Evaluation and Explanation of Heterogeneous Graph Neural Networks

Jona Becher

Referent: Prof. Dr. Markus Döhring | Korreferentin: Prof. Dr. Antje Jahn
Darmstadt University of Applied Sciences

h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Introduction

Insurance fraud causes significant financial damage, with approximately 10% of all claims being fraudulent and annual damages reaching €5 billion in Germany alone [1]. While current fraud detection methods employ machine learning techniques, they often analyze claims in isolation, missing crucial patterns across related entities. Therefore this research investigates:

1. How heterogeneous Graph Neural Networks (GNNs) improve fraud detection by modeling relationships between claims, customers, and vehicles?
2. How GNN explainability and visualization approaches can enhance the interpretability of fraudulent networks?

The proposed approach combines network-based fraud detection with interpretable machine learning, enabling fraud detection experts to understand and investigate suspicious patterns more effectively. This study evaluates both the predictive performance of heterogeneous GNNs against traditional methods and applies explainability techniques for practical application.

Graph Construction

Data in most companies is stored in relational databases and accessed using SQL queries. To leverage modern graph-based machine learning methods, this relational data needs to be transformed into a graph structure. The transformation follows a systematic process: each table row becomes a node, with node types corresponding to their source tables. Primary and foreign key relationships define the initial edge structure, while additional edges can be created based on shared attribute values between entities. Each node retains its original table attributes as feature vectors, preserving the rich information contained in the database [2].

In this insurance fraud detection application, this process transforms claims, customer, and vehicle data into a heterogeneous graph with 12 million nodes and 26 million edges. The nodes represent individual claims, customers, and vehicles, while edges capture both direct relationships (e.g., customer files claim) and implicit connections (e.g., shared addresses between customers). This representation allows for the detection of suspicious patterns across the entire network of insurance relationships rather than analyzing claims in isolation.

Graph Neural Networks

Neural message passing forms the foundation of GNNs, where nodes iteratively exchange and aggregate information with their neighbors to learn representations that capture both node features and graph structure. In each iteration, a node collects feature vectors from its neighbors, combines them through an aggregation function, and updates its own representation, enabling the network to capture increasingly larger neighborhood contexts [3]. Four key architectures are employed in this work:

- Graph Convolutional Networks (GCN) utilize a degree-normalized aggregation that combines neighborhood features through weighted summation [3].
- Graph Attention Networks (GAT) introduce learnable attention mechanisms to dynamically weight the importance of neighboring nodes, enabling adaptive information aggregation [3].
- GraphSAGE employs neighborhood sampling and concatination of node and neighborhood features to efficiently handle large graphs while preserving node-specific information [3].
- Heterogeneous Graph Attention Network (HAN) employs dual-level attention, combining node-level attention for neighbor importance with semantic-level attention for meta-path weighting, to capture both local node relationships and global semantic structures in heterogeneous graphs [4].

Model Evaluation

Heterogeneous GNNs were compared against Neural Networks and LightGBM as baselines, as well as homogeneous GNNs. Both baseline models process claims in isolation, while homogeneous GNNs operate on a graph containing only claim nodes and their relationships.

Category	Model	F1-Score
Baseline	LightGBM	0.593
	Neural Network	0.601
Homogeneous GNNs	GraphSage	0.624
	GAT	0.612
Heterogeneous GNNs	GCN	0.599
	GAT	**0.681**
	GraphSage	0.677
	HAN	0.643

Table 1. Performance comparison of heterogeneous GNNs, homogeneous GNNs, and traditional approaches.

Experimental results, as detailed in table 1, demonstrate the superiority of heterogeneous graph architectures for fraud detection with the GAT variant achieving an F1-score of 0.681. While homogeneous GNNs show moderate improvements over traditional methods, incorporating multiple node types and their relationships substantially enhanced model performance. The GAT architecture's attention mechanism effectively leverages the rich structural information between claims, customers, and vehicles, enabling the identification of complex fraud patterns.

