Highlights

DeXposure-FM: A Time-series, Graph Foundation Model for Credit Exposures and Stability on Decentralized Financial Networks

Aijie Shu, Wenbin Wu, Gbenga Ibikunle, Fengxiang He

- Financial economics foundation model: We build DeXposure-FM, the first time-series, graph foundation model for decentralized finance.
- Empirical verification on machine learning benchmarks: DeXposure-FM significantly outperforms strong competitors in machine learning benchmarks of multi-step forecasts of edge-level exposures and network-level statistics, such as concentration, density, and sector-to-sector links.
- Financial economics tools: DeXposure-FM produces tools that support macroprudential monitoring and scenario-based DeFi stress testing, by enabling protocol-level systemic-importance scores, sector-level spillover, and concentration measures via a forecast-then-measure pipeline.
- Open-source efforts and reproducibility: Both model and code have been publicly available. We will continue our open-source efforts in future developments.

DeXposure-FM: A Time-series, Graph Foundation Model for Credit Exposures and Stability on Decentralized Financial Networks

Aijie Shu[a], Wenbin Wu[b], Gbenga Ibikunle[c], Fengxiang He[d]

[a]*Independent Scholar, Edinburgh, United Kingdom*
[b]*Cambridge Centre for Alternative Finance, Judge Business School, University of Cambridge, Cambridge, United Kingdom*
[c]*Business School, University of Edinburgh, Edinburgh, United Kingdom*
[d]*School of Informatics, University of Edinburgh, Edinburgh, United Kingdom*

Abstract

Credit exposure in Decentralized Finance (DeFi) is often implicit and token-mediated, creating a dense web of inter-protocol dependencies. Thus, a shock to one token may result in significant and uncontrolled contagion effects. As the DeFi ecosystem becomes increasingly linked with traditional financial infrastructure through instruments, such as stablecoins, the risk posed by this dynamic demands more powerful quantification tools. We introduce DeXposure-FM, the first time-series, graph foundation model for measuring and forecasting inter-protocol credit exposure on DeFi networks, to the best of our knowledge. Employing a graph-tabular encoder, with pre-trained weight initialization, and multiple task-specific heads, DeXposure-FM is trained on the DeXposure dataset that has 43.7 million data entries, across 4,300+ protocols on 602 blockchains, covering 24,300+ unique tokens. The training is operationalized for credit-exposure forecasting, predicting the joint dynamics of (1) protocol-level flows, and (2) the topology and weights of credit-exposure links. The DeXposure-FM is empirically validated on two machine learning benchmarks; it consistently outperforms the state-of-the-art approaches, including a graph foundation model and temporal graph neural networks. DeXposure-FM further produces financial economics tools that support macroprudential monitoring and scenario-based DeFi stress testing, by enabling protocol-level systemic-importance scores, sector-level spillover

Email address: F.He@ed.ac.uk (Fengxiang He)

and concentration measures via a forecast-then-measure pipeline. Empirical verification fully supports our financial economics tools. The model and code have been publicly available.
Model: `https://huggingface.co/EVIEHub/DeXposure-FM`.
Code: `https://github.com/EVIEHub/DeXposure-FM`.

Keywords: Time-Series Analysis, Graph Tabular Model, Foundation Model, Credit Exposures, Financial Stability, Decentralized Finance, Financial Networks

---

1. Introduction

Decentralized finance (DeFi) has emerged as an important and fast-evolving financial ecosystem, offering lending, trading, derivatives, and payment-like services via smart contracts deployed across multiple blockchains [39, 25]. A defining feature of DeFi is composability: protocols routinely hold, lock, or accept tokens issued by other protocols as collateral, liquidity, or reserves. As a result, credit exposure is often implicit and token-mediated rather than expressed as bilateral contracts, creating a dynamic network of inter-protocol dependencies [47, 9]. When a token that is widely used as collateral or liquidity experiences a shock, through a price collapse, governance failure, exploit, or liquidation event, losses can propagate to protocols that rely on that token, potentially generating amplification and contagion across chains [25, 5]. Policy authorities have highlighted the vulnerabilities of DeFi networks and underscored persistent gaps in quantitative tools that limit effective monitoring for the DeFi ecosystem [4]. This measurement gap is increasingly salient as links between crypto markets and the broader financial system deepen [21].

Despite the importance, empirical work has been constrained by limited data and tools. DeFi exposures are not reported in standardized balance sheets; they must be reconstructed from on-chain states and transactions. Even when exposures can be estimated, they are (1) *high-dimensional*: there are millions of potential directed links, (2) *nonstationary*, in terms of protocol upgrades, new assets, changing incentives, and (3) *strongly heteroskedastic*: regime shifts are frequent between calm and crisis. Classical time-series tools such as vector autoregression (VAR) [41] and low-rank factor methods [42] provide strong baselines, but they suffer from restricted capacity to represent data features, and struggle to represent the coupled evolution of *node-level*

states (e.g., total locked value in a protocol), *edge-level exposures* (e.g., who is economically linked to whom, and with what weight), and *meso-structure* (e.g., sectoral organization and its rewiring under stress). Meanwhile, modern graph neural networks (GNNs) [28, 44] and temporal graph models for dynamic interaction data [38] have shown impressive performance in wide areas; yet they are typically trained for a narrow task on a limited dataset, rather than serving as reusable infrastructure across multiple tasks.

To address this gap, we train DeXposure-FM, the first time-series, graph foundation model for measuring and forecasting inter-protocol credit exposures in global decentralized financial networks, to the best of our knowledge. The model is trained on more than 43.7 million weekly inter-protocol exposure data entries from 2020-2025, stored in the DeXposure dataset, which constructs credit exposures across over 4,300 protocols on 602 blockchains and covers 24,300+ unique tokens [47]. DeXposure constructs inter-protocol exposures by linking protocol-level balance-sheet proxies (e.g., total value locked, TVL [4]) to token- and sector-level structure, enabling a unified view of volumes, topology, and compositional risk. The dataset captures an evolving, weighted, directed multigraph whose nodes are protocols and whose edges represent economically meaningful exposure links inferred from token composition and valuation dynamics. DeXposure-FM follows the foundation-model paradigm: it pretrains a general representation on large-scale data and reuses it across diverse downstream tasks, a strategy that has proven effective in wide areas [10].

DeXposure-FM employs a large-size graph-tabular encoder (GraphPFN [20] in this paper) to extract representations from input data, followed by multiple heads to forecast (1) edge existence and edge weights on the graph, and (2) node-level TVL changes at multiple horizons. In optimization, the encoder is initialized by open-source pre-trained weights, while the heads are initialized randomly. Time-based walk-forward splits [8], early stopping [34], and gradient clipping [33] are employed to improve robustness under non-stationary environments. In training, DeXposure-FM is optimized by Adam [27] for the joint tasks of (1) edge existence classification, (2) edge weight regression, and (3) node-level TVL-change prediction. These outputs map naturally to policy-relevant statistics: concentration, density, sector connectivity, and directional spillovers. In crises, the relevant question is not only "what is the expected TVL tomorrow?", but "which parts of the network amplify shocks, and along which pathways does distress travel?" as well.

We evaluate DeXposure-FM on two machine learning benchmarks, tai-

lored for economic measurement and risk analytics: (1) multi-step forecasting of edge-level exposures and network-level statistics (e.g., concentration, density, and sector connectivity), and (2) predictive stress testing, where we forecast a future exposure network and compare simulated system losses on predicted versus realized graphs under fixed counterfactual shock definitions. Across tasks, we benchmark against strong competitors, including Graph-PFN, ROLAND [49] (a temporal graph neural network), and a persistence baseline that directly uses the current values as prediction.

DeXposure-FM provides a range of financial economics tools for macroprudential monitoring in decentralized financial infrastructure. We use the trained model to construct dynamic measures of protocol-level systemic importance, sector level spillover, and early-warning indicators based on shifts in network concentration and dependence on fragile collateral. These model-based measurements support DeFi-specific stress testing under counterfactual shocks. These tools are all validated in experiments. As the first time-series, graph foundation model for measuring and forecasting inter-protocol credit exposure on DeFi networks, DeXposure-FM advances the measurement toolkit needed to understand systemic risk in DeFi and its interaction with the wider financial system.

DeXposure-FM will be continuously updated and extended along five directions: (1) enlarging the training data pool; (2) improving exposure measurement by moving beyond raw TVL; (3) addressing the continuous model drift through periodic retraining on new data, drift monitoring by sector/chain, and versioned releases tied to specific data vintages; (4) innovating the architecture; and (5) fostering an open-source community by maintaining public code and model weights, running competitions, etc.

2. Related Works

This section reviews the literature in relevant areas.

2.1. Credit Risk Modeling in Financial Markets

In traditional finance, credit risk modeling spans both individual-default models and systemic network approaches. Classical structural models and reduced-form models quantify default probabilities of single entities, while network contagion models examine how defaults propagate through financial institutions. For example, Dolfin et al. apply network epidemic models to credit contagion, highlighting how interconnected portfolios can amplify

systemic credit risk [18]. In decentralized finance, credit risk manifests differently, as loans are typically over-collateralized and mediated by smart contracts. Bertomeu et al. propose simple aggregate risk measures for DeFi lending protocols, using only total deposits and borrowings, and report that systemic fragility surged around mid-2021 during crypto market turmoil [9]. Conceptual studies have begun to map TradFi and DeFi risks together: Aufiero et al. review systemic risk mechanisms, finding that while basic risk types (leverage, liquidity shocks, correlated exposures) are common to both systems, DeFi's algorithmic execution and composability lead to unique propagation channels [5].

2.2. Machine Learning for Systemic Risk, Contagion, and Network Analysis

Machine learning techniques are increasingly applied to systemic risk detection and contagion modeling [6, 26]. Graph neural networks (GNNs) in particular leverage network structure to improve risk prediction [28, 44]. Balmaseda et al. show that a GNN-based model dramatically outperforms traditional ML for classifying systemic importance of banks in simulated networks [6]. Gonon et al. integrate GNNs with explicit interbank liability networks, computing systemic risk measures as the minimum capital needed to secure the system [26]. Their framework effectively learns contagion channels (e.g., Eisenberg-Noe clearing) by incorporating graph-structured data into the risk aggregation function.

