Graph Neural Reinforcement Learning for Cross-Market Contagion-Aware Portfolio Risk Control Under Extreme Financial Stress

Mohan Khanna
Department of Computer Science, University of Houston, Houston, TX, USA.
mohankhanna95@uh.edu

Danjia Tian
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
danjia.tian@unr.edu

Devide Greene
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
davide461@ku.edu

Casper Carpenter
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
casper1986@buffalo.edu

Abstract

The intensification of cross-market financial linkages during periods of extreme stress has exposed fundamental limitations in conventional portfolio risk control frameworks, which typically treat markets as isolated entities and rely on linear correlation structures that collapse under crisis dynamics. This paper proposes a graph neural reinforcement learning (GNRL) architecture that directly encodes the latent network topology of global financial markets into the decision process of an adaptive portfolio agent, enabling it to internalize contagion pathways and adjust exposures before systemic propagation crystallizes. The system constructs a dynamic graph with nodes representing individual assets or market indices and weighted edges capturing time-varying dependencies such as volatility spillovers, tail dependencies, and information flow. A graph neural network encoder processes this evolving structure to generate state representations that preserve higher-order contagion signatures. These representations feed a policy network trained via actor-critic reinforcement learning, where the reward function integrates risk-adjusted returns with drawdown penalties and stress-sensitive adjustments derived from composite stress indices. The architecture is examined not only through its algorithmic design but also through the lens of system-level tradeoffs, including latency constraints in live trading pipelines, resilience to concept drift during regime shifts, and the governance challenges posed by autonomous rebalancing in illiquid environments. We analyze the implications of deploying such a system at scale, focusing on fairness across market participants, the potential for unintended amplification of systemic risk due to homogeneity in learned policies, and the interpretability requirements demanded by prudential regulators. Infrastructure considerations such as cloud-native orchestration, continuous model validation under realistic walk-forward protocols, and the incorporation of leakage-safe early warning signals are discussed in depth. By synthesizing advances in graph representation learning, deep reinforcement learning, and financial stress analytics, this study provides a comprehensive blueprint for building next-generation portfolio control systems that are both contagion-aware and operationally viable under the most severe market conditions.

Keywords

Graph neural networks; reinforcement learning; portfolio risk management; financial contagion; extreme market stress; cross-market dependencies; systemic risk; interpretability.

1. Introduction

Portfolio risk management in the presence of extreme financial stress remains one of the most consequential challenges facing institutional investors, central banks, and regulatory bodies. The global financial crisis of 2007–2009 and the COVID-19 market dislocations of 2020 demonstrated that cross-market contagion can propagate within hours, rendering traditional diversification and static hedging strategies ineffective precisely when they are needed most. Conventional risk models, including those based on mean-variance optimization [1] and value-at-risk frameworks, rely heavily on historical correlation matrices that exhibit severe instability during tail events. In these circumstances, assets that previously appeared decorrelated converge in their loss trajectories, and the topology of financial interdependence shifts dramatically. The inability of linear, backward-looking models to capture such non-stationary network effects has motivated a growing body of research into dynamic, data-driven approaches that can adapt allocation decisions in real time [2,3]. Reinforcement learning (RL) has emerged as a particularly promising paradigm for sequential decision-making under uncertainty in financial domains, with attention-enhanced architectures demonstrating notable success in portfolio optimization tasks [4]. However, most existing RL-based portfolio managers treat asset returns as independent streams or rely on pairwise relationships encoded in a fixed feature vector, ignoring the structural contagion channels that amplify losses during systemic events. Parallel advances in graph neural networks (GNNs) have proven highly effective at modeling relational data in domains as diverse as drug discovery, traffic prediction, and social network analysis, but their application to financial risk control, especially under extreme stress, remains nascent and fragmented [5,6]. The core proposition of this paper is that the integration of graph neural encoding with reinforcement learning creates a powerful hybrid architecture capable of internalizing the latent network of cross-market dependencies and using that knowledge to proactively adjust portfolio exposures ahead of cascading dislocations. Such a system must confront not only algorithmic complexity but also a multitude of system-level concerns: real-time data ingestion from heterogeneous market sources, resilience to data quality degradation during turbulent episodes, compliance with regulatory standards for explainability, and the potential for feedback loops if multiple large agents deploy similar learned strategies. This paper develops a comprehensive systems research perspective on graph neural reinforcement learning for contagion-aware portfolio risk control, addressing architecture design, deployment infrastructure, governance, sustainability, and fairness while anchoring the discussion in insights from recent advances in stress indexing and early warning design.

