metodología** describe datos, protocolo temporal, modelos e inferencia. El **alcance** delimita el tipo de conocimiento que se produce. Los **resultados** presentan la evidencia de clasificación, parsimonia, ablación, score y portafolios. La **discusión** interpreta los

hallazgos frente a la literatura; las **limitaciones** acotan lo que el diseño permite afirmar; y las **conclusiones** (con trabajo futuro) sintetizan la respuesta al problema. Completan el documento las **referencias** y los **anexos**.

3 Problema de Investigación

En mercados accionarios, la importancia sistémica relativa de un instrumento puede cambiar en el tiempo a medida que se reconfiguran las interdependencias de retornos. Una clasificación contemporánea aporta diagnóstico, pero entrega menor utilidad anticipatoria que predecir la clase futura bajo un protocolo temporal estricto. En la literatura reciente, contrastes entre Graph Neural Networks (GNN) y Machine Learning tabular para clasificar importancia sistémica se han desarrollado sobre todo en dominios bancarios y redes de exposiciones (Balmaseda et al., 2023). En la literatura revisada no se identificaron trabajos que combinen, en un mismo protocolo, instrumentos accionarios, redes dinámicas de correlación y predicción fuera de muestra de una clase relativa futura.

El problema se formula como clasificación supervisada a nivel de nodo. En cada fecha de ventana t se observan variables del instrumento y la estructura de la red financiera G_t . A partir de esa información se busca predecir la clase de **importancia sistémica futura relativa** en un horizonte posterior $t + h$, definida a partir del *strength* futuro (conectividad ponderada agregada) y su partición ordinal dentro del universo vigente. La relación empírica central vincula atributos y estructura en t con la clase ordinal en $t + h$. La etiqueta es relativa a la composición del universo: no representa un nivel absoluto de riesgo sistémico, ni probabilidad de crisis, ni un pronóstico de precios o retornos.

La dificultad central es que la red es dinámica, el horizonte introduce incertidumbre temporal y los modelos tabulares no procesan de forma nativa las relaciones entre nodos. Cuando incorporan la red, suelen hacerlo mediante ingeniería de centralidades (Kou et al., 2019). Por ello se evalúa si GCN, GraphSAGE y GAT presentan, cada una, un desempeño predictivo superior o

competitivo respecto de Machine Learning clásico cuando clásicos y modelos de grafo reciben el mismo vector tabular y esas arquitecturas procesan además la adyacencia. La evaluación es fuera de muestra en el tiempo sobre el **Universo Global 2015** (598 instrumentos accionarios listados en Estados Unidos).

La justificación es analítica y metodológica. Anticipar la conectividad relativa futura puede apoyar la priorización del monitoreo por gestores de riesgo, analistas cuantitativos e inversionistas institucionales. En el plano teórico, el planteamiento permite comparar el desempeño de modelos que incorporan la adyacencia frente a especificaciones solo tabulares bajo paridad de variables, y examinar hasta qué punto esa comparación admite una lectura de utilidad vía exposición sistémica relativa en portafolio.

3.1 Preguntas de investigación

1. **Pregunta principal.** ¿En qué medida GCN, GraphSAGE y GAT resultan, cada una, competitivas o superiores frente a modelos clásicos de Machine Learning en la clasificación temporal fuera de muestra de la importancia sistémica futura relativa en el Universo Global 2015?
2. **Pregunta secundaria.** Bajo paridad tabular (Bloques 1 y 2), ¿los modelos que procesan explícitamente la estructura del grafo presentan un desempeño superior o competitivo frente a modelos clásicos que reciben la misma información tabular?
3. **Pregunta terciaria.** ¿La incorporación de variables de contexto de mercado y métricas de red tabulares aporta valor incremental suficiente para justificar 47 variables frente a la base de 24 variables (región-v2 y sector)?

4. **Pregunta cuaternaria.** ¿El desempeño predictivo es sensible a la composición geográfica del universo, evaluada mediante una ablación que filtra instrumentos sobre el Bloque 2?
5. **Pregunta quinta.** ¿La señal de importancia relativa, vía un score ordinal y su uso en portafolio, permite modular la exposición sistémica relativa sin empeorar de forma material el perfil de riesgo financiero observado?

La pregunta principal compara cada arquitectura (GCN, GraphSAGE y GAT) con los clásicos de referencia (OE1). La secundaria compara modelos de grafo y clásicos bajo paridad tabular (OE1–OE2). La terciaria orienta la parsimonia 24 frente a 47 (OE4). La cuaternaria examina la sensibilidad geográfica (OE3). La quinta articula la extensión al score (OE5) y a las variantes de portafolio (OE6).

4 Objetivos

4.1 Objetivo general

Evaluar el desempeño predictivo de GCN, GraphSAGE y GAT, cada una frente a modelos clásicos de Machine Learning, en la clasificación fuera de muestra de la importancia sistémica futura relativa de instrumentos accionarios del Universo Global 2015. El criterio primario es el quadratic weighted kappa (QWK), bajo paridad tabular y un protocolo temporal reproducible.

4.2 Objetivos específicos

1. **OE1.** Comparar el desempeño de GCN, GraphSAGE y GAT, cada una, con modelos clásicos (regresión logística, Random Forest, XGBoost, k -NN y MLP) en el Bloque 1 (seis variables individuales), bajo paridad tabular.
2. **OE2.** Comparar el desempeño de GCN, GraphSAGE y GAT, cada una, con los mismos modelos clásicos en el Bloque 2 (24 variables: región-v2 y sector), bajo paridad tabular.
3. **OE3.** Examinar la sensibilidad del desempeño predictivo a la composición geográfica del universo mediante una ablación geográfica por filtrado de instrumentos sobre el Bloque 2.
4. **OE4.** Contrastar la especificación de 47 variables (Bloque 3) con la de 24 variables, en una lectura de parsimonia.
5. **OE5.** Definir un score ordinal $s = (3p_0 + 2p_1 + p_2)/3$ a partir de las probabilidades de clase, bajo la especificación de 24 variables, como puente entre la clasificación y las aplicaciones de portafolio.

6. **OE6.** Evaluar dos variantes de portafolio que incorporan la señal de 24 variables: Variante A (mínima varianza penalizada, fuera de muestra) y Variante B (mapa retrospectivo sobre la frontera media–varianza).

5 Marco teórico

Esta sección organiza el sustento conceptual de la memoria: primero el riesgo sistémico y su distinción respecto de la **importancia sistémica futura relativa**; luego la formalización de redes de correlación dinámicas y del proxy vía *strength*; después el aprendizaje supervisado, las Graph Neural Networks y los criterios de evaluación; a continuación, la relación con el antecedente metodológico de referencia; y, finalmente, una síntesis que vincula este marco con los objetivos empíricos. Los procedimientos, datos, particiones y decisiones experimentales se presentan en el capítulo de Metodología.

5.1 Riesgo sistémico financiero

El riesgo sistémico se entiende como la posibilidad de que el deterioro de una entidad, o de un conjunto reducido de nodos, genere efectos adversos sobre el conjunto del sistema financiero a través de interconexiones (Acemoglu et al., 2015; Bardoscia et al., 2021). A diferencia del riesgo idiosincrásico, que puede concentrarse en un activo particular, el componente sistémico depende de la estructura de relaciones y de la respuesta agregada del sistema ante perturbaciones.

La literatura ha enfatizado que no solo importan las instituciones demasiado grandes para quebrar (*too big to fail*), sino también aquellas suficientemente interconectadas como para transmitir fragilidad al resto de la red. En este sentido, una red puede operar como amortiguador cuando reparte el impacto entre múltiples contrapartes, pero, más allá de ciertos umbrales, los mismos enlaces pueden actuar como mecanismo de contagio. Esta dualidad se resume en la idea de sistemas *robust-yet-fragile* (Haldane, 2009; Acemoglu et al., 2015).

En paralelo, se han desarrollado métricas econométricas de riesgo sistémico, entre ellas ΔCoVaR y SRISK (Adrian y Brunnermeier, 2016; Brownlees y Engle, 2017), así como medidas

de contribución a la subcapitalización sistémica esperada (SES) (Acharya et al., 2017). Dichas medidas son útiles para rankings y diagnóstico desde precios o balance, pero responden a objetos distintos de la **importancia sistémica futura relativa** adoptada aquí. Trabajos posteriores combinan múltiples indicadores sistémicos para evaluar su vínculo con la macroeconomía (Giglio et al., 2016) y muestran, además, que especificaciones lineales de CoVaR pueden quedar limitadas frente a enfoques no lineales que incorporan una lectura de red (Keilbar y Wang, 2022).

En el contexto accionario no siempre existen exposiciones bilaterales contractuales observables. Aun así, la literatura de importancia sistémica sigue siendo pertinente: ciertas acciones concentran comovimientos y ocupan posiciones estructurales relevantes para el estado agregado de la red. No se estima un indicador regulatorio de contagio bancario ni un nivel absoluto de riesgo sistémico. El objeto de esta memoria es la **importancia sistémica futura relativa**, un proxy operativo basado en conectividad ponderada futura y en una partición por cuantiles dentro de cada ventana.

5.2 Redes financieras y modelación de interconexiones

Una red financiera es una representación en la cual los nodos corresponden a entidades (bancos, firmas, acciones u otros agentes) y las aristas capturan relaciones entre ellas: exposiciones, dependencia estadística u otros vínculos relevantes (Newman, 2010; Bardoscia et al., 2021). Modelar el mercado como red permite estudiar atributos individuales junto con estructura y centralidad.

Los enfoques de red aportan información que no siempre queda capturada por métricas puramente econométricas, especialmente respecto de la posición estructural de los nodos. Las redes pueden construirse desde exposiciones reconstruidas, correlaciones u otros vínculos; la elección

depende de la disponibilidad de datos y del objeto de estudio (Giudici et al., 2020; Bardoscia et al., 2021). En finanzas y seguros, la literatura de *connectedness* ha mostrado que medidas de conectividad derivadas de series de mercado permiten caracterizar interdependencias sistémicas sin exigir un mapa contractual completo (Billio et al., 2012).

En el presente trabajo, la red no se interpreta como un mapa literal de préstamos interbancarios, sino como una estructura de interdependencia empíricamente observable a partir de retornos accionarios. Con ello se mantiene el espíritu de análisis estructural, adaptado a un mercado de acciones y al objeto de **importancia sistémica futura relativa**.

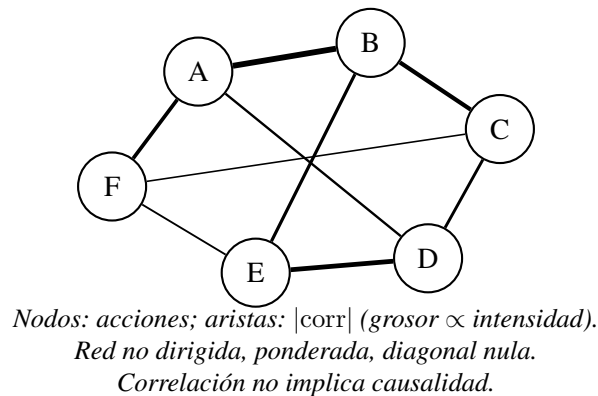


Figura 1: Representación conceptual de una red financiera accionaria basada en correlaciones de retornos.

Fuente: elaboración propia.

5.3 Redes de correlación en mercados accionarios

Cuando las exposiciones bilaterales no son públicas, una alternativa frecuente consiste en construir redes a partir de series de retornos. En este esquema, cada nodo es una acción y el peso de la arista resume la intensidad del comovimiento en una ventana temporal. La idea de transformar correlaciones en una estructura jerárquica o topológica del mercado tiene una tradición sólida en la econofísica y en el análisis de mercados financieros (Mantegna, 1999). Trabajos posteriores han

explorado además filtrados de la matriz de correlación, por ejemplo mediante árboles de expansión mínima (*minimum spanning tree*, MST) u otras redes filtradas, con el fin de resaltar vínculos estructuralmente relevantes y reducir ruido (Tumminello et al., 2005).

Esas construcciones filtradas se invocan aquí como **contexto** de la tradición de correlaciones y árboles de expansión mínima; el diseño empírico principal privilegia una red densa de correlaciones. Más recientemente, se han contrastado redes de exposición directa con redes basadas en correlación, mostrando que ambas fuentes de conectividad aportan información útil para el análisis sistémico (Giudici et al., 2020), en línea con medidas econométricas de conectividad en sectores financieros (Billio et al., 2012).

Desde el punto de vista interpretativo, es crucial distinguir **correlación** de **causalidad**. Una arista basada en correlación indica comovimiento estadístico entre retornos, no una dirección de contagio contractual ni un canal causal identificado. En consecuencia, la red de correlación no debe leerse como prueba de que un shock “se transmite” mecánicamente de un nodo a otro, sino como evidencia de dependencia empírica útil para caracterizar importancia estructural relativa.

En el diseño conceptual adoptado, el peso de arista se define a partir de la correlación absoluta de retornos, de modo que la intensidad de asociación no dependa del signo del comovimiento. La red resultante se interpreta como **no dirigida**, **ponderada** y con **diagonal nula** ($w_{ii}(t) = 0$). La formalización correspondiente se presenta a continuación, junto con el carácter dinámico de las ventanas.

5.4 Redes financieras dinámicas y ventanas temporales

A diferencia de un grafo estático único, el mercado se conceptualiza aquí como una **secuencia de redes** $\{G_t\}_{t=1}^T$ construidas sobre ventanas móviles de retornos. En cada ventana se

estiman correlaciones, se actualizan pesos de aristas y, con ello, se recalculan métricas de centralidad. En el diseño empírico principal (red densa) la conectividad es esencialmente completa y lo que varía de forma sistemática son los **pesos** de correlación; cambios de topología (existencia o no de aristas) corresponden a variantes esparsificadas, no al protocolo principal. Aun así, la estructura ponderada que resume el pasado no necesariamente coincide con la que caracterizará el futuro, lo cual justifica formular la tarea como predicción de una clase de **importancia sistémica futura relativa**, y no solo como clasificación contemporánea de una red fija.

Los detalles operativos (longitud de ventana, solapamiento, horizonte de predicción y criterios de inclusión de activos) pertenecen al capítulo de metodología. Aquí basta fijar el principio: la unidad temporal relevante es la ventana, y el objeto estructural es la red G_t asociada a dicha ventana.

5.5 Formalización mínima: retornos, aristas, *strength* y clases

Sea $P_{i,\tau}$ el precio de la acción i en el tiempo τ . El retorno logarítmico se define como

$$r_{i,\tau} = \ln P_{i,\tau} - \ln P_{i,\tau-1}. \quad (1)$$

Dada una ventana W_t de retornos, el peso de la arista entre los nodos i y j se define, en el marco conceptual adoptado, mediante la correlación absoluta

$$w_{ij}(t) = \begin{cases} |\text{Corr}(r_i, r_j \mid W_t)| & \text{si } i \neq j, \\ 0 & \text{si } i = j. \end{cases} \quad (2)$$

Así, $w_{ij}(t) = w_{ji}(t) \in [0, 1]$ resume intensidad de comovimiento sin dirección de transmisión, y

la diagonal nula evita autoenlaces artificiales. La red en t queda caracterizada por el conjunto de nodos y por estos pesos. El diseño empírico principal es la **red densa** de correlaciones absolutas; filtrados por umbrales, árboles de expansión mínima u otras redes esparcidas pertenecen al contexto de la literatura de correlaciones y MST y **no** organizan el protocolo reportado en Resultados.

En una red ponderada no dirigida, el *strength* del nodo i en la ventana t es la suma de pesos incidentes (Newman, 2010; Barrat et al., 2004):

$$s_i(t) = \sum_{j \neq i} w_{ij}(t). \quad (3)$$

Un *strength* alto indica, bajo esta definición, fuerte conectividad agregada con el resto del sistema en esa ventana.

La etiqueta de interés no es el *strength* contemporáneo como variable continua, sino una **clase futura** obtenida por cuantiles. Ordenando $\{s_i(t+h)\}$ en la ventana futura (horizonte h) y particionando en cuartiles, se asigna

$$y_i^{(h)}(t) \in \{0, 1, 2, 3\}, \quad (4)$$

donde $y_i^{(h)}(t)$ se calcula a partir de $s_i(t+h)$, pero debe predecirse con información disponible en t . La convención de lectura se resume en la Tabla 1. El problema predictivo consiste en aproximar $y_i^{(h)}(t)$ a partir de variables de nodo y, eventualmente, de la estructura de G_t .

Elaboración propia. Las clases se recalculan en cada ventana; no son umbrales absolutos de riesgo.

Tabla 1: Convención de clases ordinales a partir del *strength* futuro.

Clase	Lectura
0	Mayor importancia sistémica futura relativa
1	Importancia relativa alta
2	Importancia relativa baja
3	Menor importancia sistémica futura relativa

5.6 Métricas de centralidad e importancia sistémica futura relativa

Las métricas de centralidad resumen la posición de un nodo dentro de la red. El *degree* cuenta el número de vecinos (o de enlaces incidentes) y es especialmente informativo en redes no ponderadas. El *strength*, ya formalizado, generaliza esa idea a redes ponderadas sumando intensidades de conexión (Newman, 2010; Barrat et al., 2004). La centralidad de eigenvector incorpora no solo cuántos vecinos tiene un nodo, sino cuán centrales son dichos vecinos. El PageRank puede entenderse como una variante de difusión que redistribuye importancia a lo largo de caminos, con un componente de “teletransporte”. El coeficiente de clustering, por su parte, describe la tendencia local a formar tríadas o vecindarios densos alrededor de un nodo.

Estas medidas son complementarias: capturan dimensiones distintas de la estructura. El proxy principal de **importancia sistémica futura relativa** para el target es el *strength* futuro de correlaciones absolutas. Dicho proxy resume intensidad agregada de dependencia estadística: el valor absoluto no distingue cobertura de comovimiento positivo ni identifica dirección causal. En una red densa puede estar influido por factores comunes de mercado, y los cuartiles fuerzan una partición relativa incluso cuando las diferencias absolutas entre nodos sean pequeñas. Por ello, no equivale a DebtRank (Battiston et al., 2012; Bardoscia et al., 2015), CoVaR, SRISK, pérdida sistémica ni probabilidad de crisis; su validación frente a otros proxies queda como extensión futura. Ello no implica descartar otras centralidades como variables explicativas descriptivas o de

control; su uso concreto se define en metodología.

5.7 Machine Learning supervisado para clasificación financiera

La predicción de la clase de **importancia sistémica futura relativa** se formula como aprendizaje supervisado: se busca una función que, a partir de información en t , asigne una etiqueta $y_i^{(h)}(t) \in \{0, 1, 2, 3\}$. El objeto de predicción **no** es el retorno ni el precio; es esa clase ordinal dentro del universo.

Los modelos pueden explotar solo variables tabulares del nodo o, además, la estructura relacional de G_t . Un punto recurrente en la literatura de riesgo sistémico y redes es que el Machine Learning tradicional trata con frecuencia cada observación como independiente y, cuando incorpora la red, lo hace con frecuencia mediante resúmenes tabulares de centralidad, lo cual limita el uso granular de la adyacencia (Balmaseda et al., 2023; Kou et al., 2019). Las GNN se motivan precisamente como alternativa para procesar esa estructura de forma nativa. Esta distinción fundamenta la pregunta secundaria de la memoria: si una GNN puede mantener desempeño competitivo **sin** recibir métricas de red precalculadas como variables de entrada, aprovechando en su lugar la adyacencia.

La validez empírica del ejercicio depende de un protocolo temporal estricto (entrenamiento, validación y prueba sin fuga de información; evaluación fuera de muestra), desarrollado en metodología.

5.8 Modelos clásicos de referencia

Como referencias tabulares se consideran familias ampliamente usadas en clasificación. La regresión logística ofrece una referencia lineal interpretable. Los ensambles basados en árboles,

en particular Random Forest y XGBoost, capturan no linealidades e interacciones entre variables y suelen constituir referencias competitivas en problemas tabulares (Breiman, 2001; Chen y Guestrin, 2016). El clasificador k -NN depende de proximidad local en el espacio de atributos y es sensible a escala y dimensionalidad. Un multilayer perceptron (MLP) aporta una referencia de aprendizaje profundo *sin* estructura de grafo: procesa el vector de variables del nodo, pero no la matriz de adyacencia.

