

CATNet: A Geometric Deep Learning Approach for CAT Bond Spread Prediction

Dixon Domfeh¹ Saeid Safarveisi²

¹College of Computing, Georgia Institute of Technology, Atlanta, United States

²Actuarial Research Group, AFI, Faculty of Economics and Business, KU Leuven,
Leuven, Belgium

June 10, 2026

Outline

1. Motivation and Background
2. CAT Bond Market as a Relational Graph
3. Network Topology and Centrality Analysis
4. CATNet Architecture
5. Predictive Performance and Feature and Edge Importance

Motivation: CAT Bonds Are Relational Instruments

- ▶ CAT bonds transfer natural catastrophe risk from insurers/reinsurers to capital market investors.
- ▶ Pricing these instruments is challenging because each deal connects several entities:
 - ▶ cedents, underwriters, perils, regions, triggers, and risk modelers.
- ▶ Existing prediction models often treat each bond as an independent tabular observation.
- ▶ But CAT bond contracts are naturally connected through shared market participants, perils, and geographies.

Key question: Are we losing pricing information by ignoring these connections?

Background: From Linear Models to Machine Learning

- ▶ Early CAT bond pricing studies relied mainly on regression-based models.
- ▶ More recent studies show that machine learning methods such as Random Forest and XGBoost improve spread prediction by capturing nonlinearities and interactions.

Why Traditional Tabular Models Struggle

- ▶ CAT bond data contain many high-cardinality categorical variables.
- ▶ Standard approaches rely on:
 - ▶ aggregation,
 - ▶ or grouping categories together.
- ▶ These approaches introduce several limitations:
 - ▶ curse of dimensionality,
 - ▶ loss of granular information,
 - ▶ inability to represent relationships directly,
 - ▶ limited handling of dependencies across contracts.
- ▶ Example:
 - ▶ Two contracts covering the same peril and region may be strongly related,
 - ▶ but tabular models treat them as independent observations.
- ▶ Our approach:
 - ▶ represent the market directly as a graph structure.

Our approach: CATNet

- ▶ We introduce **CATNet**, a geometric deep learning framework for CAT bond spread prediction.
- ▶ The CAT bond primary market is represented as a **heterogeneous multi-relational graph**.
- ▶ Nodes represent entities such as:
 - ▶ contracts, cedents, underwriters, perils, countries, states/provinces, and risk modelers.
- ▶ Edges represent the economic and contractual relationships between them.
- ▶ We use a **Relational Graph Convolutional Network (R-GCN)** to learn from this structure.

Thesis: Network connectivity is a key determinant of CAT bond spreads.

Dataset Overview

- ▶ Dataset contains **803 CAT bond contracts** issued between 1999 and 2021.
- ▶ Data collected and merged from:
 - ▶ Lane Financial Reports
 - ▶ Artemis Deal Directory

Contract Features

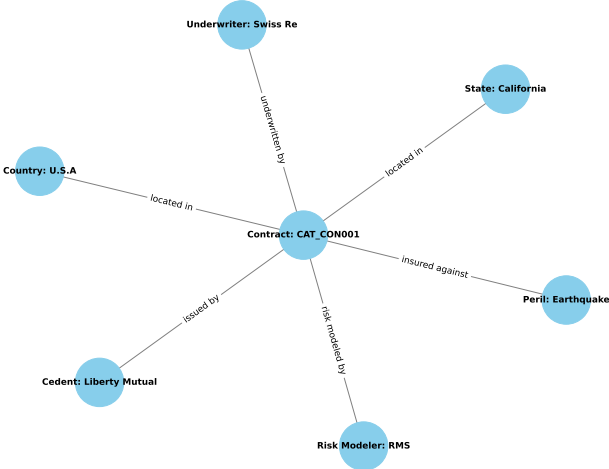
- ▶ Risk premium
- ▶ Expected loss
- ▶ Probability of loss
- ▶ Issue size and issuance year
- ▶ Trigger type
- ▶ Geographic coverage
- ▶ Risk modelers

High-cardinality categorical features

- ▶ 21 perils
- ▶ 32 countries
- ▶ 73 states/provinces
- ▶ 129 cedents
- ▶ 32 underwriters

Graph Representation of the CAT Bond Market

We construct a **heterogeneous multi-relational graph** $(\mathcal{V}(G), \mathcal{E}(G))$:



Graph Representation of the CAT Bond Market

Nodes represent entities, while edges encode relationships between them through the adjacency matrix:

$$(\mathbf{A}_r)_{uv} = \begin{cases} 1, & \text{if } (u, r, v) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases}$$

The resulting graph captures:

- ▶ structural dependencies,
- ▶ market connectivity,
- ▶ and interaction patterns between entities.