2. The Contagion Challenge and the Limits of Current Approaches

Financial markets have evolved into a densely interlinked global system in which shocks to one asset class or geography can propagate rapidly through multiple channels. These channels include direct ownership linkages, common factor exposures, liquidity spirals arising from margin calls, and information-based herding behavior among market participants. Network-based analyses of volatility connectedness have provided compelling evidence that during periods of market turmoil the structure of interdependence becomes both denser and more synchronized, leading to a sharp increase in systemic risk [7]. In such an environment, a portfolio manager who relies solely on marginal risk metrics without accounting for the higher-order dependencies captured by the full market graph will systematically underestimate the probability of simultaneous drawdowns. Traditional risk-parity and minimum-variance portfolios, while robust under normal conditions, often exacerbate losses when correlations jump toward unity because they fail to differentiate between benign diversification and fragile interdependence born of latent contagion channels. Even more sophisticated dynamic conditional correlation models exhibit significant estimation delays and can generate noisy allocation signals during rapid regime transitions.

The application of deep reinforcement learning to portfolio management has produced encouraging results by learning policies directly from market data without explicit distributional assumptions. Models that incorporate attention mechanisms, such as transformer-based architectures, can weigh the relevance of different assets dynamically [4]. However, attention applied to a flat vector of asset characteristics does not natively encode the multi-edge relational structure that characterizes modern financial networks, where the influence of one market on another is mediated through multiple overlapping layers of credit, equity, currency, and commodity linkages. A portfolio exposed to emerging market equities, for instance, may experience amplified losses not only because of direct correlation with developed market indices but because of an intermediate shock transmitted through commodity currencies and sovereign credit spreads. Capturing these indirect pathways requires a relational inductive bias that graph neural networks are uniquely positioned to provide. Graph convolutional and graph attention networks can learn node embeddings that summarize not only an asset's own state but also its position within the broader network topology, thereby reflecting second-order and third-order contagion potential [8,9]. Yet building a trading system that leverages such embeddings for real-time portfolio rebalancing under stress demands careful attention to the entire data-to-decision pipeline, including the construction of the graph itself, the choice of message-passing mechanism, and the alignment of the reinforcement learning objective with stress-aware reward signals.

3. System Architecture for Graph Neural Reinforcement Learning

We propose a layered system architecture consisting of a dynamic graph construction module, a GNN-based encoder, a reinforcement learning agent, and an execution and monitoring layer. The graph construction module ingests high-frequency data from multiple asset classes—equities, bonds, currencies, and commodities—and builds a heterogeneous graph with nodes corresponding to individual securities or representative indices. Edges are initialized based on pre-crisis balance sheet linkages and continuously updated using rolling windows of realized volatility spillovers, tail dependence coefficients, and nonlinear Granger causality measures. During periods of elevated stress, edges are reweighted dynamically to reflect the increased probability of contagion, ensuring that the graph structure remains aligned with the prevailing market regime. This adaptive rewiring is informed by composite stress indicators that combine principal component analysis and arbitrage pricing theory decompositions, providing a multidimensional view of market strain that goes beyond simple volatility measures [10]. The resulting graph serves as the input to a graph attention network encoder that computes asset embeddings by iteratively aggregating information from neighboring nodes. Through a multi-head attention mechanism, the encoder learns to assign greater weight to connections that are historically more predictive of spillover risk during crises, effectively implementing a learned form of contagion screening.

