Systemic Risk in the Global Banking System: A Graph Neural Network Surrogate Approach

Jim Gantsidis
Department of International and European Economic Studies
Thesis supervisor: Vasia Panousi

Abstract

This paper develops a Graph Neural Network surrogate model to evaluate systemic risk across all 29 Global Systemically Important Banks. To bypass the confidentiality of regulatory data, a multi-channel proxy interbank network is reconstructed using a structural gravity model calibrated on Bank for International Settlements statistics and public disclosures. A sequential cascade engine then simulates 1,449 stress scenarios to capture how financial distress propagates through the system. The Graph Transformer surrogate precisely replicates these complex cascade dynamics out of sample, achieving an R^2 of 0.980 for aggregate system losses and 0.938 for secondary default counts. The resulting high-speed evaluation framework reveals that systemic risk is heavily concentrated within Chinese megabanks, led by ICBC with a simulated loss footprint of $995 billion. These ordinal risk rankings are highly robust against an alternative maximum-entropy topology ($\rho = 0.877$). Finally, a significant negative rank correlation ($\rho = -0.400$) with a market-implied ΔCoVaR benchmark demonstrates that structural balance-sheet contagion and market-priced risk perceptions capture distinct, complementary dimensions of systemic centrality.

Email: jim.gantsidis@gmail.com. I would like to thank my thesis supervisor, Professor Vasia Panousi, for her guidance, support, and constructive feedback throughout this thesis. I hereby declare that this thesis is my own original work and that all external sources have been properly acknowledged and cited.

1 Introduction

The structural stability of the global financial system depends critically on the dense topological network of Global Systemically Important Banks (G-SIBs). When an individual systemic institution experiences acute distress, the resulting balance-sheet shocks can propagate non-linearly across international borders. This contagion travels through multi-layered transmission channels—including direct bilateral credit exposures, common asset overlaps, and wholesale funding market dependencies—potentially triggering cascading insolvencies with severe macroeconomic consequences. The Global Financial Crisis (GFC) of 2007–2009 demonstrated that traditional microprudential regulatory frameworks, which analyze banks in isolation, were ill-equipped to capture these network amplification mechanisms.

This paper addresses the fundamental empirical challenge of assessing systemic risk within the global interbank network when granular, bilateral counterparty exposures are shielded by regulatory confidentiality. To circumvent this information asymmetry, this study offers three primary contributions to the financial stability literature. Firstly, a proxy global interbank network is mapped for all 29 G-SIBs identified by the Financial Stability Board (FSB), utilizing a structural gravity model calibrated against Bank for International Settlements (BIS) Consolidated Banking Statistics and publicly available high-level regulatory disclosures. Secondly, a multi-round, sequential cascade simulation engine is deployed to generate a high-dimensional dataset of 1,449 labeled stress scenarios, systematically mapping the systemic risk surface across diverse shock configurations, localized severities, and Loss Given Default (LGD) regimes. Thirdly, a deep Graph Neural Network (GNN) surrogate architecture is trained on these simulation manifolds, enabling the ultra-low-latency evaluation of system-wide losses and establishing a data-driven, topologically aware systemic importance ranking of individual G-SIBs.

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3.3 details the data and sources. Section 4 formalizes the network reconstruction methodology, the systemic contagion engine, and the Graph Transformer surrogate architecture. Section 5 discusses the empirical results, focusing on surrogate model performance, the cross-sectional systemic importance rankings, joint-shock scenario analysis, and structural robustness checks against alternative network topologies and market benchmarks. Finally, Section 6 provides concluding remarks and outlines relevant macroprudential policy implications alongside prospective avenues for future research.

2 Related Literature

This paper relates to four distinct strands of literature. First, the literature on systemic risk and financial contagion emphasizes the role of network structure in determining the resilience of the banking system. Allen and Gale (2000) demonstrated that the architecture of interbank connections can either mitigate or amplify shocks, depending on the degree of diversification across institutions. Eisenberg and Noe (2001) subsequently developed a clearing framework for interconnected financial systems, studying default propagation through financial networks. Building on these foundations, Nier et al. (2007) showed that contagion outcomes are highly sensitive to network connectivity, capitalization, and concentration. Soramaki et al. (2007) provided empirical evidence on the topology of interbank payment systems, documenting structural features that influence

the transmission of financial distress. Battiston et al. (2012) introduced DebtRank, a recursive measure of systemic importance that captures the indirect effects of distress propagation beyond immediate defaults. Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015) further demonstrated that, while dense financial networks can absorb small shocks efficiently, they may become highly fragile when exposed to sufficiently large disturbances. Together, these studies establish the importance of modeling interconnected balance-sheet exposures when assessing systemic risk.

A second literature addresses the challenge that bilateral interbank exposures are generally unavailable to researchers and often inaccessible even to regulators. As a result, numerous approaches have been proposed to reconstruct hidden financial networks from aggregate information. Early studies frequently relied on maximum-entropy methods to infer bilateral exposures consistent with observed balance-sheet constraints. While computationally convenient, these approaches tend to produce overly diversified networks and may underestimate contagion risk. Subsequent work has therefore explored alternative reconstruction techniques that generate sparser and more realistic topologies. The broader network reconstruction literature has highlighted the sensitivity of systemic risk estimates to assumptions regarding the underlying structure of financial linkages. The approach adopted in this paper contributes to this literature by employing a structural gravity framework calibrated using BIS statistics and publicly available disclosures to infer a multi-channel network among all G-SIBs. The comparison with an alternative maximum-entropy topology further allows an assessment of the robustness of resulting systemic risk rankings.

The paper also contributes to the growing literature that applies machine-learning methods to financial networks and systemic risk analysis. Traditional contagion simulations are computationally intensive because each stress scenario requires the iterative propagation of losses through a network of interconnected institutions. Recent advances in graph representation learning have created new opportunities for approximating such complex dynamics. Graph Neural Networks are specifically designed to exploit relational information and have demonstrated strong performance across a wide range of network prediction tasks. Comprehensive surveys by Wu et al. (2021) and Hamilton (2020) document the rapid development of graph learning methods and their increasing relevance for economic and financial applications. Although machine learning has been widely applied to credit risk, asset pricing, and financial forecasting, its use as a surrogate model for systemic risk simulations remains relatively limited. The present paper extends this emerging literature by training a Graph Transformer architecture to emulate the outcomes of a large-scale contagion engine. Rather than replacing the underlying economic model, the surrogate learns its mapping from network characteristics and initial shocks to systemic outcomes, thereby enabling high-speed evaluation of thousands of stress scenarios.

Finally, this paper relates to the literature on market-based measures of systemic importance. One of the most influential approaches is CoVaR, introduced by Adrian and Brunnermeier (2016), which measures the contribution of a financial institution to system-wide tail risk. Related measures include the Marginal Expected Shortfall framework developed by Acharya et al. (2017) and the SRISK methodology proposed by Brownlees and Engle (2017). These indicators infer systemic importance from market prices and investor expectations rather than from direct balance-sheet exposures. While market-based measures have proven useful for identifying institutions perceived as systemically important by investors, they do not explicitly model the channels through which losses propagate across financial networks. The significant divergence between the rankings produced by

the contagion framework in this paper and those implied by CoVaR is therefore consistent with the broader view that structural interconnectedness and market-perceived risk capture distinct but complementary dimensions of systemic importance.

Taken together, this paper bridges several strands of literature by combining network reconstruction, contagion modeling, and graph-based machine learning within a unified framework. It contributes to the systemic risk literature by developing a global network of all G-SIBs, introducing a highly accurate Graph Transformer surrogate for contagion simulations, and providing new evidence that structural and market-based measures of systemic centrality capture different aspects of financial vulnerability.

3 Data

This section details the empirical data and corresponding institutional sources utilized to construct, simulate, and validate the systemic risk framework. Appendix A provides detailed lists of abbreviations, technical terms and mathematical notation.

3.1 Regulatory Data

The foundational empirical analysis is constructed from a granular panel of 29 Global Systemically Important Banks (G-SIBs) selected from the Financial Stability Board's (FSB) 2025 annual list. These institutions span ten sovereign jurisdictions—including the United States, China, France, Japan, the United Kingdom, Canada, Switzerland, Spain, Germany, and the Netherlands—and represent the systemic core of the global interbank market operating under the Basel III regulatory framework.

For the European G-SIBs, data are primarily extracted from the European Banking Authority (EBA) Transparency Exercise disclosures where available. This subset covers six institutions: Groupe Cr'edit Agricole, Groupe BPCE, Soci'et'e G'en'erale, Deutsche Bank, Santander, and ING. The December 2024 regulatory filings provide exact, audited Common Equity Tier 1 (CET1) ratios, Tier 1 capital ratios, leverage ratios, and total risk-weighted assets (RWA). These supervisory figures replace the baseline market-database estimates within the node feature space. Incorporating empirically observed RWA figures is vital, as it replaces a uniform RWA density assumption of 0.45 with institution-specific values. The observed densities range from approximately 0.19 for Groupe BPCE to 0.28 for Santander, a variance that materially impacts the CET1-based component of the hybrid capital buffer embedded in the contagion simulation. Notably, because Groupe BPCE lacks a public equity series and Refinitiv fundamentals, its CET1 capital, leverage ratio, and RWA density are sourced exclusively from its EBA filings.

For the United States G-SIBs, data are gathered from the Federal Reserve, which mandates that domestic bank holding companies with consolidated assets exceeding $100 billion file the FR Y-15 systemic risk report quarterly. This reporting panel provides additional, high-fidelity interconnectedness metrics for the eight US institutions in the sample. These granular filings supplement the broader Basel Committee on Banking Supervision (BCBS) indicator set and sharpen the calibration of the gravity model used to construct bilateral exposures, though they do not fully substitute for confidential supervisory bilateral exposure datasets.