Key idea: Learn spreads from both contract attributes and network topology.

The CAT Bond Market as a Scale-Free Network

- ▶ Network analysis reveals that the CAT bond market exhibits a **scale-free structure**.
- ▶ A small number of nodes act as highly connected hubs:
 - ▶ major issuers,
 - ▶ dominant underwriters,
 - ▶ frequently modeled perils.
- ▶ Most nodes have relatively few connections.
- ▶ Implications:
 - ▶ the market is highly concentrated,
 - ▶ information and influence are unevenly distributed,
 - ▶ some entities have disproportionate impact on market structure.
- ▶ This structure may also indicate potential systemic vulnerabilities.

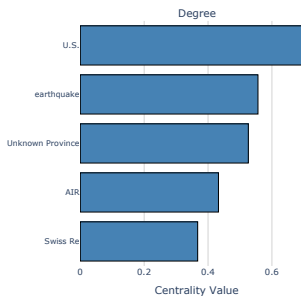
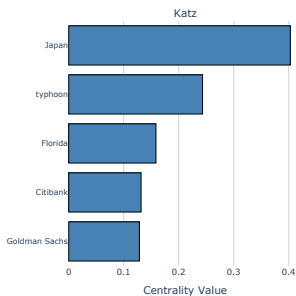
Topological Diagnostics and Centrality Measures

We use graph centrality measures to quantify the structural importance of entities within the CAT bond market network.

Measure	Standard Definition	CAT Bond Interpretation
Degree	Number of direct connections of a node	Market participation and direct connectivity
Betweenness	Frequency with which a node lies on shortest paths	Intermediary influence between market participants
Closeness	How close a given node is to most other nodes	A proxy for reputation and experience
Eigenvector	Influence weighted by neighbors' importance	Systemic importance through influential connections
Katz	Influence from both direct and indirect connections	Global market influence across the network

Topological Diagnostics and Centrality Measures

- ▶ Major issuers, underwriters, and peril regions emerge as dominant hubs.
- ▶ Unlike local centrality measures, Katz centrality also accounts for indirect connections, explaining why globally connected entities such as Japan exhibit high influence despite not always having the largest number of direct links.



CATNet Architecture: Message Passing

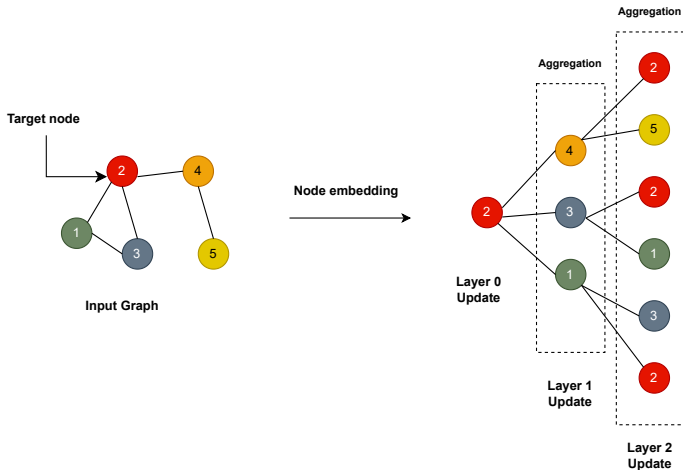
Graph Neural Networks learn node representations by iteratively aggregating information from neighboring nodes.

$$\mathbf{h}_u^{(k+1)} = \text{UPDATE}^{(k)} \left(\mathbf{h}_u^{(k)}, \text{AGGREGATE}^{(k)} \left(\{ \mathbf{h}_v^{(k)} : v \in \mathcal{N}(u) \} \right) \right)$$

- ▶ $\mathbf{h}_u^{(k)}$: embedding of node u at layer k
- ▶ $\mathcal{N}(u)$: neighborhood of node u
- ▶ **AGGREGATE**: collects messages from neighboring nodes
- ▶ **UPDATE**: updates the node embedding using aggregated information

Key idea: each node learns from its local network structure.