The reinforcement learning agent consumes the GNN-generated embeddings together with additional features such as portfolio holdings, transaction costs, and liquidity indicators. The agent's policy is parameterized by a deep neural network that outputs a vector of target portfolio weights, which are then executed subject to realistic constraints on leverage, turnover, and sector concentration. The critic network evaluates state-action pairs using a custom reward function that blends the portfolio's expected return, a penalty proportional to conditional value-at-risk, and a stress augmentation term that increases loss aversion whenever systemic stress indices breach predefined thresholds. This stress augmentation can be calibrated using leakage-safe early warning signals, as recent work has shown the importance of constructing benchmarks that do not inadvertently incorporate future information, thereby preserving the economic credibility of backtesting [11]. By integrating such early warning signals, the reward function becomes more sensitive to the drawdown risks that matter most to long-horizon institutional investors, particularly pension funds and sovereign wealth funds that face severe reputational and regulatory consequences if they breach minimum funding thresholds. The use of a graph-structured state representation enables the policy to learn defensive rotation strategies that would be invisible to a flat representation. For instance, the agent might reduce exposure to a seemingly uncorrelated asset after detecting that the asset has become a critical bridge node in the contagion network, a strategic maneuver that mirrors the intuition of experienced macro traders but is executed systematically and without cognitive biases.

4. Contagion-Aware Risk Control Mechanisms

The integration of contagion awareness into the agent's decision process is achieved through two complementary mechanisms: explicit edge reweighting during stress buildup and implicit representation learning via the GNN's attention heads. When the system detects a shock in one market, it propagates a simulated stress pulse through the graph using a variant of personalized PageRank, thereby estimating the expected impact on all other nodes before any observable price moves materialize. This precomputed contagion footprint is concatenated with the asset embeddings to provide the policy network with a forward-looking risk map. The footprint is updated at the frequency of the slowest data feed, which in practice could be on the order of minutes for liquid futures markets but may stretch to hours for certain over-the-counter instruments, imposing a tradeoff between timeliness and completeness. The system must therefore incorporate a data imputation and alignment module that can handle heterogeneity in reporting lags without introducing spurious signals. During extreme events, when some data sources may become unreliable or delayed, robust graph construction techniques that down-weight edges with high estimation uncertainty become indispensable.

A critical design choice concerns the balance between risk minimization and return degradation. Overly conservative policies that persistently underweight equities during normal times erode long-term returns and generate tracking error relative to benchmarks, making them unacceptable to asset owners. The GNRL framework addresses this by conditioning the stress augmentation term on the level of systemic risk, activating aggressive defensive measures only when composite stress indices exceed calibrated thresholds derived from historical crisis episodes. This regime-dependent activation can be further refined using interpretable machine learning models trained under realistic walk-forward constraints, ensuring that the threshold selection does not overfit to past crises [12]. When the stress indices recede, the agent rapidly re-engages risk-taking, benefiting from the rebound that typically follows acute dislocations. This dynamic risk budgeting stands in contrast to static rule-based approaches such as targeted volatility overlays, which react to realized volatility with a lag and can be whipsawed during rapid sentiment shifts.

The system's dependence on stress indices raises the question of how to construct reliable stress gauges that do not suffer from leakage or look-ahead bias. A robust solution involves the extraction of residual stress signals from asset return data after stripping out common factors identified through principal components and arbitrage pricing theory, producing a market drawdown indicator that isolates idiosyncratic stress components [13]. Separate work has developed residual-stress signals specifically designed to predict drawdown risk without introducing information leakage, making them suitable for integration into reinforcement learning reward functions [14]. Moreover, rigorous benchmark design for stress early warning systems ensures that any improvement attributed to contagion-aware policies is genuinely due to predictive power and not to inadvertent data snooping [15]. The GNRL architecture incorporates these principles by maintaining a strict separation between the data used for graph construction, stress indicator computation, and the policy training, testing, and validation windows, thereby upholding the integrity of the backtesting framework.

