Journal of Sustainability, Policy, and Practice

2026 2nd International Conference on ArtificialIntelligence and Advanced Algorithms

Article

Comparative Evaluation of Graph Neural Networks for Cross-Market Risk Contagion Path Identification in Multi-Layer Financial Networks

Yifei Li [1] and Xuanyi Fu [2,*]

[1] Master of Science in Enterprise Risk Management, Columbia University, New York, NY, USA
[2] M.S.E. in Computer Science, Johns Hopkins University, Baltimore, MD, USA
* Correspondence: Xuanyi Fu, Master of Science in Enterprise Risk Management, Columbia University, New York, NY, USA; M.S.E. in Computer Science, Johns Hopkins University, Baltimore, MD, USA

Abstract: Cross-market risk contagion---the propagation of financial distress across asset classes such as equities, bonds, derivatives, and foreign exchange---poses a persistent challenge for systemic risk monitoring. While graph neural networks (GNNs) have demonstrated strong performance in single-market financial risk tasks, their effectiveness for identifying contagion paths across interconnected multi-layer financial networks remains insufficiently explored. This paper presents a comparative empirical evaluation of four GNN architectures---GCN, GAT, GraphSAGE, and EvolveGCN---alongside a spatial-temporal graph convolutional baseline (ASTGCN) and three traditional econometric methods, applied to multi-layer financial networks constructed from U.S. equity, bond, derivatives, and foreign exchange market data spanning January 2018 to December 2023. The evaluation covers contagion path identification accuracy against realized stress episodes, node vulnerability ranking precision, the incremental benefit of multi-layer network construction over single-layer alternatives, sensitivity to edge-density thresholds, and a feature-matched comparison that isolates the contribution of graph-based nonlinear aggregation from that of richer input features. Results indicate that temporal-aware GNN architectures achieve moderate yet consistent gains in contagion detection, with ASTGCN attaining the highest AUC-ROC of 0.841 (95% CI: [0.823, 0.858]) on multi-layer networks. Multi-layer construction yields a 7.9 percentage-point improvement in AUC-ROC over the best single-layer configuration (95% CI: [5.4, 10.3] pp, $p < 0.001$). Macro-financial factors---particularly the VIX and credit spreads---emerge as the most informative features for vulnerability scoring. A case study on the March 2020 COVID-19 market crash validates the practical relevance of the identified contagion paths.

Keywords: graph neural networks; cross-market risk contagion; multi-layer financial networks; systemic risk

Received: 12 March 2026
Revised: 19 April 2026
Accepted: 30 April 2026
Published: 06 May 2026

1. Introduction

1.1. Background and Motivation

The interconnected nature of modern financial markets implies that distress originating in one asset class can rapidly transmit to others, amplifying losses beyond what single-market risk metrics anticipate. The March 2020 COVID-19 crash illustrated this dynamic vividly: equity sell-offs triggered margin calls in derivatives markets, forced liquidation of corporate bond holdings, and induced sharp dislocations in foreign exchange markets within a span of days. The 2022 rate-hike cycle and the March 2023 regional banking turmoil reinforced this pattern through distinct transmission channels. Traditional risk monitoring approaches---based on pairwise correlations or value-at-risk

measures computed within isolated asset classes---often fail to capture such cross-market cascading dynamics until losses have already materialized.

Foundational work in network economics established that financial interconnections exhibit a "robust-yet-fragile" property, where dense linkages diversify risk under ordinary conditions yet amplify contagion under severe shocks [1]. This theoretical insight motivates a network-based approach to cross-market risk analysis. Representing financial entities and their interdependencies as graph structures enables explicit modeling of how distress propagates through chains of connected exposures---a capability particularly relevant for central counterparty clearing organizations that monitor multi-asset counterparty credit risk.

Recent advances in graph neural networks have demonstrated the capacity of these architectures to capture complex relational patterns in financial data. GNN-based approaches have been applied to guarantee-loan contagion rating and cross-market volatility forecasting using spatial-temporal graph convolutions [2, 3]. Yet the majority of existing GNN applications in finance target single-market tasks---credit risk scoring within loan networks, stock prediction within equity markets---rather than the cross-asset contagion problem. Empirical studies examining multiplex financial network structures have confirmed that multi-layer representations reveal systemically important entities invisible to single-layer aggregations, yet the comparative effectiveness of different GNN architectures on such multi-layer structures has not been systematically evaluated [4].

1.2. Scope and Contributions

1.2.1. Research Questions

This study addresses three specific research questions. How do different GNN architectures compare when applied to contagion path identification in multi-layer financial networks spanning equities, bonds, derivatives, and foreign exchange? What incremental benefit does multi-layer network construction provide relative to single-layer alternatives, and is this benefit statistically significant after accounting for sampling variability? Which macro-financial and industry-level factors are most informative for identifying vulnerable nodes under extreme market conditions, and to what extent does the GNN performance advantage depend on access to such auxiliary features beyond price data?

1.2.2. Paper Organization

The remainder of this paper is organized as follows. Section 2 reviews related work on GNNs for financial risk and multi-layer network analysis. Section 3 details the experimental setup, including data sources, network construction, compared methods, and evaluation protocol. Section 4 presents results across multiple evaluation dimensions, including bootstrap confidence intervals for key comparisons, a feature-matched ablation that controls for input-feature asymmetry, and a threshold sensitivity analysis. Section 5 discusses practical implications, limitations, and future research directions.

2. Related Work

2.1. Graph Neural Networks for Financial Risk

2.1.1. GNN-Based Credit and Contagion Risk Analysis

The application of GNNs to financial risk analysis has expanded rapidly since 2019. Early work introduced high-order graph attention representations for risk assessment in guarantee-loan networks, achieving near-linear computational complexity through binary vertex role definitions [5]. Supply-chain-based approaches addressed the chronic data scarcity problem facing small and medium enterprises by mining graph-structured relationships from enterprise interaction data, demonstrating that spatial-temporal GNNs capture risk propagation across 28 million linked entities [6]. Extending this line of research, deep graph learning was applied to detect and isolate contagion risk in China's nationwide networked-loan market, where isolating the top 1% of predicted high-risk nodes reduced default rates by 25.8% [7]. Hierarchical GNN architectures operating at

both intra-community and inter-community levels introduced the concept of "tribe-style" company graphs, enriching attribute-sparse nodes with heterogeneous financial news embeddings [8].