2.3. Foundation Models in Finance and Economics

Foundation models are increasingly penetrating into financial and economic domains by shifting the workflow from task-specific tools toward reusable representations learned from broad data. Instead of fitting a separate model for each target, recent approaches pretrain large neural networks, mostly using an architecture of transformer [43] and its variants, on massive data; and then transfer the pre-trained model across domains and tasks via prompting or light adaptation [2, 36, 14]. In parallel, an emerging policy-facing literature uses large language models to extract quantitative signals from narrative text (e.g., reports, releases, and news) and incorporate them into nowcasting pipelines, suggesting that relatively small but information-dense text sources can improve real-time prediction when combined with standard indicators [11, 30, 13]. These models are also adapted to graph inputs, which augment self-attention with graph structural encodings, demonstrating the

feasibility of large pretrained models on networks that resemble webs of financial contracts or cross-asset relationships; a good example is Graphormer [48]. For tabular data, the Graph Prior-data Fitted Network (GraphPFN) [20] is pretrained on millions of synthetic tabular tasks and reports strong out-of-the-box performance on diverse benchmarks.

3. Formalizing Credit-Exposures on Decentralized Financial Networks

This section formally defines terminologies and notations.

3.1. Credit Exposure in DeFi Networks

Let $X_G(q)$ denote the set of all tokens issued or generated by protocol q, and let $X_p(t)$ denote the set of tokens held (or locked) by protocol p at time t. We call that protocol p has credit exposure to protocol q at time t if the tokens it holds include at least one that was issued by q:

$$X_p(t) \cap X_G(q) \neq \varnothing. \tag{1}$$

In other words, p holds a liability of q; if the tokens issued by q lose value or become illiquid, then p faces a corresponding loss. This token-mediated dependence generalizes the notion of counterparty credit risk in traditional finance to decentralized platforms, where exposures arise not from explicit bilateral contracts but from the possession of tokenized claims. The dynamics of these relationships can be measured by reconstructed total value locked (TVL) time series for each protocol and tracking how the composition of tokens held by each protocol evolves over time [47].

Figure 1 illustrates how a single-user transaction creates a multi-layer chain of credit exposures in DeFi. A user deposits ETH into Lido, a liquid staking protocol, and receives stETH, a rebasing token that accrues staking rewards. The stETH is then wrapped into wstETH, a non-rebasing representation that is more compatible with DeFi protocols. Finally, wstETH is deposited into Pendle, a yield-tokenization protocol that splits the asset into a principal token PT-wstETH and a yield token YT-wstETH, allowing users to trade fixed and variable yield separately.

In balance-sheet terms, Lido holds staked ETH with validators and issues stETH as liabilities, the wstETH wrapper holds stETH and issues wstETH, and Pendle holds wstETH and issues the PT and YT pair. Because each

protocol's assets consist of tokens issued by the preceding protocol, a chain of credit exposures emerges: Pendle has exposure to the wstETH wrapper, which in turn has exposure to Lido. If Lido experienced a slashing event or stETH de-pegged from ETH, the losses would cascade through wstETH and into Pendle, affecting PT and YT holders. This example demonstrates that DeFi exposures are inherently multi-layered, as liquid staking, wrapping, and yield tokenization compound into complex dependency structures that the DeXposure dataset aims to capture.

3.2. Mapping Tokens to Issuing Protocols

To operationalize eq. (1), we must determine which protocol issues each token and thus identify the direction of exposure. Let $\mathcal{P}$ denote the set of all protocols and $\mathcal{C}$ the set of blockchains. For each token σ appearing in protocol p, a mapping function $M(\sigma)$ returns the protocol q that issues or manages σ. We can construct M using a fallback procedure:

1. *Metadata lookup:* Tokens are directly matched to issuing protocols using metadata from DefiLlama [16] when available.
2. *Manual mapping:* For tokens with high TVL that lack metadata links, researchers manually assign the issuing protocol based on documentation and expert knowledge.
3. *Text-vector similarity:* For remaining tokens, the text descriptions of tokens and protocols are vectorized via the term frequency–inverse document frequency (TF-IDF) representation [40]. Cosine similarity scores between token and protocol vectors determine the best match, and tokens are mapped to the protocol with the highest similarity score exceeding a threshold θ.
4. *Primary market tokens:* If none of the above methods yields a mapping, the token is treated as its own protocol (e.g., WETH).

This mapping ensures that token flows between protocols can be attributed to changes in credit exposure with respect to the correct issuing protocol, even when tokens are bridged across chains or wrapped by multiple platforms. This procedure was implemented to curate the DeXposure dataset.

3.3. Network Representation and Value Flows

Credit exposures over a discrete time interval $\tau = t_2 - t_1$ are represented by a weighted directed graph $G_\tau = (P_\tau, E_\tau)$, where P_τ is the set of active

protocols and E_τ is the set of directed edges. Each vertex $p \in P_\tau$ is assigned a node weight equal to the value of assets it holds at the end of the interval:

$$w_\tau(p) = \sum_{\sigma \in X_p(t_1) \cap X_p(t_2)} v_{t_2,\sigma}, \tag{2}$$

where $v_{t_2,\sigma}$ is the USD value of token σ at time t_2. Nodes with negligible weight (below a threshold θ) can be pruned for clarity. To define the edge weight from protocol p to protocol q, we first compute the value flow for each token σ:

$$F^{\sigma}_{pq,\tau} = \begin{cases} \max\big(0,\, -\Delta S^{\sigma}_{p,\tau}\big), & \text{if } \Delta S^{\sigma}_{p,\tau} < 0, \\ \max\big(0,\, \Delta S^{\sigma}_{q,\tau}\big), & \text{if } \Delta S^{\sigma}_{q,\tau} \geq 0, \end{cases} \tag{3}$$

where $\Delta S^{\sigma}_{p,\tau} = v^{\sigma}_{p,t_2} - v^{\sigma}_{p,t_1}$ is the change in the USD value of token σ held by protocol p over the interval. Eq. (3) reflects the intuition that if p decreases its holdings of σ and q increases theirs, then σ has flowed from p to q; negative flows are reversed so that all flows are non-negative. The edge weight $w_\tau(e_{pq})$ is the sum of flows for all tokens that map to the issuing protocol q:

$$w_\tau(e_{pq}) = \sum_{\sigma \in X_{pq,\tau}} F^{\sigma}_{pq,\tau}, \tag{4}$$

where $X_{pq,\tau}$ denotes the set of tokens with issuing protocol q that move from p to q over τ. A positive edge weight indicates increasing exposure from p to q. For single-token protocols, the flow of that token equals the edge weight.

In summary, one aggregates token holdings at the start and end of the interval, computes node weights, discards nodes below a threshold, and then computes edge weights by summing positive flows; negative flows are reversed to ensure all weights are non-negative. The resulting time-indexed sequence of graphs captures the evolving structure of credit dependencies in the DeFi ecosystem. The curation of the DeXposure dataset follows this setting.

4. DeXposure-FM: Architecture and Optimization

This section introduces the DeXposure-FM, including its architecture and the training process.

4.1. Model Architecture

DeXposure-FM is a time-series graph foundation model designed to forecast the evolution of inter-protocol credit exposures. The architecture employs a graph-tabular encoder (GraphPFN in this case) to produce protocol embeddings from the input weekly snapshots, followed by task heads that process embeddings for tasks of (1) edge existence classification, (2) edge weight regression, and (3) node-level TVL changes prediction. Network-level statistics are then obtained by applying deterministic functionals to the predicted graph.

4.1.1. Input and Features

Let $t \in \mathcal{T}$ be discrete observation times and $\tau = (t_1, t_2)$ be a weekly interval. For each τ, we construct a weighted directed exposure graph

$$G_\tau = (\mathcal{P}_\tau, \mathcal{E}_\tau), \tag{5}$$

where nodes $p \in \mathcal{P}_\tau$ are protocols and edges $(p, q) \in \mathcal{E}_\tau$ represent token-mediated exposure from p to issuing protocol q. Node weights encode protocol size (TVL) and edge weights encode the weekly change in exposure induced by token flows and valuation updates. Each node p additionally has a tabular descriptor vector $x^{\text{tab}}_{p,\tau}$ including log-scaled TVL and token-composition summaries (e.g., number of token types held, concentration measures, entropy) and a sector/category one-hot encoding.

4.1.2. Pretrained Graph-Tabular Encoder

We adopt GraphPFN as the encoder $E(\cdot)$ in our model, which is a transformer-based graph-tabular foundation model that fuses graph structure with tabular node features via multi-head self-attention augmented with structural information [20]. Given $(G_\tau, \{x^{\text{tab}}_{p,\tau}\}_{p\in\mathcal{P}_\tau})$, the encoder produces d-dimensional node representations

$$\{h_{p,\tau}\}_{p\in\mathcal{P}_\tau} = E_{\text{GraphPFN}}\big(G_\tau, \{x^{\text{tab}}_{p,\tau}\}_{p\in\mathcal{P}_\tau}\big), \qquad h_{p,\tau} \in \mathbb{R}^d. \tag{6}$$

4.1.3. Edge-Level Prediction Head

For each forecast horizon $h \in \{1, 4, 8, 12\}$ weeks, DeXposure-FM predicts both edge existence and edge weight (i.e., exposure change) for ordered protocol pairs (p, q). We form a pairwise feature vector using a standard symmetric composition of embeddings:

$$f_{pq,\tau} = \Big[h_{p,\tau};\, h_{q,\tau};\, h_{p,\tau} \odot h_{q,\tau};\, |h_{p,\tau} - h_{q,\tau}|\Big], \tag{7}$$

and pass it through a multi-layer perceptron (MLP) head to obtain

$$(\widehat{y}_{pq}^{\text{exist}}, \widehat{r}_{pq,\tau,h}) = \text{MLP}_{\text{link}}(f_{pq,\tau}), \tag{8}$$

where $\widehat{y}_{pq}^{\text{exist}} \in (0,1)$ is the predicted probability of edge existence and $\widehat{r}_{pq,\tau,h} \in \mathbb{R}$ is a residual term. In our implementation, we adopt residual learning for edge weights: let $\tilde{w}_\tau(e_{pq}) = \log(1 + w_\tau(e_{pq}))$; then, we reconstruct the predicted log-weight as

$$\widehat{\tilde{w}}_{\tau+h}(e_{pq}) = \tilde{w}_\tau(e_{pq}) + \widehat{r}_{pq,\tau,h}, \tag{9}$$

with $\tilde{w}_\tau(e_{pq}) = 0$ when the edge is absent at time τ.