En el marco teórico, estos modelos se entienden como controles necesarios para evaluar si incorporar explícitamente la red agrega información predictiva. Aquí solo se justifica su rol conceptual; la selección concreta de referentes queda fuera de este capítulo.

5.9 Graph Neural Networks

Las Graph Neural Networks son modelos de aprendizaje profundo diseñados para datos en forma de grafo y pueden formularse mediante esquemas de *message passing* (Gilmer et al., 2017). Permiten tareas a nivel de nodo, de arista o de grafo completo. El foco aquí es la **clasificación a nivel de nodo**: predecir la clase de cada instrumento dada su representación y su vecindario en G_t .

5.9.1. Message passing y matriz de adyacencia

Un marco general es el de *message passing* (Gilmer et al., 2017). Sea x_v el vector de variables del nodo v , $N(v)$ su vecindario según la adyacencia, y $h_v^{(k)}$ su embedding en la capa k . De manera esquemática,

$$h_v^{(0)} = x_v, \quad h_v^{(k)} = U_k \left(h_v^{(k-1)}, \sum_{u \in N(v)} M_k(h_v^{(k-1)}, h_u^{(k-1)}, x_{vu}^e) \right), \quad (5)$$

donde M_k construye mensajes desde vecinos y U_k actualiza el estado del nodo. La matriz de adyacencia (o su versión ponderada) determina *quién* es vecino de *quién* y, por tanto, qué información relacional puede agregarse. Tras L capas, el embedding $h_v^{(L)}$ se proyecta a logits de clase.

La diferencia esencial respecto de un MLP es que este último transforma x_v de forma aislada, mientras que una GNN condiciona la representación de v a su contexto en la red. Si la adyacencia contiene señal útil para $y_v^{(h)}$, la GNN tiene un canal mecánico para explotarla; si la adyacencia es poco informativa respecto del target, dicho canal puede no mejorar, e incluso puede añadir ruido, frente a un buen modelo tabular.

5.9.2. Graph Convolutional Network (GCN)

La Graph Convolutional Network trata la agregación sobre el grafo como una convolución espectral aproximada: cada nodo actualiza su representación como una combinación lineal de sí mismo y de sus vecinos, con coeficientes fijos que dependen de los grados y con self-loops que incorporan el estado propio (Kipf y Welling, 2017). En notación compacta, la capa k toma la forma

$$h_v^{(k)} = \sigma \left(\sum_{u \in N(v) \cup \{v\}} c_{vu} W^{(k)} h_u^{(k-1)} \right), \quad (6)$$

donde c_{vu} es un peso de normalización determinado por la conectividad (no aprendido por par) y σ una no linealidad. El supuesto sobre el vecindario es de **mezcla simétrica y homogénea**: todos los vecinos entran con una importancia estructural dada por el grado, no por una relevancia específica para la tarea.

Esa lógica es pertinente en una red densa de correlaciones absolutas porque, cuando casi todos los pares tienen arista, la convolución por grado opera de hecho sobre un vecindario muy

amplio. El canal relacional existe, pero puede diluir señales locales al promediar comovimientos de distinta intensidad. GCN ofrece, así, un procesamiento nativo de la adyacencia con una regla de agregación parsimoniosa; no se sigue de ello que esa parsimonia sea óptima para anticipar clases de *strength* futuro.

5.9.3. GraphSAGE

GraphSAGE formula la actualización del nodo como una agregación *permutation-invariant* sobre un vecindario, combinada después con el estado propio (Hamilton et al., 2017). De manera esquemática,

$$h_v^{(k)} = \sigma \left(W^{(k)} \cdot \text{COMBINE} \left(h_v^{(k-1)}, \text{AGG}(\{h_u^{(k-1)} : u \in \tilde{N}(v)\}) \right) \right), \quad (7)$$

donde AGG puede ser media, pooling u otro operador simétrico, $\tilde{N}(v)$ es un vecindario (en la formulación original, eventualmente muestreado) y COMBINE concatena o fusiona la información local con la agregada. El supuesto deja de ser una convolución única por grado: lo que se aprende es **cómo resumir un conjunto de vecinos** de modo que la regla se aplique a nodos no necesariamente vistos en entrenamiento (carácter inductivo).

En un mercado accionario representado por correlaciones densas, el vecindario de cada instrumento es, en principio, casi el universo. GraphSAGE es conceptualmente atractivo porque separa la agregación de un conjunto de mensajes de una normalización rígida por grado y porque su énfasis inductivo encaja con grafos que se reconstruyen ventana a ventana. Ello no implica que el muestreo de vecinos sea el mecanismo empírico del protocolo, ni que esa flexibilidad se traduzca automáticamente en mejor predicción de importancia relativa futura.

5.9.4. Graph Attention Networks (GAT)

Graph Attention Networks sustituyen los coeficientes fijos de agregación por pesos de atención aprendidos entre nodos conectados (Veličković et al., 2018). Para cada par (v, u) en el vecindario se obtiene un coeficiente α_{vu} a partir de los embeddings, y la actualización es una suma ponderada

$$h_v^{(k)} = \sigma \left(\sum_{u \in N(v) \cup \{v\}} \alpha_{vu} W^{(k)} h_u^{(k-1)} \right), \quad (8)$$

con $\sum_u \alpha_{vu} = 1$ en el vecindario considerado. El supuesto es que **no todos los vecinos son igual de informativos**: la contribución de cada uno depende del dato, no solo de la topología o del grado.

Esa hipótesis es especialmente natural en redes densas de correlación, donde muchos pares están conectados con intensidad no nula y una agregación uniforme puede mezclar vínculos de mercado generales con dependencias más específicas. Un mecanismo de atención es, en teoría, una vía para reponderar ese vecindario amplio, en lugar de tratar las conexiones con un esquema uniforme o solo normalizado por grado. Ello no garantiza superioridad empírica de GAT ni equivale a afirmar que aproveche, por construcción, los pesos continuos de correlación, ni que GCN, GraphSAGE y GAT procesen de manera idéntica la adyacencia: esas diferencias de implementación se detallan en metodología. Un panorama de la familia y de sus variantes se resume en la revisión de literatura de Wu et al. (2021). La utilidad relativa de las tres debe evaluarse frente a benchmarks tabulares bajo el mismo protocolo, sin asumir de antemano un ganador.

5.10 Métricas de evaluación

La etiqueta a predecir es ordinal (0–3; Tabla 1): los cuartiles de *strength* futuro tienen un orden natural. Un acierto en la clase contigua no es equivalente a un error entre extremos, y una tasa

de aciertos agregada puede ocultar ambos fenómenos. Por ello el protocolo necesita una **métrica primaria** alineada con esa estructura ordenada y un panel complementario que informe asociación categórica, magnitud del error y balance por clase, sin disputar la función de decisión.

5.10.1. Métrica primaria: quadratic weighted kappa (QWK)

El QWK resume el acuerdo entre etiquetas observadas y predichas más allá de lo esperable por azar, asignando a cada desacuerdo un peso que crece con la distancia entre clases (Cohen, 1968). Si O es la matriz de coincidencias observadas y E la esperada bajo independencia, el índice toma la forma

$$\kappa_w = 1 - \frac{\sum_{i,j} w_{ij} O_{ij}}{\sum_{i,j} w_{ij} E_{ij}}, \quad w_{ij} = (i - j)^2, \quad (9)$$

con $i, j \in \{0, 1, 2, 3\}$. Los pesos cuadráticos hacen que confundir el cuartil más bajo con el más alto penalice más que un desliz entre cuartiles adyacentes. Esa lógica encaja con el objeto empírico: la importancia sistémica futura relativa se define como posición ordenada en la red, no como cuatro categorías nominales intercambiables.

QWK no mide utilidad de portafolio, ni calibración de probabilidades, ni el aporte causal de la adyacencia. Es un criterio de **acuerdo ordinal** para seleccionar especificaciones y para contrastar familias de modelos bajo el mismo target. Su elección como primaria se justifica por la ordinalidad de los cuartiles, no por un supuesto de superioridad de las GNN.

5.10.2. Métricas complementarias

El Matthews Correlation Coefficient (MCC) resume la matriz de confusión en una sola medida de **asociación** entre clases observadas y predichas, más equilibrada que la sola *accuracy*

cuando las clases no se interpretan de forma aislada (Matthews, 1975; Chicco y Jurman, 2020). Un valor próximo a 1 indica asociación fuerte en el conjunto de la matriz; un valor próximo a 0 es compatible con un acuerdo no mejor que el azar. En el antecedente de Balmaseda et al. (2023) el MCC es la métrica principal; aquí se conserva como **complemento**, no como criterio de selección ni de inferencia. Como el problema tiene cuatro clases, la lectura multiclase se apoya en la extensión a K categorías de Gorodkin (2004): las referencias binarias no bastan, por sí solas, para justificar el MCC de este protocolo.

El Class RMSE trata las etiquetas 0–3 como una escala numérica y reporta la raíz del error cuadrático medio. Mide **cuánto** se desplaza la predicción a lo largo del orden, no solo si el cuartil coincidió: un error entre los extremos (clases 0 y 3) pesa más que un desliz entre clases adyacentes. Un valor más bajo indica predicciones más cercanas, en esa escala, a la etiqueta observada. No selecciona especificaciones ni sustituye a QWK; se reporta para ver si el acuerdo ordinal convive con desplazamientos grandes.

El F1 macro promedia el F1 de cada clase con igual peso. Resume el compromiso entre precisión y exhaustividad **por cuartil**, de modo que un deterioro concentrado en una clase no quede oculto por el desempeño en las demás. Un F1 macro alto indica que las cuatro clases se recuperan de manera equilibrada; una caída alerta de que el acuerdo puede estar sostenido por solo algunos cuartiles. Se usa como complemento de balance, no de decisión.

La *accuracy* es la proporción de coincidencias exactas. Trata todos los errores por igual: confundir la clase 0 con la 1 cuenta lo mismo que confundir la 0 con la 3. Por ello no organiza la selección en un problema ordinal. *Accuracy*, F1 por clase y matrices de confusión se usan como diagnósticos de apoyo.

Ninguna de estas medidas sustituye a las demás: juntas cubren acuerdo ordinal, asociación

categoría, magnitud del error y balance por clase. Solo QWK organiza la selección de hiperparámetros y el contraste principal entre modelos. Cuando se cite evidencia histórica construida con MCC como primaria, se declara explícitamente. El contraste por diferencias y el remuestreo temporal coherente con ventanas solapadas se operacionalizan en Metodología y en la subsección de significancia estadística.

5.11 Significancia estadística y bootstrap por ventanas

Un valor puntual de QWK no basta para contrastar familias de modelos. Las ventanas de evaluación se solapan en el calendario, de modo que observaciones sucesivas no son independientes, y en cada fecha t los instrumentos comparten el mismo grafo: la dependencia es a la vez temporal y cruzada entre nodos. La inferencia debe respetar esa estructura; un intervalo o un contraste que la ignore puede atribuir a una arquitectura una ventaja que en realidad refleja rachas comunes del mercado o la correlación inducida por la red.

5.11.1. Remuestreo temporal y moving-block bootstrap

El bootstrap clásico remuestrea observaciones como si fueran independientes e idénticamente distribuidas, a fin de aproximar la distribución muestral de un estimador y construir intervalos empíricos (Efron y Tibshirani, 1993). Esa justificación se sostiene cuando las unidades remuestreadas son i.i.d. En series con dependencia temporal, o cuando varias unidades se observan bajo el mismo estado de mercado, un remuestreo i.i.d. trata como independientes observaciones que no lo son y, con ello, tiende a subestimar la variabilidad real del estimador.

El moving-block bootstrap (MBB) responde a ese problema remuestreando **bloques continuos** de fechas de ventana en lugar de fechas aisladas: se conservan rachas locales de dependencia

que un sorteo i.i.d. rompería. La versión circular de bloques de longitud fija (Künsch, 1989; Politis y Romano, 1992) es el marco conceptual del protocolo. Esa lógica es pertinente aquí porque, en cada t , los instrumentos no constituyen observaciones independientes: comparten la misma estructura de red y el mismo contexto de mercado. Remuestrear ventanas (o bloques de ventanas) preserva esa dependencia cruzada; remuestrear nodos como si fueran i.i.d. la disolvería.

5.11.2. Multiplicidad y alcance de la inferencia

En cada escenario se contrastan tres arquitecturas GNN frente al mismo referente tabular. Esos tres contrastes forman una **familia** de pruebas, no tres experimentos aislados: declarar superioridad en cualquiera de ellos eleva la probabilidad de un falso positivo. El procedimiento de Holm (Holm, 1979) controla ese exceso ajustando los valores p de la familia. Holm no es el valor p de un solo contraste: es una corrección sobre un conjunto.

En este trabajo se aplica dentro de cada escenario, no como corrección global del estudio: no se ajusta entre bloques de 6, 24 y 47 variables ni entre ablaciones o escenarios geográficos.

La inferencia así obtenida es evidencia bajo el remuestreo adoptado, no una demostración de superioridad en todo mercado o período. El intervalo de confianza del protocolo se construye al 95 % y resume un rango de valores de Δ coherente con ese remuestreo, no una prueba de que dos modelos sean equivalentes. Cuando el intervalo contiene el cero, no hay evidencia suficiente de superioridad: el contraste **empata** en ese sentido conceptual, y no constituye una prueba de equivalencia ni afirma que las arquitecturas sean idénticas. Un valor p de Holm alto (por ejemplo, próximo a 1) tampoco afirma igualdad: indica que, tras el ajuste por familia, no hay apoyo a la hipótesis de superioridad de esa GNN. El valor p no se lee como la cola del bootstrap de Δ ; se obtiene por un test de aleatorización coherente con bloques temporales, cuyo detalle se reserva a

Metodología.

El contraste Δ , el intervalo, Holm y la prueba pareada se operacionalizan en Metodología. La significancia estadística es, en este sentido, condicional al diseño experimental y al esquema de remuestreo.

5.12 Trabajos relacionados y contraste con el problema de investigación

En la línea de contrastes entre GNN y modelos tabulares para clasificar importancia sistémica, Balmaseda et al. (2023) estudian entidades bancarias en redes de exposiciones, con etiquetas basadas en DebtRank, discretización en cuartiles y evaluación predominantemente contemporánea o transductiva. Compiten modelos tabulares (regresión logística, Random Forest, k -NN y redes *fully connected*) con GCN, GraphSAGE y GAT, y privilegian MCC como métrica principal. A diferencia de ese dominio, el presente trabajo se sitúa en acciones, redes dinámicas de correlación, predicción temporal fuera de muestra y un proxy de *strength* futuro. DebtRank, C2R y la red interbancaria pertenecen al diseño de aquel estudio y no constituyen el método operativo del presente trabajo.

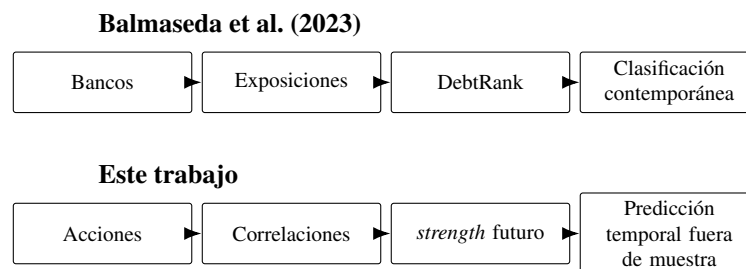


Figura 2: Contraste conceptual entre el dominio de Balmaseda et al. (2023) y el problema estudiado aquí.

Fuente: elaboración propia a partir de Balmaseda et al. (2023).

5.12.1. Elementos compartidos en la lógica de investigación

Se comparte la lógica de (i) representar el sistema financiero como red; (ii) discretizar importancia sistémica en clases ordinales (aquí, **importancia sistémica futura relativa**); (iii) contrastar GNN contra Machine Learning tabular; (iv) reportar un panel de métricas con remuestreo. Por la ordinalidad del target, aquí se privilegia QWK como métrica primaria y se conserva MCC como complemento. También se retoma el argumento de que el Machine Learning tabular no explota de forma nativa la adyacencia, mientras que las GNN pueden usarla junto con variables de nodo (Balmaseda et al., 2023).

5.12.2. Elementos que no se trasladan

No se replica el eje de contagio por DebtRank ni la red interbancaria de exposiciones. No se adopta C2R como componente central. La selección de hiperparámetros se basa en validación temporal, no en desempeño de prueba. Los márgenes reportados en SymNet/EurNet no se asumen trasladables al Universo Global 2015.

5.12.3. Brechas que aborda el presente estudio

Respecto de ese contraste y de un diseño solo tabular, el aporte se organiza en evaluación temporal estricta, paridad tabular en amplitudes de 6, 24 y 47 variables, atributos regionales y sectoriales, una ablación geográfica de sensibilidad, una prueba de parsimonia 24 frente a 47, y un puente a score ordinal y portafolios. La operacionalización corresponde a Metodología y Resultados.

5.13 Síntesis del marco teórico

Cuatro ideas sostienen el planteamiento. Primero, el riesgo sistémico depende de interconexiones, pero el objeto empírico aquí es la **importancia sistémica futura relativa**, no una medida regulatoria absoluta. Segundo, en mercados accionarios esa importancia puede aproximarse mediante redes de correlación dinámicas no dirigidas y ponderadas. Una arista de correlación resume comovimiento estadístico, no contagio causal. Tercero, el *strength* futuro, discretizado en clases ordinales, define la etiqueta de importancia sistémica futura relativa a predecir con información disponible en t . Cuarto, las GNN se motivan porque permiten incorporar la adyacencia de forma nativa, frente a modelos que tratan observaciones como independientes o que resumen la red en centralidades calculadas previamente. La utilidad relativa de GCN, GraphSAGE y GAT, y la extensión a score y portafolio, se evalúan en el protocolo empírico.

6 Metodología

Este capítulo organiza el protocolo en tres etapas. En clasificación se comparan GCN, GraphSAGE y GAT, cada una, con modelos clásicos, bajo tres conjuntos de variables de amplitud creciente (B1, B2 y B3). Sobre esa clasificación se define un score ordinal y dos variantes de portafolio.

El diseño se diferencia del antecedente de Balmaseda et al. (2023), situado en redes interbancarias y evaluación transductiva: aquí el dominio es accionario, la red es dinámica de correlación y la predicción es temporal fuera de muestra.

6.1 Universo y período

El universo único es el Universo Global 2015, compuesto por 598 instrumentos accionarios listados en Estados Unidos.

Las series de precios y volúmenes se obtienen desde Yahoo Finance mediante el paquete `yfinance` (Aroussi, 2026; Yahoo, . f.), con precios ajustados (*Adj Close*) en el panel diario, con lo cual dividendos y splits quedan incorporados en la serie de precio. El corte de elegibilidad es 2015-01-01. Tras filtros de historia mínima, cobertura de precios (*ffill* causal acotado) y liquidez se pasa de un inventario inicial de historia a 598 activos finales (Anexo 12.3); el inventario es fijo tras esos filtros, lo que introduce posible sesgo de supervivencia. Las fechas de ventana del protocolo adoptado van desde 2015-06-30 (primera de entrenamiento) hasta 2026-01-07 (última de prueba); la construcción del target a $H = 126$ exige retornos posteriores al cierre de cada ventana. Valores faltantes se tratan en la etapa de elegibilidad y de panel, antes de estimar correlaciones. La correlación de Pearson en cada ventana usa las observaciones disponibles del par en esa ventana, tras esa limpieza. El detalle operativo consta en la Tabla 15.

6.2 Ventanas, grafo y variable objetivo

Se emplean ventanas históricas de $W = 126$ días hábiles, paso $\Delta = 21$ y horizonte del target $H = 126$ días (diseño principal). Para una fecha de cierre de ventana t , los atributos y el grafo usan solo la ventana histórica $[t - W + 1, \dots, t]$; la etiqueta usa la ventana futura $[t + 1, \dots, t + H]$, estrictamente posterior (Figura 3). En cada t se estiman correlaciones de Pearson sobre log-retornos y se define $A_{ij}^{(t)} = |\rho_{ij}^{(t)}|$ para $i \neq j$, con diagonal nula. El grafo es no dirigido, ponderado y denso. Ese protocolo denso constituye el diseño empírico principal; filtrados del tipo MST u otros esparsimientos no organizan los experimentos reportados en Resultados.