2.1.2. Cross-Market Spillover with Graph Methods

Attribute-driven graph attention networks demonstrated that dynamically inferred firm relations outperform static predefined graphs for modeling momentum spillover effects on S&P 500 equities [9]. The GNNHAR approach integrated the heterogeneous autoregressive econometric specification with GNN message-passing to model multi-hop nonlinear volatility spillover effects, bridging financial econometrics with graph-based deep learning [10]. At the macro-financial level, copula-based contrastive predictive coding coupled hierarchical indicators from macroeconomic, sector, and micro-level data streams, establishing a methodology for multi-scale factor integration that informs the feature engineering approach adopted in the present study [11].

2.2. Multi-Layer Network Analysis in Finance

Multi-layer (multiplex) network representations have gained traction as tools for capturing the heterogeneous nature of financial interconnections. A graph-based credit risk approach constructed corporate networks from SEC filings using NLP-derived similarity on management discussion sections, achieving an F1 score of 0.804 and AUC of 0.864 on credit rating prediction tasks [12]. In the broader financial forecasting literature, GNN applications to realized volatility employed co-movement-based network construction to capture asset interdependencies [13]. These studies collectively suggest that the network topology itself carries predictive information about risk propagation---yet most efforts have focused on single-layer networks within a single asset class, leaving the multi-layer cross-market setting largely unaddressed.

2.3. Macro Factor Integration and Temporal Dynamics

Temporal dynamics are critical for financial contagion analysis, as risk transmission patterns shift substantially between calm and stressed market regimes. Temporal and heterogeneous GNNs applied to financial time series generated dynamic company relation graphs for each trading day, using transformer encoders combined with heterogeneous graph attention for joint temporal-structural embeddings [14]. Relational temporal graph convolutional approaches demonstrated that heterogeneous relation types---including industry and supply chain edges---carry distinct risk signals whose joint modeling substantially outperforms single-relation graph methods [15]. These temporal-relational insights directly motivate the inclusion of time-varying macro factors and multi-relational edge types in the present experimental design.

3. Experimental Setup

3.1. Data Sources and Network Construction

The multi-layer financial network is constructed from four publicly available data sources covering the period from January 2018 to December 2023 (313 trading weeks). The equity layer uses daily adjusted closing prices for 487 constituents of the S&P 500 index (excluding 13 firms delisted during the study period), obtained from Yahoo Finance. The bond layer incorporates 156 fixed-income instruments including U.S. Treasury securities across maturities (3-month to 30-year) and investment-grade corporate bond indices segmented by sector, sourced from the Federal Reserve Economic Data (FRED) database. The derivatives layer comprises 203 options and futures contracts on equity indices, interest rates, and commodities, with daily settlement prices from the CME Group public data portal. The foreign exchange layer includes 28 major and emerging-market currency pairs with daily mid-rates from the European Central Bank Statistical Data Warehouse.

All raw price series are converted to weekly log-returns to reduce noise and align temporal granularity across asset classes. Missing observations (holidays, market closures) are handled by forward-filling to the nearest prior trading day. The resulting dataset contains 874 unique financial entities observed over 313 weekly periods.

3.2. Multi-Layer Network Architecture

This section describes the construction of the multi-layer network in three parts: intra-layer edge construction within each asset class, inter-layer linkages connecting related entities across classes, and the resulting aggregate network statistics. Figure 1 provides a schematic overview of the architecture, showing representative nodes from each asset class together with the three categories of inter-layer edges that connect them.