4.1.4. Node-Level Prediction Head and Network Statistics

To capture protocol-level dynamics, we predict the log-change in protocol TVL using a node head:

$$\begin{aligned}\widehat{\Delta}_{p,\tau+h} &= \text{MLP}_{\text{node}}\Big([h_{p,\tau};\, h_{p,\tau}^{\text{in}};\, h_{p,\tau}^{\text{out}}]\Big),\\ \widehat{\Delta}_{p,\tau+h} &= \log(1 + w_{\tau+h}(p)) - \log(1 + w_\tau(p)),\end{aligned}$$

for any protocol $p \in \mathcal{P}_\tau \cap \mathcal{P}_{\tau+h}$, where $h_{p,\tau}^{\text{in}}$ and $h_{p,\tau}^{\text{out}}$ denote simple weighted aggregations of incoming/outgoing neighbor embeddings at time τ (using the current-week edge weights), which inject local flow context into the TVL-change prediction.

Network-level statistics (e.g., density, concentration measures, and sector connectivity) are computed as deterministic functionals of the predicted graph $\widehat{G}_{\tau+h}$, enabling evaluation of the model both at the edge level and at the level of aggregate stability indicators.

4.2. Training Pipeline

This section describes how we construct supervised training pairs from weekly DeXposure snapshots, define the multi-task learning objective, and optimize DeXposure-FM.

4.2.1. Data Preparation

DeXposure provides weekly snapshots of the inter-protocol exposure network. For each discrete time index τ, we construct a weighted directed graph $G_\tau = (\mathcal{P}_\tau, \mathcal{E}_\tau)$ with node weights $w_\tau(p)$ (protocol TVL) and edge weights

$w_\tau(e_{pq})$ (exposure from p to q). Training examples are formed by an *anchor* snapshot and a forecast horizon:

$$G_\tau \longrightarrow G_{\tau+h}, \qquad h \in \{1, 4, 8, 12\}, \tag{10}$$

so that the model predicts the future graph state at horizon h conditional on the current snapshot.

For each protocol $p \in \mathcal{P}_\tau$ we provide tabular node features $x_{p,\tau}^{\text{tab}}$ (including log-scaled TVL, token counts, concentration measures, and sector/category indicators) together with the graph structure G_τ.

Table 1 summarizes the dataset characteristics. The network exhibits high edge persistence (mean overlap ratio 98.5%), reflecting the stable nature of DeFi protocol interconnections.

Time-based expanding window splits. We adopt an expanding-window walk-forward evaluation protocol to prevent look-ahead bias: all validation and test targets occur strictly after the corresponding training window, and the training window expands over time while validation and test windows remain fixed.

Negative sampling. Edge existence is highly imbalanced. For each horizon h, we construct a labeled edge set $\mathcal{S}_{\tau+h}$ consisting of all positive edges in $\mathcal{E}_{\tau+h}$ plus uniformly sampled negative edges from the complement, using a negative-to-positive ratio of 5:1. This sampled set is used for the edge existence loss, while the edge weight loss is computed only on positive edges.

4.2.2. Loss Functions

DeXposure-FM is trained with a multi-task objective combining (1) edge existence classification, (2) edge weight regression, and (3) node-level TVL-change prediction.

Edge existence (link prediction). For each candidate pair $(p, q) \in \mathcal{S}_{\tau+h}$ we predict $\widehat{y}_{pq}^{\text{exist}} \in (0, 1)$ and minimize a weighted binary cross-entropy loss:

$$\mathcal{L}_{\text{exist}} = \sum_{(p,q)\in\mathcal{S}_{\tau+h}} \text{BCE}\big(\widehat{y}_{pq}^{\text{exist}},\, y_{pq}^{\text{exist}};\, w_{\text{pos}}\big)\,, \tag{11}$$

where w_{pos} upweights positives to match the negative sampling ratio. Here, binary cross-entropy is defined as below.

Definition 1 (binary cross-entropy (BCE)). *For a binary label $y \in \{0,1\}$ and a predicted probability $\hat{y} \in (0,1)$, the binary cross-entropy loss is*

$$\mathrm{BCE}(\hat{y}, y) = -\Big(y \log(\hat{y}) + (1-y) \log\big(1-\hat{y}\big)\Big). \tag{12}$$

Equivalently, if $\hat{y} = \sigma(z)$ is produced from a logit $z \in \mathbb{R}$ via the sigmoid $\sigma(z) = 1/(1+e^{-z})$, then

$$\mathrm{BCE}(\sigma(z), y) = \log\big(1+e^{z}\big) - yz, \tag{13}$$

which is numerically stable and is often referred to as the logistic loss.

Edge weights (residual learning). For positive edges $(p,q) \in \mathcal{E}_{\tau+h}$, we predict a residual term relative to the previous observed week. Let $\tilde{w}_\tau(e_{pq}) = \log(1 + w_\tau(e_{pq}))$ and $\tilde{w}_{\tau+h}(e_{pq}) = \log(1 + w_{\tau+h}(e_{pq}))$. The link head outputs $\widehat{r}_{pq,\tau,h}$ and we apply a robust Smooth L1 loss:

$$\mathcal{L}_{\text{weight}} = \sum_{(p,q)\in\mathcal{E}_{\tau+h}} \mathrm{SmoothL1}\Big(\widehat{r}_{pq,\tau,h} - \big(\tilde{w}_{\tau+h}(e_{pq}) - \tilde{w}_\tau(e_{pq})\big)\Big). \tag{14}$$

Equivalently, we reconstruct $\widehat{\tilde{w}}_{\tau+h}(e_{pq}) = \tilde{w}_\tau(e_{pq}) + \widehat{r}_{pq,\tau,h}$ and minimize $\mathrm{SmoothL1}(\widehat{\tilde{w}}_{\tau+h}(e_{pq}) - \tilde{w}_{\tau+h}(e_{pq}))$. We set $\tilde{w}_\tau(e_{pq}) = 0$ when the edge is absent at time τ. The $\log(1+\cdot)$ transform stabilizes training under the heavy-tailed exposure distribution. Here, smooth L1 loss is defined as follows.

Definition 2 (smooth L1 loss). *For a scalar prediction $\hat{u} \in \mathbb{R}$ and target $u \in \mathbb{R}$, let the residual be $r = \hat{u} - u$. The Smooth L1 (Huber) loss with threshold $\delta > 0$ is defined as*

$$\mathrm{SmoothL1}_\delta(r) = \begin{cases} \frac{1}{2}\frac{r^2}{\delta}, & \text{if } |r| < \delta, \\ |r| - \frac{1}{2}\delta, & \text{otherwise.} \end{cases} \tag{15}$$

Equivalently, written directly in terms of $(\hat{u}, u)$:

$$\mathrm{SmoothL1}_\delta(\hat{u}, u) = \begin{cases} \frac{1}{2}\frac{(\hat{u}-u)^2}{\delta}, & \text{if } |\hat{u}-u| < \delta, \\ |\hat{u}-u| - \frac{1}{2}\delta, & \text{otherwise.} \end{cases} \tag{16}$$

Smooth L1 behaves like an ℓ_2 loss near zero (encouraging small residuals) and like an ℓ_1 loss for large residuals (reducing sensitivity to heavy tails and outliers). A common choice is $\delta = 1$, when SmoothL1_δ is briefly written as SmoothL1.

Node-level TVL dynamics. We also predict the log-change in protocol TVL, $\Delta_{p,\tau+h} = \log(1 + w_{\tau+h}(p)) - \log(1 + w_\tau(p))$, using Smooth L1:

$$\mathcal{L}_{\text{node}} = \sum_{p \in \mathcal{P}_\tau \cap \mathcal{P}_{\tau+h}} \text{SmoothL1}\Big(\widehat{\Delta}_{p,\tau+h} - \Delta_{p,\tau+h}\Big), \tag{17}$$

where $\Delta_{p,\tau+h} = \log(1 + w_{\tau+h}(p)) - \log(1 + w_\tau(p))$ is the ground-truth log-change in TVL.

Multi-task objective. The total loss is a weighted sum:

$$\mathcal{L} = \lambda_{\text{exist}}\mathcal{L}_{\text{exist}} + \lambda_{\text{weight}}\mathcal{L}_{\text{weight}} + \lambda_{\text{node}}\mathcal{L}_{\text{node}}, \tag{18}$$

with $\lambda_{\text{exist}} = 2.0$, $\lambda_{\text{weight}} = 0.5$, and $\lambda_{\text{node}} = 20.0$. These weights are calibrated to balance the gradient contributions across tasks despite their different loss magnitudes: edge existence loss $\mathcal{L}_{\text{exist}} \approx 0.22$, edge weight loss $\mathcal{L}_{\text{weight}} \approx 2.4$, and node loss $\mathcal{L}_{\text{node}} \approx 0.05$ on average. They also reflect the primary importance of link prediction while using weight and node supervision to encourage economically meaningful representations.

4.2.3. Optimization

Initialization. We adopt the open-source weights of GraphPFN as initialization of the encoder.

Optimizer. We use Adam [27] with hyperparameters $\beta_1 = 0.9$, $\beta_2 = 0.999$ and learning rate $\eta = 5 \times 10^{-4}$.

Training details. We train for up to 20 epochs with early stopping (patience = 3) based on validation AUPRC. Gradient clipping ($\|\nabla\|_2 \leq 1.0$) is applied for stability.

Optional sharpness-aware training. To further improve generalizability, an optional variant wraps the base optimizer with Sharpness-Aware Minimization (SAM), which seeks parameters with lower local sharpness by optimizing against adversarial weight perturbations [23].

5. Empirical Evaluations on Machine Learning Benchmarks

We evaluate DeXposure-FM on two machine learning benchmarks: (1) multi-step forecasting of edge existence, edge weights, and node TVL changes;

and (2) predictive contagion stress testing, where we forecast future simulator-implied stress-test losses under fixed counterfactual shocks. The following Section 6 provides economic interpretation and applications of these predictive outputs.

