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

Aijie Shu [1] **Wenbin Wu** [2] **Gbenga Ibikunle** [3] **Fengxiang He** [1]

Abstract

Credit exposure in Decentralized Finance (DeFi) is implicit and token-mediated, so shocks to widely held tokens can cascade across protocols and chains. We introduce DeXposure-FM, a domain-specific time-series graph foundation model (GFM) for forecasting inter-protocol credit exposure. A GraphPFN graph-tabular encoder is fine-tuned end-to-end on 43.7M DeXposure observations covering 4,300+ protocols, 602 chains, and 24,300+ tokens, and exposes the learned representation through lightweight multi-horizon forecasting heads whose predicted graphs feed a shared downstream measurement layer. On multi-step forecasting and predictive contagion stress testing, DeXposure-FM improves edge-existence prediction and reduces magnitude RMSE over neural baselines, while persistence remains competitive on stable magnitudes. Forecasted graphs further support macroprudential indicators, including systemic-importance scores, sector spillovers, concentration, and scenario stress losses. Gains concentrate in structural-change and tail-error regimes where carry-forward assumptions are least reliable.

Artifacts. Code is available on GitHub; model checkpoints are available on Hugging Face.

1. Introduction

Decentralised Finance (DeFi) has become a fast-evolving financial ecosystem offering lending, trading, derivatives and payment-like services through smart contracts on hundreds of blockchains (Schär, 2021; Gogol et al., 2024). A defining feature is *composability*: protocols routinely hold tokens issued by other protocols as collateral, liquidity, or reserves, so credit exposure is implicit and token-mediated rather than expressed as bilateral contracts (Wu et al., 2025; Bertomeu et al., 2024). When a token used widely as collateral suffers a shock, losses can cascade through every protocol that depends on it, often across chains (Aufiero et al., 2025). Policy authorities have repeatedly highlighted the vulnerabilities of these networks and stressed the persistent shortage of quantitative tools that would enable systematic monitoring (Saggese et al., 2025; European Systemic Risk Board, 2025).

Network primitives. A DeFi *protocol* is a smart-contract application providing a financial service such as lending, automated market making or liquid staking (Schär, 2021; Gogol et al., 2024). Each protocol holds an inventory of *tokens*—fungible on-chain assets whose USD value aggregates into Total Value Locked (TVL) (Saggese et al., 2025; DefiLlama, 2026). Because tokens issued by one protocol are routinely held by another, the cross-protocol state at each timestep forms a directed weighted graph (Wu et al., 2025): nodes are protocols, edges are token-mediated USD exposures, and node features are tabular descriptors (TVL, token composition, sector). The *credit exposure* of p to q is the value of q-issued tokens held by p, equivalently the weight of edge $p \to q$; *systemic risk* corresponds to shock propagation along multi-hop paths in this graph as nodes, edges and weights evolve weekly (Eisenberg & Noe, 2001; Battiston et al., 2012; Aufiero et al., 2025).

Empirical work on DeFi systemic risk is constrained by data and tools. Exposures are not reported in standardised balance sheets and must be reconstructed from on-chain states; even when reconstructed, they are *high-dimensional* (millions of potential directed edges), *non-stationary* (protocol upgrades, new assets, changing incentives), and *strongly heteroskedastic* (regime shifts between calm and crisis). Classical time-series tools (Sims, 1980; Stock & Watson, 2002) offer strong baselines but cannot represent the coupled evolution of node states, edge exposures and meso-structure. Modern graph and temporal-graph neural networks (Kipf & Welling, 2017; Veličković et al., 2018; You et al., 2022; Rossi et al., 2020) are expressive but typically train one

[1]School of Informatics, University of Edinburgh, United Kingdom [2]Judge Business School, University of Cambridge, United Kingdom [3]Business School, University of Edinburgh, United Kingdom. Correspondence to: Fengxiang He <F.He@ed.ac.uk>.

The Workshop on Graph Foundation Models at the 43rd International Conference on Machine Learning (ICML 2026), Seoul, South Korea.

model per task per dataset, rather than serving as reusable infrastructure across the ecosystem.

The *foundation-model* paradigm offers an alternative: pretrain a general representation on broad data and reuse it across tasks via light adaptation (Bommasani et al., 2021). Recent work in this direction has produced graph and tabular foundation models (Ying et al., 2021; Hollmann et al., 2025; Eremeev et al., 2025b;a), time-series foundation models (Ansari et al., 2024; Rasul et al., 2024; Das et al., 2024), and domain-specific GFMs in molecules, knowledge graphs and recommendation. *Finance, and DeFi in particular, has so far had no systematic graph-foundation-model treatment*, despite being a textbook case for one: the DeFi exposure network is heterogeneous, dynamic, tabular-rich, and reconstructable at scale.

DeFi credit exposure networks instantiate several themes in domain-specific GFMs. They are *temporal graphs* because exposures evolve weekly; *heterogeneous financial graphs* because protocols, tokens, chains, and sectors induce multiple semantic layers; and *graph-tabular systems* because each protocol carries tabular states such as TVL, token composition, and sector descriptors. Unlike generic graph benchmarks, the downstream objects are not only link accuracy but also stress-test losses, concentration measures, and systemic-importance rankings. This makes DeFi a natural finance-domain testbed for studying how GFMs can be adapted, evaluated, and deployed under non-stationarity.

Contributions. We make the following contributions to domain-specific GFMs for finance.

(i) *Domain formulation.* We formulate DeFi credit exposures as dynamic graph-tabular financial networks whose nodes are protocols, directed weighted edges are token-mediated exposures, and tabular descriptors encode protocol size, token composition, and sector structure.

(ii) *Domain-adapted graph-tabular GFM.* We build DeXposure-FM around a GraphPFN-based encoder that jointly ingests each weekly exposure graph and protocol-level tabular descriptors, then fine-tune it end-to-end with multi-task, multi-horizon heads for edge existence, edge weight and node TVL change at $h \in \{1, 4, 8, 12\}$ weeks. A single encoder is shared across all heads and the downstream measurement tasks of Section 6.

(iii) *Strict temporal evaluation.* On a 2025 hold-out from a 43.7M-observation DeFi benchmark, DeXposure-FM achieves strong edge-existence performance (AUROC 0.993–0.995, AUPRC 0.967–0.973) and consistently lower RMSE than all baselines (including persistence) on edge-weight and node-ΔTVL targets across all four horizons; MAE on these magnitude targets matches persistence within $\sim$0.1 log-units, reflecting the network's 98.5% week-to-week edge overlap.