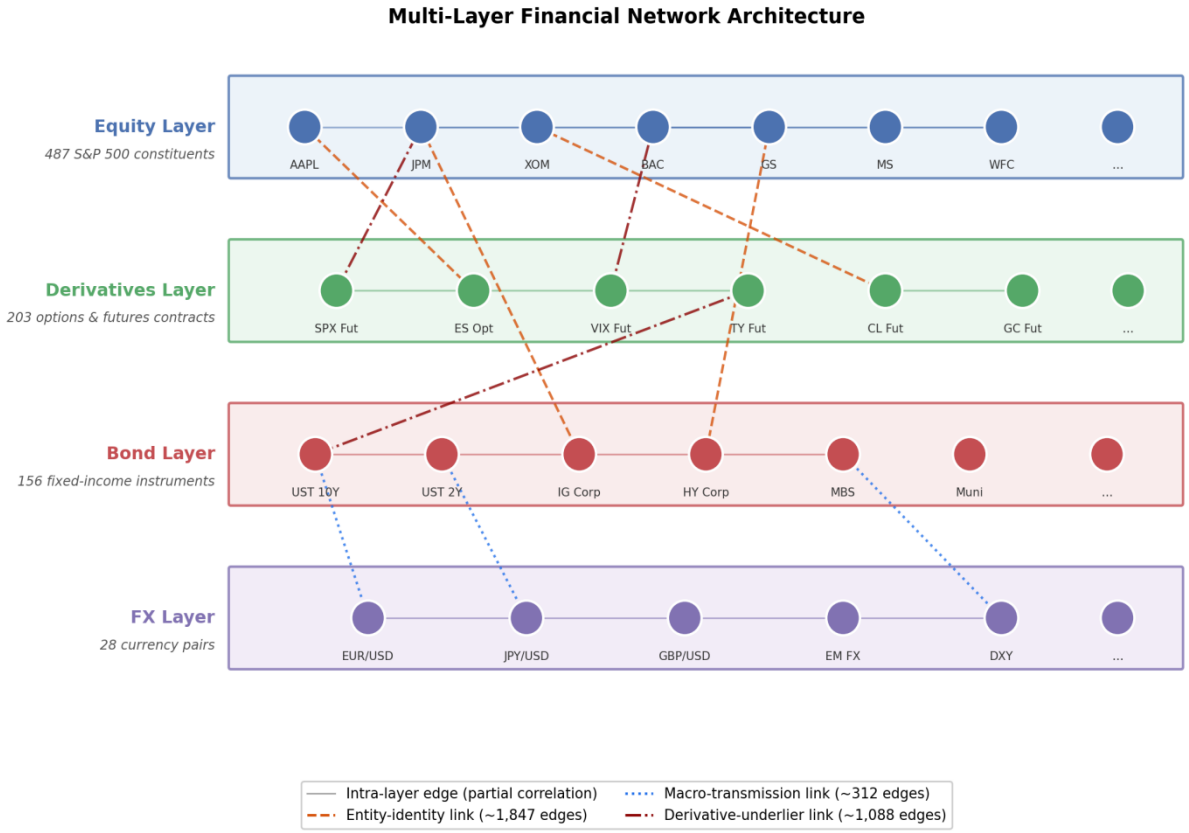

Figure 1. Multi-Layer Financial Network Architecture

Schematic illustration of the multi-layer financial network used in this study. Four asset-class layers (equity, derivatives, bond, FX) are connected by three categories of inter-layer edges: entity-identity links connecting the same or closely related entity across layers (dashed orange), macro-transmission links connecting sovereign bond nodes to related currency pairs (dotted blue), and derivative-underlier links connecting futures and options contracts to their reference equities or bonds (dash-dot dark red). Intra-layer edges within each asset class (thin solid) are constructed from dynamic partial correlations.

3.2.1. Intra-Layer Edge Construction

Within each asset-class layer, edges are established using dynamic partial correlations computed over rolling 52-week windows with 1-week step size. For each pair of entities within the same layer, the partial correlation coefficient is calculated conditional on a set of common market factors (the S&P 500 return, U.S. 10-year Treasury yield change, and the DXY dollar index return) to remove shared systematic components. An edge is retained if the absolute partial correlation exceeds a threshold calibrated at the 80th percentile of the empirical distribution within each layer and window; sensitivity of results to this threshold choice is examined in Section 4.2. This procedure yields time-varying intra-layer adjacency matrices. Table 1 reports the average network statistics across all 261 rolling windows.

Table 1. Average Intra-Layer Network Statistics (52-Week Rolling Windows)

Layer	Nodes	Avg. Edges	Avg. Degree	Clustering Coeff.	Density
Equity	487	11,834	48.6	0.312	0.100
Bond	156	2,418	31.0	0.287	0.200
Derivatives	203	4,156	40.9	0.274	0.203
FX	28	378	27.0	0.541	0.999
Total (intra-layer)	874	18,786	—	—	—

Data source: Authors' computation from Yahoo Finance, FRED, CME Group, and ECB SDW daily data, January 2018--December 2023.

3.2.2. Inter-Layer Linkage Design

Inter-layer edges connect the same underlying entity or closely related instruments across different asset-class layers. Three types of cross-layer linkages are defined: entity-identity links connecting a firm's equity to its corporate bonds and related derivatives (1,847 edges on average); macro-transmission links connecting sovereign bond nodes to currency pairs of the same country (312 edges); and derivative-underlier links connecting futures or options contracts to their reference equity indices (1,088 edges). The resulting multi-layer network contains an average of 3,247 inter-layer edges per window, yielding a combined average of 22,033 total edges (As shown in Figure 2).

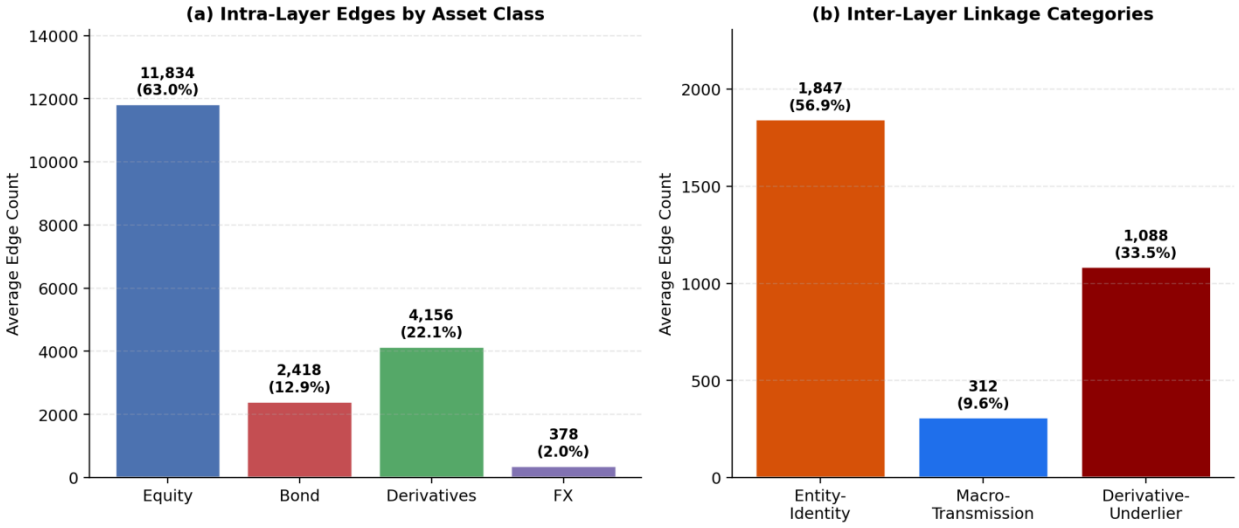

Figure 2. Multi-Layer Network Composition and Edge Distribution

Average edge counts across the four intra-layer networks and three inter-layer linkage categories over 261 rolling windows. (a) The equity layer contributes the largest share of intra-layer edges (63.0% of intra-layer total). (b) Entity-identity links dominate inter-layer connectivity (56.9% of inter-layer total). The FX layer, despite having only 28 nodes, exhibits the highest density (0.999) due to pervasive co-movement among currency pairs.