5.1. Baseline and Competitors

We compare our model with three approaches: GraphPFN: pre-trained GraphPFN encoder with frozen weights; only the following task-specific prediction heads are trained on DeXposure. In contrast, DeXposure-FM uses GraphPFN weights as initialization and fine-tunes the encoder and task heads end-to-end; (2) ROLAND: a representative temporal graph neural network [49], trained from scratch on DeXposure; and (3) a persistent baseline that directly uses the current value at time t as the prediction for a future timing $t + h$, where h is the horizon.

They use the same optimizer, learning rate, and training schedule as DeXposure-FM.

5.2. Task I: Multi-step Forecasting

We first evaluate DeXposure-FM on task of multi-step forecasting.

5.2.1. Implementation Details

We evaluate multi-step forecasts at horizons $h \in \{1, 4, 8, 12\}$ weeks on three targets: (1) *edge existence*, a binary classification task of whether an edge (p, q) is present at $t + h$, (2) *edge weight*, a regression task on log-scaled exposure weight for edges that exist, and (3) *node TVL change*, a regression task. We report AUROC [22] and AUPRC [15] for edge existence and MAE and RMSE for edge weights and ΔTVL (all on the log scale).

5.2.2. Empirical Results

Table 3 reports the forecasting results on the strict 2025 hold-out test set. Despite the strong persistence of the exposure network (Table 1), DeXposure-FM substantially outperforms other methods for edge existence prediction, edge weights, and node TVL changes.

5.3. Task II: Predictive Contagion Stress Testing

We then evaluate our model on whether forward-looking network forecasts translate into accurate forecasts of *stress-test losses* generated by a fixed contagion simulator. Specifically, we compare system losses computed on

the model-predicted future network $\hat{G}_{t+h}$ with losses on the realized future network G_{t+h}, under fixed counterfactual shock definitions.

5.3.1. Implementation Details

We train DeXposure-FM on the pre-2025 training period (Section 4.2) and evaluate on the strict 2025 hold-out set. Each test week provides an origin t, yielding multiple out-of-sample pairs $(t, t+h)$ for horizons $h \in \{1, 4, 8, 12\}$. To handle protocol entry/exit, all comparisons are computed on the common node set $V_t \cap V_{t+h}$.

Stress-test losses are computed using a DebtRank-style contagion simulator [19, 7] under three scenarios: (1) *top protocol shock*: 50% TVL loss to the largest protocol, (2) *top-5 protocols shock*: 30% TVL loss to the top-5 protocols, and (3) *bridge sector shock*: 100% TVL loss to bridge protocols. For each $(t, t+h)$, we run the same simulator on three graphs: the observed network at t (persistence baseline), the predicted network $\hat{G}_{t+h}$, and the realized network G_{t+h}.

Because the DeFi exposure network is highly persistent (Table 1), the persistence baseline is difficult to beat on average. We therefore report both (1) overall performance and (2) a *post-hoc stratified evaluation* on the **worst 20%** of test cases under persistence, defined by the largest baseline absolute errors within each horizon (pooled across shock scenarios). This stratification is computed on the held-out test set and is not used for model selection. We summarize improvement with $\Delta\text{MAE} = \text{MAE(baseline)} - \text{MAE(model)}$, where positive values indicate lower error than persistence.

5.3.2. Empirical Results

Figure 2 shows comparison between three methods, and Figure 3 visualizes ΔMAE in both the overall population and the worst-20% tail. Table 4 reports horizon-level summary statistics.

On average, persistence remains a strong baseline ($\Delta\text{MAE(all)} < 0$, i.e., persistence achieves lower error than the model), reflecting the high week-to-week stability of the DeFi exposure network. The value of a learned forecaster emerges in the tail: among the 20% of test cases where persistence errors are largest, DeXposure-FM achieves positive ΔMAE with win rates of 83–100%. This tail regime, where network structure shifts most from the previous week, is precisely where forward-looking stress tests matter for macroprudential monitoring. The improvement is most pronounced under

bridge-sector shocks (Figure 3), consistent with the sensitivity of cross-chain contagion to evolving network topology.

6. Financial Economics Tools and Experimental Verifications

We now translate our model into financial-economics tools and provide experimental verifications. DeXposure-FM is directly relevant to macroprudential monitoring of decentralized financial infrastructure, addressing the data gaps highlighted.

Concretely, let G_t denote the observed weekly exposure graph and let $\hat{G}_{t+h}$ denote the model's h-step-ahead forecast. We treat each monitoring "tool" as a deterministic functional of a graph (a measurement map) that can be applied either descriptively on G_t or predictively via the *forecast-then-measure* pipeline:

$$G_t \xrightarrow{\text{DeXposure-FM}} \hat{G}_{t+h} \xrightarrow{\text{tools}} \widehat{\mathcal{M}}_{t+h} \equiv \mathcal{M}(\hat{G}_{t+h}), \tag{19}$$

where $\mathcal{M}(\cdot)$ includes protocol-level systemic-importance rankings (SIS), cross-sector spillover concentration, and scenario-based stress-test losses. This framing aligns with the macroprudential objective of producing *actionable, time-t signals* about *time-t + h fragility* (e.g., watchlists, risk channels, and scenario backstops), while remaining fully transparent about the underlying network objects.

6.1. Systemic Risk Measurement: SIS and Spillovers

We first introduce SIS and spillovers as systemic risk measurement.

6.1.1. Implementation

We compute protocol-level and network-level indicators on each weekly exposure snapshot. For protocol p, the SIS combines interconnectedness, exposure concentration, and scale:

$$\text{SIS}_p = \alpha \cdot \widetilde{\text{PageRank}}_p + \beta \cdot \text{TailExposure}_p + \gamma \cdot \log(\widetilde{1 + \text{TVL}_p}), \tag{20}$$

where TailExposure_p is the share of p's top-k outgoing exposures (we use $k = 5$), and α, β, γ are non-negative weights that sum to 1 (default $\alpha = \beta = \gamma = 1/3$). We normalize PageRank [12] and $\log(1 + \text{TVL})$ to comparable scales (denoted by tildes). In parallel, we aggregate edge weights into a

sector-to-sector spillover matrix $\mathbf{S} \in \mathbb{R}^{K \times K}$ where S_{ij} is total exposure from sector i to j, and compute a scale-invariant spillover concentration index as the Herfindahl-Hirschman Index (HHI) [37] of off-diagonal entries:

$$\text{SpilloverIndex} = \text{HHI}\left(\{S_{ij} : i \neq j\}\right). \tag{21}$$

These measures operationalize the network-theoretic insight that interconnectedness, not just size, determines systemic fragility [1, 24]. Weekly SIS rankings help identify protocols whose distress would propagate broadly, while spillover monitoring can highlight sector pairs that act as risk conduits (e.g., stablecoins and bridges) [29, 21, 17].

SIS and spillovers are standard descriptive measures when computed on the observed network G_t. The contribution of DeXposure-FM is to make them *forward-looking*: we first forecast a future exposure graph $\hat{G}_{t+h}$ and then compute

$$\widehat{\text{SIS}}_{t+h}(p) = \text{SIS}_p(\hat{G}_{t+h}),$$

$$\widehat{\text{SpilloverIndex}}_{t+h} = \text{SpilloverIndex}(\hat{G}_{t+h}).$$

This yields actionable forecasts such as (1) a predicted watchlist of systemically important protocols at $t+h$, and (2) predicted cross-sector risk channels.

At each week t, a supervisor computes the current dashboard on G_t and the forecast dashboard on $\hat{G}_{t+h}$. Large predicted changes in the top-K SIS list or spikes in predicted spillover concentration motivate targeted stress scenarios (Section 6.2) and deeper qualitative review (e.g., governance risk and code audit status of newly elevated protocols).

6.1.2. Empirical Validation

Empirically, we validate this "forecast-then-measure" workflow at the level of aggregate risk metrics (including mean SIS and spillover concentration) by comparing values computed on $\hat{G}_{t+h}$ vs. G_{t+h} (Figure 4) and by event-window case studies (Figure 5).

6.2. Stress Testing and Contagion Assessment

Stress testing maps a counterfactual shock scenario into a system-wide loss measure by propagating distress through the exposure network, providing a scenario-specific analogue of macro stress tests in a DeFi setting.

6.2.1. Implementation

We implement a DebtRank-style contagion simulation to assess systemic loss propagation, following the loss-allocation logic [19, 7]. Given an initial shock to protocol p_0 with loss ratio δ_0, we run the following procedure:

1. *Initialize:* $\text{Loss}_{p_0} = \delta_0 \cdot \text{TVL}_{p_0}$.
2. *Propagate:* When a distressed debtor p has loss Loss_p, we allocate its loss to its creditors q proportionally to their exposures E_{qp}:

$$\Delta\text{Loss}_q \;=\; \text{Loss}_p \cdot \frac{E_{qp}}{\sum_{q'} E_{q'p}}. \tag{22}$$

3. *Iterate:* If $\text{Loss}_q > \tau \cdot \text{TVL}_q$ (distress threshold $\tau = 0.1$), protocol q becomes distressed and propagates losses to its creditors. Losses are capped at TVL_q.
4. *Terminate:* When no new protocols become distressed.

We report total system loss (as a percentage of total TVL), contagion depth (propagation rounds), and affected-protocol counts. The three stress scenarios used throughout are: top protocol shock (50% TVL loss), top-5 protocols shock (30%), and bridge sector shock (100%).

6.2.2. Empirical Verification

Section 5.3 evaluates *forward-looking* stress testing; here, we run the same simulator on G_t (persistence baseline), $\hat{G}_{t+h}$ (model forecast), and G_{t+h} (realized future), and then compare the resulting system loss. Because the DeXposure network is highly persistent week-to-week, the persistence baseline is strong *on average*. However, in the subset of weeks where realized losses deviate most from persistence (the worst 20% baseline-error regime), DeXposure-FM provides consistent improvements (Table 4), with gains concentrated in bridge-sector shocks (Figure 3). This supports a macroprudential interpretation: the model is most valuable as a *backstop* for scenario analysis under non-stationary reallocation of exposures.

6.3. Forward-looking Risk Metric Forecasting

Aggregate risk metrics (e.g., concentration, density, spillovers) serve as dashboard-style indicators of evolving fragility. Forecasting these aggregates provides a practical way to monitor trend shifts and calibrate the forecast-then-measure pipeline.