CATNet Architecture: Message Passing



CATNet Architecture: Relational GCN

R-GCN extends message passing to heterogeneous graphs with multiple relation types.

For relation type r , messages are aggregated as:

$$\mathbf{m}_{\mathcal{N}_r(u)}^{(k)} = \sum_{v \in \mathcal{N}_r(u)} \frac{1}{c_{uvr}} \mathbf{W}_r^{(k)} \mathbf{h}_v^{(k)}$$

Node embeddings are then updated by:

$$\mathbf{h}_u^{(k+1)} = \sigma \left(\sum_{r \in \mathcal{R}} \mathbf{m}_{\mathcal{N}_r(u)}^{(k)} + \mathbf{W}_0^{(k)} \mathbf{h}_u^{(k)} \right)$$

- ▶ $\mathbf{W}_r^{(k)}$: relation-specific transformation matrix
- ▶ c_{uvr} : normalization constant
- ▶ $\sigma(\cdot)$: nonlinear activation function (ReLU)
- ▶ Final embedding ($\mathbf{z}_u = \mathbf{h}_u^{(K)}$) predicts CAT bond spread:

$$\hat{y}_u = \mathbf{w}^\top \mathbf{z}_u + b$$

Benchmark Models and Experimental Setup

- ▶ CATNet is evaluated against strong tabular machine learning benchmarks:
 - ▶ Random Forest
 - ▶ XGBoost
- ▶ Benchmarks were selected because they are among the best-performing models in previous CAT bond pricing literature.
- ▶ Evaluation focuses on:
 - ▶ predictive accuracy,
 - ▶ robustness,
 - ▶ and interpretability.
- ▶ Main research question:
 - ▶ Does explicitly modeling relational structure improve spread prediction?

Predictive Performance

Table 1: OOS model performance across folds

Fold	Random Forest		XGBoost		R-GCN	
	MSE	R ² (%)	MSE	R ² (%)	MSE	R ² (%)
1	0.4345	61.21	0.3588	67.97	0.2754	75.42
2	0.1680	79.74	0.1573	81.02	0.1088	86.88
3	0.3051	73.92	0.2506	78.58	0.2849	75.64
4	0.2218	73.69	0.2024	75.99	0.2472	70.67
5	0.2199	64.72	0.1422	77.19	0.1231	80.26
6	0.3610	74.09	0.2737	80.36	0.2419	82.64
7	0.2262	80.72	0.2605	77.81	0.1958	83.31
8	0.3275	69.91	0.3013	72.32	0.2743	74.80
9	0.1842	75.79	0.1101	85.52	0.1671	78.03
10	0.2549	77.07	0.2275	79.54	0.1907	82.85
Average	0.2703	73.09	0.2284	77.63	0.2109	79.05

Predictive Performance

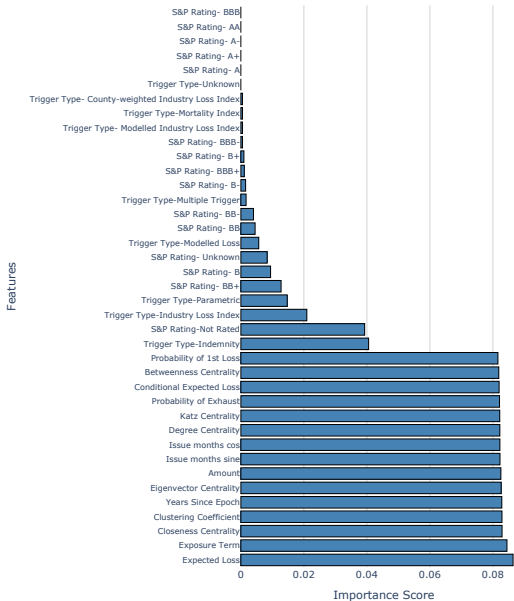
Key findings:

- ▶ CATNet achieves the strongest average out-of-time predictive performance across walk-forward splits.
- ▶ The performance gain is particularly pronounced during periods of changing market structure and elevated market uncertainty.
- ▶ Results suggest that relational dependencies and network topology contain economically meaningful information not fully captured by tabular models.