In sum, the empirical architecture utilizes BCBS systemic indicators for all 29 banks, specialized EBA supervisory disclosures for 6 European institutions, FR Y-15 reports for the 8 US institutions, and Refinitiv (LSEG) financial fundamentals and equity series for

all G-SIBs with the sole exception of the privately owned Groupe BPCE. The overall distribution of these fundamental indicators can be observed in **??**.

3.2 Market Data

To evaluate the external validity of the simulation-based findings against a market-determined benchmark, ΔCoVaR estimates are generated from daily equity return series. The ΔCoVaR metric quantifies the marginal shift in the Value-at-Risk of the financial system conditional on an individual institution experiencing extreme distress, serving as a well-established proxy for market-implied systemic risk contributions. Bilateral ΔCoVaR pairs are computed for the 28 publicly traded institutions with sufficient historical equity data, yielding 756 directed institution-pair observations. The resulting market-implied cross-sectional ranking provides an independent validation benchmark to compare against the GNN-derived risk rankings, testing the degree to which balance-sheet network topology aligns with the risk dimensions priced by liquid capital markets.Because actual bilateral exposure matrices collected under the Bank for International Settlements (BIS) reporting framework remain confidential, all institution-to-institution linkages are proxied from these public fundamentals and regulatory indicators via the network gravity model detailed in Section 4.

Bilateral exposure data, as collected by supervisors under the BIS reporting framework, remain confidential. Therefore, all bilateral exposures used in this study are proxied from publicly available fundamentals and regulatory indicators using a gravity model described in Section 4.

3.3 Variables

Given the confidentiality restrictions surrounding direct interbank exposure data, the network graph is constructed by synthesizing institution-level regulatory disclosures with aggregated country-level financial statistics. Institutional balance-sheet variables are sourced from Refinitiv (LSEG) and include total assets, total deposits, Tier 1 capital, and selected accounting line items. Where databases report aggregate Tier 1 capital rather than explicit CET1 figures, the latter is approximated via the structural prior of $0.93 \times$ Tier 1, assuming Additional Tier 1 (AT1) instruments comprise approximately seven percent of the core structure.Baseline risk-weighted assets are initially modeled as 45 percent of total assets. A structural caveat to this uniform proxy is that Chinese banking institutions historically exhibit lower risk-weighted asset densities, typically around 35 to 40 percent. This uniform proxy may systematically overstate their capital buffers within the simulation, thereby understating their marginal contribution to system-wide losses. Conversely, United States investment banking entities often exhibit higher densities of 55 to 65 percent, meaning the benchmark proxy understates their structural buffers and overstates their simulated loss contribution. For the six European G-SIBs captured by the EBA Transparency exercise, these stylized assumptions are replaced with empirical, bank-specific densities.Additionally, daily closing equity prices spanning a five-year historical window are extracted from Refinitiv for the 28 publicly traded entities, with the exception of the privately owned Groupe BPCE. These market series are necessary for the statistical estimation of ΔCoVaR as well as for computing the 90-day annualized equity volatility, 30-day returns, and 90-day returns utilized across subsequent modeling stages.The primary structural network linkages are driven by indicators sourced from the Basel Committee

on Banking Supervision, which requires G-SIBs to disclose twelve high-level systemic risk metrics annually. Eight of these metrics are map-compiled into the proxy network architecture. The total exposures indicator represents the comprehensive Basel III leverage exposure and establishes the primary scale proxy for each node. Intra-financial-system assets, which capture all claims on other financial institutions, dictate each bank's aggregate outbound interbank exposure capacity. Conversely, intra-financial-system liabilities, representing funding obligations to other financial counterparties, anchor the marginal constraints of the alternative maximum-entropy benchmark network.International connectivity is governed by cross-jurisdictional claims, which dictate the cross-border proportion of each bilateral linkage, and cross-jurisdictional liabilities, which reflect international funding profiles used in the exposure-channel decomposition. The gross OTC derivatives notional determines the proportional weight of each bank's derivative-channel linkages. Finally, trading and available-for-sale securities measure mark-to-market asset pools used to calibrate liquidity-drain sensitivities, while Level 3 assets capture highly illiquid fair-value instruments that, alongside trading and available-for-sale securities, parameterize the fire-sale contagion channel.All monetary values are standardized and denominated in USD. Refinitiv fundamentals are extracted in their native USD configurations. BCBS G-SIB indicators and FR Y-15 metrics, typically reported in millions of USD, are rescaled to absolute USD levels. EBA Transparency data contribute solely dimensionless regulatory ratios, exempting that specific data stream from currency translation effects.

4 Methodology

The central analytical challenge in this study stems from the fact that granular, bilateral interbank exposures among Global Systemically Important Banks (G-SIBs) are shielded by regulatory confidentiality. While supervisory datasets—such as the European Banking Authority's asset quality reviews or the Federal Reserve's Dodd-Frank Act Stress Testing disclosures—provide valuable systemic insights, they are invariably aggregated by broad counterparty sectors or withheld from public disclosure entirely. To circumvent this empirical constraint, the strategy adopted here is to systematically reconstruct a proxy exposure network. This synthetic topology is designed to replicate the defining empirical regularities of empirical interbank networks, specifically structural sparsity, a distinct core-periphery architecture, and pronounced geographic clustering, utilizing the publicly available consolidated indicators detailed in Section 3.3.

This network-based framework is executed across four interconnected methodological phases. The architecture begins with the construction of the proxy interbank network via a structural gravity model calibrated directly against Bank for International Settlements consolidated statistics. Once this baseline topology is established, it is subjected to a multi-round network contagion simulation, which systematically propagates financial distress to generate a large, labeled dataset of systemic shock scenarios. These simulation outputs subsequently serve as the training targets for a Graph Neural Network surrogate model, which is optimized to approximate the highly non-linear dynamics of the contagion engine at a fraction of the computational cost. Finally, the framework concludes with a series of validation and robustness protocols that evaluate the model's predictive accuracy by comparing the network-derived systemic rankings against both an alternative maximum-entropy network topology and an independent, market-implied risk benchmark.

4.1 Interbank Network Construction

Bilateral exposures between banks are estimated using a network gravity model, which posits that the financial flow between two institutions is directly proportional to their respective economic sizes and inversely related to bilateral frictions.[1] To calibrate the country-level parameters, the model utilizes the Bank for International Settlements (BIS) Consolidated Banking Statistics (CBS). For each reporting jurisdiction, the predicted share of cross-border claims allocated to a counterparty country is modeled as a function of the destination country's total inbound size, scaled by geographic friction multipliers consisting of a same-country bias, a same-region bias, and a cross-region reference baseline normalized to 1.0. Because the BIS CBS data excludes domestic claims, the same-country bias cannot be directly identified; it is therefore fixed at a conservative prior of 2.0, consistent with the empirical international banking literature on home bias avdjiev2012. Conversely, the same-region bias is endogenously estimated at 2.25 by minimizing the mean squared error between predicted and observed bilateral cross-border claim shares via Nelder-Mead optimization. At the idiosyncratic level, bank-specific connectivity scaling uses the square root of total assets as the size proxy, and a cross-border floor of 0.2 is imposed as a minimum exposure share for all cross-regional pairs to prevent structural disconnection between distant jurisdictions.

To capture distinct systemic risk vulnerabilities, the aggregate bilateral exposure matrix is decomposed into five distinct transmission channels, enabling the contagion engine to apply channel-specific Loss Given Default (LGD) rates. Given data constraints, this decomposition relies on a combination of bank-specific indicators and fixed structural priors. The cross-border claims channel is calculated using the bank's Basel III cross-jurisdictional claims indicator relative to its total outbound exposure, while the derivatives channel is calibrated from the bank's over-the-counter (OTC) derivatives nominal volume scaled by total assets. Conversely, the payment links and common overlaps channels are fixed uniformly across all node pairs at 2 percent and 5 percent, respectively, due to the absence of public bilateral data for these specific networks. The interbank loans channel absorbs the remaining residual exposure.

To account for the uncertainty inherent in this estimation technique, small random perturbations are introduced into the estimated connection weights.[2] The baseline calibrated topology yields a directed network of 29 nodes and 331 directed edges, where degree centrality ranges from 3 to 11 outbound counterparties per bank. Finally, as a structural robustness check, an alternative dense network is constructed via maximum-entropy optimization, constrained such that the row and column sums of the estimated exposure matrix exactly replicate each institution's aggregate intra-financial-system assets and liabilities.

4.2 Contagion Simulation

The systemic contagion framework is formalized as a multi-round, sequential cascade model designed to endogenously track the non-linear erosion of equity across the interbank network. Each simulation run begins by subjecting a designated subset of G-SIBs to an initial exogenous macroeconomic shock, which directly reduces a specified fraction of their total asset value. This asset impairment propagates through the network topology

[1] This approach is directly analogous to the structural gravity models used in international trade.

[2] A fixed random seed is utilized to ensure these perturbations remain deterministic across simulations, preserving exact reproducibility.

via bilateral linkages. If the cumulative mark-to-market losses inflicted on a counterparty exceed the structural capital buffers implied by its Common Equity Tier 1 (CET1) ratio, scaled by the channel-specific Loss Given Default (LGD) rates of its exposures, the institution triggers a default state, subsequently initiating further downstream asset devaluation cascades.

Each bank's CET1-based equity buffer is constructed as:

$$\text{Buffer}_{\text{CET1}} = \text{Total Assets} \times \text{CET1 ratio} \times \text{RWA density} \tag{1}$$

RWA density defaults to 0.45 (the G-SIB average) but is replaced by bank-specific values from EBA Transparency disclosures for the six European G-SIBs. The leverage-based buffer is:

$$\text{Buffer}_{\text{lev}} = \text{Total Assets} \times \text{Leverage ratio} \tag{2}$$

Under the "hybrid" capital mode, the simulated buffer is the simple average of these two definitions.