5. Infrastructure and Deployment Considerations

Deploying a GNRL-based portfolio controller in a live institutional setting demands a robust technological infrastructure capable of ingesting, cleaning, and aligning multi-market data streams with sub-second latency for the most liquid instruments, while simultaneously training graph neural encoders and reinforcement learning agents on historical data in a parallel offline pipeline. In practice, this necessitates a hybrid cloud architecture in which real-time inference is executed on co-located servers near exchange gateways, while computationally intensive graph construction and policy gradient updates are performed on scalable GPU clusters during non-trading hours. The online component must incorporate circuit breakers that revert portfolios to a predefined safe-haven allocation if the model starts producing erratic outputs, a scenario that can arise when market microstructure noise overwhelms the GNN's message passing during flash events. These circuit breakers are designed to be progressive, initially imposing a reduction in leverage and an increase in cash allocation before escalating to a complete trading halt if anomalous behavior persists. The deployment must also account for model versioning and canary testing, wherein a new policy is first run in shadow mode against live data for a probationary period before receiving capital allocation authority.

The continuous evolution of financial market structures introduces concept drift that can rapidly degrade the performance of a static model. The system therefore implements an online learning loop that periodically retrains the graph encoder and the RL policy using a sliding window of recent data, with a particular emphasis on incorporating crisis periods from the window to prevent catastrophic forgetting. The retraining schedule is governed by drift detection algorithms that monitor the distribution of GNN embeddings and trigger a full retraining cycle when a significant divergence from the reference distribution is observed. Walk-forward validation protocols, as formalized in recent work on volatility forecasting under realistic constraints, serve as the gold standard for model evaluation, ensuring that the agent's performance estimates do not suffer from the overoptimism that plagues conventional cross-validation in time series settings [16]. This validation regime is computationally expensive, requiring multiple sequential retraining passes, but it is essential for building trust with risk officers and regulators.

Data governance constitutes a critical cross-cutting concern. The system consumes proprietary market data, broker quotes, and alternative data such as news sentiment and supply chain satellite imagery, all of which must be subject to rigorous provenance tracking and licensing compliance. The graph construction module must be transparent about the provenance of each edge weight, allowing compliance teams to audit whether a specific trade decision was driven by a sanctioned data source or by an unverified signal. This auditability is becoming a regulatory requirement in jurisdictions that have adopted the European Union’s Artificial Intelligence Act or similar frameworks. Furthermore, the system must be designed to respect data localization requirements when operating across multiple regulatory jurisdictions, which can impose constraints on cross-border data flows for the graph aggregation step.

6. Governance, Fairness, and Policy Implications

The deployment of autonomous portfolio rebalancing agents that incorporate cross-market contagion signals raises profound questions of market fairness and systemic stability. From a fairness perspective, institutions that possess the technological and financial resources to implement sophisticated GNRL systems may gain an asymmetric advantage in stress anticipation, enabling them to front-run dislocations before smaller participants can react. This asymmetry could exacerbate inequality in market access and contribute to a two-tier market structure in which the gap between technologically advanced and traditional investors widens during crises. Policymakers may need to consider whether certain types of preemptive de-risking based on non-public graph analytics should be subject to disclosure requirements or position limits, analogous to the debates surrounding high-frequency trading a decade ago. Moreover, the widespread adoption of contagion-aware policies could introduce new forms of herding if many agents independently train on similar graph structures and converge to comparable defensive rotations. The very mechanism that protects an individual portfolio could, in the aggregate, accelerate capital outflows from vulnerable markets, deepening the downturn and creating a self-fulfilling prophecy of contagion. This emergent systemic risk highlights the need for regulatory stress tests that simulate the coordinated behavior of GNRL agents, not merely the response of individual institutions.