(iv) *Forecast-then-measure downstream tasks.* We use predicted future graphs to compute systemic-importance scores, sector-spillover concentration, and DebtRank-style stress-test losses. Under DeFi's 98.5% week-to-week edge overlap, persistence is a structurally strong system-level baseline; the GFM's macroprudential value emerges *conditionally*—in the worst-20% of weeks where persistence is least reliable (win rates 83–100%) and in bridge-sector shocks where topology shifts most.

2. Related Work

Graph and tabular foundation models. Transformer-based graph models with structural encodings (Ying et al., 2021) and tabular prior-fitted networks (Hollmann et al., 2025) demonstrate the feasibility of pretrained models for structured data. GraphPFN (Eremeev et al., 2025b) extends the prior-data fitted network paradigm to attributed graphs by jointly encoding tabular node features and graph structure in a single transformer; a related line of work shows that tabular foundation models can themselves be turned into graph foundation models (Eremeev et al., 2025a). We use the open-source GraphPFN weights as initialisation and fine-tune on DeFi.

Time-series foundation models for finance. Chronos (Ansari et al., 2024), Lag-Llama (Rasul et al., 2024) and TimesFM (Das et al., 2024) pretrain on broad time-series corpora; in the financial domain a parallel literature uses LLMs for narrative-based nowcasting (de Bondt & Sun, 2025; Kwon et al., 2024; Carriero et al., 2025). These models target univariate or panel time series, not typed credit-exposure graphs.

Machine learning for systemic risk. GNNs improve systemic-importance classification on simulated banking networks (Balmaseda et al., 2023) and have been integrated with explicit interbank liability networks to compute systemic-risk capital (Gonon et al., 2025). Temporal GNNs such as ROLAND (You et al., 2022) and TGNs (Rossi et al., 2020) learn dynamic graph representations but are trained per task. In DeFi, simple aggregate exposure measures (Bertomeu et al., 2024) and microscopic risk maps (Aufiero et al., 2025) demonstrate the need for richer network-level tools.

To our knowledge, no prior work delivers a domain-specific GFM for DeFi that jointly models the graph, tabular features, and temporal evolution of inter-protocol credit exposures.

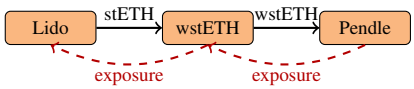

Figure 1. Multi-layer credit exposure in DeFi. A user stakes ETH with Lido (issuing stETH), wraps it into wstETH, and deposits into Pendle. Each protocol holds the previous one's liability, chaining exposures from Pendle back to Lido. A slashing event or stETH de-peg cascades through all three.

3. Credit Exposures on DeFi Networks

Credit exposure. Let $X_p(t)$ be the set of tokens held by protocol p at time t, and $X_G(q)$ the set issued or generated by q. We say p has *credit exposure* to q at t if

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

i.e., p holds a tokenised liability of q (Wu et al., 2025). This generalises bilateral counterparty risk to settings where exposures arise purely from token possession.

Figure 1 illustrates a single user transaction inducing a three-layer chain of credit exposures: a slashing event at Lido or a stETH de-peg cascades into wstETH and Pendle, hitting PT/YT holders downstream. DeFi exposures are intrinsically multi-layered through liquid staking, wrapping and yield-tokenisation; capturing them at scale requires reconstructing token holdings and issuance.

Token-to-protocol mapping. Each token σ is attributed to an issuing protocol $M(\sigma)$ via a DefiLlama-metadata-first fallback procedure (manual mapping for high-TVL exceptions, TF-IDF similarity (Shahzad Qaiser, 2018), primary-market default; full procedure in (Wu et al., 2025)).

Network and value flows. Over a weekly interval $\tau = (t_1, t_2)$ we form a weighted directed graph $G_\tau = (\mathcal{P}_\tau, \mathcal{E}_\tau)$. Node weights are end-of-period USD holdings over tokens observed across the interval, $w_\tau(p) = \sum_{\sigma \in X_p(t_1) \cap X_p(t_2)} v_{t_2,\sigma}$. The edge weight from p to q aggregates token-level value flows $F^{\sigma}_{pq,\tau}$:

$$w_\tau(e_{pq}) \;=\; \sum_{\sigma:\, M(\sigma)=q} F^{\sigma}_{pq,\tau}, \tag{2}$$

where $F^{\sigma}_{pq,\tau}$ is the non-negative portion of σ's movement from p to q over τ (loss for p matched to gain for q). A positive $w_\tau(e_{pq})$ indicates increasing exposure of p to q.

4. DeXposure-FM Model

We use "foundation model" in the domain-specific sense: a reusable graph-tabular encoder is adapted to a large financial graph domain and exposed through task heads that serve multiple downstream forecasting and measurement tasks. Time enters through weekly graph snapshots and multi-horizon forecasting pairs $G_\tau \to G_{\tau+h}$ rather than through a separate sequence encoder. The goal is not only to fit a single supervised target, but to learn representations of DeFi exposure networks that transfer across edge, node, and aggregate risk-measurement tasks.

4.1. Graph-Tabular Encoder

DeXposure-FM follows an *encoder–multi-head* design. A pretrained GraphPFN graph-tabular transformer (Eremeev et al., 2025b), originally trained as a prior-data fitted network for node-level classification and regression on attributed graphs, consumes each weekly graph G_τ together with tabular node descriptors $\{x^{\text{tab}}_{p,\tau}\}$, including log-scaled TVL, token-composition summaries and sector one-hots. It outputs d-dimensional protocol embeddings $\{h_{p,\tau}\}$ via structurally-augmented self-attention. We discard GraphPFN's original node-prediction head, initialise the encoder from the open-source weights, and fine-tune it end-to-end together with the new multi-task heads in Section 4.2.

4.2. Multi-Task Multi-Horizon Heads

For each forecast horizon $h \in \{1, 4, 8, 12\}$ weeks the model attaches three heads.

Edge head. For an ordered pair (p, q) the symmetric pairwise feature $f_{pq,\tau} = [h_p; h_q; h_p \odot h_q; |h_p - h_q|]$ is fed to an MLP that outputs an existence probability $\widehat{y}^{\text{exist}}_{pq}$ and a log-space residual $\widehat{r}_{pq,\tau,h}$ used to reconstruct the predicted log-weight as $\widehat{\tilde{w}}_{\tau+h}(e_{pq}) = \tilde{w}_\tau(e_{pq}) + \widehat{r}_{pq,\tau,h}$, with $\tilde{w}_\tau(\cdot) = \log(1 + w_\tau(\cdot))$ and $\tilde{w}_\tau(e_{pq}) = 0$ when the edge is absent at τ.

Node head. A second MLP predicts the log-change in protocol TVL from each protocol's embedding concatenated with weighted in/out-neighbour aggregates under current-week edge weights, injecting local flow context.