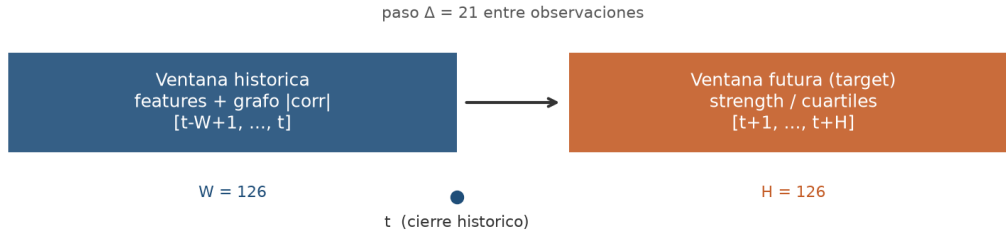


Figura 3: Esquema temporal de una observación instrumento–ventana: atributos y grafo en la ventana histórica y clase futura a horizonte H .

La Figura 4 ilustra esa red en una ventana de prueba del Universo Global 2015. El dibujo aplica una esparsificación **solo visual** (árbol de expansión mínima más el 1,5 % de aristas de mayor $|\text{corr}|$); el protocolo de entrenamiento y de evaluación permanece denso. El color identifica la región en la taxonomía de ocho grupos del congelamiento del universo, distinta de la región-v2 del Bloque 2; el tamaño del nodo es el *strength*. Los recortes que destacan cada región sobre el mismo layout constan en el Anexo 12.4.

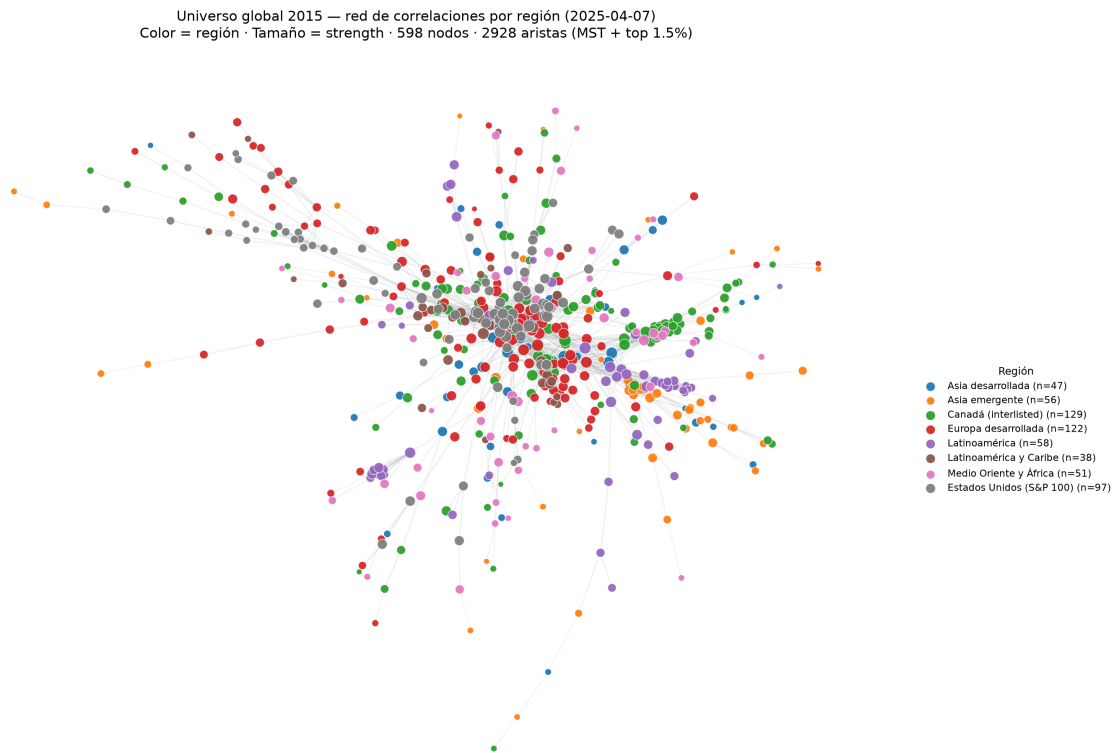


Figura 4: Universo Global 2015 en una ventana de prueba (2025-04-07). Color: región (ocho grupos del congelamiento); tamaño: *strength*. Esparsificación solo visual; el protocolo experimental usa la red densa. No identifica causalidad regional ni sustituye los contrastes de Resultados.

Fuente: elaboración propia a partir de la red de correlaciones del Universo Global 2015.

El *strength* del nodo i en una ventana es $s_i = \sum_{j \neq i} A_{ij}$. La clase futura se obtiene *dentro de cada fecha de ventana* (y dentro del universo o subconjunto vigente): se ordena el *strength* de mayor a menor, resolviendo empates por orden de aparición, y se parte ese ranking en cuatro grupos de tamaño similar, de modo que las clases queden balanceadas.¹ Convención: clase 0 = mayor *strength* relativo; clase 3 = menor.

¹Operativamente, el ranking usa `rank` con método `first` (descendente) y la partición en cuartiles, `qcut`.

6.3 Uso del grafo en GCN, GraphSAGE y GAT

Las tres GNN operan sobre tensores densos. A partir de $A^{(t)}$ se construyen normalizaciones específicas: $\hat{A} = D^{-1/2}(A + I)D^{-1/2}$ para GCN; normalización por filas de A (sin sumar I) para GraphSAGE. GAT recibe una máscara de conectividad $(A_{ij} > 0) \vee (i = j)$; en el protocolo denso esa máscara es prácticamente completa, de modo que un resultado favorable de GAT **no** demuestra el uso de pesos continuos de correlación. GCN y GraphSAGE sí utilizan intensidades ponderadas. El procedimiento de entrenamiento de GAT en el universo denso (mini-lotes y grilla) se especifica en la subsección de modelos; no se reserva a la interpretación de Resultados.

6.4 Variables por bloque

Región y sector se incorporan como *dummies* (variables indicadoras 0/1; en Machine Learning, *one-hot*). Se conservan todas las categorías; en lo sucesivo se denominan dummies.

Tabla 2: Conjuntos de variables del diseño experimental.

Bloque	Contenido	n
B1	Seis individuales: retorno medio anualizado, volatilidad, drawdown máximo, volumen log, porcentaje de días sin transar, beta	6
B2	6 individuales + 6 dummies de región (región-v2) + 12 dummies de sector (11 sectores tipo GICS + Otros)	24
B3	24 anteriores + 18 de contexto de mercado + 5 métricas de red tabulares *_actual	47

La taxonomía **región-v2** fusiona Asia desarrollada y Asia emergente en **Asia**, y Latinoamérica con Latinoamérica y Caribe en **Latinoamérica**. Las seis categorías resultantes son: Asia; Canadá (interlisted); Europa desarrollada; Latinoamérica; Medio Oriente y África; Estados Unidos (S&P 100). Una especificación previa con ocho regiones (26 columnas) y su ampliación a 49 columnas se conserva solo como **especificación preliminar** para puentes descriptivos; no organiza

el diseño principal.

Las cinco métricas de red de la ventana histórica (solo B3) son: `strength_actual`, `corr_promedio_actual`, `eigenvector_actual`, `pagerank_actual` y `clustering_actual`. Región y sector son dummies estáticos por ticker (sin *leakage* temporal). La región es una taxonomía geográfica/listado (no GICS); el sector se homologa a once sectores del esquema GICS más la categoría Otros, a partir de etiquetas de mercado documentadas en el congelamiento del universo. En cada bloque, clásicos y GNN reciben el mismo vector tabular; las GNN reciben además el grafo. El diccionario operativo de variables consta en el Anexo 12.5. La correspondencia entre bloques narrativos y el protocolo operativo se resume en el Anexo 12.2.

6.5 Ablación geográfica (filtrado de instrumentos)

Para examinar la **sensibilidad** del protocolo B2 a la composición geográfica del universo se reentrena el mismo diseño de 24 columnas sobre subconjuntos de tickers: se filtran instrumentos; las dummies de regiones ausentes quedan en cero. En esta ablación se define operativamente una partición de las seis dummies de región-v2. El **núcleo** comprende Canadá (interlisted), Europa desarrollada y Estados Unidos (S&P 100). La **periferia** comprende Asia, Latinoamérica y Medio Oriente y África. Esa partición es una convención del diseño, no una categoría teórica de mercados. Al modificar el conjunto de instrumentos se reconstruyen la red, el ranking de *strength* y las etiquetas relativas (cuartiles); por ello la ablación evalúa la sensibilidad *integral* del problema predictivo y **no** identifica un efecto causal atribuible a una región específica.

Tabla 3: Casos de ablación geográfica sobre el Bloque 2 (24 variables).

ID	Caso	Definición	<i>N</i>
<i>Presencia del núcleo</i>			
—	Escenario base	Universo Global completo	598
f	Solo núcleo	Canadá + Europa desarrollada + S&P 100	348
a	Sin S&P 100	Excluye Estados Unidos (S&P 100)	501
<i>Sin núcleo</i>			
b	Sin núcleo	Solo periferia (Asia, Latinoamérica y Medio Oriente y África)	250
<i>Regiones añadidas al núcleo</i>			
d	Núcleo + Asia	Canadá + Europa + S&P 100 + Asia	451
e	Núcleo + MEA	Canadá + Europa + S&P 100 + Medio Oriente y África	399
c	Núcleo + LatAm	Canadá + Europa + S&P 100 + Latinoamérica	444

Elaboración propia a partir del protocolo de ablación geográfica B2 (filtrado de instrumentos).

En cada caso se **repite** la búsqueda de hiperparámetros y la selección del mejor clásico por QWK de validación del mismo subconjunto (no se reutiliza la configuración del escenario base Global). El contraste Δ QWK frente a ese clásico se calcula por escenario, con Holm sobre las tres GNN *dentro* del caso y **sin** corrección entre a–f; el objeto de la ablación en Resultados es el movimiento del nivel de QWK entre composiciones, no ese contraste. El detalle de desempeño se reserva a Resultados.

6.6 Modelos y entrenamiento

Clásicos: regresión logística, Random Forest, XGBoost (Breiman, 2001; Chen y Guestrin, 2016), k -NN y MLP. **GNN:** GCN, GraphSAGE y GAT (Kipf y Welling, 2017; Hamilton et al., 2017; Veličković et al., 2018), con optimizador AdamW, pérdida de entropía cruzada, hasta 300 épocas y *checkpoint* por mejor QWK de validación (*early stopping* por checkpoint, no por paciencia adicional distinta).

Tabla 4: Criterio de selección, semillas y agregación según modelo.

Modelo	Criterio de selección	Semillas	Agregación
Clásicos	Máximo QWK de validación	42 (fija)	Una corrida por configuración de la grilla
GCN y GraphSAGE	Máximo de la media del QWK de validación entre semillas	42, 43 y 44	Media en la grilla; contrastes con <i>hard vote</i> (moda)
GAT	Máximo QWK de validación	42 (una en la grilla)	Grilla a una semilla por costo computacional; contrastes con <i>hard vote</i> sobre 42/43/44

La selección de hiperparámetros usa solo QWK de validación. En GAT, por el grafo denso ($N = 598$), la grilla se evalúa además con mini-lotes de cuatro ventanas por paso. La media $\pm$ desviación por semilla se reserva a robustez o anexo. El mejor clásico de cada bloque o caso de ablación es el de mayor QWK de validación del mismo conjunto; el líder del conjunto de prueba, si difiere, es solo descriptivo. Detalle de hiperparámetros seleccionados: Anexo 12.6.

6.7 Partición temporal y *embargo*

Se introduce un *embargo* temporal. Primero se adoptan cortes cronológicos nominales 70/15/15 (entrenamiento / validación / conjunto de prueba) sobre las fechas de ventana; a continuación se excluyen $E = \lceil H/\Delta \rceil$ ventanas en cada frontera, de modo que las particiones no compartan retornos del horizonte del target (para $H = 126$, $E = 6$). El conjunto de prueba no interviene en la selección de hiperparámetros. La Tabla 5 resume el reparto verificado del protocolo adoptado.

El preprocesamiento tabular es un *z-score* (`StandardScaler` en clásicos; media/desviación de entrenamiento en GNN) ajustado **solo** en entrenamiento y aplicado a **todas** las columnas del vector del bloque, incluidas las dummies de región y sector. Clásicos y GNN comparten la misma lista de variables tabulares; no hay un pipeline distinto que deje las dummies sin escalar.

Tabla 5: Partición temporal del Universo Global 2015 (fechas de ventana).

Partición	N.º ventanas	Fecha inicial	Fecha final
Entrenamiento	82	2015-06-30	2022-03-31
Embargo 1	6	2022-05-02	2022-09-30
Validación	13	2022-10-31	2023-11-01
Embargo 2	6	2023-12-01	2024-05-03
Prueba	20	2024-06-04	2026-01-07

Elaboración propia a partir de `split_audit_universo_global_2015_daily.md` y del dataset supervisado (etiquetas de split por fecha de ventana).

6.8 Métricas e inferencia estadística

La métrica primaria de selección e inferencia en clasificación es QWK. Como complementarias se reportan MCC, Class RMSE y F1 macro (Chicco y Jurman, 2020; Gorodkin, 2004; Matthews, 1975).

Cada GNN (*hard vote*) se contrasta contra el mejor clásico según validación del bloque (o del caso de ablación) mediante $\Delta\text{QWK} = \text{QWK}_{\text{GNN}} - \text{QWK}_{\text{ref}}$ en el conjunto de prueba. La incertidumbre de Δ se obtiene por **moving-block bootstrap** (MBB) sobre fechas de ventana: se remuestran bloques contiguos de longitud fija $L = 6$ con remuestreo circular (wrap-around) de las ventanas del conjunto de prueba, se concatenan las filas de esos bloques y se recalcula el QWK *pooled* (agregado); $B = 10\,000$ réplicas y semilla 42 (Künsch, 1989; Politis y Romano, 1992). El IC del 95 % es el intervalo percentil empírico (cuantiles 2,5 % y 97,5 % de $\{\Delta^{*(b)}\}$).

El valor p **no** se calcula como proporción de réplicas MBB con $\Delta^* \leq 0$. Se obtiene por una **prueba de aleatorización pareada** bajo H_0 de ausencia de diferencia a favor de la GNN: en cada réplica se intercambian, por ventana (en bloques de longitud L), las predicciones de la GNN y del clásico, se recalcula Δ^* y se cuenta $k = \#\{b : \Delta^{*(b)} \geq \Delta_{\text{obs}}\}$; entonces $p = (k + 1)/(B + 1)$

(unilateral, $H_1: \Delta > 0$). La adaptación por bloques temporales corresponde al diseño con ventanas solapadas de esta memoria; no se identifica con un algoritmo de “sign-flip” de la literatura bajo ese nombre. Sobre las tres arquitecturas GNN del *mismo* escenario se aplica Holm (Holm, 1979) ($m = 3$). El valor p ajustado se reporta en Resultados como p_{Holm} . No se corrige Holm entre bloques de variables distintos ni entre escenarios geográficos a–f.

Regla de lectura del contraste primario (conjunto de prueba completo), alineada a la implementación:

- *supera*: el IC del 95 % queda enteramente por encima de cero ($lo > 0$);
- *empata*: el IC del 95 % contiene al cero;
- *pierde*: el IC del 95 % queda enteramente por debajo de cero ($hi < 0$).

Un IC que excluye el cero puede coexistir con un p ajustado por Holm no significativo al 5 %: en ese caso se reportan ambos y **no** se afirma superioridad confirmada por el p ajustado por Holm. El contraste unilateral de superioridad **no** demuestra por sí solo inferioridad; el veredicto *pierde* proviene del IC enteramente negativo. Los cortes del horizonte calendario son exploratorios (IC del 95 % sin ajuste de Holm) y no sustituyen el contraste del conjunto de prueba completo. Solo se considerará evidencia de superioridad estadísticamente significativa cuando se cumplan los criterios inferenciales previamente establecidos. Tampoco se mezclan en un mismo contraste especificaciones de 6, 24 y 47 variables, ni especificaciones preliminares de 26 o 49 variables.

6.9 Horizonte calendario de evaluación

Además del horizonte del target H , se examina el **horizonte calendario de evaluación**: meses desde la primera fecha de ventana del conjunto de prueba (t_0). Se distinguen tres objetos,

alineados a la implementación:

- **QWK por ventana** QWK_t : QWK en la fecha de ventana t (figuras temporales).
- **Media de Δ QWK** en un prefijo: promedio, sobre las ventanas del prefijo, de $QWK_t^{\text{GNN}} - QWK_t^{\text{ref}}$ (figuras de Δ y tablas exploratorias a $\approx 12 / 18$ meses).
- **QWK absoluto a cortes**: QWK *pooled* sobre todas las predicciones del prefijo $[t_0, \dots, t_k]$ (tablas a $\approx 6 / 12 / 18$ meses).

Este análisis es distinto de H y se interpreta con cautela cuando el prefijo contiene pocas ventanas.

6.10 Diagnósticos de valor incremental y parsimonia

Tras los bloques de paridad tabular, se evalúa si el contexto de mercado y las métricas de red tabulares justifican abandonar la especificación de 24 variables a favor de 47. El contraste GraphSAGE 24 vs 47 es descriptivo (no MBB/Holm entre bloques) y constituye la evidencia de parsimonia reportada en Resultados.

Como material auxiliar de trazabilidad se conservan diagnósticos lineales regularizados y ablaciones por grupos bajo especificaciones históricas de 26 y 49 variables (taxonomía de ocho regiones). Ese material no reproduce GraphSAGE del protocolo vigente ni identifica el aporte causal del grafo; su detalle numérico consta en el Anexo 12.12, etiquetado como histórico.

6.11 Sensibilidad del horizonte del target

Como robustez metodológica se evalúa $H \in \{63, 84, 126\}$. Se mantienen fijos el universo, $W = 126$ y $\Delta = 21$. El embargo **sí** depende del horizonte: $E = \lceil H/\Delta \rceil \in \{3, 4, 6\}$. Cambia la

longitud de la ventana futura del target y, con ella, las etiquetas. La corrida documentada utiliza la especificación preliminar de **26** variables (región-8), no el B2 de 24 variables: se interpreta como robustez **histórica** etiquetada, no como reentrenamiento bajo región-v2. El detalle numérico no se reporta en Resultados; se remite a Limitaciones y, cuando corresponda, al material histórico del Anexo.

6.12 Score ordinal y exposición sistémica relativa

Para vincular la clasificación ordinal con la aplicación de portafolio se define un score probabilístico en $[0, 1]$:

$$s = \frac{3p_0 + 2p_1 + p_2}{3}, \quad (10)$$

donde p_c es la probabilidad de clase c (la clase 3 no aporta al numerador). El score vale cerca de 1 cuando el modelo asigna mayor probabilidad a las clases de alta importancia sistémica relativa, y cerca de 0 cuando esa probabilidad se concentra en las clases de menor importancia. El score principal se obtiene de GraphSAGE con *soft vote* (promedio de probabilidades entre semillas 42/43/44), distinto del *hard vote* de los contrastes de clasificación, a partir de la señal de 24 variables; como referencia tabular se documenta además un score de MLP bajo las mismas 24 columnas.

La **validación del score** (antes de interpretarlo como puente económico) usa, en el protocolo del score ordinal de 24 variables: correlación de Spearman entre s y el *ranking/strength* futuro; Ranked Probability Score (RPS) (Epstein, 1969); y Brier agregado sobre clases (Brier, 1950). Esas métricas diagnostican ordenación y calibración ordinal; **no** demuestran utilidad de portafolio. El score no es un pronóstico de retorno.