Governance frameworks for such systems must therefore encompass both internal risk controls and external oversight. Internally, model risk management committees should review the graph encoder’s edge attribution maps on a regular basis to ensure that the system is not inadvertently amplifying exposures due to spurious correlations, such as those induced by synchronous extreme moves that lack causal economic linkages. Externally, central banks and financial stability boards are beginning to explore the establishment of certification regimes for AI-driven trading algorithms, including requirements for kill switches, interpretability dashboards, and mandatory participation in industry-wide stress tests. Interpretable machine learning techniques can be employed to extract simple decision rules from the GNN embeddings that approximate the agent’s behavior, providing human supervisors with an intuitive understanding of why the system is reducing exposure to a particular region or sector [12]. These explanations are essential for maintaining the trust of asset owners and for satisfying the fiduciary duty of care that governs institutional investment management.

7. Sustainability and Robustness

The computational demands of training graph neural networks on large-scale, multi-asset financial graphs with frequent retraining cycles give rise to significant energy consumption and carbon footprint concerns. While a single training run may be manageable, the walk-forward validation and online retraining regime described earlier can multiply energy usage by a factor of ten or more, placing sustainability at odds with the need for timely model adaptation. Several mitigation strategies warrant exploration. Knowledge distillation can transfer the learned contagion detection capability from a large teacher GNN to a compact student network that is suitable for real-time inference with drastically lower power consumption. Pruning techniques applied to the graph edges can reduce the message-passing overhead by removing spurious connections that contribute little to prediction accuracy. From a hardware perspective, the use of specialized accelerators such as graph processing units and spiking neural network chips may eventually offer orders-of-magnitude improvements in energy efficiency for this class of workloads.

Robustness extends beyond energy considerations to encompass the system's resilience to adversarial manipulation. A sophisticated actor aware of the GNRL agent's reliance on volatility spillover edges could attempt to influence the graph structure by placing orders designed to alter realized covariances during the edge-weight calculation window, thereby misleading the agent into executing a portfolio shift that benefits the adversary. Adversarial training, wherein the GNN encoder is exposed to perturbed graphs during the training phase, can harden the policy against such attacks. Additionally, the incorporation of causal discovery algorithms into the graph construction pipeline offers a path to distinguish genuine causal contagion channels from ephemeral statistical associations, reducing the attack surface available to adversaries. From a data robustness standpoint, the system must gracefully handle missing or delayed data from markets that experience trading halts or circuit breaker activations during extreme stress, ensuring that the graph does not disintegrate. Graph imputation methods based on matrix completion and temporal neighborhood aggregation can fill in missing edge weights while quantifying the associated uncertainty, which can then be fed into the policy network as an additional signal that moderates aggressiveness.

8. Case Illustrations and Cross-Domain Comparisons

Consider a stylized scenario inspired by the events of March 2020, when a global pandemic shock triggered a simultaneous sell-off in equities, a flight to US Treasury bonds, a collapse in oil prices, and acute stress in corporate credit markets. A GNRL agent trained on a graph that included edges linking airline equities, energy sector bonds, high-yield credit indices, and currency pairs would have observed a rapid thickening of edges in the weeks preceding the peak panic as volatility spillovers and tail dependencies intensified. The GNN embeddings of airline stocks and energy bonds would have moved closer in the latent space, signaling elevated co-movement risk. In response, the RL agent could have preemptively reduced exposure to these sectors and increased allocations to safe-haven currencies and gold before the worst of the dislocation occurred, significantly mitigating drawdowns. By contrast, a traditional risk-parity portfolio that mechanically rebalanced to equalize risk contributions would have been forced to sell assets across the board as volatility spiked, contributing to the very market turmoil it sought to avoid. This illustrative example underscores the value of relational reasoning under stress.