Network statistics. Density, concentration, sectoral connectivity and other network-level quantities are obtained as deterministic functionals of the predicted graph $\widehat{G}_{\tau+h}$. This separation lets us reuse the same model for forecasting and for the macroprudential measurements in Section 6.

4.3. Training

Walk-forward splits. Training pairs $G_\tau \to G_{\tau+h}$ are formed by anchor and horizon. We use expanding-window walk-forward splits with strict pre-2025 training and a 2025 hold-out, preventing look-ahead bias under non-stationarity.

Negative sampling. Edge existence is highly imbalanced ($\sim$30k positives among millions of candidate pairs). For each (τ, h) we sample negatives at a 5:1 ratio relative to positives; the existence loss is computed on this set, while the weight loss is restricted to positive edges.

Multi-task objective. The total loss combines weighted binary cross-entropy on edge existence, Smooth-L1 on edge log-weight residuals, and Smooth-L1 on node log-TVL change:

$$\mathcal{L} = \lambda_{\text{exist}}\mathcal{L}_{\text{exist}} + \lambda_{\text{weight}}\mathcal{L}_{\text{weight}} + \lambda_{\text{node}}\mathcal{L}_{\text{node}}, \qquad (3)$$

with λ_{exist}=2.0, λ_{weight}=0.5, λ_{node}=20.0, calibrated to balance gradient magnitudes across tasks. Optimisation uses Adam (η=5×10^{-4} for heads; 5×10^{-5} for the GraphPFN backbone), 20 epochs, early stopping on validation AUPRC (patience 3), and gradient clipping at $\|\nabla\|_2 \leq 1.0$.

5. Empirical Evaluation

We evaluate DeXposure-FM on two ML benchmarks: (I) multi-step forecasting of edges and node TVL; (II) predictive contagion stress testing, where forecast graphs are fed to a fixed contagion simulator and the resulting system losses are compared to those on realised graphs.

Baselines. (1) *GraphPFN (frozen)*: pretrained GraphPFN with frozen weights and a small trainable MLP probe head. (2) *ROLAND* (You et al., 2022): a representative temporal GNN trained from scratch. (3) *Persistence*: a naive baseline carrying the last observed graph forward, $\widehat{A}_{t+h} = A_t, \widehat{W}_{t+h} = W_t$. Trained baselines share Adam, η=5×10^{-4} and the training schedule with DeXposure-FM. The strict 2025 hold-out (33 weeks, 2025-01-06 to 2025-08-18) is never used during training or model selection.

5.1. Task I: Multi-step Forecasting

We forecast (i) edge existence (binary), (ii) log-edge-weight (regression on positive edges), and (iii) node Δlog-TVL. We report AUROC (Fawcett, 2006) and AUPRC (Davis & Goadrich, 2006) for existence, and MAE/RMSE on the log scale for the regression targets.

Table 1 summarises the results. Despite the network's strong week-to-week persistence (98.5% mean edge overlap), DeXposure-FM substantially improves *edge existence* prediction: AUROC 0.993–0.995 and AUPRC 0.967–0.973 across horizons, versus AUPRC 0.58–0.64 for persistence and 0.936–0.940 for the frozen GraphPFN probe. On *edge weights*, DeXposure-FM has lower errors than the learned neural baselines (GraphPFN-frozen and ROLAND); persistence remains competitive because last week's exposure magnitude is itself a strong predictor. On *node ΔTVL*, MAE is close to persistence at short horizons and slightly higher at longer horizons, while RMSE is lower than persistence across all four horizons; frozen GraphPFN and ROLAND remain competitive on some MAE entries. The pattern is consistent with the domain structure: graph-foundation representations are most useful for predicting edge formation and decay, while highly persistent magnitude targets leave less room over a carry-forward baseline.

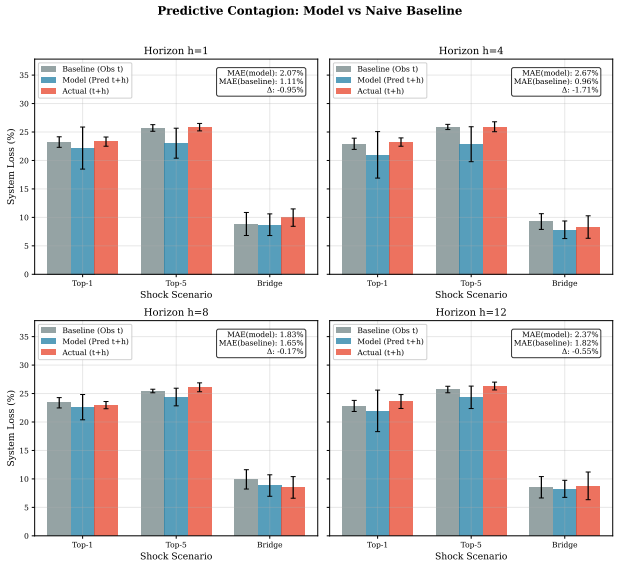

Figure 2. Three-way predictive contagion stress testing: system-loss percentage under identical shock scenarios on the persistence graph G_t, the model forecast $\widehat{G}_{t+h}$, and the realised graph G_{t+h}. Because DeFi exposure networks have 98.5% week-to-week edge overlap, the persistence graph is itself a near-perfect input to the simulator on the largest exposure links that drive contagion—an unusually demanding baseline at the system-aggregate level. The model bars accordingly sit consistently below the realised bars on the full sample.

5.2. Task II: Predictive Contagion Stress Testing

We run a DebtRank-style contagion simulator (Eisenberg & Noe, 2001; Battiston et al., 2012) on three graphs — the persistence baseline G_t, the model forecast $\widehat{G}_{t+h}$, and the realised future G_{t+h} — under three counterfactual shocks: (i) 50% TVL loss to the largest protocol; (ii) 30% to the top-5 protocols; (iii) 100% to bridge-sector protocols. Comparisons are restricted to the common node set $V_t \cap V_{t+h}$ to handle entry/exit. Because the network is highly persistent, persistence is a strong baseline *on average*; we therefore additionally report a post-hoc stratification on the 20% largest persistence-baseline errors per horizon (pooled across shocks), where forward-looking forecasts matter most. We summarise improvement with $\Delta\text{MAE} = \text{MAE}(\text{baseline}) - \text{MAE}(\text{model})$.