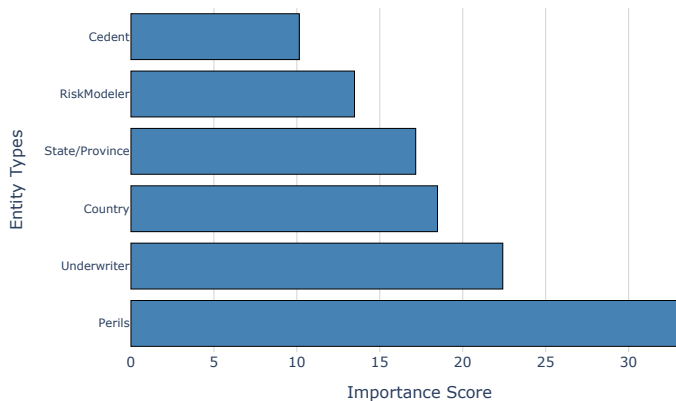
Table 2: OOT model performance across walk-forward splits

Train / Val Split	Test Year	Random Forest		XGBoost		R-GCN	
		MSE	R ² (%)	MSE	R ² (%)	MSE	R ² (%)
1999–2014 — Val 2015	2016	0.1722	74.92	0.1205	82.45	0.0820	88.05
1999–2015 — Val 2016	2017	0.1578	73.39	0.1149	80.63	0.0668	88.73
1999–2016 — Val 2017	2018	0.2335	62.01	0.2860	53.48	0.1227	80.04
1999–2017 — Val 2018	2019	0.2295	69.13	0.2010	72.97	0.2741	63.14
1999–2018 — Val 2019	2020	0.3252	45.01	0.3237	45.27	0.2853	51.77
1999–2019 — Val 2020	2021	0.1851	73.23	0.1489	78.46	0.0758	89.03
Average	–	0.2172	66.28	0.1992	68.88	0.1511	76.79

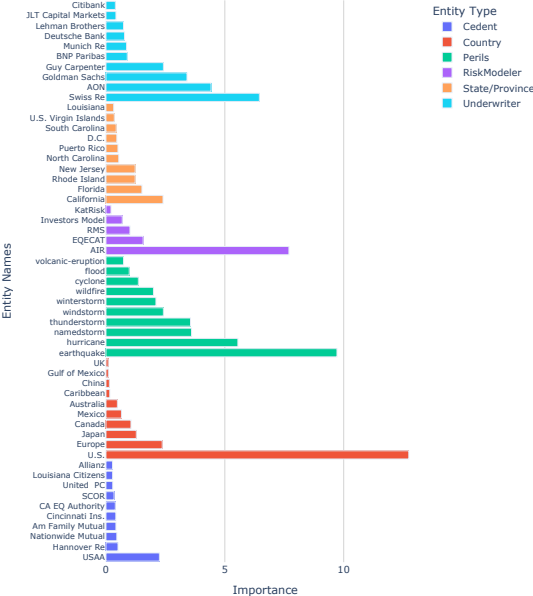
Ranked node feature importance



Ranking of group level edge importance by entity type



Ranking the edge importance of the top 10 entities by entity type



Some References



Braun, A. (2019).

Pricing in the CAT bond market: new empirical evidence.

Journal of Risk and Insurance.



Götze, A., Gürtler, M., and Winkler, C. (2020).

Artificial intelligence in CAT bond pricing.

European Journal of Operational Research.



Bronstein, M., Bruna, J., LeCun, Y., Szlam, A., and Vandergheynst, P. (2017).

Geometric Deep Learning: Going beyond Euclidean data.

IEEE Signal Processing Magazine.



Schlichtkrull, M. et al. (2018).

Modeling Relational Data with Graph Convolutional Networks.

ESWC.

Thank You!

Questions?