Systemic defaults propagate through the topology via a sequential loss-absorption mechanism, wherein each creditor institution absorbs a balance-sheet impairment proportional to its specific bilateral exposure to the defaulted entity, scaled by the channel-specific Loss Given Default (LGD) rate. If these secondary counterparty losses entirely exhaust a creditor's residual capital buffers, that institution triggers a subsequent default state, propagating the distress vector further downstream until the network reaches a stable equilibrium where no additional insolvencies are induced. To systematically evaluate the structural resilience of these institutions under varying macroeconomic assumptions, the framework evaluates three distinct channel-specific LGD regimes—Benign, Base, and Stress—as summarized in Table 4, where the parameter hierarchy across transmission channels reflects the empirical seniority and collateralization structures typical of each exposure class.

Beyond direct credit losses, surviving institutions are simultaneously subjected to secondary liquidity and market-risk spillovers. First, funding-stress haircuts are imposed on surviving banks, endogenously calibrated as a function of each institution's short-term wholesale funding dependency, its Liquidity Coverage Ratio (LCR) shortfall relative to the 1.0 regulatory baseline, and its 90-day historical equity volatility. Second, asset fire-sale dynamics inflict mark-to-market losses proportional to the asset-weighted volume of concurrent liquidations and each bank's baseline allocation to trading-book and Level 3 illiquid assets. Any institution whose core equity is driven to or below zero by these joint credit and liquidity channels is structurally categorized as defaulted, acting as a renewed source of contagion in the subsequent iteration. This cascading process terminates when the network achieves structural stabilization with zero incremental failures, or upon reaching a hard computational horizon of 15 rounds.

To construct a comprehensive training target space for the downstream Graph Neural Network surrogate model, a high-dimensional combinatorial grid yields a total of 1,449 distinct simulation scenarios. This dataset is generated by systematically permuting the structural configurations of the initial stress vector: the cardinality and specific identity of the initially shocked nodes—spanning from single-institution idiosyncratic failures to highly correlated multi-bank systemic shocks—the localized shock severity expressed as a continuous fraction of total assets, and the prevailing global LGD regime. Each discrete simulation run records three primary target variables mapping the systemic risk surface: aggregate financial system losses denominated in USD, the absolute count of induced

secondary defaults, and the temporal duration quantified as the number of rounds required to reach network stabilization.

4.3 Graph Neural Network Surrogate

To accelerate the evaluation pipeline, a Graph Neural Network (GNN) surrogate model is introduced to directly approximate the non-linear mapping between the initial shock parameters and the downstream contagion metrics. By treating the physical simulation engine as a black box, this surrogate architecture serves as an ultra-low-latency representation of the cascade dynamics, bypassing the computationally intensive multi-round iterative loops of the physical contagion simulation. The neural network takes the identical calibrated baseline network topology as input and output-maps the latent structural representations directly into the three aggregate system-level targets. The foundational architecture is formalized as a four-layer Graph Transformer network operating with four parallel attention heads. This mechanism allows each node within the financial network to dynamically weight and aggregate features from its immediate neighborhood, where the localized aggregation paths are governed by learned, anisotropic attention coefficients. To stabilize gradient propagation throughout this deep graph architecture, the model maintains a 128-dimensional hidden representation augmented with layer normalization and residual skip connections at each message-passing layer.

At the input layer, the graph's structural properties are dense-encoded through specialized node and edge feature spaces. Each bank node is parameterized by a 19-dimensional vector encompassing critical idiosyncratic balance-sheet and market indicators, such as total assets, CET1 capital ratios, Risk-Weighted Asset density, and intra-financial-system asset shares. Crucially, this node feature space includes an explicit shock indicator vector, which allows the model to naturally generalize to concurrent, multi-bank stress events by encoding the localized severity of the initial exogenous impairment. Simultaneously, connectivity linkages are defined by a standardized multidimensional edge vector containing the five channel-specific bilateral exposure weights, including interbank loans, over-the-counter derivatives, cross-border claims, common asset overlaps, and payment links, alongside an aggregated total exposure proxy derived from the structural gravity model.

Following the graph attention layers, the latent node embeddings are processed by a multi-task decoding architecture consisting of two highly specialized prediction heads. The first is a localized, node-level classification head that applies a sigmoid activation function to project individual node states into discrete, bank-specific default probabilities. The second is a global, graph-level regression head that condenses the entire permutation network via a permutation-invariant global mean pooling operation. The resulting graph embedding is passed through a deep Multi-Layer Perceptron to output the three global target variables, which are total financial system loss, absolute secondary default count, and the temporal duration of the cascade.

The empirical optimization space comprises the 1,449 simulation scenarios, which combine single-bank idiosyncratic failures across five severity thresholds and three LGD regimes with randomly sampled multi-bank simultaneous shocks, spanning two to five concurrent node failures, and country-level regional clusters. The deliberate inclusion of joint-probability multi-bank shock vectors provides a vital regularizing constraint because without this structural representation in the training manifold, the GNN exhibits an out-of-distribution extrapolation bias, overestimating systemic financial losses during joint

stress events by up to 37 percent. The dataset is strictly partitioned into a 70 percent training set consisting of 1,014 scenarios, a 15 percent validation set consisting of 217 scenarios, and a 15 percent out-of-sample held-out test set consisting of 218 scenarios.

Model parameters are optimized via the AdamW algorithm utilizing decoupled weight decay to enforce structural regularization. The model is initiated with a learning rate of 5×10^{-4} and a weight decay parameter of 10^{-4}. Given the multi-task learning configuration, the composite loss function optimizes a joint objective, minimizing a binary cross-entropy loss for the node-level classification task and a robust Huber loss, which mitigates the influence of extreme systemic tail risks, for the graph-level regression targets. The optimization trajectory employs a dynamic learning rate scheduler that reduces the learning rate by a factor of 0.5 upon validation loss stagnation, regulated by an early-stopping patience threshold of 50 epochs to prevent over-fitting. Final generalization performance is validated against the held-out test set utilizing the coefficient of determination and the mean absolute error.

4.4 Systemic Importance Ranking and Validation

The GNN surrogate model enables a granular, computationally precise assessment of idiosyncratic vulnerability and systemic centrality across the interbank network. To quantify each node's systemic importance, the model formalizes a systemic vulnerability score defined as the aggregate financial system loss induced under a simulated full idiosyncratic liquidation event, parameterized by a shock fraction of 1.0, the stress LGD regime, and the integration of hybrid capital buffers. This marginal contribution metric operates analogously to cooperative game-theoretic frameworks, such as Shapley-value-inspired risk indicators established in the systemic risk literature, by endogenously isolating the network amplification mechanisms that drive cascade dynamics. Consequently, the score evaluates vulnerability as a joint function of balance-sheet scale and topological placement. Thus, a large but structurally insulated bank will generate a significantly lower contagion signature than a counterparty of comparable magnitude that is deeply integrated into the core network architecture.

4.5 Market Validation via ΔCoVaR

To establish external empirical validity, the simulation-based systemic risk rankings are cross-validated against the market-implied ΔCoVaR metric, bridging structural balance-sheet topology with public market perceptions of distress transmission. For each idiosyncratic bank pair (i, j), the directional systemic impact $\Delta\text{CoVaR}_{i \to j}$ is formalized as the marginal shift in the conditional value-at-risk of counterparty j at the 5th percentile when institution i transitions from its median state to a state of extreme financial distress. This bivariate relationship is endogenously estimated via quantile regression utilizing a five-year rolling panel of daily asset log-returns, where any institution pairs lacking a minimum threshold of 100 concurrent non-null observations are structurally pruned to avoid statistical noise. The aggregate outward systemic risk footprint for a given bank i is subsequently derived by computing the mean absolute ΔCoVaR across the remaining institutional universe ($j \neq i$). By organizing the sample into descending cross-sectional ordinals based on both this empirical market-implied vector and the structural simulation-based outputs, the framework directly quantifies the degree of alignment between network-driven cascade vulnerabilities and the real-time forward pricing of systemic risk by financial markets.

4.6 Robustness and Sensitivity Analysis

To establish the structural stability of the simulation findings, the framework deploys two primary robustness protocols designed to test sensitivity across parametric and topological dimensions.

First, the entire network contagion architecture is systematically re-executed across nine distinct structural permutations, capturing the complete combinatorial space of three alternative capital buffer definitions—Common Equity Tier 1, leverage ratio, and a hybrid capital mode—and the three Loss Given Default regimes. The cross-sectional consistency of the resulting institutional risk rankings is rigorously evaluated across these alternative configurations to ensure the baseline findings are not idiosyncratic artifacts of specific regulatory or loss-absorption assumptions.

Second, to assess the model's sensitivity to the underlying network reconstruction methodology, the systemic rankings derived from the baseline gravity network are compared against those generated by executing the identical cascade engine on an alternative maximum-entropy network topology. The structural alignment between the rankings produced by these two fundamentally different network architectures is formally quantified using a Spearman rank correlation coefficient.

5 Results

5.1 Surrogate Model Performance

Before turning to the systemic importance rankings, it is necessary to confirm that the GNN surrogate reliably approximates the contagion simulation. The GNN surrogate demonstrates strong predictive accuracy on the test set of 218 scenarios. On the held-out test set, the surrogate achieves a coefficient of determination of 0.980 for total systemic loss, 0.938 for the number of secondary defaults, and 0.906 for the number of cascade rounds. Performance is highly consistent across data splits, indicating that the surrogate generalises well outside the training distribution.