Figure 2 shows the three-way comparison of raw simulator outputs across horizons and shock scenarios; Figure 3 summarises the same data as ΔMAE. On the full hold-out, persistence is hard to beat ($\Delta\text{MAE} < 0$ on average), reflecting week-to-week stability: with 98.5% mean edge overlap, the carry-forward graph is itself a near-perfect input to the simulator on the largest exposure links that drive contagion, so any imperfect forecast is at a structural disadvantage at

Model	h	Edge Existence		Edge Weight		Node ΔTVL	
		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 (frozen)	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 1. Multi-step forecasting on the 2025 hold-out at $h \in \{1, 4, 8, 12\}$ weeks. DeXposure-FM = fine-tuned graph-tabular foundation model; GraphPFN (frozen) = pretrained encoder with a trainable MLP probe; Persistence carries forward A_t, W_t. All metrics are computed on log-scale exposures. Bold marks the best value in each metric/horizon.

the system-aggregate level. The value of a learned forecaster therefore emerges where this carry-forward assumption breaks down. In the worst-20% tail, DeXposure-FM achieves consistently positive $\Delta\mathrm{MAE}$ with win rates of 83–100%; gains are most pronounced under bridge-sector shocks, consistent with cross-chain contagion being sensitive to evolving network topology. The macroprudential reading is therefore selective: learned forecasts are most useful as a backstop for scenario analysis when exposures reallocate non-stationarily, i.e., in high-error regimes where a carry-forward graph is least reliable.

6. Forecast-then-Measure Downstream Tasks

A domain-specific GFM is most useful when its predicted representations can be reused by downstream decision tools. Section 5.2 evaluated contagion losses as a *forecasting metric*: forecast graphs were scored against realised graphs through a fixed contagion simulator. This section instead uses the same forecast graphs as *inputs to a measurement layer*: predicted graphs are translated into protocol watchlists, sectoral risk channels and scenario losses that regulators or risk managers can inspect before the realised network is observed. The financial-economics contribution of DeXposure-FM is therefore not only better edge prediction, but a forward-looking macroprudential measurement layer for digital financial infrastructure. We treat each macroprudential indicator as a deterministic functional $\mathcal{M}(\cdot)$ of a graph and apply it to either the observed snapshot G_t (descriptive) or the model forecast $\widehat{G}_{t+h}$ (forward-looking),

$$G_t \xrightarrow{\text{DeXposure-FM}} \widehat{G}_{t+h} \xrightarrow{\mathcal{M}(\cdot)} \widehat{\mathcal{M}}_{t+h}. \tag{4}$$

This *forecast-then-measure* pipeline yields actionable signals: predicted watchlists of systemically important protocols, predicted cross-sector risk channels, and predicted stress-test losses. Its value is deliberately interpreted as *conditional*: when the exposure network is stable, persistence is a strong monitoring baseline; when exposures reallocate, especially in tail-error weeks and bridge-sector shocks, the learned forecast provides the additional signal needed for macroprudential backstops.

Systemic-importance score (SIS). For each protocol p,

$$\mathrm{SIS}_p = \alpha\widetilde{\mathrm{PR}}_p + \beta\,\mathrm{TailExp}_p + \gamma\log(\widetilde{1+\mathrm{TVL}_p}), \tag{5}$$

combining normalised PageRank (Brin & Page, 1998), the share of p's top-k outgoing exposures (we use k=5), and log-TVL with default weights α=β=γ=1/3. Replacing G_t by $\widehat{G}_{t+h}$ produces a forward-looking systemic watchlist.

Spillover concentration. Aggregating edge weights into a sector-to-sector matrix $\mathbf{S} \in \mathbb{R}^{K\times K}$ gives a Herfindahl–Hirschman concentration of off-diagonal flows, $\mathrm{SpilloverIndex} = \mathrm{HHI}(\{S_{ij}: i \neq j\})$ (Rhoades, 1993). Figure 5 (Appendix D) visualises one such snapshot. Among the forecast-based indicators, network density and mean SIS are well-calibrated against realised values and support quantitative comparisons; HHI-based concentration is informative for directional changes rather than absolute levels (see Appendix D for the level-vs-direction breakdown).

Stress test. A DebtRank-style propagation under distress threshold τ=0.1 outputs system-wide loss, contagion depth, and affected-protocol counts under the three shocks defined

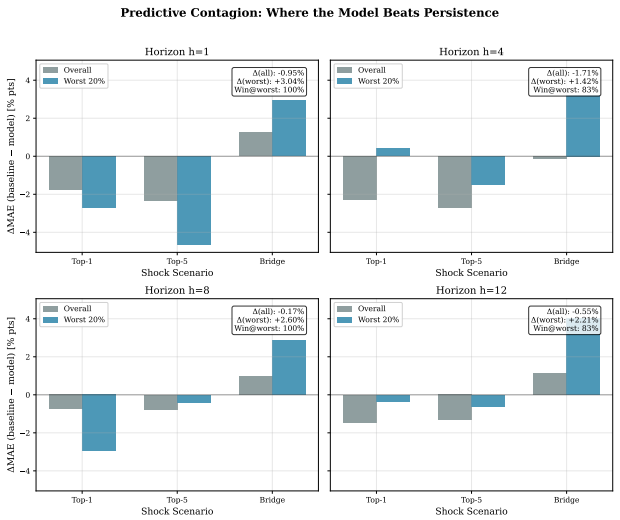

Figure 3. Predictive contagion stress testing (Task II). Bars report $\Delta\mathrm{MAE} = \mathrm{MAE}(\text{persistence}) - \mathrm{MAE}(\text{model})$ on system-loss percentage points. "Overall" is the full 2025 hold-out; "Worst 20%" restricts to the 20% largest persistence-baseline errors per horizon (pooled over shocks). Positive values indicate DeXposure-FM outperforms persistence; gains concentrate in the tail and under bridge-sector shocks where network structure shifts most.

in Section 5.2. Applied to $\widehat{G}_{t+h}$ this yields predicted losses and the worst-20% tail-regime gains reported in Figure 3.

Case study: Terra/Luna collapse (May 2022). We retrain an event-specific forecaster using only snapshots strictly prior to the event window and produce one-step-ahead forecasts within it. Between the last pre-event week and the final event week, observed TVL dropped from \$261.9 B to \$166.3 B (-36.5%) while the edge count *rose* by $+9.3\%$, a flight-to-quality pattern in which capital reorganised across surviving protocols. The event study compares forecast-based HHI and predicted contagion loss against realised trajectories, illustrating how forecast-then-measure outputs can complement descriptive monitoring during regime changes.

Limitations. Model-based measures are only as reliable as their training conditions; the four most material caveats are: (i) *data coverage* — DeXposure-FM is trained exclusively on on-chain DeFi protocols and does not observe centralised-exchange balances, OTC derivatives, or off-chain stablecoin reserves; (ii) *weekly resolution* — intraday flash-loan attacks (Qin et al., 2021) and rapid liquidation cascades (Warmuz et al., 2023) are smoothed out; (iii) *TVL as proxy* (Saggese et al., 2025) aggregates heterogeneous assets without risk weights, so the indicators are most useful for *relative* monitoring; (iv) the tools *complement, not replace* qualitative judgement on smart-contract audits, governance and oracles (Aspembitova & Bentley, 2023).