Dado un vector de pesos w (long-only, $\sum_i w_i = 1$), se define la exposición sistémica relativa

$$ESR_t^{\text{pred}} = w_t^\top s_t, \quad ESR_t^{\text{real}} = w_t^\top s_t^{\text{fut}}, \quad (11)$$

donde s_t es el score (o percentil) disponible en t y s_t^{fut} es el percentil transversal de `strength_futura` en la fecha de rebalanceo (definición del protocolo de Variante A). ESR **no** es Expected Shortfall financiero (CVaR). En la optimización entra ESR^{pred} , la evaluación primaria fuera de muestra usa ESR^{real} .

6.13 Portafolio Variante A (mínima varianza penalizada, fuera de muestra)

La Variante A incorpora la señal como penalización en un problema de mínima varianza relativa (Markowitz, 1952):

$$\min_w \left[\frac{w^\top \hat{\Sigma}_t w}{w_{\text{EW}}^\top \hat{\Sigma}_t w_{\text{EW}}} + \lambda w^\top s_t \right] \quad \text{s.t.} \quad \sum_i w_i = 1, \quad 0 \leq w_i \leq w_{\text{máx}}. \quad (12)$$

Parámetros operativos del protocolo documentado: $w_{\text{máx}} = 0,01$; grilla $\lambda \in \{0, 0,01, 0,03, 0,1, 0,3, 1, 3, 10\}$; covarianza Ledoit–Wolf (Ledoit y Wolf, 2004) con ventana histórica de 126 días hábiles; rebalanceo cada 21 días hábiles; costo de transacción de referencia 10 pb sobre turnover. Cuando $\lambda = 0$ se recupera MV bajo las mismas restricciones.

Selección de λ (solo validación): se elige el menor λ de la grilla tal que la media temporal de $ESR_\lambda^{\text{real}} - ESR_{\text{MV}}^{\text{real}}$ sea negativa y el Sharpe neto (Sharpe, 1966) cumpla $\text{Sharpe}_\lambda \geq \text{Sharpe}_{\text{MV}} - \text{SE}(\text{Sharpe}_{\text{MV}})$; si ningún $\lambda > 0$ cumple, se adopta $\lambda = 0$. El conjunto de prueba no interviene en esa elección.

Se distinguen dos relojes: el **horizonte financiero** del rebalanceo (21 días hábiles entre t

y $t + 21$) y el **horizonte estructural** del target ($H = 126$), usado para construir s^{fut} . La métrica primaria de contraste es ΔESR^{real} (con MBB), no la superioridad en retorno medio.

6.14 Portafolio Variante B (mapa retrospectivo sobre la frontera media–varianza)

La Variante B se fija en esta memoria como **mapa diagnóstico retrospectivo**, no como experimento fuera de muestra de utilidad económica ni como estrategia de trading. En una primera etapa se construye un menú discreto de portafolios en la frontera media–varianza (Markowitz, 1952) (sin el clasificador). En una segunda, la señal s evalúa $ESR(w) = w^T s$ en cada punto y permite ordenar o colorear la frontera. El modelo **no genera** la frontera.

El diseño documentado emplea estimación de μ y Σ con la muestra completa, con fines ilustrativos: la frontera es retrospectiva y no constituye un backtest fuera de muestra. Esa limitación se declara al presentar Resultados.

Se utiliza Python con PyTorch, scikit-learn, pandas, NetworkX y XGBoost. El diccionario de variables y los hiperparámetros seleccionados se remiten a los Anexos 12.2, 12.5 y 12.6.

7 Alcance

En términos de conocimiento, el estudio es predominantemente **correlacional–predictivo**: examina el grado de asociación y el desempeño relativo entre especificaciones de modelos (GNN frente a Machine Learning tabular) y la clase futura de conectividad relativa, bajo un protocolo temporal fuera de muestra. Incorpora componentes **descriptivos**, en particular paneles de métricas y el mapa de exposición sistémica relativa sobre una frontera media–varianza. No es un estudio explicativo–causal: no identifica el efecto causal del grafo ni de regiones geográficas sobre el desempeño.

El fenómeno estudiado es la predicción de la **importancia sistémica futura relativa** de instrumentos accionarios, no la predicción de precios, retornos, crisis o riesgo sistémico absoluto. El análisis se limita al **Universo Global 2015** (598 instrumentos listados en Estados Unidos), con datos diarios y un horizonte de predicción de seis meses hábiles en el diseño principal. Quedan fuera usos regulatorios o de trading.

La inferencia se interpreta como predictiva y comparativa entre modelos bajo un mismo protocolo, no como explicación causal de la estructura de red.

8 Resultados

Este capítulo reporta, en el Universo Global 2015 (598 instrumentos), los tres bloques de clasificación, el score y las dos variantes de portafolio. En cada bloque se contrastan GCN, GraphSAGE y GAT, cada una, con el clásico elegido en validación; la lectura del intervalo y de Holm es la de Metodología.

8.1 Bloque 1: seis variables individuales

8.1.1. Diseño

Clásicos y GNN reciben las mismas seis variables individuales (`ret_medio_anual`, `vol_anual`, `drawdown_max`, `volumen_prom_log`, `pct_dias_sin_transar`, `beta`). Las GNN reciben además el grafo denso `|corr|`; las predicciones se agregan por *hard vote* (voto mayoritario) entre semillas 42/43/44. El modelo clásico de referencia, seleccionado por QWK en validación, es el MLP (QWK val = 0,647).

8.1.2. Resultados en el conjunto de prueba

La Tabla 6 resume el panel fuera de muestra de clásicos y GNN. GraphSAGE lidera el bloque en QWK; el MLP es la referencia del contraste primario.

Tabla 6: Bloque 1 (seis variables). Panel en el conjunto de prueba.

Modelo	QWK	MCC	Class RMSE	F1 macro
<i>Clásicos</i>				
MLP (referencia por validación)	0,601	0,305	0,972	0,486
Random Forest	0,582	0,293	1,014	0,472
XGBoost	0,582	0,297	1,015	0,474
k -NN	0,560	0,273	1,044	0,457
Regresión logística	0,525	0,262	1,105	0,442
<i>GNN</i>				
GraphSAGE	0,609	0,322	1,021	0,477
GAT	0,457	0,224	1,234	0,377
GCN	0,194	0,121	1,627	0,194

El QWK de esa tabla es el valor *pooled* (agregado) sobre las 20 fechas de ventana del conjunto de prueba. La Figura 5 descompone el mismo periodo en el QWK por ventana QWK_t , alineado al horizonte calendario de evaluación (meses desde la primera fecha del conjunto de prueba). El ejercicio es complementario y exploratorio: no sustituye el panel ni el contraste ΔQWK de la subsección siguiente. En el trazado, GraphSAGE permanece cerca del MLP y del grupo de clásicos. GAT se desacopla a la baja tras el primer tercio del conjunto de prueba y GCN colapsa hacia valores próximos a cero. Varios modelos comparten un pico temprano y un descenso posterior, de modo que la variación temporal no es exclusiva de una arquitectura. Esa lectura muestra que el orden del panel no equivale a una ventaja constante fecha a fecha, y motiva el intervalo y Holm sobre el conjunto completo.

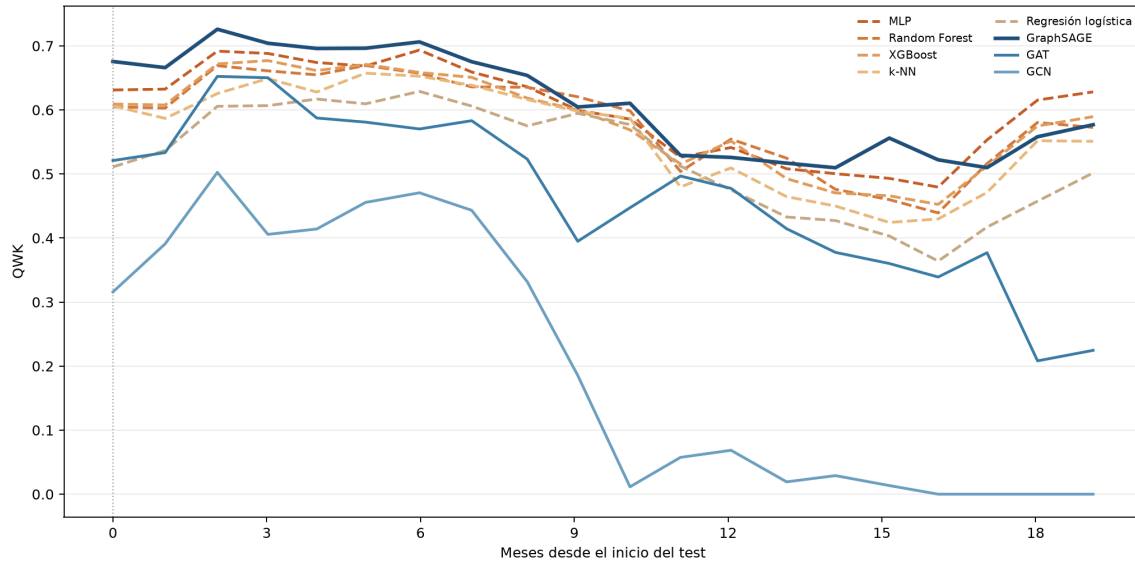


Figura 5: Bloque 1. QWK por ventana del conjunto de prueba (panel temporal; complementario al QWK *pooled*).

8.1.3. Inferencia

La Tabla 7 reporta ΔQWK frente al MLP. La columna p_{Holm} es el valor p ajustado por Holm. GraphSAGE **empata** (IC contiene el cero; Holm no rechaza). GAT y GCN **pierden** por IC enteramente negativo. La Figura 6 ilustra los intervalos.

Tabla 7: Bloque 1. Inferencia ΔQWK (GNN *hard vote* vs MLP).

GNN	ΔQWK	IC95 %	p_{Holm}	Lectura
GraphSAGE	+0,008	$[-0,003, 0,020]$	1,000	empata
GAT	-0,144	$[-0,204, -0,093]$	1,000	pierde
GCN	-0,407	$[-0,498, -0,287]$	1,000	pierde

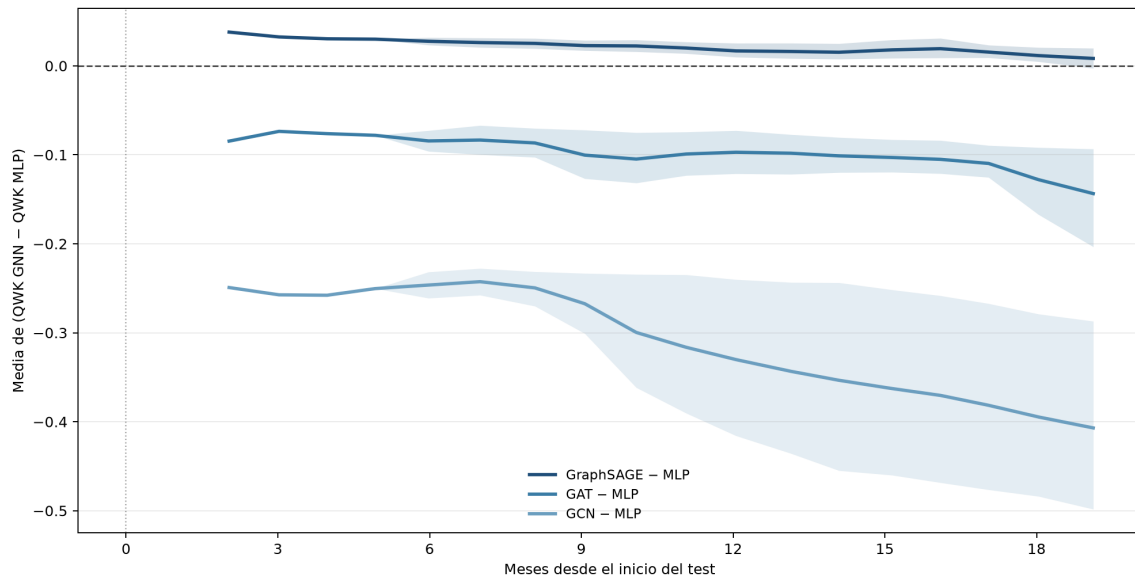


Figura 6: Bloque 1. Δ QWK vs MLP con IC95 % (MBB).

Con seis variables individuales, GraphSAGE lidera el panel en QWK y empata al MLP por IC; GAT y GCN pierden.

8.2 Bloque 2: región-v2, sector (24 variables) y ablación geográfica

8.2.1. Diseño

Se incorporan seis dummies de región (taxonomía región-v2) y doce de sector, elevando el conjunto a **24** columnas bajo la misma paridad tabular. El clásico de referencia del escenario base Global es el MLP (QWK val = 0,656). Sobre ese diseño se define además una ablación geográfica: filtrar instrumentos, reconstruir red y etiquetas, y reentrenar en cada recorte, para examinar cómo se mueve el nivel de QWK al cambiar la composición (núcleo y periferia).

8.2.2. Resultados en el conjunto de prueba

La Tabla 8 resume el panel fuera de muestra de clásicos y GNN. GraphSAGE obtiene el mejor QWK del conjunto de prueba (0,629). XGBoost supera al MLP en QWK puntual del conjunto de prueba, pero la referencia del contraste primario sigue siendo el MLP seleccionado en validación.

Tabla 8: Bloque 2 (24 variables). Panel en el conjunto de prueba.

Modelo	QWK	MCC	Class RMSE	F1 macro
<i>Clásicos</i>				
MLP (referencia por validación)	0,610	0,314	0,971	0,488
XGBoost	0,614	0,326	0,979	0,494
Random Forest	0,589	0,308	1,010	0,482
Regresión logística	0,570	0,302	1,044	0,472
<i>k</i> -NN	0,559	0,281	1,050	0,460
<i>GNN</i>				
GraphSAGE	0,629	0,323	0,981	0,485
GAT	0,606	0,324	1,017	0,483
GCN	0,462	0,211	1,233	0,386

8.2.3. Inferencia

La Tabla 9 reporta Δ QWK frente al MLP. GraphSAGE **empata** ($\Delta = +0,019$; IC cruza cero; Holm = 0,551). GAT empata; GCN pierde. La Figura 7 ilustra los intervalos.

Tabla 9: Bloque 2. Inferencia Δ QWK (GNN *hard vote* vs MLP).

GNN	Δ QWK	IC95 %	p_{Holm}	Lectura
GraphSAGE	+0,019	[−0,007, 0,042]	0,551	empata
GAT	−0,005	[−0,025, 0,012]	1,000	empata
GCN	−0,149	[−0,202, −0,081]	1,000	pierde

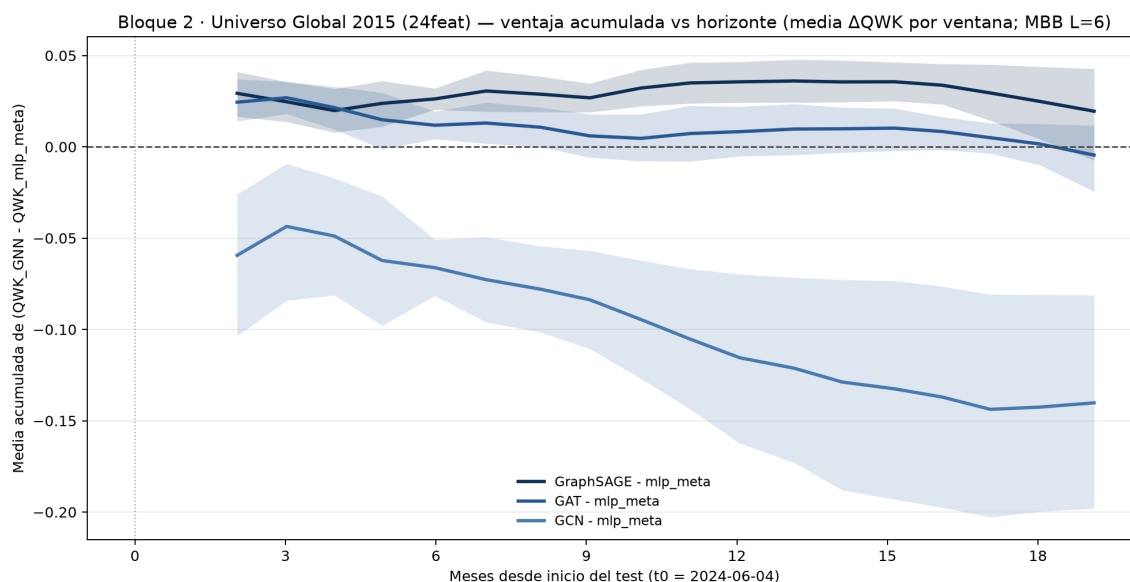


Figura 7: Bloque 2. Δ QWK vs MLP con IC95 % (MBB).

En prefijos de ≈ 6 , 12 y 18 meses la media de Δ QWK de GraphSAGE es positiva (+0,026, +0,036 y +0,025); en el conjunto de prueba completo el IC vuelve a contener el cero, coherente con Holm.

Con 24 variables, GraphSAGE obtiene el mejor QWK del bloque y empata al MLP por IC.

8.2.4. Ablación geográfica a-f

La ablación examina cómo se mueve el **nivel** de QWK cuando cambia la composición geográfica del universo, no si una arquitectura gana a otra en cada recorte. El núcleo se define operativamente como Canadá, Europa desarrollada y S&P 100: es una convención del diseño, no una categoría teórica. Filtrar instrumentos reconstruye red, ranking de *strength* y cuartiles; la lectura es de sensibilidad integral, no de efecto causal de una región. La Tabla 10 y la Figura 8 agrupan los escenarios según si el núcleo está presente, ausente o extendido con regiones añadidas al núcleo.

Tabla 10: Bloque 2. Ablación geográfica: nivel de QWK en el conjunto de prueba según la composición del universo.

ID	Caso	N	QWK GraphSAGE	QWK clásico
<i>Presencia del núcleo</i>				
—	Escenario base Global	598	0,629	0,610
f	Solo núcleo	348	0,592	0,569
a	Sin S&P 100	501	0,614	0,614
<i>Sin núcleo</i>				
b	Solo periferia	250	0,660	0,626
<i>Regiones añadidas al núcleo</i>				
d	Núcleo + Asia	451	0,646	0,630
e	Núcleo + MEA	399	0,609	0,581
c	Núcleo + LatAm	444	0,573	0,544

El universo solo-núcleo (f) queda por debajo del escenario base Global en GraphSAGE y en el clásico del conjunto de prueba. La periferia sola (b) eleva ambos niveles, con $N = 250$. Quitar únicamente el S&P 100 (a) deja el QWK cerca del Global. El núcleo, por sí solo, no reproduce el desempeño del universo mixto.

Sobre el núcleo, añadir Asia (d) sube el nivel por encima del Global; MEA (e) queda intermedio; LatAm (c) produce el QWK más bajo de la ronda. GraphSAGE y el clásico se desplazan en el mismo sentido: cambia la dificultad del problema relativo, no un ranking GNN.

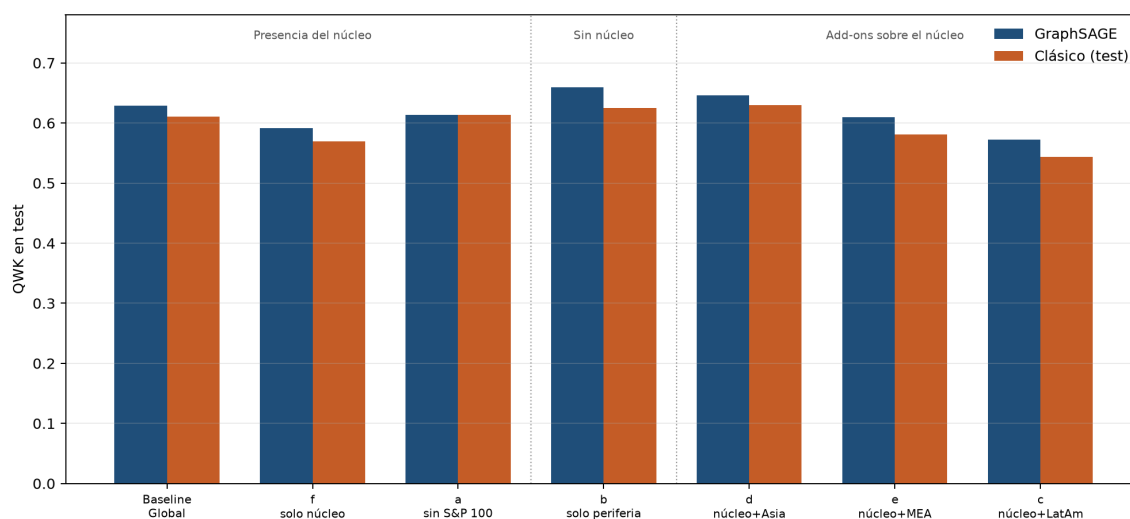


Figura 8: Bloque 2. QWK del conjunto de prueba por composición geográfica (GraphSAGE y clásico del conjunto de prueba).