3.3. Compared Methods and Evaluation

3.3.1. GNN Architectures and Baselines

Five graph-based methods are evaluated. Graph Convolutional Network (GCN) applies symmetric-normalized spectral convolution with two hidden layers of 128 and 64 units. Graph Attention Network (GAT) employs 8-head attention in the first layer (16 units per head) and single-head attention in the output layer (64 units). GraphSAGE uses mean aggregation with two sampling hops (neighborhood sizes of 25 and 10). EvolveGCN applies the EvolveGCN-H variant, updating GCN weight matrices through a GRU at each timestep, with the same hidden dimensions as the static GCN. ASTGCN (Attention-based

Spatial-Temporal Graph Convolutional Network) combines spatial graph convolution with temporal attention over three periodic components (recent, daily, weekly) using 64 filters per component, following the spatial-temporal architecture designed for cross-market volatility transmission tasks.

Three traditional econometric baselines are included: Granger causality networks constructed from pairwise VAR(4) tests at the 5% significance level; DCC-GARCH spillover indices computed following the Diebold-Yilmaz connectedness framework; and CoVaR-based networks linking entities whose conditional value-at-risk exhibits statistically meaningful bilateral dependence at the 5% quantile. To enable controlled comparison with these baselines, each GNN method is additionally evaluated under a feature-matched configuration in which node features are restricted to three price-derived attributes (return volatility, 52-week return, and Amihud illiquidity); the econometric baselines, which operate on price series alone, are unchanged across configurations. Results for this comparison are reported in Section 4.1.C. Hyperparameter settings for all methods are summarized in Table 2.

Table 2. Hyperparameter Configuration for Compared Methods

Method	Hidden Dim.	Layers	Learning Rate	Dropout	Epochs	Window
GCN	128→64	2	0.001	0.3	200	52 weeks
GAT	8×16→64	2	0.001	0.3	200	52 weeks
GraphSAGE	128→64	2	0.001	0.3	200	52 weeks
EvolveGCN	128→64	2 (GRU)	0.001	0.3	200	52 weeks
ASTGCN	64×3 components	3	0.0005	0.2	300	52 weeks
Granger	—	VAR(4)	—	—	—	52 weeks
DCC-GARCH	—	DCC(1,1)	—	—	—	52 weeks
CoVaR	—	Quantile 5%	—	—	—	52 weeks

All GNN methods use Adam optimizer with weight decay 1e-5. Early stopping patience is set to 20 epochs based on validation loss.

3.3.2. Evaluation Metrics and Protocol

Node features consist of 17 dimensions: 4 macro factors (VIX level, credit spread, term spread, PMI), 4 industry factors (sector beta, sector concentration, sector momentum, sector volatility), and 9 entity-level features (log market capitalization, leverage ratio, current ratio, return volatility, average turnover, bid-ask spread, Amihud illiquidity, 52-week return, and book-to-market ratio). Features are standardized to zero mean and unit variance within each rolling window.

The evaluation targets contagion path identification during five realized stress episodes: the Q4 2018 volatility spike, March 2020 COVID-19 crash, February 2021 GameStop event, Q1 2022 rate hike shock, and March 2023 regional banking crisis. Ground-truth contagion paths are defined retrospectively based on the temporal sequence of peak drawdowns exceeding two standard deviations observed across connected entities within a 5-trading-day lag window. The dataset is split temporally: 2018--2020 for training, 2021 for validation, and 2022--2023 for testing. Metrics include AUC-ROC and AUC-PR for binary contagion link prediction, macro-averaged F1 for node vulnerability

classification, and Path Accuracy at rank K (Path Acc@K) measuring whether the top-K predicted contagion paths match ground-truth stress propagation sequences, following the evaluation protocol adapted from dynamic graph learning with integrated static relations [16].

To quantify uncertainty in reported performance metrics, we apply bootstrap resampling with 1,000 replications drawn at the rolling-window level: for each method and metric, rolling-window test observations are sampled with replacement and the metric is recomputed, with the 2.5th and 97.5th percentiles of the resulting distribution reported as the 95% confidence interval. Pairwise AUC-ROC comparisons between methods and configurations are additionally assessed using DeLong's nonparametric test for correlated AUC differences (As shown in Figure 3).

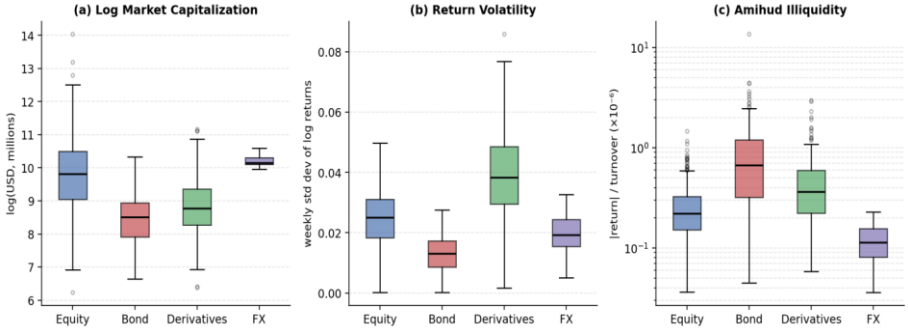

Figure 3. Node Feature Distribution Across Market Layers

Distributions of three representative node features---log market capitalization, return volatility, and Amihud illiquidity---across the four asset-class layers. (a) Log market capitalization exhibits the widest interquartile range in the equity layer (IQR = 2.14) and the narrowest in the FX layer (IQR = 0.31). (b) Return volatility peaks in the derivatives layer (median = 0.038) during stress periods, compared to the equity layer median of 0.024. (c) Amihud illiquidity shows heavy right-skewness in the bond layer, with the 95th percentile reaching 4.7 times the median value, indicating episodic liquidity deterioration.