The corresponding scatter plot of simulated versus predicted values across all three targets is provided in **??**. The cascade-rounds target achieves the lowest test-set score among the three outputs, reflecting the discrete, threshold-driven nature of this metric, where small continuous errors in loss prediction can shift a bank over its default threshold. Performance is nonetheless consistent across all three data splits, confirming that the model recovers most of the systematic variation. For the purpose of systemic importance ranking, system loss and default count are the economically relevant outputs, and the surrogate performs accurately on both.

The residual distributions for all three targets are shown in **??**. Residuals are approximately mean-zero and tightly concentrated near zero across the three outputs, with a mild left-skew in the system-loss residuals indicating a slight tendency to under-predict the most extreme, low-probability tail losses.

5.2 Systemic Importance Ranking and Country-Level Concentration

With the surrogate validated, the analysis turns to the systemic importance of individual G-SIBs. Each of the 29 banks is subjected to an idiosyncratic failure shock, defined as full

equity depletion under a stress LGD regime, and the resulting system-wide loss is used as the ranking criterion.

The top of the systemic importance hierarchy is dominated by Chinese megabanks, as illustrated in the comprehensive cross-sectional ranking in **??**. The Industrial and Commercial Bank of China ranks first with a simulated system loss of $995 billion, followed by China Construction Bank, Agricultural Bank of China, Bank of China, and JP Morgan Chase. HSBC ranks sixth, BNP Paribas seventh, and Bank of America eighth.

While balance-sheet size is the main driver, it is not the sole determinant. Citigroup ranks fifteenth, placing it below the smaller Barclays which ranks tenth, reflecting Citigroup's lower cross-jurisdictional claims ratio and stronger funding profile. Similarly, several large US institutions rank lower than their size would suggest due to stronger capital buffers and lower reliance on wholesale funding that collectively dampen the funding-stress channel during a cascade.

Aggregating these results reveals a profound concentration of systemic risk by country. China accounts for over $3.6 trillion in total simulated system loss across its five G-SIBs, roughly three times the per-bank mean for US institutions. The four French G-SIBs place France third in aggregate and second on a per-bank basis. Single-institution countries like Spain and Switzerland post a per-bank systemic footprint comparable to the US average, while the Netherlands falls clearly below, reflecting the smaller cross-jurisdictional reach of its institution.

This regional risk concentration is conservative under the modelling assumptions used here. The proxy RWA-density applied to Chinese G-SIBs overstates their CET1-based equity buffer relative to the actual densities these institutions typically report. A correctly calibrated buffer would thin their loss-absorbing capacity, allowing larger losses to propagate before defaults are triggered, and would therefore widen the systemic gap between Chinese G-SIBs and the rest of the sample.

5.3 Stress Scenario Analysis

To complement the single-bank ranking, eight stress scenarios are evaluated to capture the network's response to coordinated multi-bank shocks. The scenarios span a wide range of severity, from a benign idiosyncratic shock to an extreme simultaneous failure of the five largest G-SIBs.

Results range from a lower bound of $9.3 billion to an upper bound of $3.31 trillion for the Global Top-5 scenario, visualised side-by-side with surrogate predictions in **??**. The Global Top-5 event is the only scenario in which the cascade extends beyond the initially shocked banks, producing cascading secondary defaults as demonstrated in Figure 9.

A key macroprudential insight is that aggregate monetary loss and cascade depth are distinct dimensions of systemic risk that do not necessarily move together. The China Banking Crisis scenario generates massive simulated system loss but produces zero secondary defaults. Because the regional shock does not fully deplete each bank's capital buffer, no institution crosses the absolute insolvency threshold required to ignite further sequential contagion rounds. Cascades materialise exclusively under the most extreme structural assumptions, where multiple of the largest global institutions fail simultaneously and completely.

Concerning out-of-sample surrogate accuracy, the named scenarios constitute a vital test for the surrogate since they correspond to plausible real-world stress configurations. Agreement is incredibly close on the moderate and high-loss scenarios, all falling within

roughly five percent of the physically simulated value.

5.4 Robustness Analysis

The findings presented above rest on modelling choices that require explicit robustness testing. Firstly, regarding network topology, the full single-bank exercise is repeated on the maximum-entropy alternative network to assess whether the rankings depend strictly on the gravity parameters. The two sets of rankings yield a highly significant Spearman correlation of 0.877, visualised fully in Figure 11. The exact set of the top five banks is perfectly preserved across both network architectures.

Secondly, concerning the risk measure, the simulation ranking is compared against the market-derived ΔCoVaR ranking, with the direct institutional comparison shown in **??**. The Spearman correlation between the two rankings is a statistically significant negative relationship. The most divergent institution is China Construction Bank, ranking near the absolute top by simulation but near the bottom by ΔCoVaR. This statistically detectable divergence is theoretically expected rather than indicative of a model failure. The simulation captures systemic importance through direct bilateral credit linkages and balance-sheet topology, whereas the market metric reflects real-time perceived risk, incorporating sentiment contagion and equity-market integration. The two measures therefore capture complementary, interacting dimensions of systemic risk.

5.5 Sensitivity Analysis

To test parameter sensitivity, the simulation is rerun for the most interconnected G-SIBs under all nine combinations of capital buffer definitions and LGD regimes. The relative ordering of these core banks is strictly preserved across all configurations. Each structural step up in LGD severity escalates mean system loss by approximately 17 to 19 percent, regardless of the underlying capital specification. Across all combinations, the maximum number of secondary defaults from any idiosyncratic shock remains tightly bounded at one, confirming that the network topology is fundamentally resilient to single-institution failures under any plausible calibration of the prevailing capital parameters.

5.6 Descriptive Regression

An ordinary least squares regression is estimated with simulated system loss as the dependent variable and twelve bank-level fundamentals as regressors. This exercise is strictly descriptive because the system size renders standard errors and inference metrics uninformative. Nevertheless, the coefficient signs are broadly consistent with theoretical expectations. Total assets and intra-financial-system assets enter with positive signs, indicating that larger and more interconnected institutions generate higher systemic footprints. Capital-ratio variables enter with negative signs, confirming that robustly capitalised institutions successfully absorb their own initial shock with minimal downstream spillover. The descriptive regression therefore serves as an internal consistency check, affirming that the model effectively bridges fundamental balance-sheet scale with systemic outcome.

6 Conclusions

The primary objective of this paper is to assess systemic risk within the global interbank network under the empirical constraint that granular, bilateral counterparty exposures among Global Systemically Important Banks (G-SIBs) remain confidential. To circumvent this information asymmetry, the analytical framework synthesizes three interconnected methodological layers: a gravity-calibrated proxy network, a multi-round network contagion simulation, and a Graph Neural Network (GNN) surrogate model. Each layer is engineered to address specific technical limitations inherent in traditional macroprudential risk measurement.

The structural gravity model, calibrated on BIS Consolidated Banking Statistics from nine reporting jurisdictions, achieves a coefficient of determination (R^2) of 0.861 on observed bilateral country-level market shares while reproducing realistic topological properties. The estimated architecture highlights a dense transatlantic core linking US, UK, and French institutions, whereas Chinese and Japanese banks form a distinct but highly interconnected regional cluster. The calibrated same-country bias of 2.0 and same-region bias of 2.25 align with the well-documented home-bias anomalies in international banking literature, validating the credibility of the proxied bilateral exposures.

The downstream GNN surrogate model, utilizing a four-layer Graph Transformer architecture, precisely replicates the non-linear dynamics of the physical contagion engine. It achieves an R^2 of 0.980 for aggregate financial system losses and 0.938 for secondary default counts on an out-of-sample test set of 218 scenarios. This high-fidelity performance establishes the surrogate as an efficient, rapid scenario-evaluation tool. While the underlying physical simulation relies on computationally intensive, iterative cascade calculations, the machine-learning surrogate delivers equivalent risk estimates instantaneously. This computational efficiency unlocks the high-speed evaluation of expansive counterfactual stress configurations that would otherwise be computationally prohibitive.

A prominent empirical finding across the analysis is the high concentration of systemic risk centered within Chinese megabanks. Under the stress Loss Given Default (LGD) regime, Industrial and Commercial Bank of China (ICBC) generates the largest simulated system loss at approximately \$995 billion, followed closely by China Construction Bank (\$880 billion) and Agricultural Bank of China (\$848 billion). At the sovereign level, the five Chinese G-SIBs account for an aggregate simulated system loss of \$3.633 trillion. This represents a per-bank average of \$727 billion, which is more than three times the corresponding US baseline of \$226 billion across eight institutions. Crucially, this risk concentration remains a conservative baseline under the prevailing modeling assumptions. The uniform 45% risk-weighted asset (RWA) density proxy overstates the Common Equity Tier 1 (CET1) capital buffers of Chinese banks relative to their empirically reported densities of 35–40%. Consequently, a fully granular capital calibration would amplify rather than diminish the observed concentration of risk within the Chinese banking sector.

Importantly, the analysis demonstrates that balance-sheet scale, while a dominant driver, is not the sole determinant of systemic importance. Citigroup, with \$2.66 trillion in total assets, ranks 15th in systemic impact—placing it below the smaller Barclays (\$2.08 trillion), which ranks 10th. This ordinal divergence reflects Citigroup's lower cross-jurisdictional claims ratio and a more robust funding profile. Similarly, Morgan Stanley ranks 25th despite managing \$1.42 trillion in assets, illustrating how lower structural interconnectedness substantially dampens an institution's downstream contagion footprint. The resulting systemic risk ranking thus captures a multi-dimensional composite

of balance-sheet size, topological network centrality, and liability structures that cannot be captured by any single accounting metric.