Comparisons with deep RL architectures that do not use graph structure reveal a more nuanced picture. Standard attention-enhanced RL models can learn to attend to assets that exhibit leading indicator properties, but they lack an explicit mechanism for modeling the indirect propagation of shocks through intermediate nodes. When the contagion pathway is long and involves multiple asset classes, attention weights may fail to capture the full chain of dependencies, especially when training data is limited. GNRL's ability to aggregate information over multiple hops in the graph endows it with a structural advantage in such settings, although this advantage must be weighed against the additional complexity of graph construction and the risk of introducing noisy edges that degrade performance. Cross-domain analogies further illuminate the architecture's potential. In epidemiology, graph neural networks have been employed to model the spread of infectious diseases through transportation networks, with policies optimized via RL to allocate vaccines and impose travel restrictions. The translation of these methods to financial contagion is direct, suggesting that lessons from public health crisis management—particularly with respect to the timing of interventions and the allocation of limited resources—can inform the design of portfolio defense mechanisms.

9. Conclusion

This paper has presented a system-level analysis of graph neural reinforcement learning as a framework for cross-market contagion-aware portfolio risk control under extreme financial stress. By embedding the dynamic network structure of global markets into the state representation and reward design of a reinforcement learning agent, the GNRL architecture addresses a critical blind spot in contemporary portfolio management: the failure to anticipate and mitigate losses arising from contagion amplification. We have detailed the layered design of such a system, spanning dynamic graph construction, graph neural encoding, stress-adaptive reward shaping, and robust deployment infrastructure. The discussion has extended beyond algorithmic novelty to engage with the structural tradeoffs that govern real-world adoption, including latency management, concept drift resilience, interpretability, fairness, systemic risk implications, and environmental sustainability. The integration of leakage-safe stress indicators, walk-forward validation protocols, and interpretable policy distillation emerges as essential for building trust with regulators and asset owners. While significant challenges remain, particularly around adversarial robustness and the governance of emergent collective behavior, the convergence of graph representation learning and reinforcement learning opens a promising frontier for constructing next-generation financial risk control systems that are as vigilant to invisible network threats as they are to visible market moves.

References

1. Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1), 77–91.

2. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533.

3. Jiang, Z., Xu, D., & Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. arXiv preprint arXiv:1706.10059.

4. Xue, P., & Ye, Y. (2026). Attention-enhanced reinforcement learning for dynamic portfolio optimization. Intelligent Systems with Applications, 200622.

5. Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. Proceedings of the International Conference on Learning Representations.

6. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph attention networks. Proceedings of the International Conference on Learning Representations.

7. Diebold, F. X., & Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. Journal of Econometrics, 182(1), 119–134.

8. Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. IEEE Transactions on Neural Networks, 20(1), 61–80.

9. Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric deep learning: going beyond Euclidean data. IEEE Signal Processing Magazine, 34(4), 18–42.

10. Liu, T. (2026). PCA-APT Stress Index for Market Drawdowns.

11. Liu, T. (2026). Beyond volatility: A leakage-safe residual-stress signal for drawdown risk monitoring. Available at SSRN 6503179.

12. Liu, T. (2026). Interpretable Machine Learning for Volatility Forecasting Under Realistic Walk-Forward Constraints.

13. Hu, L., & Shen, Y. (2026). A predictive analytics approach for forecasting global stock index returns using deep learning techniques. Decision Analytics Journal, 100685.

14. Cont, R., & Bouchaud, J. P. (2000). Herd behavior and aggregate fluctuations in financial markets. Macroeconomic Dynamics, 4(2), 170–196.

15. Brunnermeier, M. K., & Pedersen, L. H. (2009). Market liquidity and funding liquidity. The Review of Financial Studies, 22(6), 2201–2238.

16. Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. (2017). Measuring systemic risk. The Review of Financial Studies, 30(1), 2–47.

17. Liu, T. (2026). Leakage-Safe Benchmark Design for Market-Stress Early Warning: An Economically Credible Evaluation.

18. Dacrema, M. F., Cremonesi, P., & Jannach, D. (2019). Are we really making much progress? A worrying analysis of recent neural recommendation approaches. Proceedings of the 13th ACM Conference on Recommender Systems, 101–109.

19. Hendricks, J. (1996). Evaluation of value-at-risk models using historical data. Federal Reserve Bank of New York Economic Policy Review, 2(1), 39–69.

20. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.

21. Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in Neural Information Processing Systems, 30.