El contraste ΔQWK frente al clásico de cada caso no organiza esta lectura: GraphSAGE empata en casi todos los escenarios. Holm no confirma superioridad en ninguno; el único *supera* por IC es b, con $N = 250$. GAT es frágil en a, d y f. El puente descriptivo de 26 variables frente a 24 consta en el Anexo 12.1.

8.3 Bloque 3: conjunto ampliado (47 variables) y parsimonia

8.3.1. Diseño

Sobre la base de 24 se añaden 18 variables de contexto de mercado y 5 métricas de red tabulares $*_{actual}$, para un total de **47** columnas, en el mismo Universo Global 2015. El clásico de referencia es XGBoost (QWK val = 0,682). El contraste 24 vs 47 es descriptivo (no MBB/Holm entre bloques).

8.3.2. Resultados en el conjunto de prueba

La Tabla 11 resume el panel fuera de muestra de clásicos y GNN. GraphSAGE lidera el bloque en QWK (0,616). Random Forest supera a XGBoost en QWK puntual del conjunto de prueba, pero la referencia del contraste primario sigue siendo el XGBoost seleccionado en validación.

Tabla 11: Bloque 3 (47 variables). Panel en el conjunto de prueba.

Modelo	QWK	MCC	Class RMSE	F1 macro
<i>Clásicos</i>				
XGBoost (referencia por validación)	0,608	0,318	0,978	0,493
Random Forest	0,611	0,317	0,960	0,495
Regresión logística	0,600	0,318	1,000	0,488
MLP	0,541	0,262	1,038	0,446
k -NN	0,529	0,274	1,097	0,452
<i>GNN</i>				
GraphSAGE	0,616	0,321	0,983	0,487
GAT	0,475	0,256	1,189	0,417
GCN	0,274	0,120	1,444	0,284

El QWK de esa tabla es el valor *pooled* (agregado) sobre las 20 fechas de ventana del conjunto de prueba. La Figura 9 descompone el mismo periodo en el QWK por ventana QWK_t , alineado al horizonte calendario de evaluación (meses desde la primera fecha del conjunto de prueba). El ejercicio es complementario y exploratorio: no sustituye el panel ni el contraste ΔQWK de la subsección siguiente. En el trazado, GraphSAGE permanece cerca del XGBoost y del grupo de clásicos en el primer tramo. GAT se desacopla a la baja más adelante y GCN se mantiene volátil y por debajo. Varios modelos comparten un descenso posterior, de modo que la variación temporal no es exclusiva de una arquitectura. Esa lectura muestra que el orden del panel no equivale a una ventaja constante fecha a fecha, y motiva el intervalo y Holm sobre el conjunto completo.

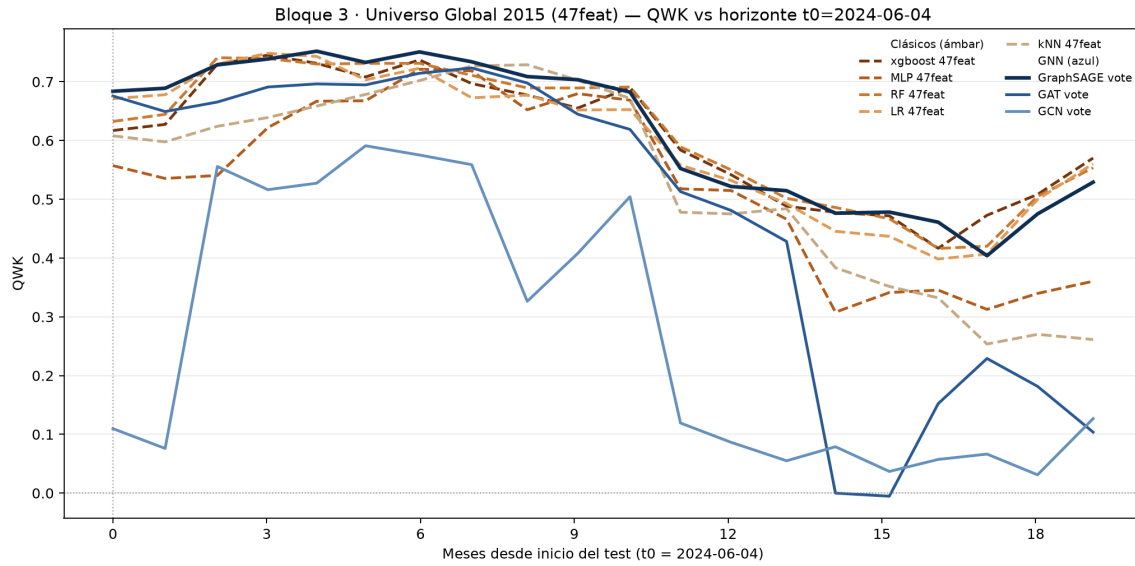


Figura 9: Bloque 3. QWK por ventana del conjunto de prueba (panel temporal; complementario al QWK *pooled*).

8.3.3. Inferencia

La Tabla 12 reporta ΔQWK frente a XGBoost. GraphSAGE **empata** ($\Delta = +0,008$; Holm = 0,925). GAT y GCN **pierden** por IC enteramente negativo. La Figura 10 ilustra los intervalos.

Tabla 12: Bloque 3. Inferencia ΔQWK (GNN *hard vote* vs XGBoost).

GNN	ΔQWK	IC95 %	p_{Holm}	Lectura
GraphSAGE	+0,008	$[-0,005, 0,022]$	0,925	empata
GAT	-0,133	$[-0,267, -0,021]$	1,000	pierde
GCN	-0,333	$[-0,421, -0,238]$	1,000	pierde

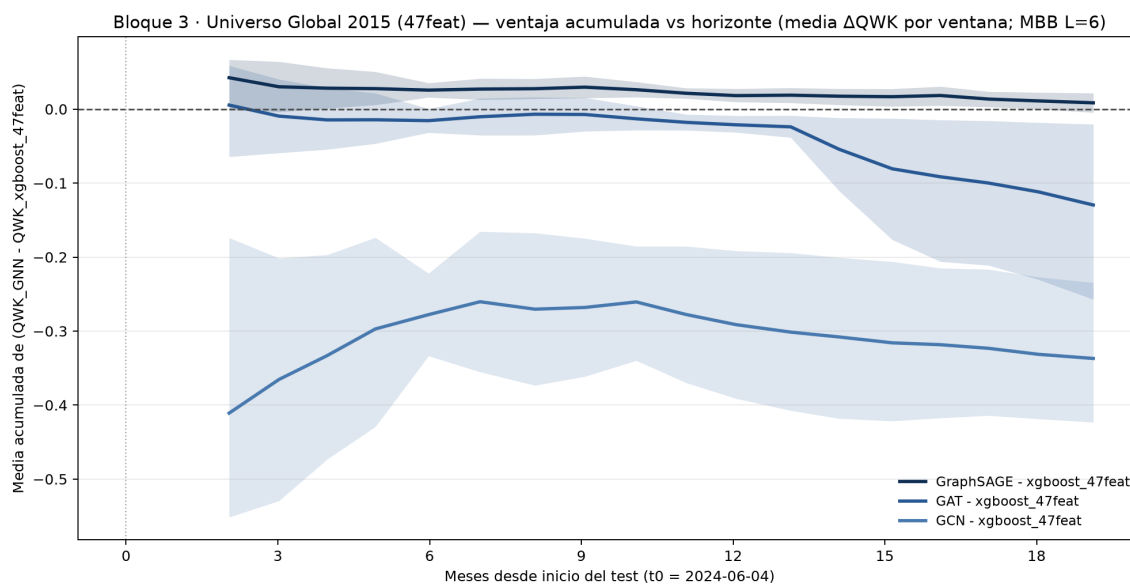


Figura 10: Bloque 3. Δ QWK vs XGBoost con IC95 % (MBB).

La media de Δ GraphSAGE vs XGBoost es positiva en prefijos cortos y empata en el plazo completo, alineada con Holm.

Con 47 variables, GraphSAGE lidera el panel en QWK y empata a XGBoost por IC; GAT y GCN pierden.

8.3.4. Parsimonia 24 vs 47

El QWK del conjunto de prueba de GraphSAGE baja de 0,629 (24 variables) a 0,616 (47 variables): con esas cifras la diferencia es $-0,013$. Con valores sin redondear (0,6155 frente a 0,6293) vale $-0,0138$, que redondea a $-0,014$. Bajo 47 variables, GraphSAGE empata a XGBoost por IC; el QWK del conjunto de prueba (0,616) queda por debajo del de B2 (0,629).

8.4 Score ordinal (24 variables)

Se exporta un score continuo

$$s = \frac{3p_0 + 2p_1 + p_2}{3} \in [0, 1] \quad (13)$$

desde GraphSAGE con *soft vote* (semillas 42/43/44), más un score MLP de referencia bajo las mismas 24 columnas. En el conjunto de prueba, Spearman(s , percentil de *strength*) $\approx 0,705$, con monotonidad por quintiles y estabilidad entre semillas. El score ordena conectividad relativa futura y se usa como entrada de portafolio (Figura 11).

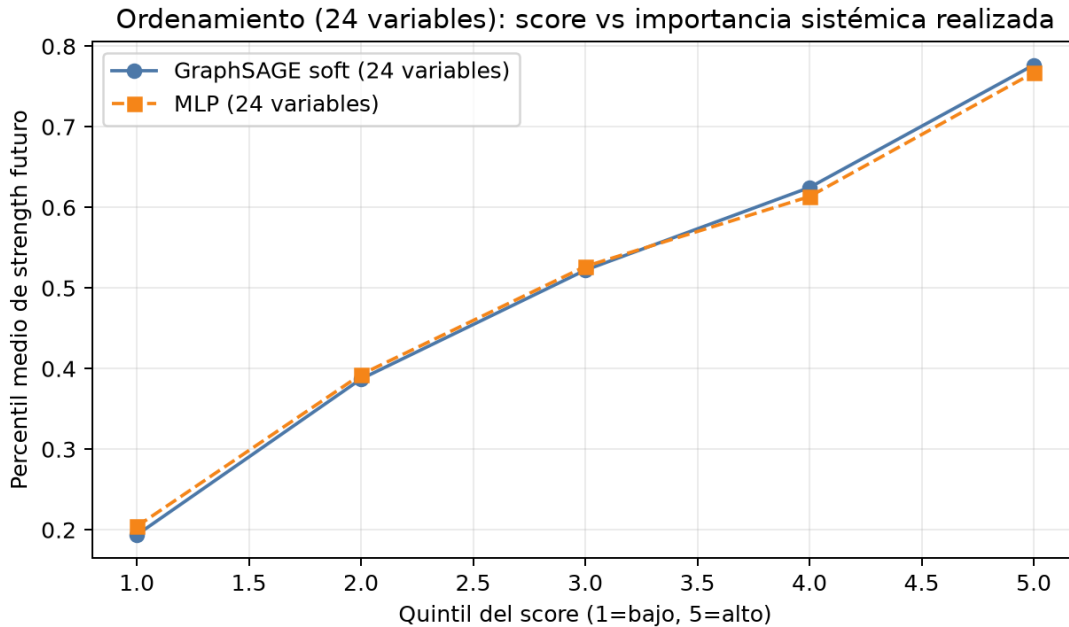


Figura 11: Score de 24 variables: quintiles vs *strength* en el conjunto de prueba.

La Figura 11 agrupa los instrumentos del conjunto de prueba en quintiles del score, del más bajo al más alto, y en cada grupo reporta el percentil medio de *strength* futuro. GraphSAGE con *soft vote* y el MLP de las mismas 24 variables recorren una trayectoria creciente: a mayor s

corresponde, en promedio, más conectividad relativa observada después. El gráfico no sostiene que el canal de grafo supere al MLP. Muestra que s ordena el target y que, con esa lectura, puede entrar como señal de portafolio.

8.5 Portafolio Variante A (mínima varianza penalizada, fuera de muestra, 24 variables)

8.5.1. Diseño

Se optimiza mínima varianza relativa con penalización $\lambda w^\top s$, covarianza Ledoit–Wolf (126 d), long-only, $w_{\text{máx}} = 1\%$, rebalanceo cada 21 días hábiles y costo de 10 pb. El parámetro λ se elige solo en validación; el test se evalúa una vez. Señales: P4 GraphSAGE, P3 MLP, P2 *strength* contemporánea, P1 MV y P0 equal-weight; P5 oráculo solo diagnóstico. En todos los casos con señal, $\lambda = 0,01$.

8.5.2. Resultados en el conjunto de prueba

La Tabla 13 resume métricas del conjunto de prueba. P4 reduce *ESR* respecto de MV con magnitud pequeña, sin deteriorar el Sharpe observado. P5 no es implementable.

Tabla 13: Variante A. Métricas del conjunto de prueba (costo 10 pb).

Estrategia	λ	<i>ESR</i>	Sharpe	Ret. acum.	Vol	MaxDD
P0 EW	0	0,5008	1,6032	0,533	0,169	−0,170
P1 MV	0	0,3242	1,9153	0,408	0,111	−0,119
P2 Actual	0,01	0,3229	1,9163	0,409	0,111	−0,119
P3 MLP	0,01	0,3227	1,9167	0,409	0,111	−0,119
P4 GraphSAGE	0,01	0,3225	1,9186	0,409	0,111	−0,119
P5 Oráculo	0,01	0,3210	1,9221	0,409	0,110	−0,118

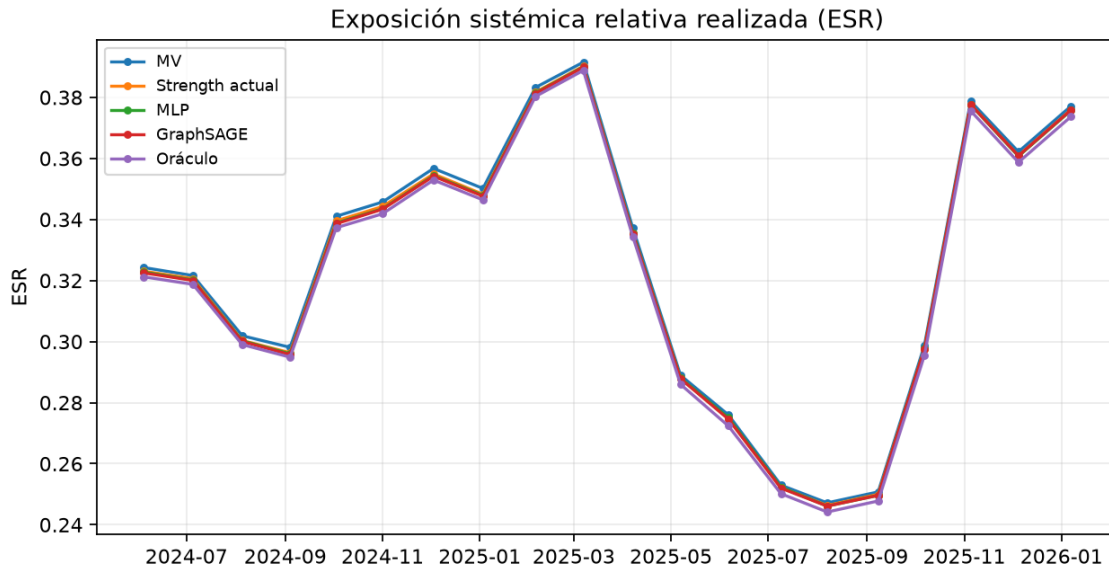


Figura 12: Variante A. Exposición sistémica relativa (ESR) en el tiempo (conjunto de prueba).

En la Figura 12, P1–P4 se mueven juntas: la reducción de ESR es de nivel y de magnitud pequeña, no un régimen distinto. Las curvas de riqueza de ese grupo se superponen; constan en el Anexo 12.9.

8.5.3. Inferencia

La Tabla 14 reporta ΔESR de P4 frente a las referencias. P4 reduce ESR respecto de MV ($\Delta ESR \approx -0,0017$; IC negativo; Holm significativo). Los contrastes frente a P2 y P3 también favorecen a P4, con efectos aún menores. El número de rebalances del conjunto de prueba es moderado (≈ 20); la inferencia MBB debe leerse con esa cautela.

Tabla 14: Variante A. Inferencia ΔESR (P4 vs referencias). La última columna indica si el contraste es significativo tras el ajuste de Holm al 5 %.

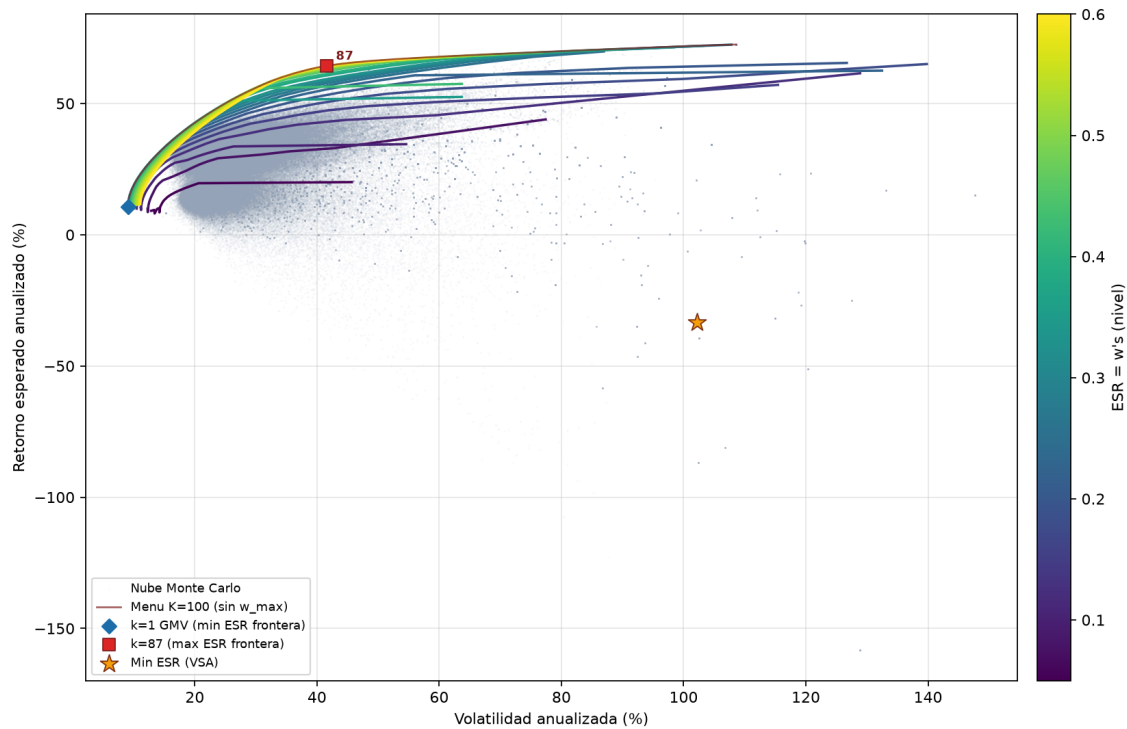
Contraste	ΔESR	IC95 %	Holm
P4 vs P1 MV	-0,0017	[-0,0022, -0,0012]	sí
P4 vs P2	-0,0004	[-0,0006, -0,0002]	sí
P4 vs P3 MLP	-0,0002	[-0,0002, -0,0001]	sí

Con la señal de 24 variables, P4 reduce ESR frente a MV en el conjunto de prueba ($\Delta ESR \approx -0,0017$; IC negativo; Holm significativo) sin deteriorar el Sharpe observado.