A two-dimensional robustness analysis confirms the stability of these structural findings. First, the cross-sectional systemic rankings exhibit high stability across alternative network topologies; the Spearman rank correlation between the gravity-based baseline and a maximum-entropy reconstruction is $\rho = 0.877$ ($p < 0.001$), with the top five risk positions preserved across both architectures. Second, a sensitivity analysis spanning nine distinct combinations of capital definitions and LGD regimes confirms that the ordinal ranking is highly stable. While the mean systemic loss fluctuates between \$336 billion and \$491 billion across these parametric configurations, the maximum number of secondary defaults triggered by any single-bank idiosyncratic shock remains capped at one. This indicates that no single institutional failure is capable of triggering a systemic collapse under baseline capitalization levels.

The stress scenario analysis offers an additional macroprudential insight: aggregate monetary losses and network cascade depth constitute distinct dimensions of systemic risk that do not necessarily covary. For instance, a simulated Chinese Banking Crisis—modeled as a simultaneous 80% equity depletion across all five Chinese G-SIBs—generates a massive \$1.52 trillion system-wide loss but triggers zero secondary defaults, as no counterparty fully crosses the threshold required to propagate further cascade rounds. Cascading failures materialize exclusively in the extreme Global Top-5 scenario (\$3.31 trillion in losses, six secondary defaults), where the five largest global institutions default simultaneously. This distinction carries meaningful regulatory implications, suggesting that supervisory frameworks focusing narrowly on cascade probabilities systematically underestimate the economic damage associated with severe, non-cascading systemic stress events.

Despite these insights, several empirical boundaries remain. Direct access to supervisory bilateral exposure data—even in an anonymized or aggregated format—would permit direct validation of the gravity-calibrated baseline and refine the precision of the systemic risk estimates. Furthermore, expanding the underlying contagion simulation to endogenously incorporate dynamic balance-sheet adjustments, liquidity hoarding spirals, and fire-sale feedback loops—which are currently parameterized as stylized haircuts—would bring the framework closer to the behavioral mechanics of real-world crises. Finally, the Graph Transformer surrogate model could be adapted for real-time macroprudential monitoring. By feeding the neural network with quarterly regulatory filings, supervisory authorities could maintain continuously updated systemic risk rankings and conduct immediate what-if policy evaluations, bridging the gap between slow-moving periodic stress tests and the rapid operational speed of modern financial markets.

Appendix A: Abbreviations, Glossary, and Notation

Table 1: List of abbreviations

Abbreviation	**Definition**
AFS	Available-for-Sale securities
AT1	Additional Tier 1 capital
BCBS	Basel Committee on Banking Supervision
BIS	Bank for International Settlements
CBS	Consolidated Banking Statistics (BIS)
CET1	Common Equity Tier 1
DFAST	Dodd–Frank Act Stress Test
EBA	European Banking Authority
FR Y-15	Federal Reserve Systemic Risk Report (Form Y-15)
FSB	Financial Stability Board
FX	Foreign Exchange
GNN	Graph Neural Network
G-SIB	Global Systemically Important Bank
LCR	Liquidity Coverage Ratio
LEI	Legal Entity Identifier
LGD	Loss Given Default
LSEG	London Stock Exchange Group (parent company of Refinitiv)
MAE	Mean Absolute Error
MLP	Multi-Layer Perceptron
OLS	Ordinary Least Squares
OTC	Over-the-Counter
RWA	Risk-Weighted Assets
VaR	Value-at-Risk

Table 2: List of technical terms

Term	**Definition**
Cascade (Contagion Cascade)	Sequential transmission of distress through the interbank network, whereby losses from one institution's failure are passed to its creditors and may trigger further defaults.
Core–Periphery Topology	Network structure characterised by a densely interconnected core of institutions surrounded by a sparsely connected periphery.
Fire-Sale Spillover	Losses incurred by surviving institutions when distressed banks liquidate assets at depressed prices, reducing the market value of similar holdings elsewhere in the system.
Funding-Stress Haircut	Additional equity reduction applied during a contagion round, calibrated using a bank's wholesale funding dependence, liquidity shortfall, and equity volatility.
Global Mean Pooling	Graph-level aggregation operation that averages node embeddings to produce a fixed-dimensional representation of the entire network.
Gravity Model	Network reconstruction approach in which bilateral exposures increase with institution size and decrease with economic or financial frictions.
Graph Transformer	Graph neural network architecture that employs multi-head self-attention to weight information received from neighbouring nodes.
Hybrid Capital Mode	Capital-buffer specification defined as the average of CET1-based and leverage-based capital buffers.
Huber Loss	Robust loss function that is quadratic for small errors and linear for large errors, balancing sensitivity and outlier resistance.
Maximum-Entropy Network	Network reconstruction method that allocates bilateral exposures as evenly as possible while matching observed row and column totals.
Message Passing	Fundamental GNN operation in which node representations are iteratively updated using information aggregated from neighbouring nodes.
RWA Density	Ratio of risk-weighted assets to total assets, used to translate regulatory capital ratios into absolute capital amounts.
Wholesale Funding	Funding obtained from institutional and market-based sources, such as interbank borrowing and commercial paper, rather than retail deposits.

Table 3: Mathematical Notation

Symbol	**Definition**
ΔCoVaR	Difference between the Value-at-Risk of the financial system conditional on an institution being in distress and the corresponding value at its median state.
R^2	Coefficient of determination; measures the proportion of variation in the target variable explained by the model.
ρ	Spearman rank correlation coefficient.
q	Tail probability used in the estimation of ΔCoVaR (set to 0.05 in this study).

Appendix B: Tables and Figures

Table 4: Channel-Specific Loss-Given-Default (LGD) Assumptions by Scenario

Contagion Channel	**Benign**	**Base**	**Stress**
Interbank Loans	40%	60%	80%
Derivatives	15%	30%	50%
Cross-Border Claims	30%	50%	70%
Common Asset Overlap	8%	15%	30%
Payment Links	2%	5%	10%

Table 4 reports the channel-specific loss-given-default (LGD) assumptions used in the contagion simulations under three stress regimes: benign, base, and stress. LGD values increase monotonically across scenarios to reflect progressively adverse market and funding conditions. Interbank loans and cross-border claims are assigned the highest LGDs owing to their direct credit exposure, whereas payment links and common asset overlaps receive lower LGDs, reflecting their comparatively indirect transmission of losses. These assumptions enable the model to capture heterogeneity in loss transmission across different contagion channels.

Table 5: Performance of the GNN surrogate model across data splits

Target Variable	**Training** ($n = 1{,}014$)	**Validation** ($n = 217$)	**Test** ($n = 218$)
System Loss	0.984 ($27.7B)	0.970 ($38.3B)	0.980 ($28.4B)
Default Count	0.939 (0.18 banks)	0.935 (0.22 banks)	0.938 (0.16 banks)
Cascade Rounds	0.930 (0.09 rounds)	0.941 (0.08 rounds)	0.906 (0.09 rounds)

Entries report the coefficient of determination (R^2). Mean absolute error (MAE) is reported in parentheses. Sample sizes for the training, validation, and test datasets are shown in the column headings.

Table 6: Top 15 Systemically Important Banks

Rank	Bank	Country	Assets ($bn)	System Loss ($bn)
1	Industrial and Commercial Bank of China	CN	7,650	995
2	China Construction Bank	CN	6,530	880
3	Agricultural Bank of China	CN	6,980	848
4	Bank of China	CN	5,490	632
5	JP Morgan Chase	US	4,420	485
6	HSBC	UK	3,230	462
7	BNP Paribas	FR	3,280	415
8	Bank of America	US	3,410	404
9	Mitsubishi UFJ FG	JP	2,720	303
10	Barclays	UK	2,080	280
11	Bank of Communications (BoCom)	CN	2,220	277
12	Groupe Crédit Agricole	FR	2,790	276
13	Santander	ES	2,190	269
14	Groupe BPCE	FR	1,960	260
15	Citigroup	US	2,660	249

Table 6 presents the fifteen institutions with the highest systemic importance scores under the stress LGD scenario and hybrid capital specification. Chinese banks dominate the upper end of the ranking, with Industrial and Commercial Bank of China generating the largest simulated system-wide loss, followed by China Construction Bank and Agricultural Bank of China. Large globally active banks from the United States, United Kingdom, France, Japan, and Spain also feature prominently, highlighting the concentration of systemic risk among the world's largest and most interconnected banking institutions. Rankings are based on simulated system-wide losses generated by the failure of a single institution under the stress LGD scenario and hybrid capital-buffer specification. Total assets are reported in billions of US dollars. System loss denotes the aggregate loss incurred by the financial system following the initial bank default.

Table 7: Country-Level Concentration of Systemic Risk

Country	**Banks**	**Aggregate Loss ($bn)**	**Mean per Bank ($bn)**
CN	5	3,633	727
US	8	1,810	226
FR	4	1,153	288
UK	3	812	271
JP	3	781	260
CA	2	376	188
ES	1	269	269
CH	1	230	230
DE	1	215	215
NL	1	74	74

Table 7 summarises the cross-country distribution of systemic risk under the stress scenario by aggregating simulated system-wide losses at the jurisdiction level. The results reveal a pronounced concentration of risk in a small number of major banking systems. China exhibits the highest aggregate loss, followed by the United States, reflecting both large balance sheet sizes and strong interconnectedness. European jurisdictions, particularly France and the United Kingdom, also contribute substantially to global systemic losses. In contrast, smaller banking systems display lower aggregate exposure in absolute terms, although some exhibit elevated mean losses per institution, indicating heterogeneity in individual bank systemic importance. Aggregate loss measures the total simulated system-wide loss attributable to banks headquartered in each country under the stress scenario. Mean per bank is computed as the aggregate loss divided by the number of banks in each jurisdiction.