7. Conclusion

DeXposure-FM frames DeFi credit exposures as a domain-specific GFM problem for dynamic graph-tabular financial networks. Initialising a GraphPFN-based encoder on weekly exposure graphs and protocol descriptors yields strong gains for edge-existence forecasting and improves edge-weight prediction over learned neural baselines, while persistence remains a demanding comparator for highly stable magnitude targets. The forecast-then-measure pipeline restores the financial-economics objective: predicted graphs become forward-looking measurements of systemic importance, sector spillovers, and stress-test losses, especially in tail regimes where the carry-forward assumption is least reliable. Future work includes zero- and few-shot transfer to other financial graphs such as interbank exposures, supply-chain credit, and cross-CEX flows; risk-weighted exposure proxies beyond TVL; higher-frequency data integration; drift-aware continuous retraining; and distributional graph forecasting for scenario analysis.

Impact Statement

DeXposure-FM gives regulators, central banks and DeFi participants new quantitative tools for monitoring inter-protocol credit exposures, including forward-looking systemic-importance scores, sectoral spillover indices and scenario-based stress-test losses. Made open, these tools could materially improve macroprudential surveillance of an ecosystem whose risk reporting infrastructure is otherwise sparse. They also carry risks: model outputs could be misread as definitive verdicts, encouraging overconfident policy or trading actions in regimes outside the model's training distribution; bad actors could in principle use the spillover map to design attacks targeting structurally important protocols; and reliance on TVL as a measurement proxy can encode market conventions that themselves shift under stress. Releases will therefore include a model card with validity boundaries, a recommendation that the indicators complement rather than replace audits, governance review and intelligence on smart-contract risk, and an explicit statement that the tools are designed for relative monitoring, not absolute loss accounting.

LLM Usage Disclosure

In accordance with the workshop policy, we disclose that large language models (LLMs) were used for language polishing, LaTeX editing assistance, and minor code-editing assistance. All scientific claims, experimental design choices, analyses, numerical results, and final text were reviewed and verified by the human authors, who remain fully responsible for the submission.

References

Anadu, K., Azar, P. D., Cipriani, M., Eisenbach, T. M., Huang, C., Landoni, M., La Spada, G., Macchiavelli, M., Malfroy-Camine, A., and Wang, J. C. Runs and flights to safety: Are stablecoins the new money market funds? Staff Report 1073, Federal Reserve Bank of New York, September 2023. URL https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr1073.pdf. Revised April 2024.

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., and Wang, Y. Chronos: Learning the language of time series, 2024. URL https://arxiv.org/abs/2403.07815.

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

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

Balmaseda, V., Coronado, M., and de Cadenas-Santiago, G. Predicting systemic risk in financial systems using deep graph learning. *Intelligent Systems with Applications*, 19:200240, 2023. ISSN 2667-3053. doi: https://doi.org/10.1016/j.iswa.2023.200240. URL https://www.sciencedirect.com/science/article/pii/S2667305323000650.

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

Bertomeu, J., Martin, X., and Sall, I. Measuring defi risk. *Finance Research Letters*, 63:105321, 2024. ISSN 1544-6123. doi: https://doi.org/10.1016/j.frl.2024.105321. URL https://www.sciencedirect.com/science/article/pii/S1544612324003519.

Bommasani, R. et al. On the opportunities and risks of foundation models, 2021. URL https://arxiv.org/abs/2108.07258.

Brin, S. and Page, L. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the 7th International Conference on World Wide Web*, pp. 107–117, 1998. doi: 10.1016/S0169-7552(98)00110-X. URL http://infolab.stanford.edu/pub/papers/google.pdf.

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

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

Davis, J. and Goadrich, M. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, pp. 233–240, 2006. doi: 10.1145/1143844.1143874.

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

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

Eisenberg, L. and Noe, T. H. Systemic risk in financial systems. *Management Science*, 47(2):236–249, February 2001. doi: 10.1287/mnsc.47.2.236.9835. URL https://pubsonline.informs.org/doi/10.1287/mnsc.47.2.236.9835.

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

Eremeev, D., Platonov, O., Bazhenov, G., Babenko, A., and Prokhorenkova, L. Graphpfn: A prior-data fitted graph foundation model, 2025b. URL https://arxiv.org/abs/2509.21489.

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

Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. doi: 10.1016/j.patrec.2005.10.010.

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

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

Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S. B., Schirrmeister, R. T., and Hutter, F. Accurate predictions on small data with a tabular foundation model. *Nature*, 637(8045):319–326, January 2025. doi: 10.1038/s41586-024-08328-6. URL https://www.nature.com/articles/s41586-024-08328-6. Published online 8 Jan 2025.

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

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

Qin, K., Zhou, L., Livshits, B., and Gervais, A. 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*, pp. 3–32, Berlin, Heidelberg, 2021. Springer-Verlag. ISBN 978-3-662-64321-1. doi: 10.1007/978-3-662-64322-8_1. URL https://doi.org/10.1007/978-3-662-64322-8_1.

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., and Rish, I. Lag-llama: Towards foundation models for probabilistic time series forecasting, 2024. URL https://arxiv.org/abs/2310.08278.

Rhoades, S. A. The Herfindahl-Hirschman Index. *Federal Reserve Bulletin*, 79:188–189, March 1993. URL https://fraser.stlouisfed.org/files/docs/publications/FRB/pages/1990-1994/33101_1990-1994.pdf.

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

Saggese, P., Fröwis, M., Kitzler, S., Haslhofer, B., and Auer, R. Towards verifiability of total value locked (TVL) in decentralized finance. BIS Working Papers 1268, Bank for International Settlements, May 2025. URL https://www.bis.org/publ/work1268.htm.

Schär, F. Decentralized finance: On blockchain- and smart contract-based financial markets. *Federal Reserve Bank of St. Louis Review*, 103(2):153–174, 2021. doi: 10.20955/r.103.153-74. URL https://www.stlouisfed.org/publications/review/2021/02/05/decentralized-finance-on-blockchain-and-smart-contract-based-financial-markets.

Shahzad Qaiser, R. A. Text mining: Use of tf-idf to examine the relevance of words to documents. *International Journal of Computer Applications*, 181(1):25–29, Jul 2018. ISSN 0975-8887. doi: 10.5120/ijca2018917395. URL https://ijcaonline.org/archives/volume181/number1/29681-2018917395/.

Sims, C. A. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1912017.

