Novelty Assessment Report

Paper: LogicXGNN: Grounded Logical Rules for Explaining Graph Neural Networks

PDF URL: https://openreview.net/pdf?id=nqZe43tRY9

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-01

Abstract

Existing rule-based explanations for Graph Neural Networks (GNNs) provide global interpretability but often optimize and assess fidelity in an intermediate, uninterpretable concept space, overlooking the grounding quality of the final subgraph explanations for end users. This gap yields explanations that may appear faithful yet be unreliable in practice. To this end, we propose LogicXGNN, a post hoc framework that constructs logical rules over reliable predicates explicitly designed to capture the GNN's message-passing structure, thereby ensuring effective grounding. We further introduce data-grounded fidelity (Fid_D), a realistic metric that evaluates explanations in their final-graph form, along with complementary utility metrics such as coverage and validity. Across extensive experiments, LogicXGNN improves Fid_D by over 20% on average relative to state-of-the-art methods while being 10-100 times faster. With strong scalability and utility performance, LogicXGNN produces explanations that are faithful to the model's logic and reliably grounded in observable data.

Core Task Landscape

This paper addresses: **Explaining Graph Neural Networks with Logical Rules**

A total of **50 papers** were analyzed and organized into a taxonomy with **21 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Rule Extraction and Logic-Based Explanation Generation**
- **Concept-Based and Neuron-Level Interpretability**
- **Neural-Symbolic Integration and Rule-Guided Learning**
- **Expressivity and Theoretical Foundations**
- **Self-Explainable and Transparent GNN Architectures**
- **Explanation Evaluation and Validation Frameworks**
- **Domain-Specific Applications and Hybrid Approaches**

Complete Taxonomy Tree

- Explaining Graph Neural Networks with Logical Rules Survey Taxonomy
- Rule Extraction and Logic-Based Explanation Generation
  - Global Rule Extraction from GNN Behavior ★ (4 papers)
  - [0] LogicXGNN: Grounded Logical Rules for Explaining Graph Neural Networks (Anon et al., 2026) View paper
  - [10] Global explainability of gnns via logic combination of learned concepts (Azzolin, 2022) View paper
  - [11] Extracting Interpretable Logic Rules from Graph Neural Networks (Geng, 2025) View paper
  - [45] GraphTrail: Translating GNN Predictions into Human-Interpretable Logical Rules (Burouj Armgaan, 2024) View paper
  - Logic Formula and Symbolic Representation Extraction (3 papers)
  - [18] From latent to lucid: Transforming knowledge graph embeddings into interpretable structures (Wehner, 2024) View paper
  - [25] Explainable GNN-based models over knowledge graphs (DJT Cucala, 2022) View paper
  - [35] Contrastive Graph Representations for Logical Formulas Embedding (Qika Lin, 2023) View paper
  - Path-Based and Subgraph Rule Explanation (3 papers)
  - [29] eXpath: Explaining Knowledge Graph Link Prediction with Ontological Closed Path Rules (Ye Sun, 2024) View paper
  - [30] Explaining GNN-based Recommendations in Logic (Wenfei Fan, 2024) View paper
  - [37] A Framework for Extracting Rules from Explanation Subgraphs for Gnn-Based Knowledge Graph Link Prediction (Jun-ping Yao, 2024) View paper
  - Association Rule Mining and Pattern Discovery (2 papers)
  - [26] A New Concept for Explaining Graph Neural Networks. (Anna Himmelhuber, 2021) View paper
  - [31] ARM-Explainer--Explaining and improving graph neural network predictions for the maximum clique problem using node features and association rule mining (Bharat Sharman, 2025) View paper
- Concept-Based and Neuron-Level Interpretability
  - Global Concept Analysis and Neuron Behavior (2 papers)
  - [3] Global concept-based interpretability for graph neural networks via neuron analysis (Han Xuanyuan, 2023) View paper
  - [6] Interpretable Hierarchical Concept Reasoning through Attention-Guided Graph Learning (Barbiero, 2025) View paper
  - Hierarchical and Relational Concept Reasoning (2 papers)
  - [20] Generating explanations for conceptual validation of graph neural networks: An investigation of symbolic predicates learned on relevance-ranked sub-graphs (Bettina Finzel, 2022) View paper
  - [38] GnnXemplar: Exemplars to Explanations - Natural Language Rules for Global GNN Interpretability (Armgaan, 2025) View paper