Table 8: Sensitivity of Single-Bank Systemic Losses Across Capital Modes and LGD Regimes

Capital Mode	**LGD Regime**	**Mean Loss ($bn)**	**Max Loss ($bn)**	**Max Defaults**
CET1	Benign	336	571	1
CET1	Base	401	689	1
CET1	Stress	473	819	1
Hybrid	Benign	345	601	1
Hybrid	Base	410	718	1
Hybrid	Stress	482	848	1
Leverage	Benign	354	631	1
Leverage	Base	419	748	1
Leverage	Stress	491	878	1

Table 8 reports the sensitivity of single-bank systemic losses to alternative capital definitions and loss-given-default (LGD) regimes. Across all specifications, system-wide losses increase monotonically with LGD severity, reflecting stronger transmission of shocks through interbank exposures. Leverage-based capitalisation yields the highest systemic losses, followed by the hybrid and CET1-based regimes, indicating that less risk-sensitive capital definitions amplify contagion effects. Despite variation in loss magnitudes, the number of secondary defaults remains limited across all scenarios, suggesting that adjustments primarily affect loss intensity rather than the breadth of default cascades. Mean and maximum losses refer to simulated system-wide losses triggered by the failure of a single institution under each scenario combination. Max defaults denotes the largest observed number of secondary bank defaults; in all cases, contagion remains limited to a single additional default.

Table 9: Named stress scenario outcomes. Simulated system losses range from $9 billion (benign idiosyncratic shock) to $3.31 trillion (Global Top-5 simultaneous failure, the only cascading scenario). GNN surrogate predictions reported for comparison.

Scenario	Shock	LGD	Sim. Loss ($bn)	GNN Loss ($bn)	Defaults
Benign Baseline (State Street)	0.40	Benign	9	19	0
Idiosyncratic Failure — Deutsche Bank	1.00	Stress	215	150	1
Idiosyncratic Failure — JP Morgan Chase	1.00	Stress	485	390	1
European Sovereign-Bank Nexus	0.60	Base	422	380	0
US Systemic Event	0.70	Stress	724	685	0
Top-5 Interconnected Simultaneous Stress	0.80	Stress	952	960	0
China Banking Crisis	0.80	Stress	1,520	1,520	0
Global Top-5 by Size	1.00	Stress	3,310	2,990	6

Table 9 reports the simulated and surrogate-predicted outcomes across the eight named stress scenarios. Simulated losses range from $9 billion under a benign shock to State Street, to $3.31 trillion in the Global Top-5 scenario—the only configuration that triggers a meaningful cascade (six secondary defaults). The China Banking Crisis is instructive: it generates $1.52 trillion in losses, the second-largest figure in the table, yet produces zero secondary defaults, since the 80% equity depletion does not exhaust each bank's loss-absorbing capacity. Cascading failures materialise only when multiple of the largest G-SIBs fail simultaneously and completely, illustrating that aggregate monetary loss and cascade depth are distinct dimensions of systemic risk that do not necessarily move together. The GNN surrogate tracks the simulation closely across the full severity spectrum, with the largest deviation arising in the Global Top-5 scenario ($320 billion, roughly 10% of the simulated total)

Appendix C: Figures

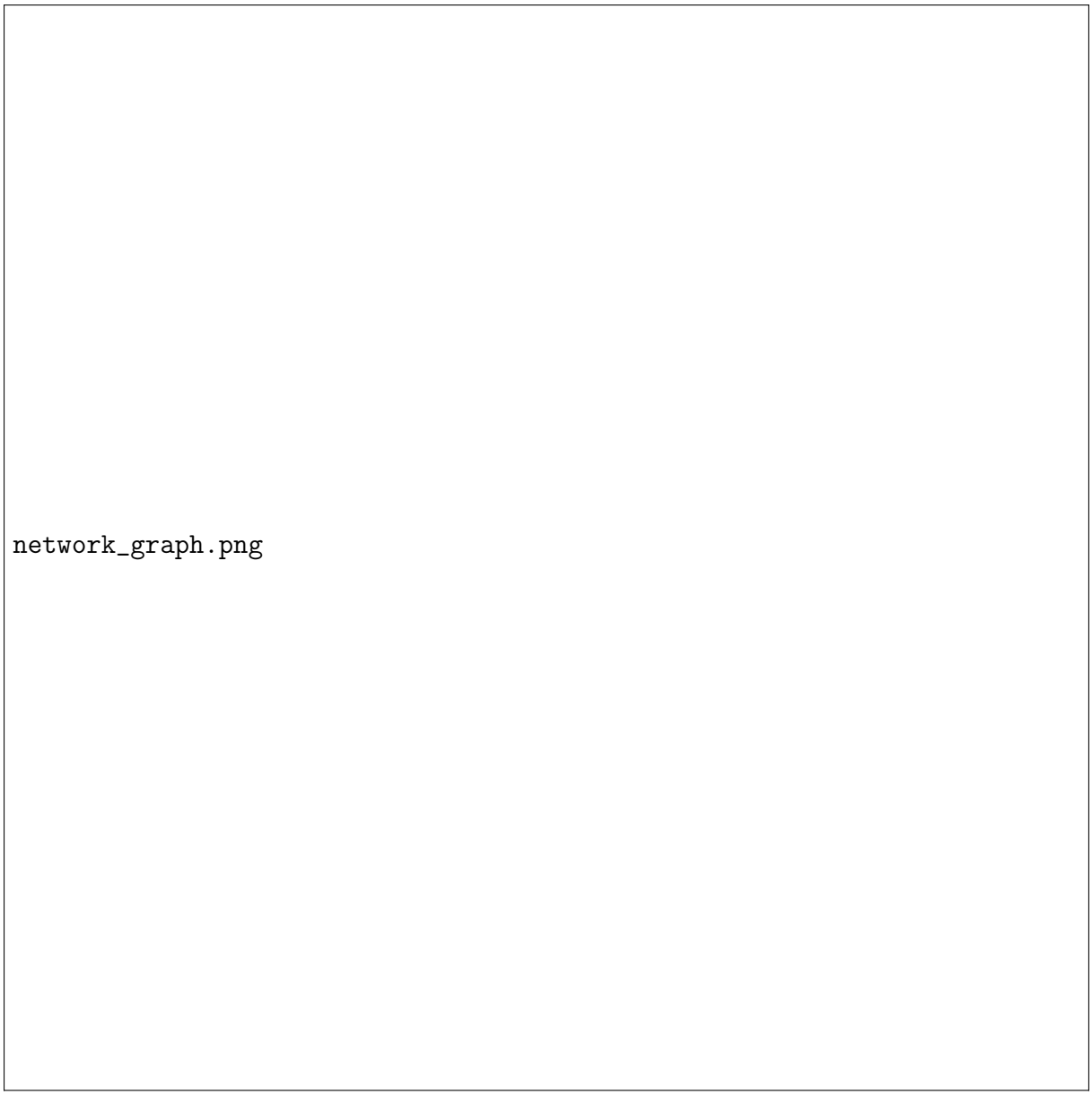

Figure 1: Proxy Interbank Network of the 29 G-SIBs

The figure presents a reconstructed interbank network of 29 global systemically important banks. Node size is proportional to total intra-financial-system assets, while edge thickness reflects estimated bilateral exposures, capturing the intensity of interbank linkages.

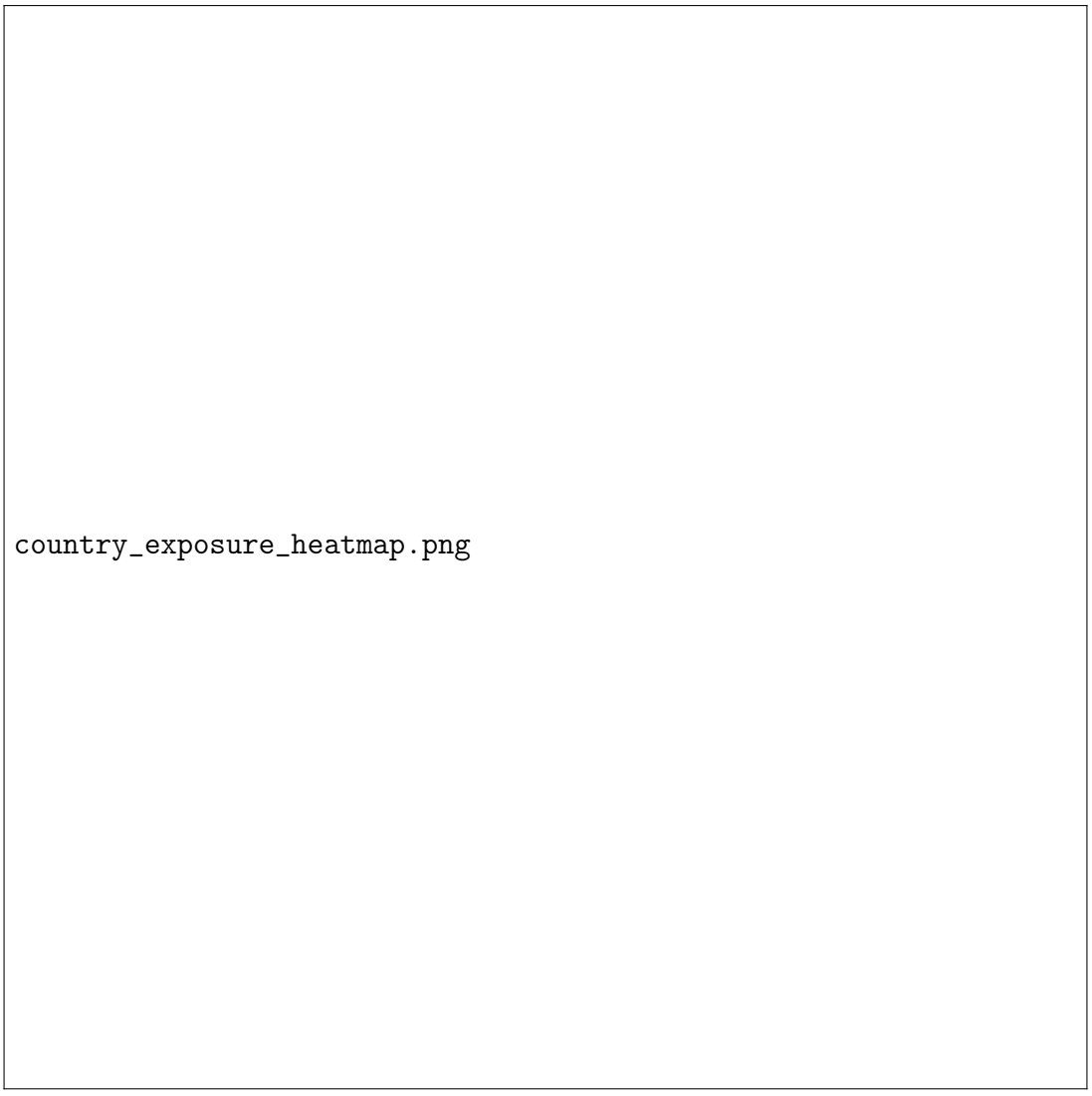

Figure 2: Country-Level Bilateral Exposure Heatmap

The heatmap reports aggregated bilateral exposures between countries, where each cell represents total exposure from the row country to the column country, highlighting cross-border financial interdependence.