Explainability: Message Passing GNNExplainer

What is GNNExplainer? GNNExplainer provides interpretability for GNNs by identifying important subgraphs and node features that influence a model's prediction. Perturbation determines relevance by masking edges or node features and measuring the impact on the prediction. To efficiently find these relevant elements, the GNNExplainer optimizes masks via gradient descent instead of testing all possible masking combinations [5]. However, the original approach suffers from result variability due to local minima convergence and high-dimensional parameter optimization [6].

Approach: The proposed message passing GNNExplainer determines edge importance in a stepwise manner, starting with edges directly connected to the target node. During each iteration, edges with importance values below a certain threshold are removed from the prediction-relevant subgraph. Edges and nodes that lose their connection to the target node through this process are excluded from the explanation without requiring explicit optimization. By utilizing message passing, this methodology accurately identifies important connections. When a node contains relevant features for predicting the target node, the connecting edges gain high importance as they facilitate the transmission of crucial information. Through this process of identifying relevant edges, the approach minimizes the number of parameters requiring optimization in the explanation.

Result: Compared to the original GNNExplainer, this method reduces the mean variance of one-hop edge importance by 35.9%, demonstrating significantly improved consistency in explanations.

Case Study

The developed approach combines GNNs with explainability methods to detect fraud patterns in insurance claims. Starting from a GNN-flagged suspicious claim, a Network Extraction Algorithm iteratively identifies connected entities and evaluates their connections and node features using importance scores from the Message Passing GNN-Explainer.

By retaining only the most relevant connections in each iteration, the algorithm extracts a focused subnetwork that highlights the key entities and relationships involved in potential fraud cases. For each identified network, a Network Report summarizes key metrics including the number of connected claims, customers, vehicles, and most important attributes. This report incorporates GNN-Explainer findings and direct links to a visualization. An interactive Network Visualization tool presents these results using color-coding and edge weighting to distinguish entity types and relationship importance.