Stock, J. H. and Watson, M. W. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, 2002. doi: 10.1198/016214502388618960. URL https://doi.org/10.1198/016214502388618960.

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

Vidal-Tomás, D., Briola, A., and Aste, T. Ftx's downfall and binance's consolidation: The fragility of centralised digital finance. *Physica A: Statistical Mechanics and its Applications*, 625:129044, 2023. ISSN 0378-4371. doi: 10.1016/j.physa.2023.129044. URL https://www.sciencedirect.com/science/article/pii/S037843712300599X.

Warmuz, J., Chaudhary, A., and Pinna, D. Toxic liquidation spirals, 2023. URL https://arxiv.org/abs/2212.07306.

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

Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., Shen, Y., and Liu, T.-Y. Do transformers really perform badly for graph representation? In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 28877–28888. Curran Associates, Inc., 2021.

URL https://proceedings.neurips.cc/paper_files/paper/2021/file/f1c1592588411002af340cbaedd6fc33-Paper.pdf.

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

A. Data and Token Mapping

Dataset statistics. Table 2 summarises the DeXposure benchmark used to train and evaluate DeXposure-FM.

Statistic	Value
Total weekly snapshots	283 weeks
Time range	2020-03-23 to 2025-08-18
Mean nodes per week	5,676
Mean edges per week	30,424
Mean week-to-week edge overlap	98.5%
Protocol categories	$\sim$15 sectors
Protocols (cumulative)	4,300+
Blockchains	602
Unique tokens	24,300+
Total exposure observations	43.7 M

Table 2. DeXposure dataset statistics.

Token-to-protocol mapping (full procedure). For each token σ a fallback chain assigns the issuing protocol $M(\sigma)$:

1. *Metadata lookup.* Direct lookup in DefiLlama (DefiLlama, 2026) when the issuer is recorded.
2. *Manual mapping.* For high-TVL tokens lacking metadata, manual assignment from documentation and expert knowledge.
3. *Text similarity.* TF-IDF (Shahzad Qaiser, 2018) cosine similarity between token and protocol descriptions, with threshold θ.
4. *Primary-market fallback.* If none of the above resolves, the token is treated as its own protocol (e.g., WETH).

Value flow derivation. For each token σ over $\tau = (t_1, t_2)$, let $\Delta S^{\sigma}_{p,\tau} = v^{\sigma}_{p,t_2} - v^{\sigma}_{p,t_1}$ be the change in USD value of σ at p. The token-level value flow from p to q is

$$F^{\sigma}_{pq,\tau} = \begin{cases} \max(0, -\Delta S^{\sigma}_{p,\tau}), & \Delta S^{\sigma}_{p,\tau} < 0, \\ \max(0, \Delta S^{\sigma}_{q,\tau}), & \Delta S^{\sigma}_{q,\tau} \geq 0, \end{cases} \tag{6}$$

i.e., a decrease at p matched against an increase at q. Equation (2) sums these over all tokens whose issuing protocol is q.

B. Training Details and Hyperparameters

Hyperparameter	Value
Data granularity	Weekly snapshots
Forecast horizons h	$\{1, 4, 8, 12\}$ weeks
Negative sampling ratio	5:1 (neg:pos)
Optimizer	Adam (β_1=0.9, β_2=0.999)
Learning rate (heads)	5×10^{-4}
Learning rate (backbone fine-tune)	5×10^{-5}
Training epochs	20
Early-stopping criterion	validation AUPRC, patience 3
Gradient clipping	$\|\nabla\|_2 \leq 1.0$
Loss weights:	
λ_{exist}	2.0
λ_{weight}	0.5
λ_{node}	20.0

Table 3. Training hyperparameters.

Walk-forward expanding-window splits. Stage 1 used 15 expanding-window folds for hyperparameter selection on pre-2025 data with ≥ 104 training weeks, 24 validation weeks, and 8 test weeks per fold; the window advances by 8 weeks per fold. Stage 2 (reported throughout the paper) is a strict hold-out: training 2020-03-23 to 2024-07-15 (225 weeks), validation 2024-07-22 to 2024-12-30 (24 weeks, used only for early stopping), test 2025-01-06 to 2025-08-18 (33 weeks).

Loss definitions. The edge-existence loss is class-weighted BCE, $\mathcal{L}_{\text{exist}} = \sum_{(p,q)\in\mathcal{S}_{\tau+h}} \text{BCE}(\hat{y}^{\text{exist}}_{pq}, y^{\text{exist}}_{pq}; w_{\text{pos}})$, with w_{pos} matching the negative ratio. The weight loss is SmoothL1 on log-space residuals over positive edges, and the node loss is SmoothL1 on Δlog-TVL over $\mathcal{P}_\tau \cap \mathcal{P}_{\tau+h}$. Loss-weight calibration: $\mathcal{L}_{\text{exist}} \approx 0.22$, $\mathcal{L}_{\text{weight}} \approx 2.4$, $\mathcal{L}_{\text{node}} \approx 0.05$ on average; the chosen λ's equalise gradient contributions.

C. Validity and Scope (Full Discussion)

The Section 6 "Limitations" paragraph summarises the four most material caveats; here we expand to all six.

Stable vs. turbulent regimes. The DeXposure dataset has 98.5% mean week-to-week edge overlap, so persistence is a strong baseline for many magnitude targets. Under structural breaks such as new chains, governance attacks, or rapid deleveraging, edge distributions can shift abruptly and forecasts should be interpreted with caution. The stress-testing results in Table 5 suggest that learned forecasts are most useful in high-error tail regimes where persistence is least reliable.

Data coverage. DeXposure-FM is trained exclusively on on-chain DeFi protocols sourced from DefiLlama. It does not observe (i) centralised exchange balances or order-book exposures, (ii) over-the-counter derivatives or off-chain lending, or (iii) off-chain reserves backing stablecoins (e.g., Treasury-backed USDC). The tools therefore measure intra-DeFi fragility and cannot directly assess DeFi–TradFi spillovers, a gap that grows as those linkages deepen (European Systemic Risk Board, 2025; Anadu et al., 2023).

Temporal resolution. Weekly aggregation smooths intra-week dynamics. Flash-loan attacks (Qin et al., 2021) and liquidation cascades (Warmuz et al., 2023) require complementary block-level monitoring tools.

Cross-chain data consistency. DeXposure aggregates 602 chains with heterogeneous indexing and oracle delays. Edge weights near reporting boundaries can reflect data artefacts rather than genuine exposure changes; users should treat boundary-week values cautiously and combine model outputs with data-quality checks.