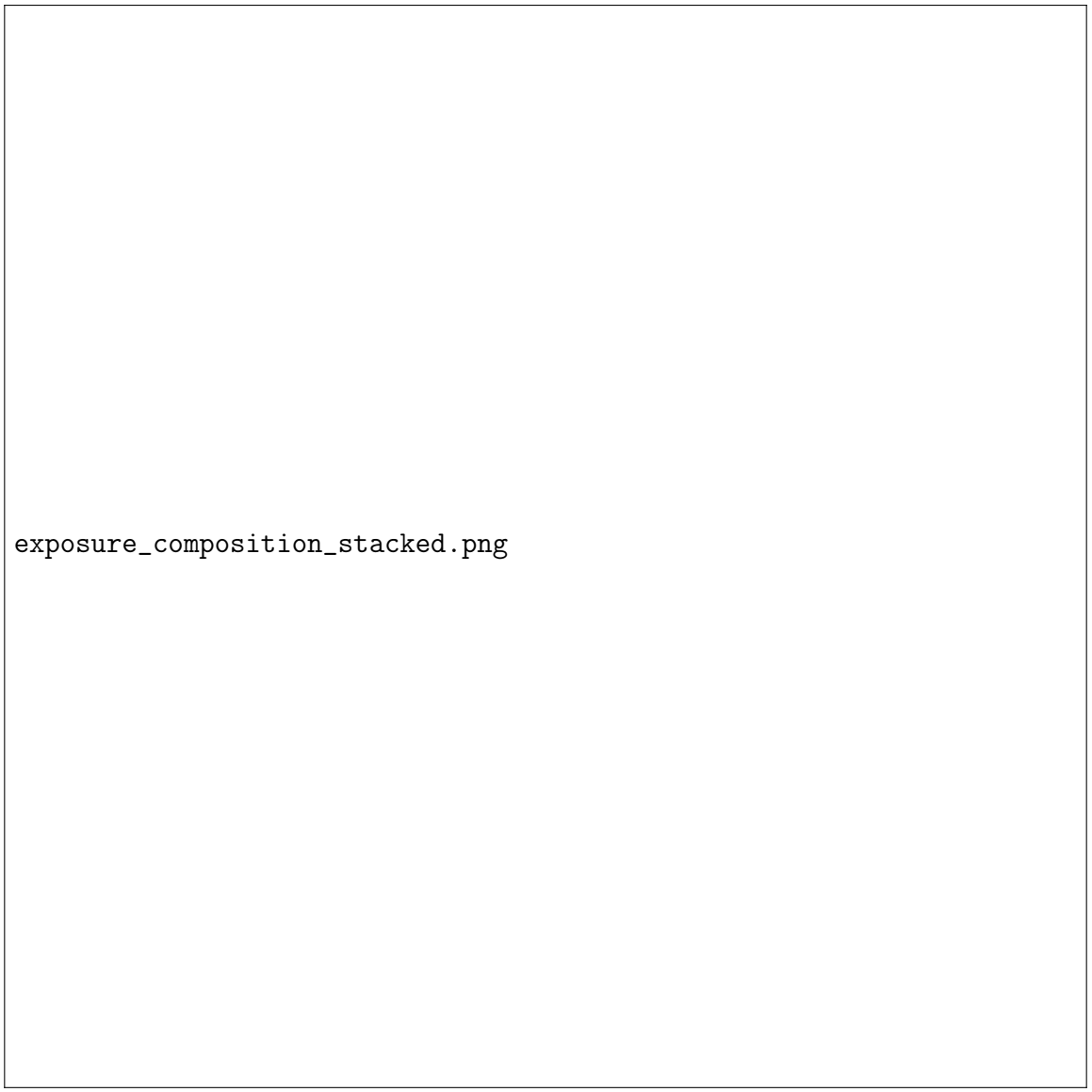

Figure 3: Bilateral Exposure Decomposition by Transmission Channel

The figure decomposes total bilateral exposures into distinct transmission channels, illustrating how financial shocks propagate through funding, credit, and indirect contagion pathways.

Figure 4: Bank-Level Fundamentals: Capitalisation vs Leverage

The scatter plot shows CET1 capital ratios versus leverage ratios for 29 G-SIBs. Bubble size represents total assets, while colour indicates country of incorporation. Data sourced from Refinitiv (LSEG) and EBA Transparency 2024.

Figure 5: GNN Surrogate Validation on the Held-Out Test Set

The figure plots simulated against surrogate-predicted values for the three target variables—system loss, secondary default count, and cascade rounds—on the held-out test set of 218 scenarios. The red dashed line denotes the 45-degree reference. Reported R^2 values quantify the surrogate's predictive accuracy.

Figure 6: Residual Distributions of the GNN Surrogate

The figure displays residual distributions (predicted minus actual) for the three target variables on the test set. Residuals are tightly concentrated around zero, indicating that the surrogate generalises well outside the training distribution.

Figure 7: Systemic Importance Ranking of All G-SIBs

The figure ranks all 29 G-SIBs by simulated system-wide loss under an idiosyncratic full-failure shock and the stress LGD regime with hybrid capital buffers. The four Chinese megabanks, together with JP Morgan Chase, dominate the upper end of the ranking.

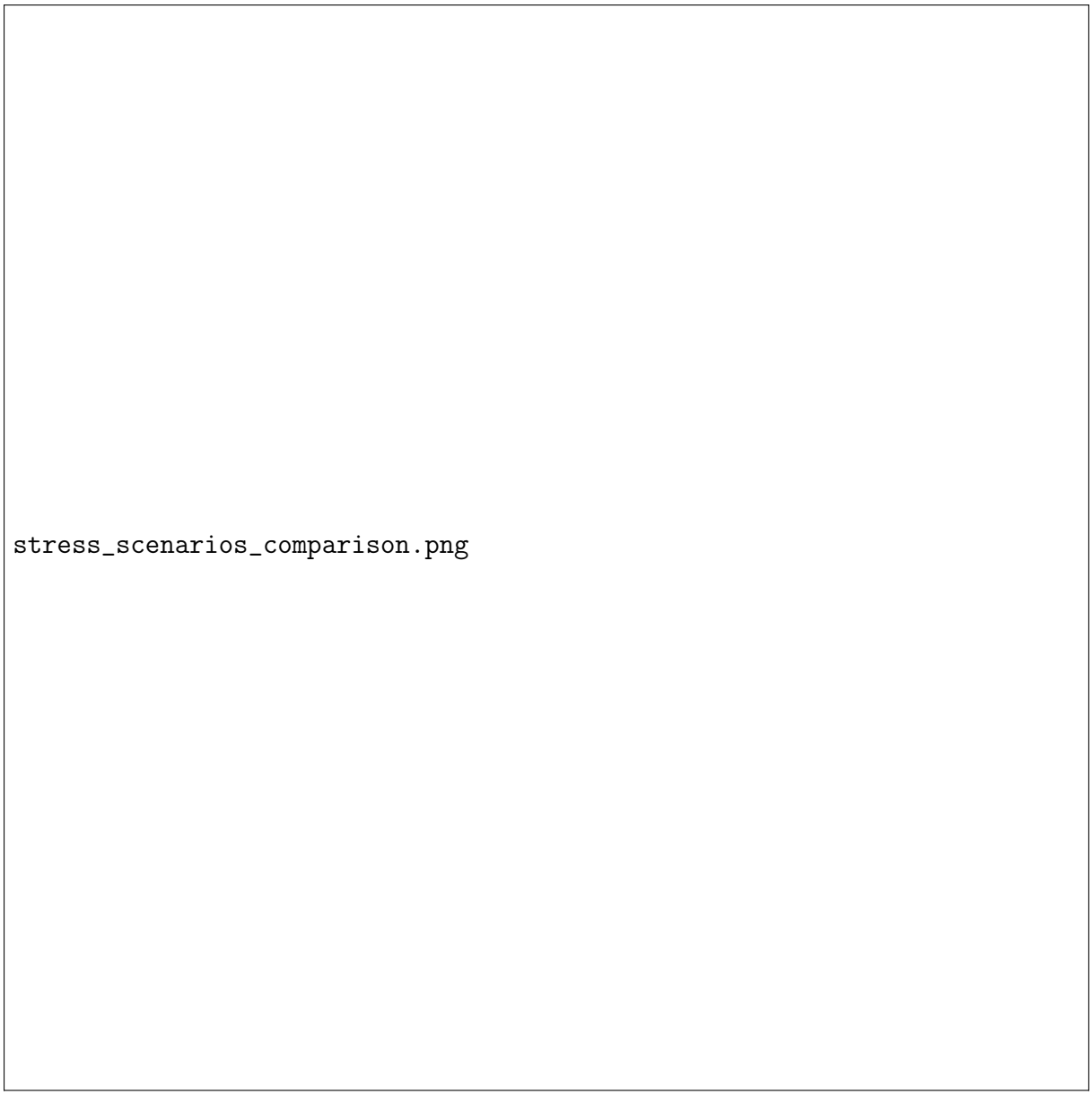

Figure 8: Named Stress Scenarios: Simulation vs GNN Surrogate

The figure compares simulated and surrogate-predicted system-wide losses across the eight named stress scenarios. Agreement is close across the full severity spectrum, with the largest deviation occurring in the Global Top-5 scenario.