6.3.1. Implementation

Beyond stress testing, we can forecast aggregate stability indicators by first predicting a future exposure graph $\hat{G}_{t+h}$ and then computing deterministic risk metrics on it (e.g., TVL concentration, edge concentration, network density, spillover index, and summary statistics of SIS). We compare these predicted metrics to their realized counterparts on G_{t+h} across the 2025 holdout set.

6.3.2. Empirical Verification

Figure 4 reports predicted vs. realized risk metrics for $h \in \{1, 4, 8, 12\}$. We interpret these plots as a *calibration and monitoring* check: even when persistence dominates levels, the predicted metrics can flag directional changes and large deviations that motivate deeper scenario analysis and qualitative review.

6.4. Early Warning Signals and Event Studies

Early warning analysis evaluates whether forecast-based measurements shift in advance of major stress events, providing event-study evidence that the forecast-then-measure tools can produce timely alerts around regime changes.

6.4.1. Implementation

We examine two major stress events: the Terra/Luna collapse (May 9, 2022) and the FTX collapse (November 7, 2022). To avoid event leakage, for each event, we train an event-specific forecaster using only snapshots strictly prior to the event window, then generate one-step-ahead forecasts within the event window. We track structural indicators and stress-test outcomes (via Section 6.2) and monitor potential early warning features:

- Changes in network connectivity (e.g., density ρ_t),
- Concentration shifts (e.g., TVL HHI), with practical threshold rules such as $\Delta\text{HHI}_t > 2\sigma$,
- Predicted stress-test losses under standard scenarios (Section 6.2).

6.4.2. Empirical Verification

Table 5 summarizes structural shifts during the event windows, and Figure 5 compares predicted vs. realized trajectories for risk concentration and stress-test loss. These case studies illustrate how model-based forecasting can complement descriptive monitoring by providing forward-looking alerts around regime changes.

6.5. Validity and Scope

Model-based risk measures are only as reliable as the conditions under which they were trained and validated. We report several boundaries that users should keep in mind.

6.5.1. Stable vs. Turbulent Regimes

The DeXposure dataset exhibits 98.5% mean edge overlap week-to-week, meaning most credit relationships persist. Under such stable conditions, GraphPFN achieves an AUROC larger than 0.91 across all forecast horizons; see Table 3. During structural breaks, such as new protocol launches, governance attacks, or rapid deleveraging, the historical edge distribution may shift abruptly, and forecasts should be interpreted with greater caution. The degradation of ROLAND illustrates how models without pre-trained representations struggle when asked to extrapolate beyond short horizons.

6.5.2. Data Coverage

DeXposure-FM is trained exclusively on on-chain DeFi protocols sourced from DeFiLlama. It does not observe:

- Centralized exchange (CEX) balances or order-book exposures,
- Over-the-counter (OTC) derivatives or lending agreements,
- Off-chain reserves backing stablecoins (e.g., Treasury holdings of USDC).

Consequently, the model captures intra-DeFi contagion but cannot directly assess spillovers to or from traditional financial institutions, a gap that matters as DeFi-TradFi linkages deepen.

6.5.3. Temporal Resolution

The training dataset aggregates to weekly snapshots. Intraday or even daily dynamics, such as flash-loan attacks [35] or rapid liquidation cascades [46], are smoothed out. For high-frequency risk monitoring, complementary tools operating on block-level data would be required.

6.5.4. Cross-chain Data Consistency

DeXposure aggregates data across 602 blockchains, each with its own indexing quirks and oracle delays. Minor inconsistencies in timestamp alignment or token-price feeds can introduce noise. Users should be aware that edge weights near reporting boundaries may reflect data artifacts rather than genuine exposure changes.

6.5.5. TVL as a Measurement Proxy

TVL aggregates heterogeneous assets without risk weights, a limitation we discuss in Section 8. Nevertheless, TVL remains the main metric for assessing DeFi protocol significance [4], and its changes over time capture the capital flows that drive contagion [50]. The tools are therefore most informative for relative monitoring (rankings, concentration changes, scenario comparisons) rather than absolute loss accounting.

6.5.6. Our aim: Complement, not replace, qualitative judgment

Model outputs such as SIS rankings and spillover indices are quantitative summaries, not definitive verdicts. A protocol may rank low on SIS yet pose idiosyncratic risks (e.g., smart-contract vulnerabilities, oracle manipulation) that network topology alone cannot capture [3, 25]. Supervisors should triangulate model signals with code audits, governance analyses, and market intelligence.

7. Future Development Plan

We plan to continuously update and improve DeXposure-FM. Below is our future development plan.

7.1. Enlarging Training Data Pool

The current DeXposure-FM is trained on weekly inter-protocol snapshots from the DeXposure dataset, using a GraphPFN encoder as pretrained initialization. While this provides broad on-chain coverage at a stable temporal resolution, it omits key off-chain channels through which DeFi is funded, hedged, and linked to traditional markets. In DeXposure-FM 2.0, we plan to expand the training pool by incorporating additional data sources, including centralized exchange (CEX) balances and order-book–implied exposures, over-the-counter (OTC) derivatives and lending agreements, and disclosures on off-chain stablecoin reserves (e.g., USDC's Treasury-backed holdings). We

also aim to add higher-frequency data (daily or intraday, where feasible), particularly around stress episodes, to better capture rapid deleveraging and liquidity dynamics that weekly aggregation smooths out.

7.2. Expanding Proxies for Credit Exposure

Total Value Locked (TVL) is a convenient but imperfect proxy for economic exposure. It aggregates heterogeneous assets into a single dollar value without adjusting for liquidity, haircuts, or collateral quality. As a result, equal TVL can imply very different loss-absorbing capacity and contagion risk; for example, $1B held in thinly traded governance tokens is not economically equivalent to $1B in USDC, and TVL can also move mechanically with prices even when underlying positions are unchanged. To address these limitations, DeXposure-FM 2.0 will incorporate asset-level risk weights and composition features (e.g., liquidity and volatility proxies, depeg risk, and observable collateral rules) to produce risk-adjusted exposure estimates and more interpretable systemic-risk indicators.

7.3. Model Drift and Continuous Update

DeFi evolves rapidly: new chains launch, token standards change, and incentive mechanisms shift. This non-stationarity implies that a model trained on historical regimes may gradually lose accuracy as exposures and flows depart from past patterns, including both gradual structural change and discrete breaks from major upgrades or new collateral rules. Periodic retraining on updated snapshots is therefore essential to maintain forecast quality, ideally paired with drift monitoring (e.g., calibration and error shifts by sector/chain) and versioned releases so that risk signals can be traced to specific data vintages and model iterations.

7.4. Innovative Model Architecture

The future DeXposure-FM 2.0 will include an innovative hybrid architecture that combines the current direct multi-horizon forecasting paradigm [36] with diffusion-based generative modeling [32]. Auto-regressive models are efficient and accurate for point and conditional forecasts, but they can struggle to represent globally coherent network trajectories over long horizons, especially under stress. Diffusion models, by contrast, provide a principled way to learn complex conditional distributions by gradually denoising samples, and have recently shown strong performance in structured generation tasks, such as in diffusion language models [31].

7.5. Open Source, Competitions, and Community Development

We endeavor to develop DeXposure-FM to become fundamental infrastructure rather than a closed research artifact. We have released an open-source version of model weights and training code. We will continue our open-source efforts when improving the model. Based on the DeXposure dataset and DeXposure-FM model, we will organize public benchmarking competitions. These challenges will be designed to encourage robust methods and will include baseline implementations, leaderboards, and model cards reporting calibration and failure modes. We also aim to build a sustained community around DeFi economic measurement. By lowering barriers to entry and making the measurement process transparent, we hope to accelerate cumulative progress on reliable, policy-relevant tools for monitoring systemic risk in decentralized financial networks.

8. Conclusions

Credit exposures in decentralized finance (DeFi) are often embedded in token holdings and smart-contract interactions, rather than recorded as explicit bilateral claims, forming a tightly coupled network of inter-protocol dependencies. In such an environment, a disturbance to a single token can propagate across protocols, generating sizable and difficult-to-contain contagion. As DeFi becomes more intertwined with traditional financial infrastructure, most prominently via stablecoins, this evolving risk landscape calls for stronger tools to measure and forecast exposures. We introduce DeXposure-FM, which is arguably the first time-series, graph foundation model, designed for DeFi. DeXposure-FM combines a graph–tabular encoder with pretrained weights as initialization and multiple task-specific heads. It is trained on the DeXposure dataset, comprising 43.7 million observations spanning 4,300+ protocols on 602 blockchains and 24,300+ unique tokens. The model is trained for credit-exposure forecasting, jointly learning the dynamics of protocol-level flows and the evolving topology and weights of exposure links, from which sector-level spillovers and other network statistics are derived. Across a suite of machine-learning benchmarks, DeXposure-FM consistently outperforms strong competitors, including a prior graph foundation model and state-of-the-art temporal graph neural networks. We also translate the model outputs into economically interpretable indicators, including time-varying measures of systemic importance, sector-level spillover,

and early-warning signals based on shifts in network concentration and reliance on fragile collateral, which illustrate applications to macroprudential monitoring, DeFi stress testing, and counterfactual policy analysis. The tools are fully supported in the empirical validations. The model and code have been publicly available.

References

[1] Acemoglu, D., Ozdaglar, A., Tahbaz-Salehi, A., 2015. Systemic risk and stability in financial networks. American Economic Review 105, 564–608. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20130456, doi:10.1257/aer.20130456.

[2] Ansari, A.F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S.S., Pineda Arango, S., Kapoor, S., Zschiegner, J., Maddix, D.C., Wang, H., Mahoney, M.W., Torkkola, K., Wilson, A.G., Bohlke-Schneider, M., Wang, Y., 2024. Chronos: Learning the language of time series. URL: https://arxiv.org/abs/2403.07815, arXiv:2403.07815.

[3] Aspembitova, A.T., Bentley, M.A., 2023. Oracles in decentralized finance: Attack costs, profits and mitigation measures. Entropy 25. URL: https://www.mdpi.com/1099-4300/25/1/60, doi:10.3390/e25010060.