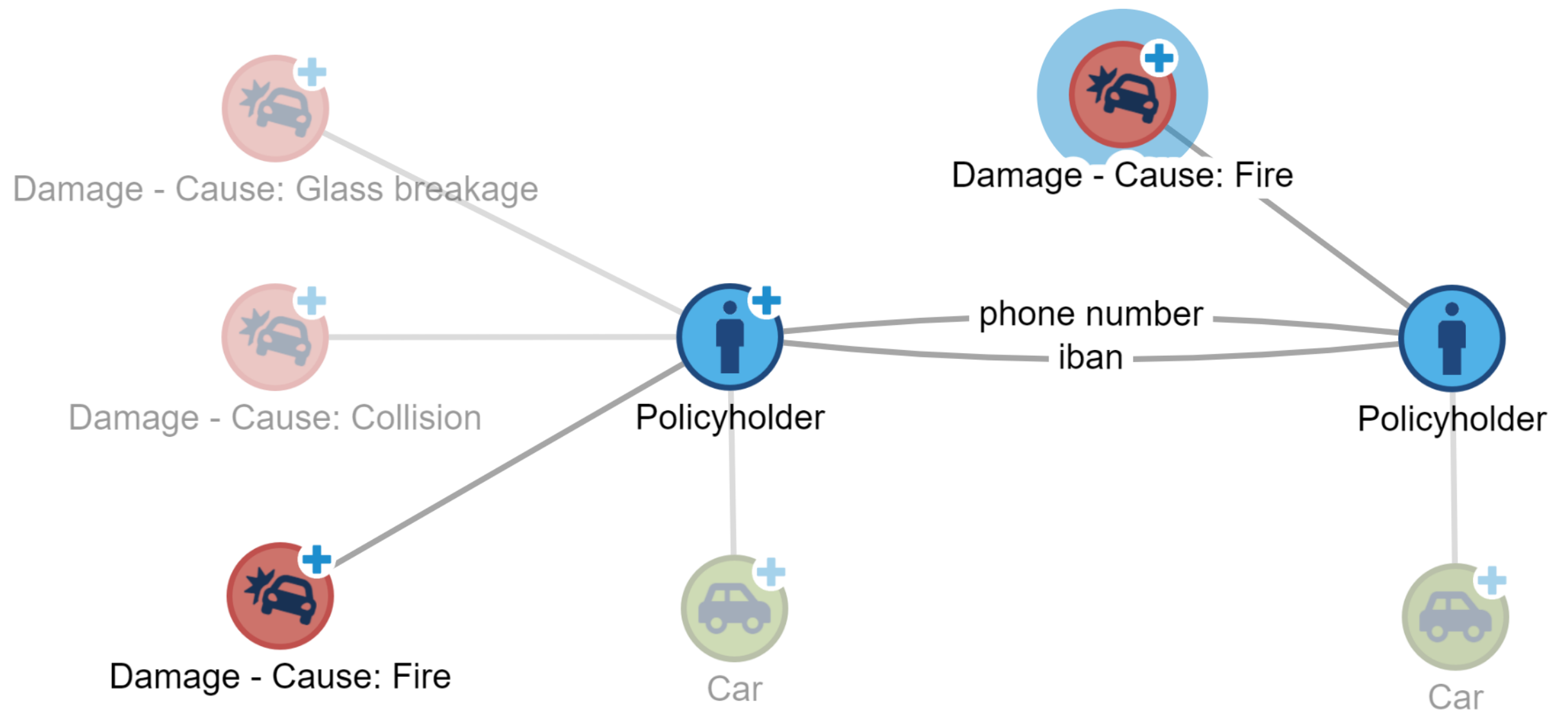

Figure 1. Network visualization highlighting suspicious claim patterns.

Figure 1 shows a network connecting insurance claims between two policyholders who share phone numbers and IBAN details, indicating a clear connection between these individuals. The GNNExplainer identified the cause of loss as the most important feature and highlighted the two fire damage claims as particularly relevant for the fraud investigation, while less important parts of the network are blurred out to reduce visual complexity. What makes this case particularly suspicious is that fire damage claims are generally considered high-risk for fraud, and here they appear between individuals who are linked through multiple shared personal identifiers. The visualization demonstrates how analyzing claims in their network context can reveal suspicious patterns that would remain hidden when examining claims in isolation, while helping fraud experts efficiently focus their investigation on the most relevant connections.

Conclusion

This research demonstrates that heterogeneous GNNs significantly enhance fraud detection in insurance claims by effectively modeling relationships between claims, customers, and vehicles. The heterogeneous GNN models outperformed traditional baselines and homogeneous GNNs, achieving higher F1-scores due to their ability to capture complex interactions within the data.

Additionally, integrating explainability techniques and network visualizations provide deeper insights into fraudulent networks, enabling focused analysis by only showing the most relevant nodes and edges. This approach provides both effective fraud detection and enhanced interpretability of the GNN predictions.

References

[1] German Insurance Association, "Insurance fraud: One in ten claims reports is suspicious," 2024. [Online]. Available: https://www.gdv.de/gdv/themen/schaden-unfall/versicherungsbetrug-jede-zehnte-schadenmeldung-ist-verdaechtig-171870 (accessed Nov. 10, 2024).

[2] M. Fey, W. Hu, K. Huang, J. E. Lenssen, R. Ranjan, J. Robinson, R. Ying, J. You, and J. Leskovec, "Position: Relational deep learning - graph representation learning on relational databases," in *Proceedings of the 41st International Conference on Machine Learning*, vol. 235 of *Proceedings of Machine Learning Research*, pp. 13592–13607, PMLR, 21–27 Jul 2024.

[3] W. L. Hamilton, *Graph Representation Learning*. Springer International Publishing, 2020.

[4] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous graph attention network," in *The World Wide Web Conference*, WWW '19, (New York, NY, USA), p. 2022–2032, Association for Computing Machinery, 2019.

[5] Z. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "Gnnexplainer: Generating explanations for graph neural networks," in *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., 2019.

[6] J. Li, M. Pang, Y. Dong, J. Jia, and B. Wang, "Graph neural network explanations are fragile," in *Proceedings of the 41st International Conference on Machine Learning*, vol. 235 of *Proceedings of Machine Learning Research*, pp. 28551–28567, PMLR, 21–27 Jul 2024.