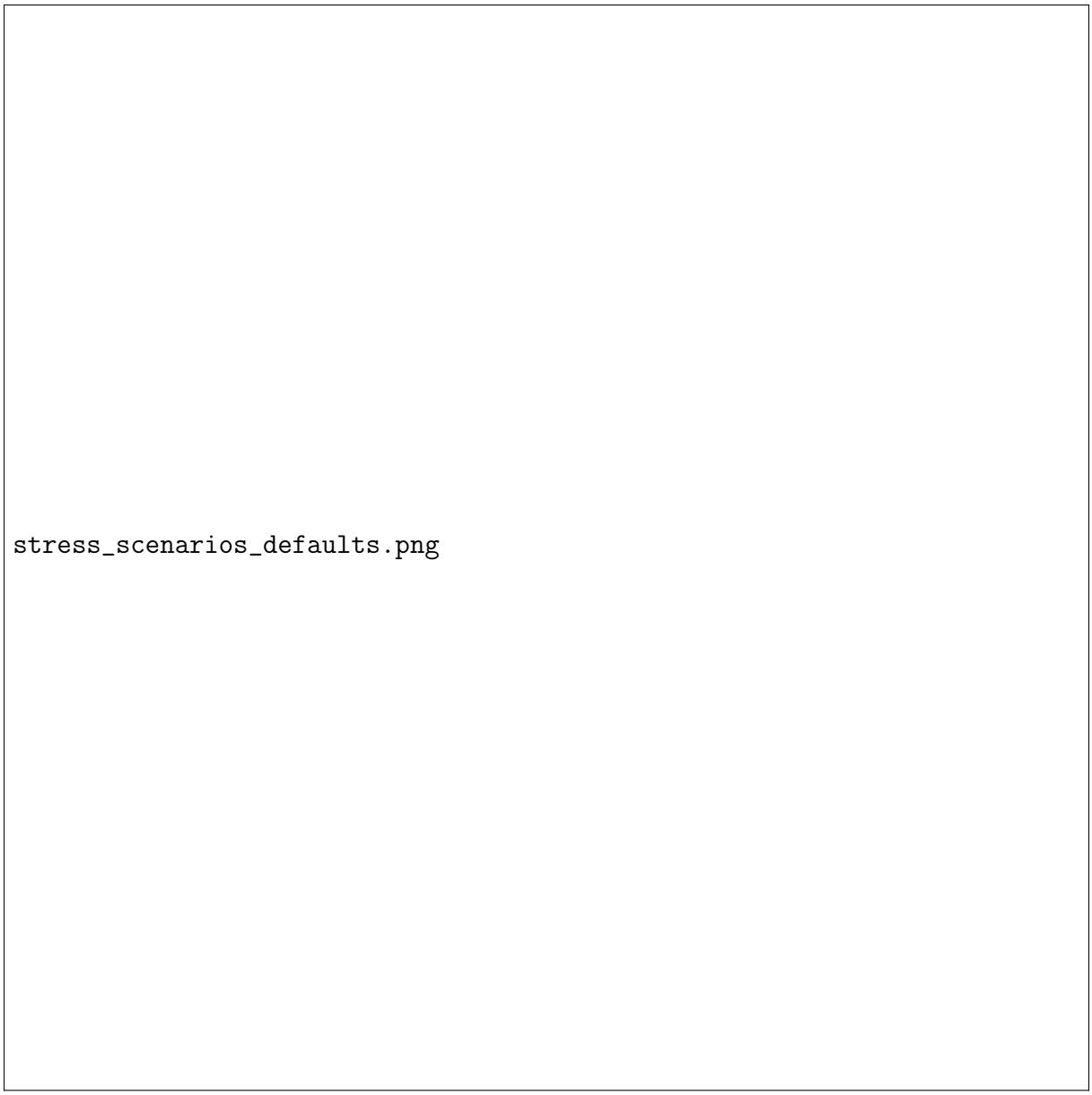

Figure 9: Secondary Defaults Across Named Stress Scenarios

The figure reports the number of secondary defaults triggered by each named stress scenario. Cascading failures materialise only under the Global Top-5 scenario, in which the five largest G-SIBs default simultaneously and completely.

Figure 10: Systemic Importance Ranking Under the Maximum-Entropy Network

The figure presents the systemic importance ranking generated by re-running the contagion engine on the alternative maximum-entropy network reconstruction. The top five positions are preserved relative to the baseline gravity network, with only minor reordering within the leading cluster.

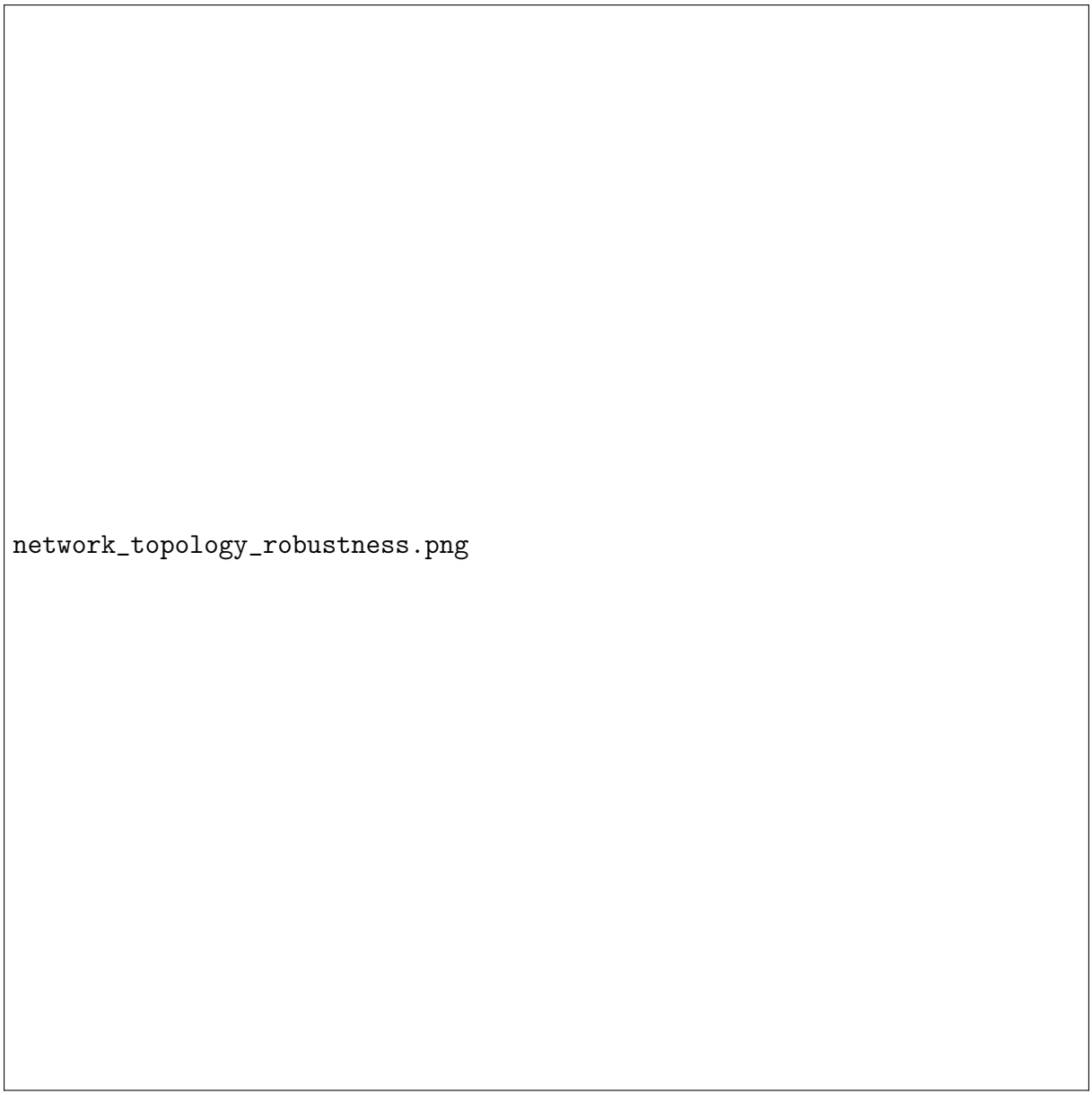

Figure 11: Network Topology Robustness: Gravity vs Maximum-Entropy Rankings

The figure compares the systemic importance ranks produced under the baseline gravity network and the alternative maximum-entropy reconstruction. The Spearman rank correlation of $\rho = 0.877$ ($p < 0.001$) and the preservation of the top five positions across both architectures indicate that the cross-sectional ranking is robust to the choice of network reconstruction methodology.

Figure 12: Simulation Ranking vs Market-Implied ΔCoVaR Ranking

The figure compares the structural simulation-based ranking with the market-implied ΔCoVaR ranking for the 28 publicly traded G-SIBs. The statistically significant negative correlation ($\rho = -0.400$, $p = 0.0352$) is consistent with the theoretical expectation that balance-sheet contagion and market-perceived risk capture complementary dimensions of systemic importance.