[4] Auer, R., Faragò, M., Turi, D., 2025. Towards Verifiability of Total Value Locked (TVL) in Decentralized Finance. BIS Working Papers 1268. Bank for International Settlements. URL: https://www.bis.org/publ/work1268.htm.

[5] Aufiero, S., Bartolucci, S., Caccioli, F., Vivo, P., 2025. Mapping microscopic and systemic risks in tradfi and defi: a literature review. URL: https://arxiv.org/abs/2508.12007, arXiv:2508.12007.

[6] Balmaseda, V., Coronado, M., de Cadenas-Santiago, G., 2023. Predicting systemic risk in financial systems using deep graph learning. Intelligent Systems with Applications 19, 200240. URL: https://www.sciencedirect.com/science/article/pii/S2667305323000650, doi:10.1016/j.iswa.2023.200240.

[7] Battiston, S., Puliga, M., Kaushik, R., Tasca, P., Caldarelli, G., 2012. Debtrank: Too central to fail? financial networks, the fed and systemic risk. Scientific Reports 2, 541. URL: https://www.nature.com/articles/srep00541, doi:10.1038/srep00541.

[8] Bergmeir, C., Benítez, J.M., 2012. On the use of cross-validation for time series predictor evaluation. Information Sciences 191, 192–213. URL: https://www.sciencedirect.com/science/article/pii/S0020025511006773, doi:10.1016/j.ins.2011.12.028. data Mining for Software Trustworthiness.

[9] Bertomeu, J., Martin, X., Sall, I., 2024. Measuring defi risk. Finance Research Letters 63, 105321. URL: https://www.sciencedirect.com/science/article/pii/S1544612324003519, doi:10.1016/j.frl.2024.105321.

[10] Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J.Q., Demszky, D., Donahue, C., Doumbouya, M., Durmus, E., Ermon, S., Etchemendy, J., Ethayarajh, K., Fei-Fei, L., Finn, C., Gale, T., Gillespie, L., Goel, K., Goodman, N., Grossman, S., Guha, N., Hashimoto, T., Henderson, P., Hewitt, J., Ho, D.E., Hong, J., Hsu, K., Huang, J., Icard, T., Jain, S., Jurafsky, D., Kalluri, P., Karamcheti, S., Keeling, G., Khani, F., Khattab, O., Koh, P.W., Krass, M., Krishna, R., Kuditipudi, R., Kumar, A., Ladhak, F., Lee, M., Lee, T., Leskovec, J., Levent, I., Li, X.L., Li, X., Ma, T., Malik, A., Manning, C.D., Mirchandani, S., Mitchell, E., Munyikwa, Z., Nair, S., Narayan, A., Narayanan, D., Newman, B., Nie, A., Niebles, J.C., Nilforoshan, H., Nyarko, J., Ogut, G., Orr, L., Papadimitriou, I., Park, J.S., Piech, C., Portelance, E., Potts, C., Raghunathan, A., Reich, R., Ren, H., Rong, F., Roohani, Y., Ruiz, C., Ryan, J., Ré, C., Sadigh, D., Sagawa, S., Santhanam, K., Shih, A., Srinivasan, K., Tamkin, A., Taori, R., Thomas, A.W., Tramèr, F., Wang, R.E., Wang, W., Wu, B., Wu, J., Wu, Y., Xie, S.M., Yasunaga, M., You, J., Zaharia, M., Zhang, M., Zhang, T., Zhang, X., Zhang, Y., Zheng, L., Zhou, K., Liang, P., 2021. On the opportunities and risks of foundation models. URL: https://arxiv.org/abs/2108.07258, arXiv:2108.07258.

[11] de Bondt, G.J., Sun, Y., 2025. Enhancing GDP nowcasts with ChatGPT: a novel application of PMI news releases. Working Paper Series 3063. European Central Bank. Frankfurt am Main, Germany. URL: https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp3063~f88c1b73fc.en.pdf, doi:10.2866/2788332. eCB Working Paper Series No. 3063; released 2025-06-30.

[12] Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems 30, 107–117. URL: https://www.sciencedirect.com/science/article/pii/S016975529800110X, doi:10.1016/S0169-7552(98)00110-X. proceedings of the Seventh International World Wide Web Conference.

[13] Carriero, A., Pettenuzzo, D., Shekhar, S., 2025. Macroeconomic forecasting with large language models. URL: https://arxiv.org/abs/2407.00890, arXiv:2407.00890.

[14] Das, A., Kong, W., Sen, R., Zhou, Y., 2024. A decoder-only foundation model for time-series forecasting. URL: https://arxiv.org/abs/2310.10688, arXiv:2310.10688. accepted at ICML 2024; model commonly referenced as TimesFM.

[15] Davis, J., Goadrich, M., 2006. The relationship between precision-recall and roc curves, in: Proceedings of the 23rd International Conference on Machine Learning, Association for Computing Machinery, New York, NY, USA. p. 233–240. URL: doi.org/10.1145/1143844.1143874, doi:10.1145/1143844.1143874.

[16] DefiLlama, 2026. DefiLlama: DeFi TVL aggregator and analytics. URL: https://defillama.com. data used: 2025; accessed: 2026-02-02.

[17] Diebold, F.X., Yilmaz, K., 2012. Better to give than to receive: Predictive directional measurement of volatility spillovers. International Journal of Forecasting 28, 57–66. URL: https://www.sciencedirect.com/science/article/pii/S016920701100032X, doi:10.1016/j.ijforecast.2011.02.006. special Section 1: The Predictability of Financial Markets Special Section 2: Credit Risk Modelling and Forecasting.

[18] Dolfin, M., Knopoff, D., Limosani, M., Xibilia, M.G., 2019. Credit risk contagion and systemic risk on networks. Mathematics 7. URL: https://www.mdpi.com/2227-7390/7/8/713, doi:10.3390/math7080713.

[19] Eisenberg, L., Noe, T.H., 2001. Systemic risk in financial systems. Management Science 47, 236–249. URL: https://pubsonline.informs.org/doi/10.1287/mnsc.47.2.236.9835, doi:10.1287/mnsc.47.2.236.9835.

[20] Eremeev, D., Bazhenov, G., Platonov, O., Babenko, A., Prokhorenkova, L., 2025. Turning tabular foundation models into graph foundation models. URL: https://arxiv.org/abs/2508.20906, arXiv:2508.20906.

[21] European Systemic Risk Board, 2025. Crypto-assets and decentralised finance: Report on stablecoins, crypto-investment products and multi-function groups. Technical Report. European Systemic Risk Board. URL: https://www.esrb.europa.eu/pub/pdf/reports/esrb.report202510_cryptoassets.en.pdf. accessed 27 December 2025.

[22] Fawcett, T., 2006. An introduction to roc analysis. Pattern Recognition Letters 27, 861–874. URL: https://www.sciencedirect.com/science/article/pii/S016786550500303X, doi:10.1016/j.patrec.2005.10.010. rOC Analysis in Pattern Recognition.

[23] Foret, P., Kleiner, A., Mobahi, H., Neyshabur, B., 2021. Sharpness-aware minimization for efficiently improving generalization, in: International Conference on Learning Representations.

[24] Glasserman, P., Young, H.P., 2016. Contagion in financial networks. Journal of Economic Literature 54, 779–831. URL: https://www.aeaweb.org/articles?id=10.1257/jel.20151228, doi:10.1257/jel.20151228.

[25] Gogol, K., Killer, C., Schlosser, M., Bocek, T., Stiller, B., Tessone, C., 2024. Sok: Decentralized finance (defi) – fundamentals, taxonomy and risks. URL: https://arxiv.org/abs/2404.11281, arXiv:2404.11281.

[26] Gonon, L., Meyer-Brandis, T., Weber, N., 2025. Computing systemic risk measures with graph neural networks. URL: https://arxiv.org/abs/2410.07222, arXiv:2410.07222.

[27] Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization, in: Proceedings of the 3rd International Conference on Learning Representations (ICLR). URL: https://arxiv.org/abs/1412.6980.

[28] Kipf, T.N., Welling, M., 2017. Semi-supervised classification with graph convolutional networks. URL: https://arxiv.org/abs/1609.02907, arXiv:1609.02907.

[29] Kosse, A., Glowka, M., Mattei, I., Rice, T., 2023. Will the Real Stablecoin Please Stand Up? BIS Papers 141. Bank for International Settlements. URL: https://www.bis.org/publ/bppdf/bispap141.pdf. bIS Papers No. 141; accessed 2025-12-27.

[30] Kwon, B., Park, T., Perez-Cruz, F., Rungcharoenkitkul, P., 2024. Large language models: a primer for economists. BIS Quarterly Review URL: https://www.bis.org/publ/qtrpdf/r_qt2412b.htm.

[31] Li, X.L., Thickstun, J., Gulrajani, I., Liang, P., Hashimoto, T.B., 2022. Diffusion-LM improves controllable text generation, in: Advances in Neural Information Processing Systems, pp. 4328–4343. URL: https://arxiv.org/abs/2205.14217.

[32] Meijer, C., Chen, L.Y., 2024. The rise of diffusion models in time-series forecasting. URL: https://arxiv.org/abs/2401.03006. arXiv:2401.03006.

[33] Pascanu, R., Mikolov, T., Bengio, Y., 2013. On the difficulty of training recurrent neural networks, in: Dasgupta, S., McAllester, D. (Eds.), Proceedings of the 30th International Conference on Machine Learning, PMLR, Atlanta, Georgia, USA. pp. 1310–1318. URL: https://proceedings.mlr.press/v28/pascanu13.html.

[34] Prechelt, L., 1998. Early Stopping - But When?. Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 55–69. URL: 0.1007/3-540-49430-8_3, doi:10.1007/3-540-49430-8_3.

[35] Qin, K., Zhou, L., Livshits, B., Gervais, A., 2021. Attacking the defi ecosystem with flash loans for fun and profit, in: Financial Cryptography and Data Security: 25th International Conference, FC 2021, Virtual Event, March 1–5, 2021, Revised Selected Papers, Part I, Springer-

Verlag, Berlin, Heidelberg. p. 3–32. URL: https://doi.org/10.1007/978-3-662-64322-8_1, doi:10.1007/978-3-662-64322-8_1.