8.6 Portafolio Variante B (mapa retrospectivo sobre la frontera media–varianza, 24 variables)

La Variante B recolora un menú de $K = 100$ portafolios en la frontera media–varianza (estimación con la muestra completa, retrospectiva; sin regenerar el QP) con $ESR(w) = w^\top s$ y s igual a la media por ticker del score GraphSAGE soft de 24 variables. El mínimo de ESR en la frontera coincide con el GMV ($k = 1$, $ESR = 0,264$); el máximo ocurre en $k = 87$ ($ESR = 0,634$), no en el extremo de máximo retorno. La correlación entre el índice del menú y ESR es 0,773 (Figura 13). Un recorte de detalle consta en el Anexo 12.9.

El ejercicio es ilustrativo y retrospectivo; no sustituye a la Variante A.



Curvas: frontera Markowitz condicionada a $ESR=w's$ (paso 0.025; score_sr media GraphSAGE 24feat). Long-only, sin w_{max} . Complemento visual full-sample.

Figura 13: Variante B. Curvas de nivel de ESR sobre la frontera (estimación con la muestra completa, retrospectiva).

8.7 Síntesis del capítulo

En el Universo Global 2015, GraphSAGE es competitivo frente a los clásicos de referencia, sin evidencia de superioridad general: GCN y GAT no reproducen ese patrón. La especificación de 24 variables queda por encima de la de 47 para GraphSAGE y se adopta como base del score y de los portafolios. La señal reduce de forma detectable la exposición sistémica relativa realizada, con magnitud pequeña.

9 Discusión

Este capítulo interpreta la evidencia empírica frente al marco teórico y a la literatura previa. No reitera paneles ni cifras en bloque: recupera solo lo necesario para sostener lecturas sobre las preguntas de investigación, delimitar lo que el diseño permite afirmar y explicitar el aporte académico.

9.1 Respuesta a la pregunta principal

La evidencia del Universo Global 2015 no respalda una superioridad universal de las Graph Neural Networks frente a modelos clásicos de Machine Learning en la clasificación fuera de muestra de la importancia sistémica futura relativa. GraphSAGE es competitivo frente al clásico de referencia seleccionado en validación en las tres amplitudes evaluadas (6, 24 y 47 variables): los intervalos de confianza de ΔQWK contienen el cero y el ajuste de Holm no confirma superioridad. GCN, y en varios contrastes también GAT, pierden o no alcanzan ese nivel de competitividad en el criterio primario. El patrón es, por tanto, de heterogeneidad arquitectónica: compartir la etiqueta “grafo” no implica un desempeño homogéneo bajo una red densa de correlaciones.

Esa lectura es coherente con la propia diversidad de la familia GNN. GCN, GraphSAGE y GAT formalizan agregaciones distintas sobre el vecindario (Kipf y Welling, 2017; Hamilton et al., 2017; Veličković et al., 2018), y los trabajos de revisión del área subrayan que el rendimiento depende de la arquitectura y del tipo de grafo (Wu et al., 2021). En este problema, tratar “las GNN” como un bloque empírico único oscurecería el hallazgo central: GraphSAGE es la arquitectura que sostiene competitividad consistente frente al mejor clásico de validación, no la familia completa.

La categoría operacional *empata* del protocolo no debe confundirse con una prueba de equivalencia estadística. Un intervalo que contiene el cero indica ausencia de evidencia suficiente

para establecer superioridad, no identidad exacta de desempeño. Con esa cautela, la respuesta a la pregunta principal es que GraphSAGE presenta un desempeño competitivo, mientras que la evidencia no permite afirmar superioridad general de las GNN.

9.2 Paridad tabular y procesamiento de la estructura relacional

Bajo paridad tabular, clásicos y GNN reciben el mismo vector de variables de nodo; las GNN procesan además la adyacencia. Ese diseño hace la comparación más informativa que un contraste entre especificaciones heterogéneas de atributos, pero no iguala por completo la información disponible: el canal relacional permanece exclusivo de las arquitecturas de grafo. El resultado recurrente (competitividad de GraphSAGE sin victoria sistemática) sugiere que procesar explícitamente la estructura relacional permite alcanzar un desempeño comparable al de un buen modelo tabular, sin garantizar una ventaja estable en QWK.

Esta interpretación se alinea con dos advertencias del marco. Primero, las redes de correlación resumen comovimiento estadístico y no identifican contagio causal (Mantegna, 1999; Billio et al., 2012; Giudici et al., 2020). Segundo, el Machine Learning tabular suele incorporar la red mediante centralidades calculadas previamente, lo que limita el uso granular de la adyacencia (Balmaseda et al., 2023; Kou et al., 2019). Las GNN se motivan precisamente para cerrar ese hueco; los hallazgos propios matizan la expectativa de que ese canal se traduzca automáticamente en superioridad predictiva bajo red densa.

Tampoco cabe atribuir causalmente el desempeño de GraphSAGE al “efecto del grafo” aislado. El contraste evalúa arquitecturas completas bajo paridad de variables tabulares, no un experimento que desconecte la adyacencia manteniendo todo lo demás fijo. Además, en el protocolo denso GAT opera sobre una máscara de conectividad prácticamente completa, de modo que un

resultado favorable suyo no demostraría el uso de pesos continuos de correlación; ese punto se desarrolla en Limitaciones. En conjunto, la pregunta secundaria se responde como competitividad asociada a GraphSAGE bajo paridad, no como identificación causal del aporte de la adyacencia.

9.3 Parsimonia: 24 frente a 47 variables

Ampliar el conjunto tabular desde 24 columnas (región-v2 y sector) hacia 47 (contexto de mercado y métricas de red contemporáneas) no eleva el QWK fuera de muestra de GraphSAGE. El nivel observado bajo 47 variables queda por debajo del de 24 en una diferencia descriptiva pequeña, y GraphSAGE empata al clásico de referencia también en el bloque ampliado. La lectura pertinente es de parsimonia: más información disponible no garantiza mayor valor predictivo para esta tarea y este protocolo.

El resultado es compatible con varias hipótesis interpretativas, ninguna identificada por el diseño. Las métricas de centralidad capturan dimensiones distintas de la estructura (Newman, 2010; Barrat et al., 2004), pero su utilidad como variables tabulares puede ser redundante respecto de lo que una GNN ya agrega desde la adyacencia, o bien difícil de aprovechar en presencia de mayor dimensionalidad. Tampoco se sigue que el contexto de mercado “perjudique” al modelo ni que exista sobreajuste demostrado a variables concretas: el contraste entre amplitudes es descriptivo y no aísla el efecto de cada grupo de atributos.

En términos del proxy adoptado, el target sigue siendo una clase relativa de *strength* futuro, no DebtRank ni una medida regulatoria de contagio (Battiston et al., 2012). Añadir resúmenes tabulares de la red contemporánea no equivale, por sí solo, a enriquecer el objeto de predicción. Por ello, conservar la especificación de 24 variables como base del score y de los portafolios es una decisión de parsimonia empírica, no un rechazo del bloque ampliado como ejercicio de sensibilidad

informativa.

9.4 Sensibilidad a la composición geográfica

La ablación sobre el diseño de 24 variables muestra que el **nivel** de QWK se mueve cuando cambia la composición geográfica y se reconstruyen red y etiquetas relativas. El núcleo, definido operativamente como Canadá, Europa desarrollada y S&P 100, no reproduce, por sí solo, el desempeño del universo Global: el QWK queda por debajo del escenario base. La periferia sola eleva el nivel, con un N menor. Quitar únicamente el S&P 100 deja el QWK cerca del Global. Sobre el núcleo, las regiones añadidas no son intercambiables: Asia sube el nivel; LatAm lo baja; MEA queda intermedio. GraphSAGE y el clásico del conjunto de prueba se desplazan en el mismo sentido.

Esa evidencia debe leerse como sensibilidad integral del problema predictivo. Filtrar instrumentos reconstruye simultáneamente el conjunto de nodos, la red, el *strength*, el ranking relativo y los cuartiles del target. Por ello no corresponde afirmar que una región “favorece” o “perjudica” a un modelo, ni que el núcleo sea más importante porque concentre conectividad. La formulación adecuada es que el desempeño es condicional a la composición del universo, en línea con el carácter relativo de la importancia sistémica definida en el problema.

El contraste ΔQWK frente al clásico de cada recorte no organiza esta lectura (Holm no confirma superioridad; el único *supera* por IC es la periferia, con $N = 250$). La pregunta cuaternaria se responde afirmando sensibilidad geográfica del *nivel* de predicción, sin convertir escenarios particulares en mecanismos regionales identificados.

9.5 Del clasificador al score ordinal

La clasificación ordinal no agota el uso de la señal. El score continuo construido a partir de las probabilidades de clase de GraphSAGE (promedio de probabilidades entre semillas, especificación de 24 variables) ordena de forma coherente la conectividad relativa futura: la asociación con el percentil de *strength* y la monotonidad por quintiles respaldan que las probabilidades contienen información ordinal más fina que la sola clase de mayor probabilidad. Esa extensión es coherente con tratar el target como estructura ordenada, la misma lógica que motiva QWK como métrica primaria (Cohen, 1968).

El score no debe interpretarse como predicción puntual del *strength*, ni como pronóstico de retorno o de precio. Su función es convertir la incertidumbre de clase en una señal continua de importancia relativa futura, utilizable como entrada de portafolio. En ese sentido, responde a la pregunta quinta en su tramo predictivo: la salida probabilística aporta ordenación útil más allá de la etiqueta discreta, sin reclamar calibración perfecta ni utilidad económica por sí sola.

9.6 Utilidad económica: Variante A (fuera de muestra)

La Variante A examina si esa señal modula la exposición sistémica relativa en un problema de mínima varianza penalizada, evaluado fuera de muestra (Markowitz, 1952; Ledoit y Wolf, 2004). Incorporar el score de GraphSAGE reduce la exposición sistémica relativa realizada respecto de la mínima varianza y también frente a referencias con *strength* contemporánea y con score tabular, con efectos de magnitud pequeña. La inferencia favorece la dirección de menor exposición sistémica relativa, sin deterioro material del Sharpe observado en el escenario reportado.

La distinción entre detectabilidad estadística e importancia económica es aquí central. Un

contraste puede ser detectable bajo el esquema de remuestreo adoptado y, al mismo tiempo, económicamente modesto. El número moderado de rebalances del conjunto de prueba limita además la potencia y la estabilidad de esa inferencia.

En síntesis, la señal permite modular marginalmente la exposición sistémica relativa sin alterar de forma material el perfil financiero observado en el test. No se afirma que la estrategia “bata” a Markowitz en retorno, ni que el efecto sea grande, ni que constituya una regla de trading superior.

9.7 Variante B como mapa retrospectivo

La Variante B asigna la misma noción de exposición sistémica relativa a un menú de portafolios en la frontera media–varianza, usando la muestra completa con finalidad diagnóstica (Markowitz, 1952). El ejercicio muestra que esa exposición varía a lo largo de la frontera y no coincide necesariamente con los extremos de retorno o volatilidad. En esa lectura, opera como una dimensión adicional para visualizar compromisos de conectividad relativa dentro de un menú ya construido.

Ese mapa no constituye evidencia fuera de muestra ni una estrategia implementable en tiempo real. No regenera el problema de optimización con información disponible hasta cada fecha, ni prueba rentabilidad futura. Su valor es conceptual: ilustrar que la señal de importancia relativa puede ordenar exposiciones sobre el espacio media–varianza sin confundirse con una evaluación fuera de muestra. La pregunta quinta, en su tramo económico, queda así respondida de forma acotada: utilidad fuera de muestra pequeña en A; mapa ilustrativo en B.

9.8 Relación con la evidencia previa y aporte académico

La literatura de redes y conectividad ha mostrado que la interdependencia de mercado puede caracterizarse sin un mapa contractual completo (Billio et al., 2012; Giudici et al., 2020), y que las GNN ofrecen un marco para tareas sobre grafos con agregaciones aprendidas (Wu et al., 2021). En la línea específica de contrastes entre GNN y modelos tabulares para clasificar importancia sistémica, Balmaseda et al. (2023) trabajan entidades bancarias, redes de exposiciones y etiquetas basadas en DebtRank, con evaluación predominantemente contemporánea o transductiva y MCC como métrica principal. El presente estudio se sitúa en otro dominio: instrumentos accionarios, redes dinámicas de correlación, predicción temporal fuera de muestra de una clase relativa de *strength* futuro, QWK como métrica primaria y extensión a score y portafolio.

Esas diferencias impiden una comparación numérica directa de márgenes. Lo que sí permiten es contrastar la lógica de investigación. Se comparte la idea de discretizar importancia en clases y de enfrentar arquitecturas relacionales a referencias tabulares (Balmaseda et al., 2023). Se diverge en el proxy, en el carácter futuro del target y en el protocolo fuera de muestra. Respecto de la expectativa de que el canal relacional aporte ventaja automática, los hallazgos propios matizan esa intuición: bajo paridad tabular y red densa, GraphSAGE es competitivo, no dominante, y la familia GNN no se comporta de manera uniforme.

El aporte académico no consiste en demostrar que “GraphSAGE gana”. Consiste en evaluar, con un protocolo temporal reproducible, la capacidad de modelos relacionales para anticipar importancia sistémica futura relativa en un universo accionario global; en documentar límites y sensibilidad de esa competitividad; en examinar parsimonia informativa entre amplitudes tabulares; y en estudiar la traducción de la señal desde clasificación ordinal hacia una aplicación financiera



de alcance acotado.

10 Limitaciones

Los resultados deben leerse a la luz de restricciones de diseño que acotan su alcance. Este capítulo no reitera el marco conceptual ni la interpretación de los hallazgos: concentra qué mide el proxy, qué lecturas quedan bloqueadas y hasta dónde cabe generalizar.

10.1 Qué mide el proxy

Como se formalizó en el marco teórico, la red se construye con correlaciones absolutas de retornos: resume comovimiento estadístico, no dirección de contagio contractual ni un mecanismo causal de transmisión. El target son cuartiles del *strength* futuro relativos al conjunto de instrumentos en cada ventana. La etiqueta mide posición estructural relativa, no riesgo sistémico absoluto, probabilidad de crisis, DebtRank, CoVaR, SRISK ni pérdidas sistémicas. En una red densa el *strength* puede absorber factores comunes de mercado, y los cuartiles fuerzan clases altas y bajas aunque las diferencias absolutas de conectividad sean pequeñas. Los resultados hablan, por tanto, de predictibilidad de conectividad relativa futura dentro de ese objeto.

La ablación geográfica filtra instrumentos y reconstruye a la vez nodos, grafo, ranking de *strength* y cuartiles. Mide la sensibilidad integral del problema predictivo a la composición del universo.

10.2 Qué no permite afirmar

El contraste bajo paridad tabular evalúa arquitecturas completas: clásicos y GNN reciben el mismo vector de nodo; las GNN procesan además la adyacencia. Esa comparación no aísla el efecto causal exclusivo del grafo. En el protocolo denso, GCN y GraphSAGE usan intensidades ponderadas, mientras que GAT recibe una máscara de conectividad prácticamente completa: su de-

sempañ no demuestra el aprovechamiento de pesos continuos de correlación, punto ya señalado en Discusión. La selección de hiperparámetros de GAT se evalúa con una sola semilla, a diferencia de la media de tres en GCN y GraphSAGE; el contraste final usa *hard vote* sobre 42/43/44, pero el presupuesto de búsqueda no es simétrico. Tres semillas aportan robustez limitada frente a la variabilidad del entrenamiento estocástico. La configuración consolidada se fijó tras la línea empírica y no se presenta como preregistro formal.

Tampoco corresponde inferir un efecto causal de regiones específicas: al reconstruir red y etiquetas relativas, la ablación no identifica mecanismos geográficos. La categoría operacional *empata*, ya precisada en Discusión, indica ausencia de evidencia suficiente de superioridad cuando el intervalo contiene el cero; no es una prueba de equivalencia. Holm se aplica solo dentro de las tres GNN de cada escenario ($m = 3$), no como corrección global entre bloques, amplitudes o casos de ablación. Los cortes de horizonte calendario son exploratorios. La sensibilidad del horizonte del target $H \in \{63, 84, 126\}$ se documentó bajo la especificación preliminar de 26 variables y es evidencia histórica, no réplica del Bloque 2 de 24 variables. No deben mezclarse en un mismo contraste primario las especificaciones de 6, 24 y 47 variables ni las preliminares de 26 o 49.

El score ordinal $s = (3p_0 + 2p_1 + p_2)/3$ se obtiene de GraphSAGE con *soft vote* sobre 24 variables, con un score MLP de referencia bajo las mismas columnas. Ese objeto es distinto del *hard vote* de los contrastes de clasificación: las inferencias del clasificador no se trasladan automáticamente. El score ordena importancia relativa predicha; no es el *strength* futuro continuo ni un pronóstico de retorno. Adoptar 24 variables como base operativa del score y de los portafolios se apoya en un contraste descriptivo entre amplitudes, no en una validación independiente de esa selección.

En la Variante A, la optimización penaliza $ESR^{\text{pred}} = w^\top s$ y la evaluación primaria fuera

de muestra usa ESR^{real} (percentiles de *strength* futuro realizados). ESR no es Expected Shortfall financiero (CVaR). El número moderado de rebalances en el conjunto de prueba (del orden de 20) limita potencia y estabilidad de la inferencia de ΔESR^{real} . El efecto es detectable bajo el esquema adoptado, de magnitud pequeña y condicional a supuestos operativos (long-only, $w_{\text{máx}} = 1\%$, covarianza Ledoit–Wolf, costos de 10 pb, grilla de λ). La Variante B estima la frontera con la muestra completa y es un mapa retrospectivo: no entrega desempeño implementable fuera de muestra. El oráculo (P5) es diagnóstico y no implementable.

10.3 Hasta dónde se puede generalizar

El Universo Global 2015 es un panel fijo de 598 instrumentos listados en Estados Unidos, seleccionado con criterios de historia, cobertura y liquidez. Esa construcción facilita grafos temporales comparables, pero puede introducir sesgo de supervivencia o de selección: los resultados describen el conjunto elegible bajo esos filtros, no necesariamente toda la población de acciones en cada fecha histórica. La diversidad geográfica amplía la cobertura respecto de un mercado doméstico, sin representar de forma exhaustiva el mercado accionario mundial.

Las series proceden de Yahoo Finance mediante `yfinance`, con precios ajustados y filtros de cobertura. El estudio depende de una fuente secundaria de mercado; revisiones o diferencias de clasificación del proveedor pueden introducir variación respecto de bases institucionales alternativas, sin afirmar error sistemático de los datos.

El diseño usa ventanas de $W = 126$ días hábiles, paso $\Delta = 21$ y horizonte del target $H = 126$. El solapamiento induce dependencia temporal entre observaciones consecutivas. El embargo entre particiones y el moving-block bootstrap mitigan parte de ese riesgo, pero no transforman las ventanas en observaciones independientes. El conjunto de prueba contiene 20 fechas de ventana,

lo que limita la potencia, la estabilidad de los intervalos y el análisis de subperíodos o regímenes por separado. Los resultados son además condicionales a esa frecuencia diaria y a ese horizonte estructural.

En conjunto, los hallazgos valen dentro del problema, del Universo Global 2015, de la representación densa de correlaciones y del protocolo temporal evaluados. Extenderlos a otros mercados, frecuencias, construcciones de red, proxies de importancia sistémica o regímenes históricos exige replicación adicional. Esas restricciones condicionan la interpretación causal, la validez externa y la magnitud de la evidencia económica; no impiden responder las preguntas predictivas planteadas dentro de este marco.

11 Conclusiones

Esta memoria evaluó, bajo un protocolo temporal fuera de muestra, el desempeño predictivo de Graph Neural Networks frente a modelos clásicos de Machine Learning en la clasificación de la importancia sistémica futura relativa de instrumentos del Universo Global 2015. La evidencia no respalda una superioridad general de las GNN. GraphSAGE presenta un desempeño competitivo frente a los clásicos de referencia en las amplitudes de 6, 24 y 47 variables, sin que los intervalos de confianza ni el ajuste de Holm permitan establecer superioridad. Esa lectura de *empate* no constituye una prueba de equivalencia estadística; el detalle queda en Discusión. GCN y GAT no reproducen de manera consistente ese nivel de competitividad. La respuesta al objetivo general es, por tanto, de competitividad condicionada de GraphSAGE, no de dominio de la familia de grafos.