TVL as a measurement proxy. TVL aggregates heterogeneous collateral and liquidity profiles without risk weights (Bertomeu et al., 2024; Saggese et al., 2025); \$1 B in governance tokens is not equivalent to \$1 B in USDC. The indicators are most informative for relative monitoring (rankings, concentration changes, scenario deltas), not absolute loss accounting.

Complement, not replace, qualitative judgement. A protocol may rank low on SIS yet pose idiosyncratic risks (smart-contract vulnerabilities, oracle manipulation, governance capture) that network topology cannot capture (Aspembitova & Bentley, 2023; Gogol et al., 2024). Supervisors should triangulate model signals with code audits, governance review and market intelligence.

D. Additional Case Study and Figures

FTX collapse (November 2022). We retrain an event-specific forecaster with data strictly before 2022-11-07 and produce one-step-ahead forecasts within the event window. Table 4 contrasts Terra/Luna with FTX. Note that on-chain DeFi TVL *rose* during the FTX window because our dataset captures only on-chain protocols: the collapse of a centralised exchange triggered a flight to decentralisation (Vidal-Tomás et al., 2023) that increased intra-DeFi exposure even as overall crypto markets contracted.

Sector-to-sector spillover snapshot. Figure 5 visualises the spillover matrix $\mathbf{S}$ on an observed snapshot, complementing the spillover-concentration discussion in Section 6.

Forward-looking risk-metric calibration. Figure 6 reports predicted vs. realised aggregate risk metrics (concentration, density, spillover index, summary statistics of SIS) for $h \in \{1, 4, 8, 12\}$. The metrics separate into two regimes. Network

Event	Pre-TVL ($B)	Post-TVL ($B)	ΔTVL	ΔEdges
Terra/Luna (May 2022)	261.9	166.3	−36.5%	+9.3%
FTX (Nov 2022)	119.9	175.1	+46.1%	−1.8%

Table 4. Network structural shifts during stress events, measured between the last pre-event week (anchor) and the final event-window week.

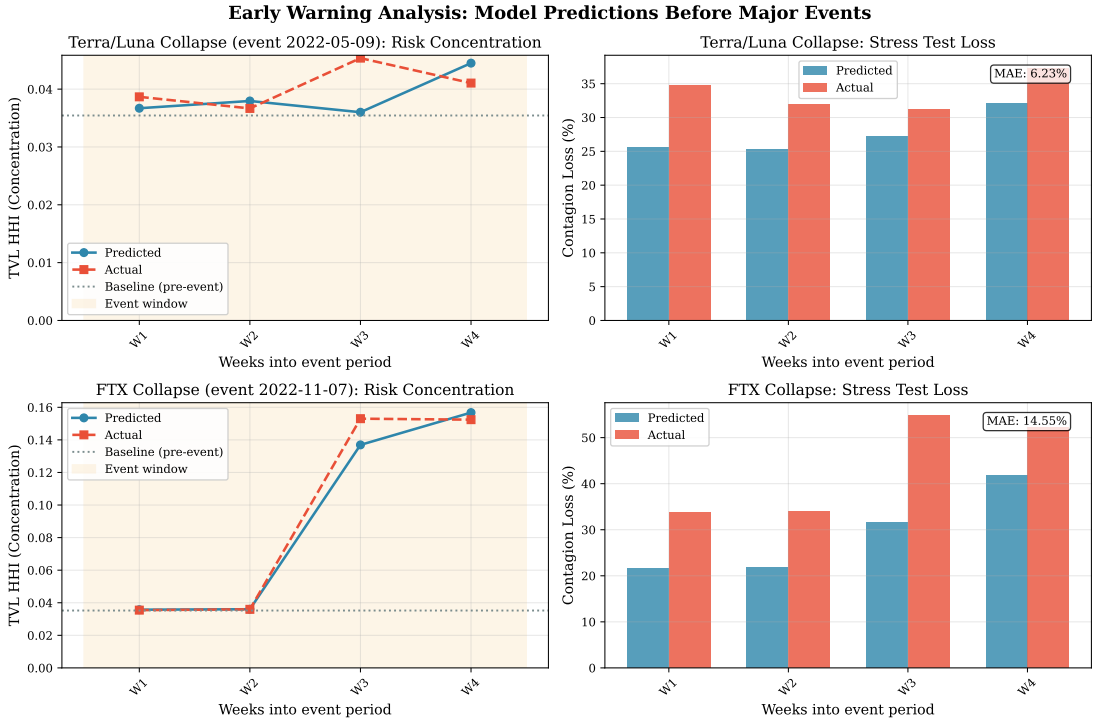

Figure 4. Early-warning event studies. For each event window, predicted vs. realised risk concentration (HHI) and predictive stress-test loss (DebtRank system loss %).

density (R^2=0.84) and mean SIS (R^2=0.70) track realised values closely and can be read as quantitative level forecasts; SIS rankings are correspondingly informative as forward-looking watchlists. The two HHI-based concentration measures, by contrast, show low level-correlation with realised values and a systematic upward bias: predicted graphs concentrate off-diagonal flows on fewer sector pairs than realised graphs do. Predicted HHI is therefore informative for directional changes—week-on-week jumps and large cross-horizon deviations—rather than for absolute concentration levels, and is most meaningful when read as a delta against the descriptive HHI on G_t.