[36] Rasul, K., Ashok, A., Williams, A.R., Ghonia, H., Bhagwatkar, R., Khorasani, A., Bayazi, M.J.D., Adamopoulos, G., Riachi, R., Hassen, N., Biloš, M., Garg, S., Schneider, A., Chapados, N., Drouin, A., Zantedeschi, V., Nevmyvaka, Y., Rish, I., 2024. Lag-llama: Towards foundation models for probabilistic time series forecasting. URL: https://arxiv.org/abs/2310.08278, arXiv:2310.08278.

[37] Rhoades, S.A., 1993. The herfindahl-hirschman index. Federal Reserve Bulletin 79, 188–189. URL: https://ideas.repec.org/a/fip/fedgrb/y1993imarp188-189nv.79no.3.html.

[38] Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., Bronstein, M., 2020. Temporal graph networks for deep learning on dynamic graphs. URL: https://arxiv.org/abs/2006.10637, arXiv:2006.10637.

[39] Schär, F., 2021. Decentralized finance: On blockchain- and smart contract-based financial markets. Federal Reserve Bank of St. Louis Review 103, 153–174. URL: https://www.stlouisfed.org/publications/review/2021/02/05/decentralized-finance-on-blockchain-and-smart-contract-based-financial-market doi:10.20955/r.103.153-74.

[40] Shahzad Qaiser, R.A., 2018. Text mining: Use of tf-idf to examine the relevance of words to documents. International Journal of Computer Applications 181, 25–29. URL: https://ijcaonline.org/archives/volume181/number1/29681-2018917395/, doi:10.5120/ijca2018917395.

[41] Sims, C.A., 1980. Macroeconomics and reality. Econometrica 48, 1–48. URL: http://www.jstor.org/stable/1912017.

[42] Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179. URL: https://doi.org/10.1198/016214502388618960, doi:10.1198/016214502388618960, arXiv:10.1198/016214502388618960.

[43] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I., 2017. Attention is all you need, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[44] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y., 2018. Graph attention networks. URL: https://arxiv.org/abs/1710.10903, arXiv:1710.10903.

[45] Vidal-Tomás, D., Briola, A., Aste, T., 2023. Ftx's downfall and binance's consolidation: The fragility of centralised digital finance. Physica A: Statistical Mechanics and its Applications 625, 129044. URL: https://www.sciencedirect.com/science/article/pii/S037843712300599X, doi:10.1016/j.physa.2023.129044.

[46] Warmuz, J., Chaudhary, A., Pinna, D., 2023. Toxic liquidation spirals. URL: https://arxiv.org/abs/2212.07306, doi:10.48550/arXiv.2212.07306, arXiv:2212.07306.

[47] Wu, W., Qian, K., Lui, A., Jack, C., Wu, Y., McBurney, P., He, F., Zhang, B., 2025. Dexposure: A dataset and benchmarks for inter-protocol credit exposure in decentralized financial networks. URL: https://arxiv.org/abs/2511.22314, arXiv:2511.22314.

[48] Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., Shen, Y., Liu, T.Y., 2021. Do transformers really perform badly for graph representation?, in: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc.. pp. 28877–28888. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/f1c1592588411002af340cbaedd6fc33-Paper.pdf.

[49] You, J., Du, T., Leskovec, J., 2022. ROLAND: Graph learning framework for dynamic graphs, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 2358–2366. URL: https://arxiv.org/abs/2208.07239, doi:10.1145/3534678.3539300.

[50] Zhang, S., Wang, Z., Zheng, J., Cartlidge, J., 2026. Systemic risk in DeFi: A network-based fragility analysis of TVL dynamics. arXiv preprint arXiv:2601.08540 URL: https://arxiv.org/abs/2601.08540.

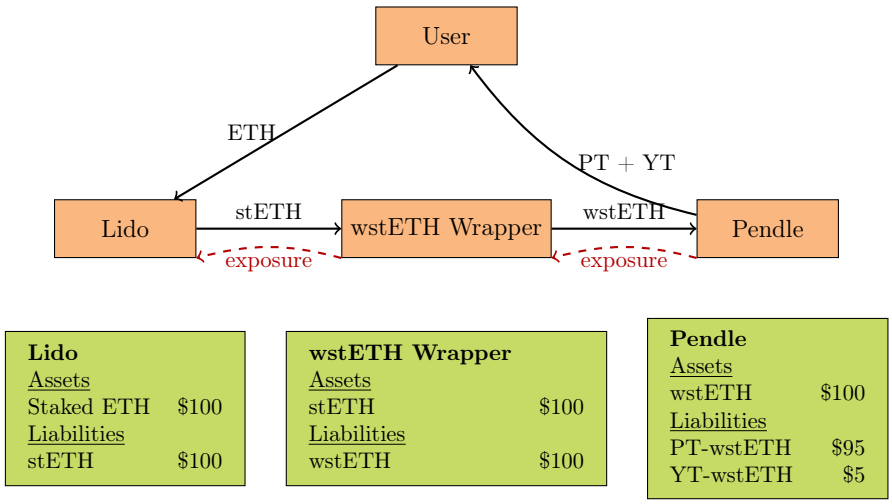

Figure 1: Multi-layer credit exposure in DeFi. A user stakes ETH with Lido, receiving stETH, which is wrapped into wstETH and deposited into Pendle. Each protocol holds the liability of the preceding one, creating a chain of credit exposures from Pendle to Lido.

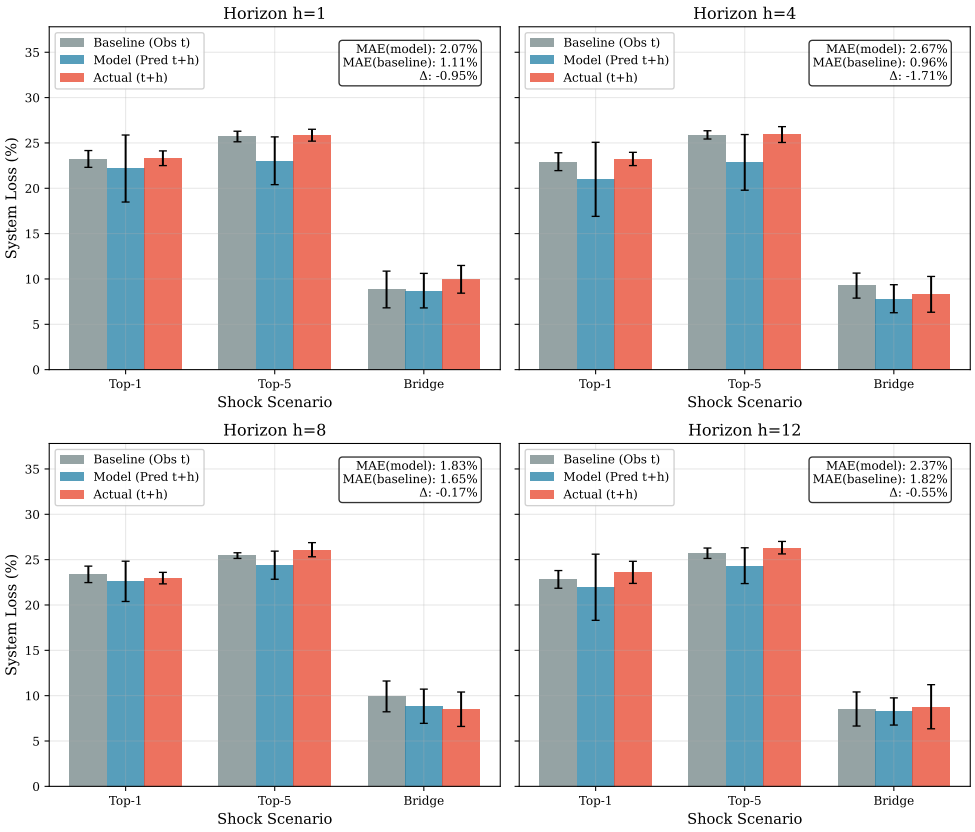

Figure 2: Predictive contagion stress testing (Task II). We compare system loss (%) under identical shocks when running the contagion simulator on the observed network at time t (persistence baseline), the model-predicted network $\hat{G}_{t+h}$, and the realized future network G_{t+h}.

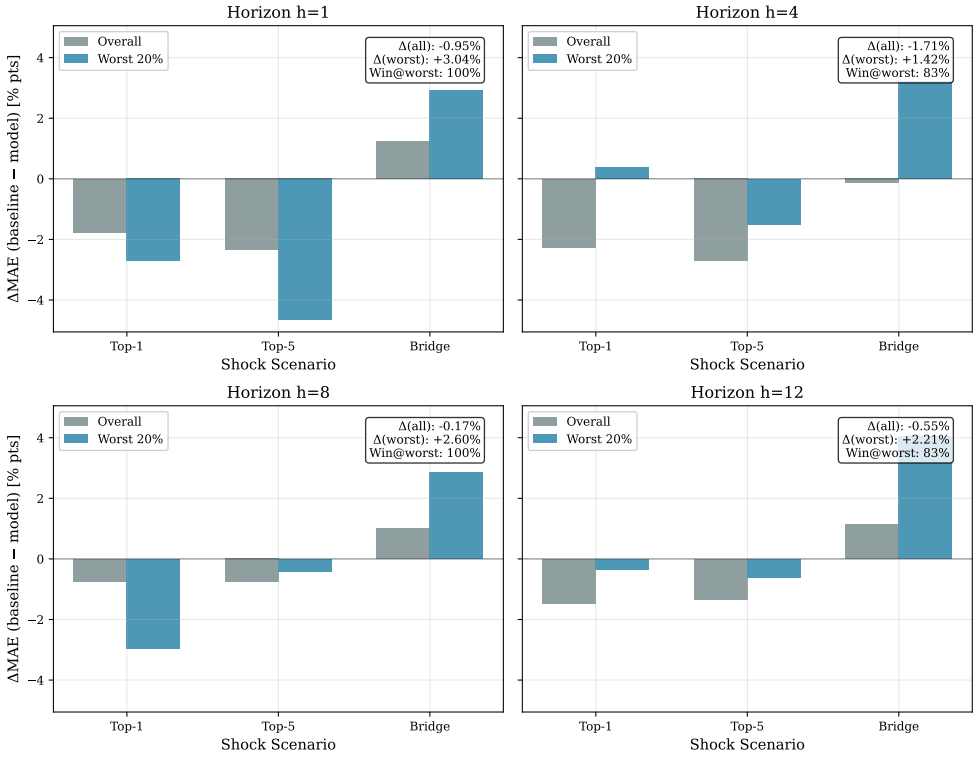

Figure 3: Advantage regimes for predictive stress testing. Bars report $\Delta\text{MAE} = \text{MAE(baseline)} - \text{MAE(model)}$ in system loss (% points). "Overall" is the full test set; "Worst 20%" restricts to the 20% largest baseline errors (tail regime). Positive values indicate DeXposure-FM outperforms persistence, highlighting where a learned forecaster adds value when the persistence assumption breaks.