Esa lectura exige distinguir arquitecturas. Compartir la etiqueta “GNN” no garantiza un mismo comportamiento empírico: el resultado depende de cómo cada modelo procesa la estructura de la red. Bajo paridad tabular, clásicos y GNN reciben las mismas variables de nodo y las GNN procesan además la adyacencia. El patrón observado respalda un desempeño competitivo de GraphSAGE en ese diseño comparativo, sin permitir atribuir causalmente el desempeño al grafo de manera aislada.

En el plano informativo, ampliar de 24 a 47 variables no entrega, en el protocolo reportado, un valor incremental suficiente para GraphSAGE frente a la especificación más compacta. La diferencia es descriptiva entre amplitudes y motiva adoptar las 24 variables como especificación descriptiva de referencia y base operativa del score y de los portafolios. En el plano transversal, el nivel de QWK es sensible a la composición geográfica: el núcleo solo no replica el universo Global, y la periferia y las regiones añadidas al núcleo mueven el desempeño en direcciones distintas.

Esa sensibilidad no autoriza conclusiones causales sobre regiones particulares.

Las probabilidades de clasificación contienen información ordinal más rica que la sola clase de mayor probabilidad. El score continuo derivado de GraphSAGE bajo 24 variables ordena de forma informativa la importancia sistémica relativa futura y habilita el puente desde la clasificación hacia una señal utilizable en decisión, sin confundirse con un pronóstico de retorno o de precio.

En la Variante A, incorporar esa señal a un problema de mínima varianza reduce de manera estadísticamente detectable la exposición sistémica relativa realizada, con magnitud pequeña y sin deterioro material del perfil financiero observado bajo el protocolo evaluado. La evidencia respalda una utilidad complementaria y acotada, no una mejora sustancial del desempeño financiero. La Variante B aporta una lectura distinta: como mapa retrospectivo, muestra que la exposición sistémica relativa puede operar como dimensión adicional sobre la frontera media–varianza, sin constituir evidencia fuera de muestra. En ambos casos, la exposición sistémica relativa no equivale al Expected Shortfall de pérdidas financieras.

El aporte del estudio no consiste en demostrar que GraphSAGE “gana”. Consiste en integrar, dentro de un protocolo temporal reproducible, la predicción ordinal de importancia sistémica futura relativa en un universo accionario global con redes dinámicas de correlación; en documentar que esa competitividad es condicional a la arquitectura, a la especificación de variables y a la composición del universo; y en trasladar la señal aprendida hacia una aplicación financiera de alcance acotado. En conjunto, la utilidad del aprendizaje relacional en este problema debe evaluarse de manera condicionada: GraphSAGE constituye una alternativa competitiva para anticipar importancia sistémica relativa y ofrece una señal que puede complementar criterios financieros tradicionales sin sustituirlos.

11.1 Trabajo futuro

Las líneas siguientes se anclan a las limitaciones ya señaladas.

- Pesos de atención que utilicen de forma explícita la intensidad de correlación absoluta, condicional a la máscara densa de GAT.
- Otros proxies de importancia sistémica y otros universos, con evaluación walk-forward, dado el solapamiento de ventanas y el tamaño del conjunto de prueba.
- Restricciones y costos de portafolio más realistas, ante la magnitud pequeña de ΔESR y el número moderado de rebalances.

Referencias

- Acemoglu D., Ozdaglar A., y Tahbaz-Salehi A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608. <https://doi.org/10.1257/aer.20130456>.
- Acharya V. V., Pedersen L. H., Philippon T., y Richardson M. (2017). Measuring systemic risk. *The Review of Financial Studies*, 30(1):2–47. <https://doi.org/10.1093/rfs/hhw088>.
- Adrian T. y Brunnermeier M. K. (2016). CoVaR. *American Economic Review*, 106(7):1705–1741. <https://doi.org/10.1257/aer.20120555>.
- Aroussi R. (2026). yfinance (Version 1.5.2) [Computer software]. PyPI. <https://pypi.org/project/yfinance/1.5.2/>.
- Balmaseda V., Coronado M., y de Cadenas-Santiago G. (2023). Predicting systemic risk in financial systems using deep graph learning. *Intelligent Systems with Applications*, 19:200240. <https://doi.org/10.1016/j.iswa.2023.200240>.
- Bardoscia M., Barucca P., Battiston S., Caccioli F., Cimini G., Garlaschelli D., Saracco F., Squartini T., y Caldarelli G. (2021). The physics of financial networks. *Nature Reviews Physics*, 3:490–507. <https://doi.org/10.1038/s42254-021-00322-5>.
- Bardoscia M., Battiston S., Caccioli F., y Caldarelli G. (2015). DebtRank: A microscopic foundation for shock propagation. *PLOS ONE*, 10(6):e0130406. <https://doi.org/10.1371/journal.pone.0130406>.
- Barrat A., Barthélemy M., Pastor-Satorras R., y Vespignani A. (2004). The architecture of complex

weighted networks. *Proceedings of the National Academy of Sciences*, 101(11):3747–3752.

<https://doi.org/10.1073/pnas.0400087101>.

Battiston S., Puliga M., Kaushik R., Tasca P., y Caldarelli G. (2012). DebtRank: Too central to fail? financial networks, the FED and systemic risk. *Scientific Reports*, 2:541. <https://doi.org/10.1038/srep00541>.

Billio M., Getmansky M., Lo A. W., y Pelizzon L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559. <https://doi.org/10.1016/j.jfineco.2011.12.010>.

Breiman L. (2001). Random forests. *Machine Learning*, 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>.

Brier G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3. [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2).

Brownlees C. y Engle R. F. (2017). SRISK: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies*, 30(1):48–79. <https://doi.org/10.1093/rfs/hhw060>.

Chen T. y Guestrin C. (2016). XGBoost: A scalable tree boosting system. En *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.

Chicco D. y Jurman G. (2020). The advantages of the Matthews correlation coefficient (MCC)

over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21:6. <https://doi.org/10.1186/s12864-019-6413-7>.

Cohen J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4):213–220. <https://doi.org/10.1037/h0026256>.

Efron B. y Tibshirani R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall.

Epstein E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, 8(6):985–987. [https://doi.org/10.1175/1520-0450\(1969\)008<0985:ASSFPF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1969)008<0985:ASSFPF>2.0.CO;2).

Giglio S., Kelly B., y Pruitt S. (2016). Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics*, 119(3):457–471. <https://doi.org/10.1016/j.jfineco.2016.01.010>.

Gilmer J., Schoenholz S. S., Riley P. F., Vinyals O., y Dahl G. E. (2017). Neural message passing for quantum chemistry. En *Proceedings of the 34th International Conference on Machine Learning*, volumen 70 de *Proceedings of Machine Learning Research*, pp. 1263–1272. PMLR. <https://proceedings.mlr.press/v70/gilmer17a.html>.

Giudici P., Sarlin P., y Spelta A. (2020). The interconnected nature of financial systems: Direct and common exposures. *Journal of Banking & Finance*, 112:105149. <https://doi.org/10.1016/j.jbankfin.2017.05.010>.

Gorodkin J. (2004). Comparing two K-category assignments by a K-category correlation coefficient.

cient. *Computational Biology and Chemistry*, 28(5–6):367–374. <https://doi.org/10.1016/j.compbiolchem.2004.09.006>.

Haldane A. G. (2009). Rethinking the financial network [Speech]. Bank of England. 28 de abril de 2009. <https://www.bankofengland.co.uk/speech/2009/rethinking-the-financial-network>.

Hamilton W. L., Ying Z., y Leskovec J. (2017). Inductive representation learning on large graphs. En *Advances in Neural Information Processing Systems (NeurIPS)*, volumen 30, pp. 1024–1034. <https://proceedings.neurips.cc/paper/2017/hash/5dd9db5e033da9c6fb5ba83c7a7e9bea9-Abstract.html>.

Holm S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Keilbar G. y Wang W. (2022). Modelling systemic risk using neural network quantile regression. *Empirical Economics*, 62(1):93–118. <https://doi.org/10.1007/s00181-021-02035-1>.

Kipf T. N. y Welling M. (2017). Semi-supervised classification with graph convolutional networks. En *International Conference on Learning Representations (ICLR)*. <https://openreview.net/forum?id=SJU4ayYgl>.

Kou G., Chao X., Peng Y., Alsaadi F. E., y Herrera-Viedma E. (2019). Machine learning methods for systemic risk analysis in financial sectors. *Technological and Economic Development of Economy*, 25(5):716–742. <https://doi.org/10.3846/tede.2019.8740>.

- Künsch H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241. <https://doi.org/10.1214/aos/1176347265>.
- Ledoit O. y Wolf M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411. [https://doi.org/10.1016/S0047-259X\(03\)00096-4](https://doi.org/10.1016/S0047-259X(03)00096-4).
- Mantegna R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B*, 11(1):193–197. <https://doi.org/10.1007/s100510050929>.
- Markowitz H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>.
- Matthews B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405(2):442–451. [https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9).
- Newman M. E. J. (2010). *Networks: An Introduction*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>.
- Politis D. N. y Romano J. P. (1992). A circular block-resampling procedure for stationary data. En LePage R. y Billard L., editores, *Exploring the Limits of Bootstrap*, pp. 263–270. John Wiley.
- Sharpe W. F. (1966). Mutual fund performance. *The Journal of Business*, 39(S1):119–138. <https://doi.org/10.1086/294846>.
- Tumminello M., Aste T., Di Matteo T., y Mantegna R. N. (2005). A tool for filtering information

in complex systems. *Proceedings of the National Academy of Sciences*, 102(30):10421–10426.

<https://doi.org/10.1073/pnas.0500298102>.

Veličković P., Cucurull G., Casanova A., Romero A., Liò P., y Bengio Y. (2018). Graph attention networks. En *International Conference on Learning Representations (ICLR)*. <https://openreview.net/forum?id=rJXMpikCZ>.

Wu Z., Pan S., Chen F., Long G., Zhang C., y Yu P. S. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24. <https://doi.org/10.1109/TNNLS.2020.2978386>.

Yahoo (s. f.). Yahoo Finance. Consultado el 10 de julio de 2026. <https://finance.yahoo.com/>.

Zou H. y Hastie T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.

12 Anexos

12.1 Puente descriptivo: especificación preliminar de 26 variables a 24 variables

Respecto de la especificación preliminar de 26 variables (ocho regiones), GraphSAGE baja de 0,642 a 0,629 en QWK del conjunto de prueba y el IC pasa de supera a empata; Holm no rechaza superioridad en ninguna de las dos especificaciones. La comparación es descriptiva entre especificaciones, no un contraste MBB entre bloques. El escenario base narrativo de Resultados es la especificación de 24 variables (región-v2).

12.2 Protocolo y artefactos

La línea experimental reportada en el cuerpo es B1 (6 variables) → B2 (24, región-v2) → B3 (47 variables). Los protocolos y manifiestos de cada corrida se mantienen en las carpetas de cada experimento. El detalle del diagnóstico lineal y de la ablación por grupos bajo las especificaciones históricas 26/49 consta en el Anexo 12.12.

- Bloque 1 y B2 (24 variables): `experiments/bloque1_global_2015/`
- Ablación geográfica a-f: `experiments/ablacion_regiones_b2/`
- Bloque 3 (47 variables): `experiments/bloque2_global_2015_meta/`

Especificaciones preliminares de 26 variables (ocho regiones) y 49 variables se conservan en carpetas de respaldo del repositorio y **no** organizan el eje del estudio.

12.3 Elegibilidad del Universo Global 2015

Tabla 15: Criterios operativos de elegibilidad del Universo Global 2015.

Criterio	Definición operativa (verificada)
Corte de elegibilidad	2015-01-01 (historia disponible desde esa fecha).
Historia mínima	Primera observación $\leq$ 2015-01-01.
Cobertura de precios	$\text{ffill causal} \leq 10$ sesiones; $\text{gap post-ffill} \leq 21$; cobertura ≥ 85 %.
Liquidez	≥ 70 % de sesiones con volumen > 0 .
Exclusiones	19 activos por liquidez; exclusión manual de TKC (Europa emergente).
Precio / ajustes	Panel diario con <i>Adj Close</i> (Yahoo Chart API).
Región / sector	Región: lista congelada del universo; sector: clasificación tipo GICS (fuente operativa documentada en el congelamiento).
n historia $\rightarrow$ final	618 (historia) $\rightarrow$ 599 (tras liquidez) $\rightarrow$ 598 (tras excluir TKC).

Elaboración propia a partir del congelamiento documentado del Universo Global 2015 y del diagnóstico de huecos 2015.

12.4 Ilustración de la red Global 2015 por región

Este anexo complementa la Figura 4. Sobre el **mismo layout** de esa ventana de prueba (2025-04-07; 598 nodos), se destaca una región y el resto del universo queda en gris. La esparsificación es solo visual (árbol de expansión mínima más el 1,5 % de aristas de mayor $|\text{corr}|$); el protocolo experimental permanece denso. La taxonomía es la de **ocho regiones** del congelamiento del universo, no la región-v2 del Bloque 2. El ejercicio es ilustrativo: no identifica un efecto causal de regiones ni sustituye la ablación geográfica de Resultados.

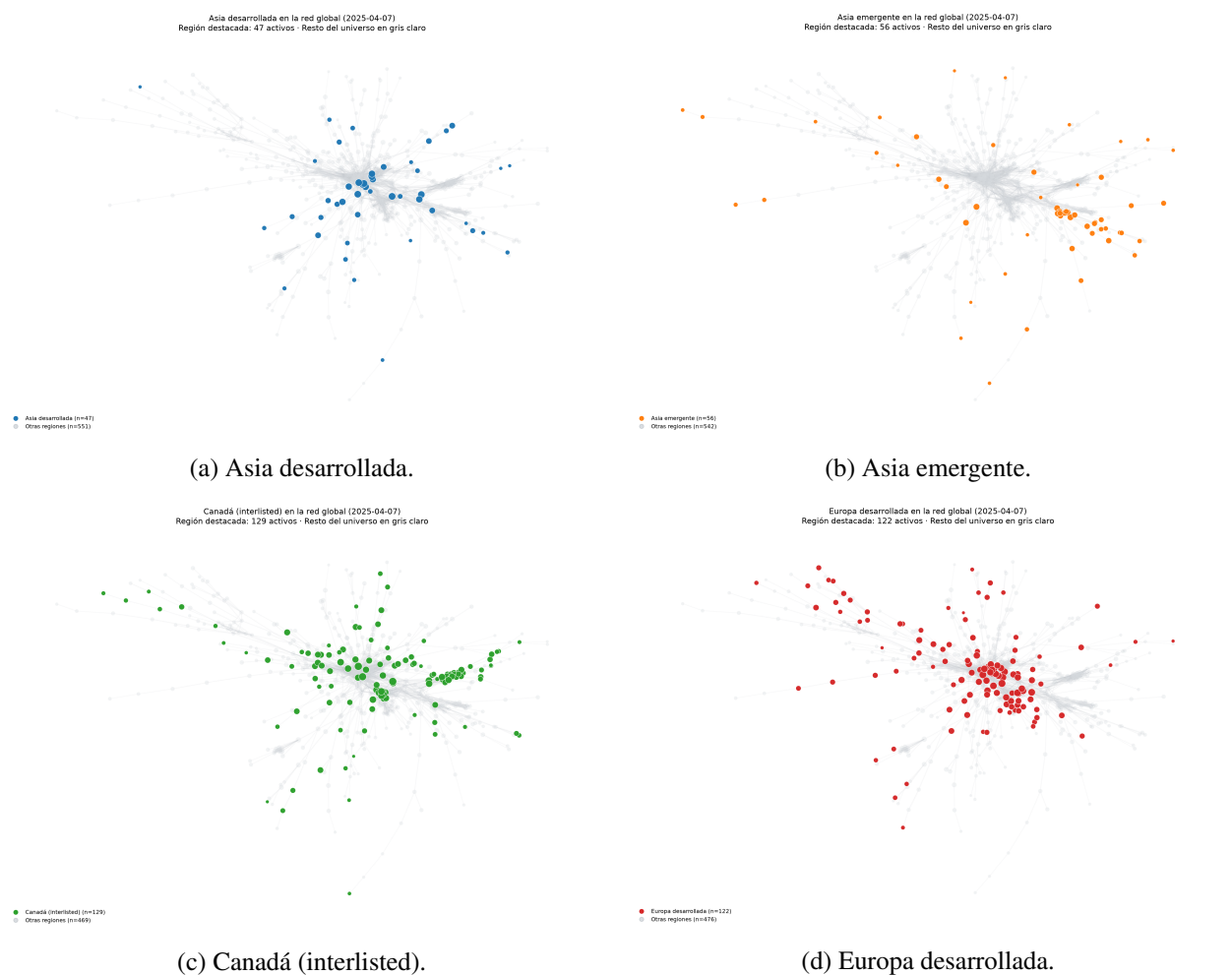


Figura 14: Red Global 2015 por región (I). Ventana de prueba 2025-04-07; layout idéntico al de Metodología. Región destacada en color; resto en gris.

Fuente: elaboración propia.

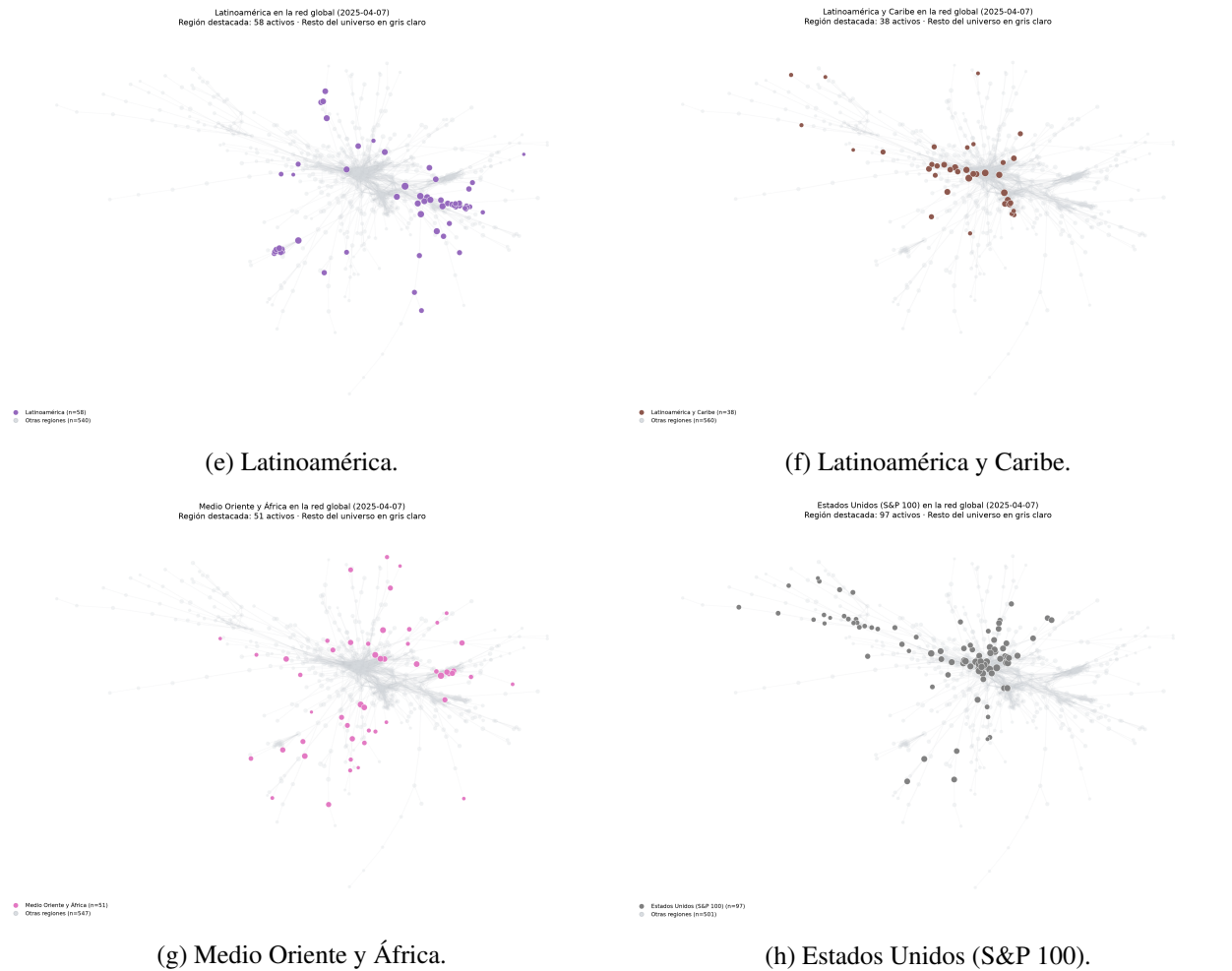


Figura 15: Red Global 2015 por región (II). Misma ventana y layout que la Figura 14.