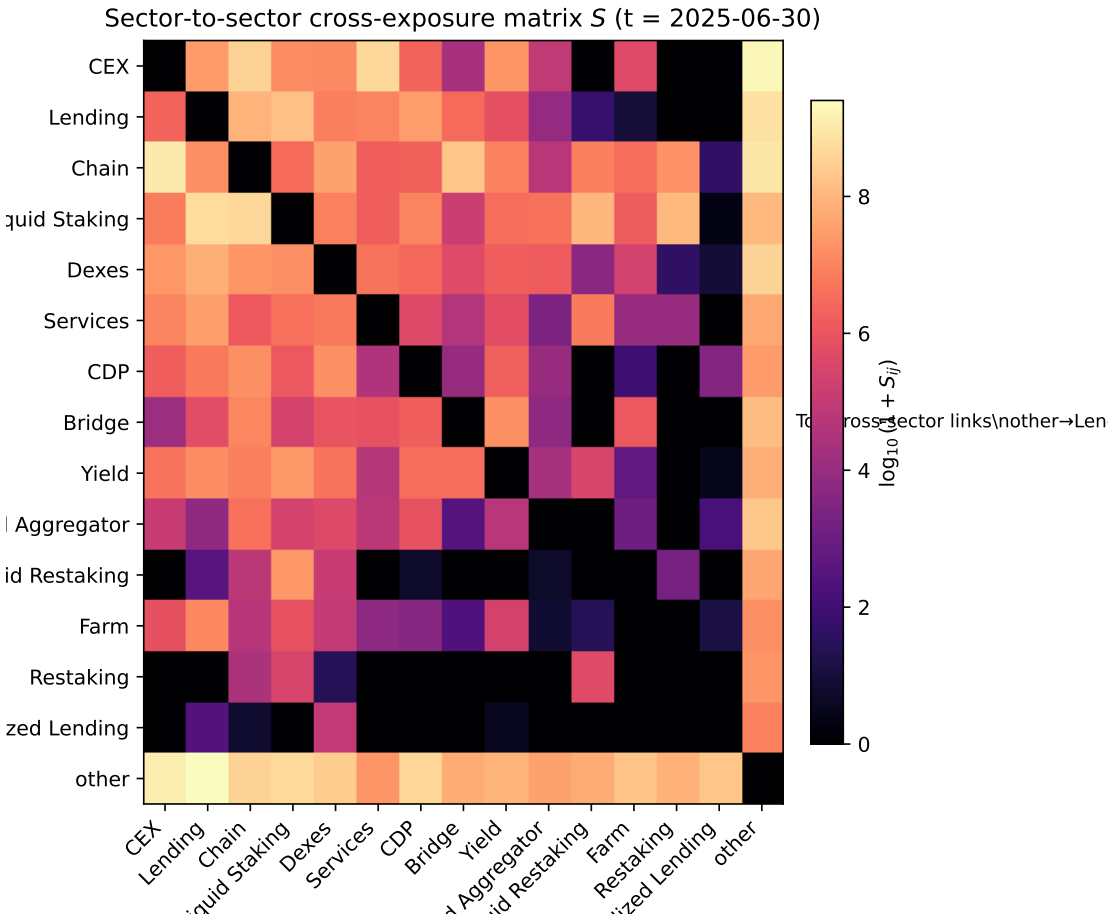

Figure 5. Sector-to-sector spillover matrix $\mathbf{S}$ on an observed snapshot G_t (t = 2025-06-30). Cell colours show $\log_{10}(1+S_{ij})$ where S_{ij} is total exposure from sector i to j. Off-diagonal flows are concentrated in a small number of sector pairs, predominantly involving Chain, Liquid Staking, Lending, CEX and the residual "other" bucket.

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. Predictive stress-testing summary on the 2025 hold-out. ΔMAE in system-loss percentage points; positive values mean DeXposure-FM beats persistence.

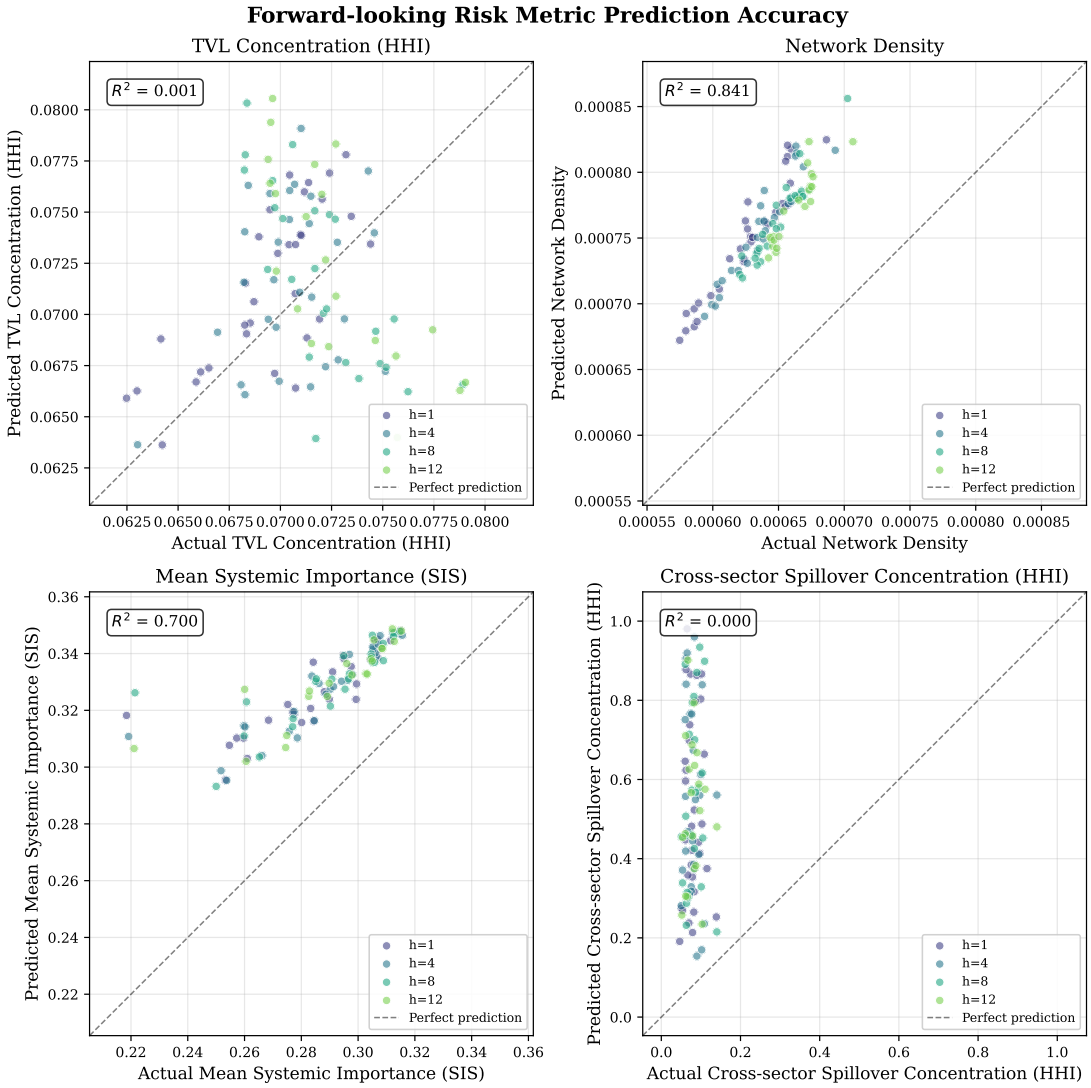

Figure 6. Forward-looking risk-metric forecasting on predicted exposure graphs. Each point compares a realised metric on G_{t+h} (x-axis) to the same metric on $\widehat{G}_{t+h}$ (y-axis), across horizons. Network density and mean SIS track realised values closely (R^2=0.84 and 0.70); the two HHI-based concentration measures show low level-correlation, with predicted HHI biased upward. Consistent with the forecast-then-measure pipeline's calibration-and-monitoring role, these indicators are interpreted as flags for directional changes and large deviations rather than as calibrated level forecasts.