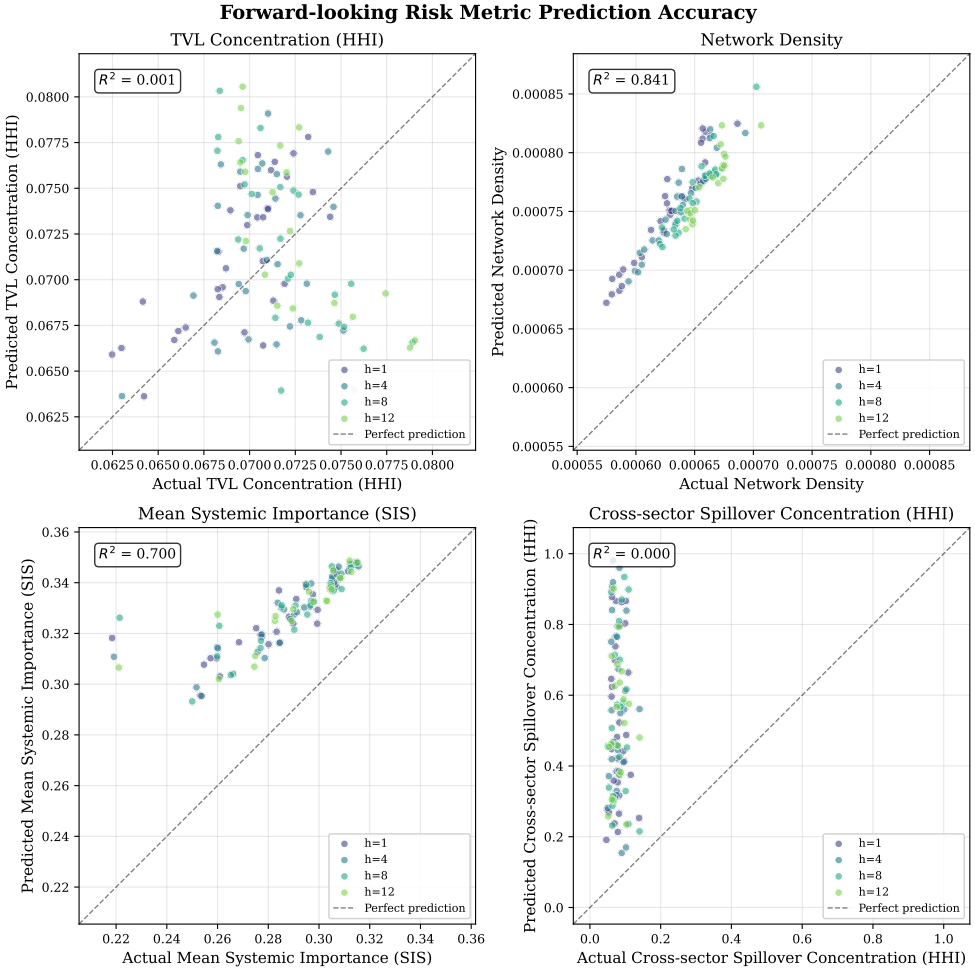

Figure 4: Forward-looking risk metric forecasting on predicted exposure graphs. Each point compares a realized metric on G_{t+h} (x-axis) to the same metric computed on the predicted graph $\hat{G}_{t+h}$ (y-axis), across horizons.

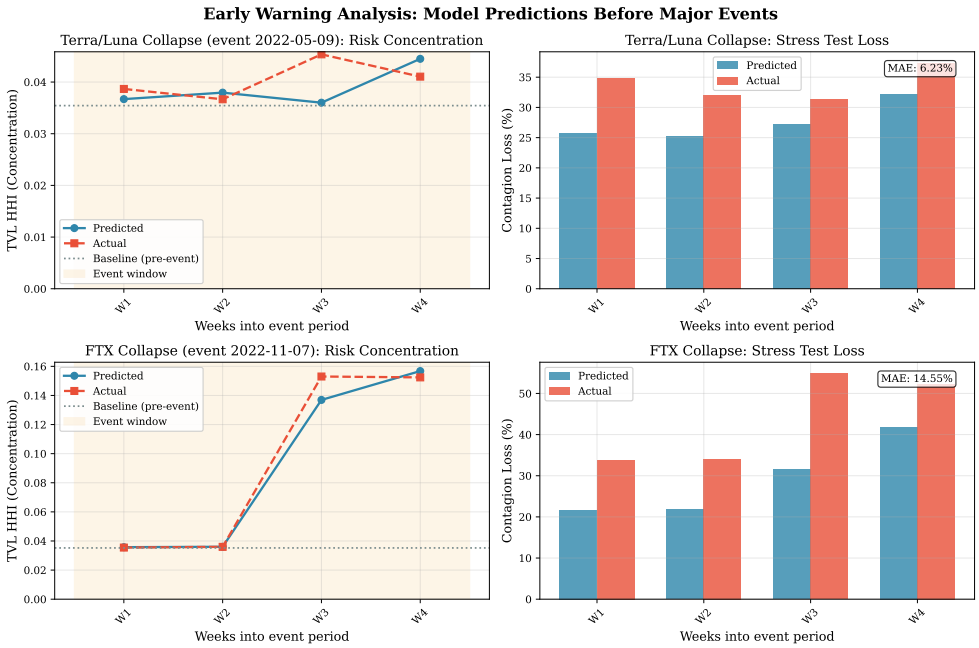

Figure 5: Early warning event studies. For each event window, we compare predicted vs. realized risk concentration (HHI) and predictive stress-test loss (contagion system loss %).

Table 1: DeXposure dataset statistics.

Statistic	**Value**
Total snapshots	283 weeks
Time range	2020-03-23 $\sim$ 2025-08-18
Mean nodes per week	5,676
Mean edges per week	30,424
Mean edge overlap ratio	98.5%
Protocol categories	$\sim$ 15 classes

Table 2: Training hyperparameters for DeXposure-FM experiments.

Hyperparameter	**Value**
Data granularity	Weekly snapshots
Forecast horizons h	$\{1, 4, 8, 12\}$ weeks
Negative sampling ratio	5:1 (neg:pos)
Optimizer	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Learning rate (heads)	5×10^{-4}
Learning rate (backbone)	5×10^{-5} (fine-tune only)
Training epochs	20 (early stopping patience=3)
Gradient clipping	$\|\nabla\|_2 \leq 1.0$
Loss weights:	
λ_{exist} (edge existence)	2.0
λ_{weight} (edge weight)	0.5
λ_{node} (node prediction)	20.0

Table 3: Multi-step forecasting results on 2025 hold-out test set at horizons $h \in \{1, 4, 8, 12\}$ weeks. DeXposure-FM = Finetuned Foundation Model, GraphPFN = Graph Prior Fitted Network with Frozen weights. *Persistence* = naive baseline predicting $\hat{A}_{t+h} = A_t$, $\hat{W}_{t+h} = W_t$.

		Edge Exist		**Edge Weight**		**Node ΔTVL**	
Model	h	AUROC	AUPRC	MAE	RMSE	MAE	RMSE
DeXposure-FM	1	**0.995**	**0.972**	**2.465**	**3.388**	**0.056**	**0.400**
	4	**0.995**	**0.973**	**2.489**	**3.424**	**0.140**	**0.680**
	8	**0.994**	**0.967**	**2.554**	**3.509**	**0.229**	**0.890**
	12	**0.993**	**0.967**	**2.648**	**3.606**	**0.286**	**1.046**
GraphPFN	1	0.988	0.938	3.260	4.383	0.059	0.401
	4	0.988	0.940	3.189	4.049	0.142	0.682
	8	0.987	0.938	3.169	4.064	0.215	0.896
	12	0.986	0.936	3.136	4.130	0.324	1.056
ROLAND	1	0.961	0.865	3.240	4.264	0.060	0.403
	4	0.962	0.868	3.242	4.162	0.141	0.684
	8	0.961	0.867	3.213	4.180	0.221	0.895
	12	0.961	0.866	3.195	4.177	0.279	1.058
Persistence	1	0.763	0.604	2.487	4.296	0.057	0.403
	4	0.782	0.635	2.372	4.138	0.138	0.685
	8	0.749	0.580	2.618	4.400	0.213	0.899
	12	0.762	0.603	2.541	4.304	0.272	1.065

Table 4: Predictive stress testing summary (2025 hold-out). ΔMAE is measured in system-loss percentage points. Worst20% selects the 20% largest persistence-baseline errors within each horizon (pooled across shock scenarios). "Win rate" is the fraction of worst20% cases where the model has lower absolute error than persistence.

Horizon h	ΔMAE(all)	ΔMAE(worst20%)	Win rate (worst20%)
1	−0.95	+3.04	100%
4	−1.71	+1.42	83%
8	−0.17	+2.60	100%
12	−0.55	+2.21	83%

Table 5: Network structure changes during stress events, comparing the last pre-event week (anchor) to the final week of the event window. Terra/Luna shows a flight-to-quality pattern with TVL collapse but edge count increase as capital reorganizes across protocols. The FTX event shows a *positive* ΔTVL because our dataset captures only on-chain DeFi protocols: the collapse of a centralized exchange triggered capital migration *into* DeFi as users withdrew funds to self-custody, consistent with the "flight to decentralization" narrative [45].

Event	**Pre-TVL ($B)**	**Post-TVL ($B)**	**ΔTVL**	**ΔEdges**
Terra/Luna	261.9	166.3	-36.5%	$+9.3\%$
FTX	119.9	175.1	$+46.1\%$	-1.8%