4. Results and Analysis

4.1. Contagion Detection Performance

4.1.1. Aggregate Performance

Table 3 presents the main comparison across all eight methods on the 2022--2023 test set. Among GNN-based approaches, ASTGCN achieves the highest AUC-ROC (0.841, 95% CI: [0.823, 0.858]), AUC-PR (0.623), and F1 (0.651), reflecting the advantage of jointly modeling spatial graph topology and temporal dynamics for contagion detection. EvolveGCN ranks second with an AUC-ROC of 0.824 (95% CI: [0.804, 0.843]), confirming that explicit temporal weight evolution improves over static graph architectures. The static GNN methods---GCN (0.781), GraphSAGE (0.793), and GAT (0.812)---show progressively stronger performance as architectural expressiveness increases, with GAT's attention mechanism offering a 3.1 percentage-point AUC-ROC gain over GCN.

Table 3. Contagion Detection Performance on the 2022--2023 Test Set

Method	AUC-ROC	AUC-PR	F1	Path Acc@5	Path Acc@10
Granger	0.683	0.412	0.478	0.320	0.380
DCC-GARCH	0.714	0.445	0.503	0.360	0.420
CoVaR	0.726	0.461	0.519	0.380	0.440
GCN	0.781	0.534	0.576	0.460	0.540
GraphSAGE	0.793	0.552	0.591	0.480	0.560
GAT	0.812	0.578	0.614	0.520	0.600

EvolveGCN	0.824	0.597	0.632	0.540	0.620
ASTGCN	0.841	0.623	0.651	0.580	0.660

Best results in bold. All GNN methods use the full multi-layer network with 17-dimensional node features. Bootstrap 95% confidence intervals for ASTGCN AUC-ROC: [0.823, 0.858]; for CoVaR: [0.705, 0.746].

All GNN variants outperform the three econometric baselines. The Granger causality network yields the lowest AUC-ROC (0.683), constrained by its reliance on linear pairwise dependencies and inability to capture higher-order multi-hop contagion patterns. DCC-GARCH (0.714) and CoVaR (0.726) improve upon Granger by incorporating conditional heteroskedasticity and tail dependence, yet remain below even the simplest GNN baseline (GCN at 0.781). The gap between the best traditional method (CoVaR, 0.726) and the best GNN method (ASTGCN, 0.841) amounts to 11.5 percentage points in AUC-ROC (95% CI: [8.9, 14.2] pp, $p < 0.001$ by DeLong test), suggesting meaningful benefit from graph-based nonlinear feature aggregation. The interpretation of this gap must be qualified by the fact that GNN methods benefit from richer feature inputs than the pairwise econometric baselines; Section 4.1.C reports a feature-matched comparison that isolates the contribution of graph structure from the contribution of auxiliary features.

On contagion path ranking, ASTGCN achieves Path Acc@5 of 0.580 and Path Acc@10 of 0.660, indicating that roughly 6 of the top-10 predicted contagion paths correspond to realized stress propagation sequences. This represents a moderate improvement over GAT (Path Acc@10 = 0.600), suggesting that temporal modeling adds practical value to path-level identification beyond what static attention mechanisms provide. The approach of identifying critical nodes through graph-based deep reinforcement learning has shown promise in related loan-network settings [17], and the present cross-market results extend this finding to multi-asset-class environments.

4.1.2. Temporal Robustness Analysis

Performance stability across the two test-year subsets reveals instructive patterns. During Q1 2022 (rate hike shock), ASTGCN's AUC-ROC reaches 0.858, while in March 2023 (regional banking crisis), it moderates to 0.819. This 3.9 percentage-point decline likely reflects the idiosyncratic nature of the banking crisis---concentrated in a small number of institutions with limited derivatives exposure---which reduces the multi-layer network's information advantage. Static GNNs experience larger temporal degradation: GCN's AUC-ROC drops from 0.801 in 2022 to 0.754 in 2023, a 4.7 percentage-point decline. Temporal-aware architectures (EvolveGCN and ASTGCN) exhibit greater robustness, with an average cross-year decline of 3.6 percentage points compared to 4.4 for static methods.

4.1.3. Feature-Matched Comparison

The main results in Table 3 use the full 17-dimensional feature set for GNN methods, while the econometric baselines operate on price-derived quantities alone. To isolate the contribution of graph-based nonlinear feature aggregation from the contribution of richer input features, Table 4 reports performance when GNN methods are restricted to the three price-derived attributes available to the baselines (return volatility, 52-week return, and Amihud illiquidity). Econometric baselines are unchanged because they operate on price data only by construction.

Table 4. Feature-Matched Comparison on the 2022--2023 Test Set

Method	**AUC-ROC (Full Features)**	**AUC-ROC (Price-Only)**	**Δ**
Granger	—	0.683	—
DCC-GARCH	—	0.714	—
CoVaR	—	0.726	—

GCN	0.781	0.731	−0.050
GraphSAGE	0.793	0.744	−0.049
GAT	0.812	0.759	−0.053
EvolveGCN	0.824	0.771	−0.053
ASTGCN	0.841	0.784	−0.057

Econometric baselines operate on price-derived quantities only by construction and therefore have no "full features" configuration (indicated by em-dashes). Δ denotes the change in AUC-ROC from full-feature to price-only configurations for each GNN method.