Fuente: elaboración propia.

12.5 Diccionario de variables

12.5.1. Bloque 1 (seis individuales).

Tabla 16: Variables individuales (Bloque 1).

Nombre académico	Nombre técnico	Transformación
Retorno medio anualizado	<code>ret_medio_anual</code>	Escala (train)
Volatilidad anualizada	<code>vol_anual</code>	Escala (train)
Drawdown máximo	<code>drawdown_max</code>	Escala (train)
Volumen promedio (log)	<code>volumen_prom_log</code>	Escala (train)
Porcentaje de días sin transar	<code>pct_dias_sin_transar</code>	Escala (train)
Beta	<code>beta</code>	Escala (train)

12.5.2. Bloque 2 (región-v2 y sector; 24 columnas).

Seis dummies de región (`region__*`) y doce de sector (`sector__*`), estáticos por ticker (sin *leakage* temporal). Categorías de región-v2: Asia; Canadá (interlisted); Europa desarrollada; Latinoamérica; Medio Oriente y África; Estados Unidos (S&P 100). Fusión respecto de la taxonomía de ocho regiones: Asia desarrollada + Asia emergente → Asia; Latinoamérica + Latinoamérica y Caribe → Latinoamérica. La región es geográfica/listado (no GICS). El sector se homologa a once sectores tipo GICS más Otros (etiquetas de mercado del congelamiento). Categorías de sector: Communication Services, Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Real Estate, Utilities, Otros.

12.5.3. Bloque 3 (contexto y red tabular; 47 columnas).

Tabla 17: Variables de contexto (18) y de red tabular (5) añadidas sobre la base de 24 (total 47).

Nombre técnico	Grupo	Definición breve
ctx_*_ret, ctx_*_vol	contexto	Retorno y volatilidad de series de mercado (SPY, DX, HG, CL, ILF, VIX) en la ventana histórica.
corr_*	contexto	Correlación del activo con cada serie de mercado en la ventana histórica.
strength_actual	red	<i>strength</i> del nodo en el grafo de la ventana histórica.
corr_promedio_actual	red	Correlación promedio del nodo.
eigenvector_actual	red	Centralidad de eigenvector.
pagerank_actual	red	PageRank.
clustering_actual	red	Coefficiente de clustering.

Fuente: manifiestos del conjunto ampliado (contexto + red). Total reportado: $24 + 18 + 5 = 47$ sobre región-v2.

12.6 Hiperparámetros seleccionados por validación

Se reportan los hiperparámetros **seleccionados** por QWK de validación (no la grilla completa). Las grillas exploradas constan en los artefactos de selección por bloque. Se separan por bloque para facilitar la lectura.

Tabla 18: Hiperparámetros seleccionados por QWK de validación. Bloque 1 (6 variables).

Modelo	Seleccionados (val)
MLP (ref.)	$\text{hidden}=(32), \alpha = 10^{-4}$
RF / XGB / LR / k -NN	RF: $n = 100$, $\text{depth} = 10$, $\text{leaf} = 5$; XGB: $n = 400$, $\text{depth} = 3$, $\text{lr} = 0,1$; LR: $C = 1$; k -NN: $k = 50$
GraphSAGE / GCN / GAT	$\text{hidden} = 32$, $\text{dropout} = 0$, $\text{lr} = 0,01$

Tabla 19: Hiperparámetros seleccionados por QWK de validación. Bloque 2 (24 variables).

Modelo	Seleccionados (val)
MLP (ref. región-v2)	$\text{hidden}=[32], \alpha = 10^{-3}$
GraphSAGE / GAT / GCN	GS: $\text{hidden} = 16$, $\text{dropout} = 0$, $\text{lr} = 0,01$; GAT: $\text{hidden} = 16$; GCN: $\text{hidden} = 32$

Tabla 20: Hiperparámetros seleccionados por QWK de validación. Bloque 3 (47 variables).

Modelo	Seleccionados (val)
XGBoost (ref.)	$n = 200$, depth= 3, lr= 0,05, subsample= 0,9, colsample= 0,9
GraphSAGE	hidden= 32, dropout= 0,3, lr= 0,003
GAT / GCN	GAT: hidden= 16, dropout= 0, lr= 0,003; GCN: hidden= 32, dropout= 0,3, lr= 0,01

Fuentes: artefactos de selección de B1 (6), B2 (24, región-v2) y B3 (47).

12.6.1. Selecciones históricas (no organizan el eje del estudio).

Para trazabilidad se conservan, fuera de la tabla principal, las selecciones bajo la taxonomía de ocho regiones: especificación de 26 variables (RF de referencia; GNN con hidden= 16) y de 49 variables (XGBoost de referencia; GraphSAGE hidden= 16, dropout= 0,3, lr= 0,003). Esas filas no sustituyen a B2 (24) ni a B3 (47).

12.7 Glosario de artefactos GNN

- **Voto duro:** moda de la clase predicha entre semillas; artefacto de contrastes de clasificación.
- **Soft vote:** promedio de probabilidades entre semillas; usado para construir el score ordinal (s) hacia portafolio.
- **Media $\pm$ std:** resumen por semilla; no sustituye al *hard vote* en la misma fila de un contraste principal.
- **Lectura IC:** *supera/empata/pierde* según si el IC95 % de Δ QWK (MBB) queda sobre, contiene o bajo cero; Holm ($m = 3$) se reporta aparte.

12.8 Corrida archivada sin red tabular (histórico; no comparable con 6/24/47)

Existe una corrida tabular archivada sin métricas de red contemporáneas (`reports_sin_red/`). **No** organiza la línea reportada en el cuerpo ($6 \rightarrow 24 \rightarrow 47$).

12.9 Figuras complementarias de portafolio

Las dos figuras siguientes complementan las Variantes A y B del capítulo de Resultados. No alteran los contrastes de prueba ni sustituyen la figura de *ESR* en el tiempo ni el mapa de la frontera.

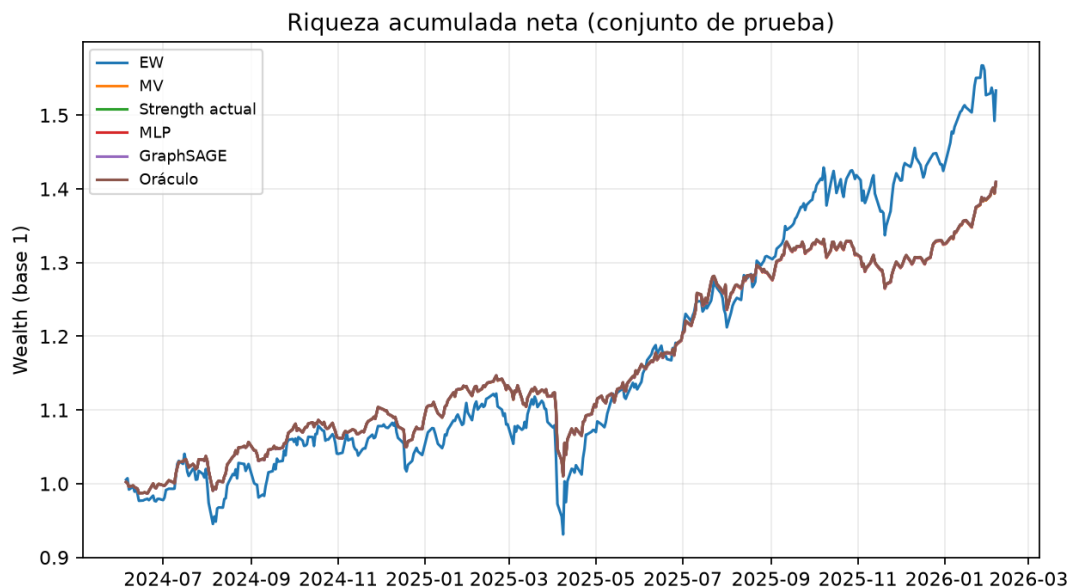


Figura 16: Variante A. Riqueza acumulada en el conjunto de prueba (complemento).

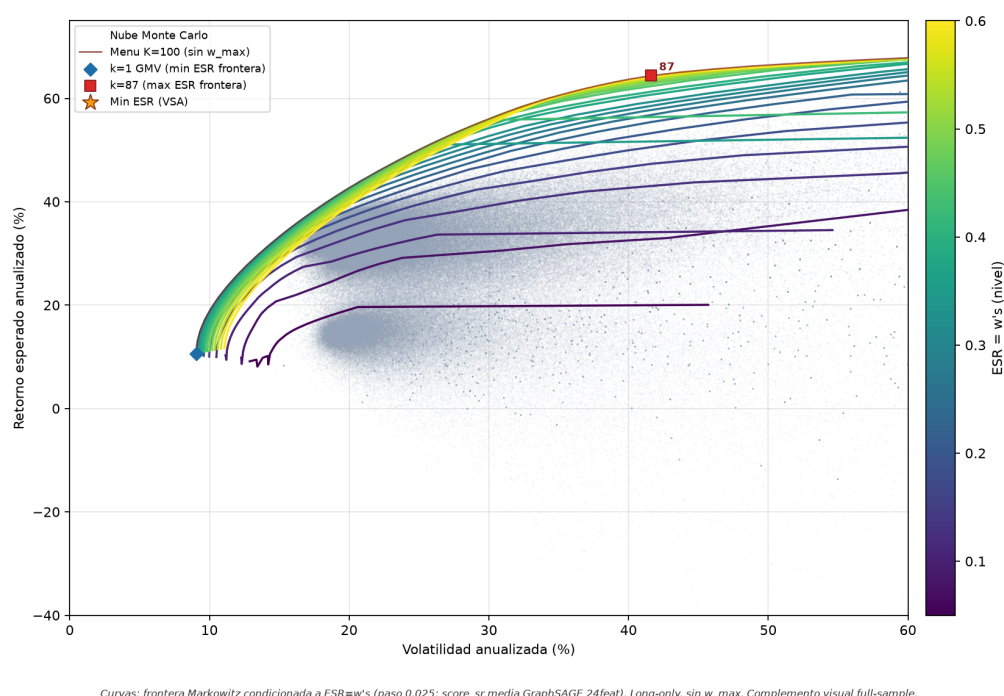


Figura 17: Variante B. Detalle (zoom) de curvas de nivel de ESR (complemento).

12.10 Inferencia extendida

Los contrastes detallados (IC, p_{Holm} , sensibilidades $L \in \{12, 18\}$) constan en los informes de cada bloque. El cuerpo reporta el contraste primario por bloque.

12.11 Notas sobre material histórico

Versiones anteriores de la redacción privilegiaban universos satélite y MCC como métrica primaria. Ese material y sus artefactos numéricos se conservan en el repositorio por trazabilidad académica, pero **no** organizan el eje del estudio (Universo Global 2015; QWK; línea 6/24/47).

12.12 Diagnóstico lineal y ablación históricos (26 y 49 variables; no comparable con 24/47)

Las cifras de este anexo corresponden a la especificación **histórica** de 26 y 49 variables (taxonomía de ocho regiones) y se conservan solo como respaldo metodológico. El eje del cuerpo

es 24 y 47 variables (región-v2). Este material **no** sustituye el contraste descriptivo de parsimonia GraphSAGE 24 frente a 47 reportado en Resultados.

Este anexo complementa la lectura de valor incremental. El logit multinomial con elastic-net (Zou y Hastie 2005) mide **señal lineal** bajo regularización; no identifica el aporte causal del grafo ni sustituye los contrastes de GraphSAGE. El ranking por variable usa la suma de $|\text{coef}|$ en las cuatro clases; el share es la fracción sobre la importancia total.

Tabla 21: Diagnóstico lineal histórico (logit elastic-net). QWK en validation y test (26/49 variables).

Especificación	n	QWK val	QWK test	log-loss test
26 variables	26	0,627	0,577	1,145
49 variables	49	0,680	0,603	1,127

En ese modelo histórico, ampliar de 26 a 49 columnas mejora el QWK del conjunto de prueba de 0,577 a 0,603 (Tabla 21).

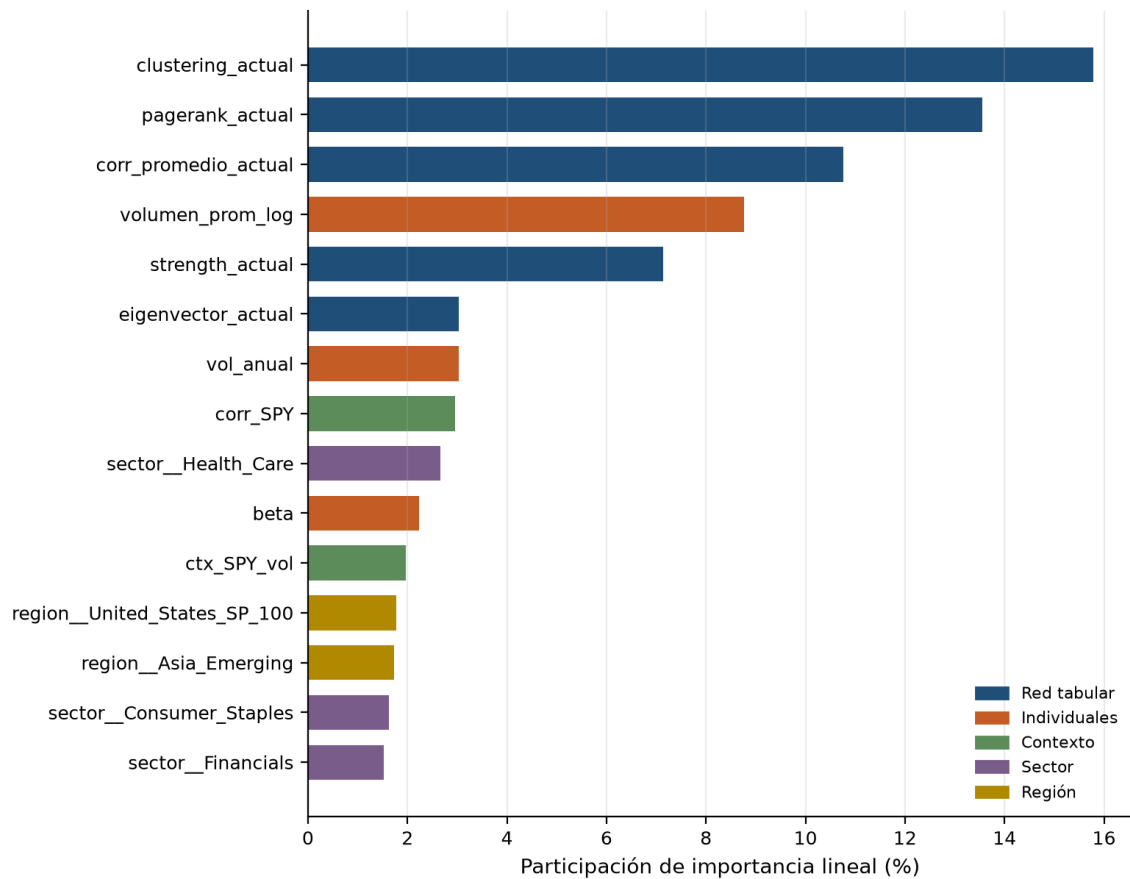


Figura 18: Quince variables con mayor participación de importancia lineal en el diagnóstico elastic-net histórico (49 variables).

Tabla 22: Ranking histórico de importancia lineal (elastic-net), 49 variables (filas 1–25). No comparable con 24/47.

Rank	Variable	Grupo	Importancia	Share
1	clustering_actual	red	3,83	15,8 %
2	pagerank_actual	red	3,29	13,6 %
3	corr_promedio_actual	red	2,61	10,8 %
4	volumen_prom_log	individuales	2,12	8,8 %
5	strength_actual	red	1,73	7,1 %
6	eigenvector_actual	red	0,73	3,0 %
7	vol_anual	individuales	0,73	3,0 %
8	corr_SPY	contexto	0,72	3,0 %
9	sector__Health_Care	sector	0,65	2,7 %
10	beta	individuales	0,54	2,2 %
11	ctx_SPY_vol	contexto	0,48	2,0 %
12	region__United_States_SP_100	región	0,43	1,8 %
13	region__Asia_Emerging	región	0,42	1,7 %
14	sector__Consumer_Staples	sector	0,39	1,6 %
15	sector__Financials	sector	0,37	1,5 %
16	sector__Utilities	sector	0,33	1,3 %
17	ret_medio_anual	individuales	0,31	1,3 %
18	region__Europe_Developed	región	0,29	1,2 %
19	corr_HG_F	contexto	0,29	1,2 %
20	corr_ILF	contexto	0,28	1,2 %
21	region__Latin_America	región	0,26	1,1 %
22	sector__Energy	sector	0,26	1,1 %
23	drawdown_max	individuales	0,22	0,9 %
24	sector__Communication_Services	sector	0,21	0,9 %
25	corr_CL_F	contexto	0,21	0,9 %

Tabla 23: Ranking histórico de importancia lineal (elastic-net), 49 variables (filas 26–37). No comparable con 24/47.

Rank	Variable	Grupo	Importancia	Share
26	ctx_HG_F_vol	contexto	0,21	0,9 %
27	sector__Materials	sector	0,21	0,9 %
28	sector__Otros	sector	0,20	0,8 %
29	ctx_HG_F_ret	contexto	0,20	0,8 %
30	ctx_ILF_ret	contexto	0,20	0,8 %
31	corr_VIX	contexto	0,18	0,7 %
32	region__Middle_East_and_Africa	región	0,14	0,6 %
33	corr_DX_Y_NYB	contexto	0,13	0,5 %
34	region__Latin_America_and_Caribbean	región	0,13	0,5 %
35	pct_dias_sin_transar	individuales	0,11	0,5 %
36	ctx_CL_F_vol	contexto	0,10	0,4 %
37	sector__Industrials	sector	0,09	0,4 %

Tabla 24: Ranking histórico de importancia lineal (elastic-net), 49 variables (filas 38–49). No comparable con 24/47.

Rank	Variable	Grupo	Importancia	Share
38	ctx_DX_Y_NYB_vol	contexto	0,08	0,3 %
39	ctx_SPY_ret	contexto	0,08	0,3 %
40	region__Asia_Developed	región	0,08	0,3 %
41	ctx_VIX_ret	contexto	0,07	0,3 %
42	ctx_ILF_vol	contexto	0,07	0,3 %
43	region__Canada_interlisted	región	0,07	0,3 %
44	sector__Real_Estate	sector	0,05	0,2 %
45	ctx_CL_F_ret	contexto	0,04	0,2 %
46	sector__Information_Technology	sector	0,04	0,2 %
47	ctx_DX_Y_NYB_ret	contexto	0,03	0,1 %
48	ctx_VIX_vol	contexto	0,03	0,1 %
49	sector__Consumer_Discretionary	sector	0,01	0,1 %

El ranking con 26 variables (sin contexto ni red) se omite por brevedad; en ese contraste histórico dominan las individuales (cerca del 57 % del share). Las figuras de ablación del cuerpo (barras y cambios de QWK) se interpretan solo de forma descriptiva.