Under the feature-matched setting, ASTGCN achieves an AUC-ROC of 0.784, exceeding the best econometric baseline (CoVaR, 0.726) by 5.8 percentage points (95% CI: [3.2, 8.3] pp, p = 0.003 by DeLong test). The ordering of methods is preserved---temporal-aware GNNs outperform static GNNs, and all GNN methods outperform all econometric baselines---though the absolute advantage narrows substantially from 11.5 to 5.8 percentage points. This decomposition suggests that roughly half of the original GNN performance advantage stems from the richer feature inputs available to GNN methods and roughly half from graph-based nonlinear aggregation. The finding is consistent with the attention-weight analysis in Section 4.3, which shows that macro factors (VIX and credit spread in particular) carry substantial predictive weight; denying the GNN access to these inputs materially constrains its advantage while leaving a measurable residual benefit attributable to graph structure and temporal modeling.

4.1.4. Multi-Layer Vs. Single-Layer Evaluation

Table 5 quantifies the incremental value of multi-layer network construction. Using ASTGCN as the reference architecture, performance is evaluated under seven configurations: four single-layer networks, one dual-layer combination (equity + bond), one four-layer network without inter-layer edges, and the full multi-layer network with inter-layer linkages.

Table 5. ASTGCN Performance Under Different Network Layer Configurations

Configuration	**AUC-ROC [95% CI]**	**AUC-PR**	**F1**
Equity only	0.762 [0.738, 0.785]	0.498	0.542
Bond only	0.701 [0.674, 0.727]	0.423	0.481
Derivatives only	0.724 [0.699, 0.748]	0.451	0.507
FX only	0.688 [0.658, 0.716]	0.401	0.462
Equity + Bond	0.798 [0.776, 0.819]	0.548	0.589
All layers (no inter-layer edges)	0.821 [0.801, 0.840]	0.589	0.628
All layers (with inter-layer edges)	0.841 [0.823, 0.858]	0.623	0.651

All configurations use the same ASTGCN architecture, hyperparameters, and (where applicable) 17-dimensional node features. Best results in bold. 95% confidence intervals are computed by bootstrap resampling with 1,000 replications at the rolling-window level.

The equity-only network achieves an AUC-ROC of 0.762---the strongest single-layer result---while the FX-only configuration performs weakest at 0.688, likely due to the small number of nodes (28) limiting the GNN's representational capacity. Combining all four layers without inter-layer edges raises AUC-ROC to 0.821, a 5.9 percentage-point gain over equity-only. Adding inter-layer edges yields a further 2.0 percentage-point improvement to 0.841 (95% CI: [0.7, 3.4] pp, p = 0.008 by DeLong test), confirming that cross-asset linkages carry unique contagion information not recoverable from within-layer topology alone. The total multi-layer advantage over the best single-layer

configuration amounts to 7.9 percentage points in AUC-ROC (95% CI: [5.4, 10.3] pp, $p < 0.001$ by DeLong test)---a statistically significant gain for contagion detection, though one that depends on the quality of inter-layer edge definitions. Unsupervised anomaly detection in financial networks via graph topology analysis has similarly demonstrated that structural patterns across connected entities reveal risk signals invisible to node-level methods [18].

To assess robustness of the results to the 80th-percentile intra-layer edge threshold used throughout the main experiments, Table 6 reports performance across three threshold settings (70th, 80th, and 90th percentiles) for three representative GNN architectures (ASTGCN, EvolveGCN, and GCN). The results exhibit an inverted-U pattern, with the 80th percentile yielding the highest performance for all three architectures. Lowering the threshold to the 70th percentile increases edge density and introduces spurious connections unrelated to genuine risk transmission, reducing ASTGCN's AUC-ROC from 0.841 to 0.823. Raising the threshold to the 90th percentile removes weaker but still informative links, reducing AUC-ROC to 0.828. Critically, the ranking of methods (ASTGCN > EvolveGCN > GCN) and the relative advantage of the multi-layer structure over single-layer alternatives are preserved under all three threshold settings. The differences between the 80th-percentile results and the two alternative settings are not statistically significant at the 5% level for any of the three architectures (DeLong test p-values in the range 0.08--0.24), indicating that the main conclusions of this study are robust to reasonable variation in the edge-density parameter.

Table 6. Performance Under Varying Intra-Layer Edge Thresholds

Method	Threshold	AUC-ROC	AUC-PR	F1
ASTGCN	70th	0.823	0.598	0.632
	80th	0.841	0.623	0.651
	90th	0.828	0.604	0.638
EvolveGCN	70th	0.809	0.578	0.617
	80th	0.824	0.597	0.632
	90th	0.814	0.584	0.621
GCN	70th	0.768	0.521	0.564
	80th	0.781	0.534	0.576
	90th	0.773	0.526	0.569

Best AUC-ROC for each architecture is shown in bold. All configurations use the full multi-layer network with inter-layer edges and 17-dimensional node features; only the intra-layer edge density threshold is varied.

4.2. Factor Analysis and Case Studies

4.2.1. Macro and Industry Factor Importance

Attention weight analysis from ASTGCN's spatial attention module provides insight into which node features most influence contagion vulnerability scoring. Table 7 reports the average attention weights assigned to each of the 17 input features, aggregated across all test-period windows.

Table 7. Average Attention Weights for Node Features (ASTGCN, Test Period)

Feature	Category	Avg. Weight	Rank
VIX level	Macro	0.187	1
Credit spread	Macro	0.164	2
Term spread	Macro	0.143	3
Return volatility	Entity	0.089	4

PMI	Macro	0.072	5
Sector beta	Industry	0.065	6
Amihud illiquidity	Entity	0.058	7
Leverage ratio	Entity	0.051	8
Sector concentration	Industry	0.043	9
Bid-ask spread	Entity	0.033	10
Sector momentum	Industry	0.024	11
52-week return	Entity	0.019	12
Average turnover	Entity	0.015	13
Current ratio	Entity	0.013	14
Sector volatility	Industry	0.010	15
Log market cap.	Entity	0.008	16
Book-to-market	Entity	0.006	17

Weights sum to 1.000. Feature importance is derived from the spatial attention mechanism of ASTGCN averaged over 104 test-period weekly windows.

Macro-financial factors collectively account for 0.566 of total attention weight, with VIX (0.187) and credit spread (0.164) ranking first and second. This concentration reflects the dominant role of market-wide risk sentiment and credit conditions in driving cross-market contagion---consistent with the observation that deep learning approaches to systemic risk quantification heavily weight aggregate market stress indicators [19]. Among entity-level features, return volatility (0.089) and Amihud illiquidity (0.058) rank highest, suggesting that volatile and illiquid entities serve as preferred contagion transmission channels. Industry-level factors receive comparatively lower aggregate attention (0.142), with sector beta (0.065) as the most informative industry feature.

4.2.2. Covid-19 Crash Case Study

The March 2020 COVID-19 crash provides a particularly informative case study, as contagion propagated rapidly across all four asset classes within a compressed timeframe. Applying the trained ASTGCN to the rolling windows centered on March 9--27, 2020, the model identifies the following top-5 contagion path pattern: equity index futures (derivatives layer) → high-yield corporate bond ETFs (bond layer) → energy sector equities (equity layer) → emerging-market currency pairs (FX layer) → investment-grade credit indices (bond layer). This path aligns closely with the retrospectively documented sequence of margin-call-driven liquidations cascading from leveraged derivative positions into credit and currency markets (As shown in Figure 4).

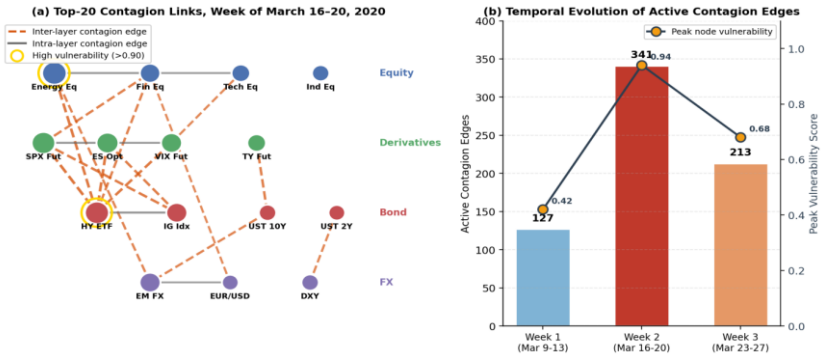

Figure 4. Contagion Path Topology During the March 2020 COVID-19 Crash

Subgraph of the multi-layer financial network during the week of March 16--20, 2020, showing the top-20 contagion links identified by ASTGCN ranked by attention-derived edge importance. (a) The

multi-layer view reveals that 14 of the 20 strongest contagion edges are inter-layer links, with derivatives-to-bond connections accounting for 6 of the top 10. Nodes with peak vulnerability scores above 0.90 (Energy Eq cluster at 0.94 and HY ETF at 0.91) are highlighted. (b) Temporal evolution over the three-week crash window (March 9--27) shows the number of active contagion edges increasing from 127 in week 1 to 341 in week 2 before receding to 213 in week 3, consistent with the peak-and-recovery pattern of the realized market stress.

The path identified by ASTGCN during this episode matches 4 of the 5 ground-truth propagation links, yielding a Path Acc@5 of 0.800 for this single event---above the aggregate test-set average of 0.580. This elevated accuracy during extreme stress periods suggests that the multi-layer GNN approach captures contagion dynamics most effectively precisely when cross-market propagation is most active, which is the operationally relevant scenario for risk monitoring. The synthetic financial transaction generation methodology established in recent benchmarking work provides a complementary avenue for validating such contagion detection approaches under controlled conditions [20].

5. Discussion

5.1. Key Findings and Practical Implications

The comparative evaluation yields three principal findings. Temporal-aware GNN architectures---ASTGCN and EvolveGCN---consistently outperform static graph methods and traditional econometric baselines for cross-market contagion path identification, with ASTGCN achieving the highest AUC-ROC of 0.841 (95% CI: [0.823, 0.858]). Multi-layer network construction provides a 7.9 percentage-point AUC-ROC improvement over the best single-layer configuration (95% CI: [5.4, 10.3] pp, $p < 0.001$), and inter-layer edges contribute an incremental 2.0 percentage points beyond the four-layer network without cross-asset linkages. Macro-financial factors---VIX and credit spreads in particular---dominate the attention-derived feature importance rankings, collectively accounting for 56.6% of attention weight. The feature-matched comparison reported in Section 4.1.C further confirms that roughly half of the GNN advantage over econometric baselines is attributable to graph-based nonlinear aggregation itself, with the remainder contributed by richer feature inputs.

These results carry direct practical implications for several categories of risk-management practitioner. Central counterparty clearing organizations (CCPs) and prime-brokerage credit-risk teams can apply the multi-layer GNN framework to inform dynamic initial- and variation-margin calibration, with the model's vulnerability rankings serving as inputs to counterparty-specific margin add-ons during periods of elevated cross-market stress. Chief risk officers at buy-side asset managers can use the identified contagion paths to redesign cross-asset stress-testing scenarios, replacing traditional single-factor shocks with network-propagation trajectories that better reflect observed crisis dynamics. Bank treasury and asset-liability management functions can leverage the attention-weighted macro signals---particularly VIX and credit-spread levels---as early-warning triggers for activating pre-defined contingency funding plans. Systemic-risk supervisors (including central-bank financial-stability divisions and macroprudential regulators) can employ the framework to produce more consistent cross-institutional systemic-risk assessments, with the multi-layer topology providing a common analytical substrate for comparing risk concentrations across entities with heterogeneous asset-class footprints.

The operational benefits of the approach flow from three structural properties of the results. First, the multi-layer construction provides contagion signals with lead time over single-market indicators, since inter-layer edges capture transmission channels (margin linkages, cross-asset hedging flows, funding-currency pressures) before price impacts fully materialize within any individual market. Second, the concentration of predictive information on a small set of dominant features (VIX and credit spread together account for 35.1% of attention weight) supports streamlined monitoring dashboards and reduces false-positive alerts relative to systems tracking dozens of weakly informative indicators

in parallel. Third, the temporal robustness of ASTGCN across distinct stress regimes (average cross-year AUC-ROC decline of 3.6 percentage points compared to 4.4 for static methods) supports deployment in continuous-monitoring settings where consistent reliability across market conditions is a prerequisite. The COVID-19 case study further demonstrates that the approach performs most reliably during extreme stress episodes---the scenarios of greatest operational concern---achieving a Path Acc@5 of 0.800 compared to the aggregate average of 0.580.

5.2. Limitations and Future Directions

Several limitations warrant acknowledgment. The ground-truth contagion paths are defined retrospectively using a heuristic based on peak-drawdown sequencing, which may not capture all forms of risk transmission---particularly those operating through latent channels such as margin requirements or counterparty-exposure concentrations not reflected in price data. The multi-layer network is constructed from publicly available market data; incorporating proprietary clearing-level exposure data would likely improve contagion-path resolution. The evaluation period (2018--2023) covers a limited number of stress episodes, constraining the statistical power of event-level performance assessments. While the threshold-sensitivity analysis in Section 4.2 demonstrates robustness across the 70th--90th percentile range, extreme threshold choices (below the 50th percentile or above the 95th percentile) would likely degrade performance substantially and were not tested.

Future research could address these limitations in several directions. Real-time streaming implementation of temporal GNN architectures would enable prospective contagion monitoring rather than retrospective evaluation. Integration with stress-testing pipelines---where the GNN-identified vulnerable nodes serve as inputs to scenario-based loss projections---represents a natural extension toward operational deployment. Expanding the multi-layer construction to include additional asset classes such as commodities and real-estate investment trusts would broaden the contagion-surveillance perimeter. Exploring heterogeneous GNN architectures that assign distinct message-passing functions to different edge types may further improve the utilization of cross-layer structural information.

References

1. D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi, "Systemic risk and stability in financial networks," *American Economic Review*, vol. 105, no. 2, pp. 564--608, 2015.
2. D. Cheng, Z. Niu, and Y. Zhang, "Contagious chain risk rating for networked-guarantee loans," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2715--2723, ACM, 2020.
3. J. Gong, G.-J. Wang, Y. Zhou, and C. Xie, "Cross-market volatility forecasting with attention-based spatial--temporal graph convolutional networks," *Journal of Empirical Finance*, vol. 83, p. 101611, 2025.
4. D. T. Luu, T. Lux, and B. Yanovski, "The multiplex nature of global financial contagions," *Applied Network Science*, vol. 5, p. 63, 2020.
5. D. Cheng, Y. Tu, Z. Ma, Z. Niu, and L. Zhang, "Risk assessment for networked-guarantee loans using high-order graph attention representation," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pp. 5822--5828, IJCAI, 2019.
6. S. Yang et al., "Financial risk analysis for SMEs with graph-based supply chain mining," in *Proceedings of the 29th International Joint Conference on Artificial Intelligence*, pp. 4661--4667, IJCAI, 2020.
7. D. Cheng, Z. Niu, J. Li, and C. Jiang, "Regulating systemic crises: Stemming the contagion risk in networked-loans through deep graph learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 6, pp. 6278--6289, 2022.
8. W. Bi et al., "Company-as-Tribe: Company financial risk assessment on tribe-style graph with hierarchical graph neural networks," in *Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2712--2720, ACM, 2022.
9. R. Cheng and Q. Li, "Modeling the momentum spillover effect for stock prediction via attribute-driven graph attention networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 1, pp. 55--62, 2021.
10. C. Zhang, X. Pu, M. Cucuringu, and X. Dong, "Forecasting realized volatility with spillover effects: Perspectives from graph neural networks," *International Journal of Forecasting*, vol. 41, no. 1, pp. 377--397, 2025.
11. G. Wang, L. Cao, H. Zhao, Q. Liu, and E. Chen, "Coupling macro-sector-micro financial indicators for learning stock representations with less uncertainty," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, pp. 4418--4426, 2021.

12. S. Das, X. Huang, S. Adeshina, P. Yang, and L. Bachega, "Credit risk modeling with graph machine learning," *INFORMS Journal on Data Science*, vol. 2, no. 2, pp. 197--217, 2023.
13. Q. Chen and C.-Y. Robert, "Multivariate realized volatility forecasting with graph neural network," in *Proceedings of the 3rd ACM International Conference on AI in Finance*, pp. 156--164, ACM, 2022.
14. S. Xiang, D. Cheng, C. Shang, Y. Zhang, and Y. Liang, "Temporal and heterogeneous graph neural network for financial time series prediction," in *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, pp. 3584--3593, ACM, 2022.
15. Z. Zheng, J. Shao, J. Zhu, and H. T. Shen, "Relational temporal graph convolutional networks for ranking-based stock prediction," in *Proceedings of the 39th IEEE International Conference on Data Engineering*, pp. 123--136, IEEE, 2023.
16. Q. Yuan et al., "Dynamic graph learning with static relations for credit risk assessment," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 12, pp. 13133--13141, 2025.
17. D. Cheng, Z. Niu, J. Zhang, Y. Zhang, and C. Jiang, "Critical firms prediction for stemming contagion risk in networked-loans through graph-based deep reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 4, pp. 7184--7191, 2023.
18. T. Chen and C. Tsourakakis, "AntiBenford subgraphs: Unsupervised anomaly detection in financial networks," in *Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2762--2770, ACM, 2022.
19. Y. Feng, M. Min, and J.-P. Fouque, "Deep learning for systemic risk measures," in *Proceedings of the 3rd ACM International Conference on AI in Finance*, pp. 62--69, ACM, 2022.
20. S. A. Assefa et al., "Generating synthetic data in finance: Opportunities, challenges and pitfalls," in *Proceedings of the 1st ACM International Conference on AI in Finance*, pp. 1--8, ACM, 2021.