- Neural-Symbolic Integration and Rule-Guided Learning
  - Rule-Guided GNN Training and Architecture Design (4 papers)
  - [2] Rule-Guided Graph Neural Networks for Explainable Knowledge Graph Reasoning (Wang Zhe, 2025) View paper
  - [8] Rule Based Learning with Dynamic (Graph) Neural Networks (Seiffarth, 2024) View paper
  - [9] Graph neural networks in neural-symbolic computing (Bikram Pratim Bhuyan, 2024) View paper
  - [42] FireGNN: Neuro-Symbolic Graph Neural Networks with Trainable Fuzzy Rules for Interpretable Medical Image Classification (Sengupta, 2025) View paper
  - Neural-Symbolic Query Execution and Logical Reasoning (3 papers)
  - [17] Neural Probabilistic Logic Learning for Knowledge Graph Reasoning (Wang Jin-yu, 2024) View paper
  - [24] Neural-symbolic models for logical queries on knowledge graphs (Zhu ZhaoCheng, 2022) View paper
  - [34] Learning reasoning strategies in end-to-end differentiable proving (Pasquale Minervini, 2020) View paper
  - Differentiable Logic and Temporal Reasoning (2 papers)
  - [7] Back to the future: Towards explainable temporal reasoning with large language models (Chenhan Yuan, 2024) View paper
  - [39] Learning Signal Temporal Logic through Neural Network for Interpretable Classification (Danyang Li, 2022) View paper
- Expressivity and Theoretical Foundations
  - Logical Characterization of GNN Architectures (3 papers)
  - [1] Graph neural networks and logic (Schwarzentruber, 2025) View paper
  - [12] Sound Logical Explanations for Mean Aggregation Graph Neural Networks (Matthew Morris, 2025) View paper
  - [44] Logical Characterizations of GNNs with Mean Aggregation (SchÃ¶nherr, 2025) View paper
  - Expressivity for Knowledge Graph Reasoning (2 papers)
  - [23] Logical Expressiveness of Graph Neural Networks on Knowledge Graphs (Xingyue Huang, 2025) View paper
  - [40] Understanding expressivity of gnn in rule learning (Qiu Haiquan, 2023) View paper
- Self-Explainable and Transparent GNN Architectures
  - Decision Tree-Based Explainable GNNs (2 papers)
  - [15] CORTEX: A Cost-Sensitive Rule and Tree Extraction Method (SaviÄ, 2025) View paper
  - [49] DT+GNN: A Fully Explainable Graph Neural Network using Decision Trees (MÃ¼ller, 2022) View paper
  - Formal Analysis of Self-Explainable GNN Properties (2 papers)
  - [13] Beyond Topological Self-Explainable GNNs: A Formal Explainability Perspective (Azzolin, 2025) View paper
  - [50] On GNN explainability with activation rules (Luca Veyrin-Forrer, 2024) View paper
- Explanation Evaluation and Validation Frameworks
  - Fidelity and Faithfulness Assessment (2 papers)
  - [27] Global explanation supervision for Graph Neural Networks (Negar Etemadyrad, 2024) View paper
  - [28] Faithfully explainable recommendation via neural logic reasoning (de Melo, 2021) View paper
  - Comparative Evaluation of Explanation Types (2 papers)
  - [21] EDGE: evaluation framework for logical vs. subgraph explanations for node classifiers on knowledge graphs (Rupesh Sapkota, 2024) View paper
  - [48] Assessing Natural Language Explanations of Relational Graph Neural Networks (Stefan Heindorf, 2025) View paper
  - Explanation Supervision and Improvement (2 papers)
  - [4] GNNBoundary: Towards explaining graph neural networks through the lens of decision boundaries (X Wang, 2024) View paper
  - [43] Ensemble interpretation: a unified framework for explanation methods (Chao Min, 2025) View paper
- Domain-Specific Applications and Hybrid Approaches
  - Knowledge Graph Completion and Link Prediction (3 papers)
  - [32] SAFRAN: An interpretable, rule-based link prediction method outperforming embedding models (Simon Ott, 2021) View paper
  - [33] Inductive knowledge graph completion with GNNs and rules: An analysis (Anil, 2024) View paper
  - [41] Generating Graph-Based Rules for Enhancing Logical Reasoning (Kai Sun, 2024) View paper
  - Medical and Healthcare Applications (2 papers)
  - [16] An interpretable model for sepsis prediction using multi-objective rule extraction (Mingzhou, 2024) View paper
  - [22] Explanation of Deep Learning Models via Logic Rules Enhanced by Embeddings Analysis, and Probabilistic Models (Victor Contreras, 2024) View paper
  - Industrial Process and Decision Support Systems (3 papers)
  - [5] Fusing logic rule-based hybrid variable graph neural network approaches to fault diagnosis of industrial processes (Min Yin, 2024) View paper
  - [14] Smart Decision Orchestration for Consumer Electronics Management Using Dynamic Neuro-Symbolic AI Fusion (Zhong Xin, 2025) View paper
  - [47] A Hybrid Transformer and GNN Framework for Interpretable Fair Value Classification in Accounting (Rahman, 2025) View paper
  - Hybrid Explainability Frameworks (2 papers)
  - [36] Fuzzy Rule-Based Explainer Systems for Deep Neural Networks: From Local Explainability to Global Understanding (Fatemeh Aghaeipoor, 2023) View paper
  - [46] GraphMind: Theorem Selection and Conclusion Generation Framework with Dynamic GNN for LLM Reasoning (Yutong Li, 2025) View paper
  - General Interpretability Surveys and Frameworks (1 papers)
  - [19] A comprehensive survey on self-interpretable neural networks (Ji Yang, 2025) View paper

Narrative

Core task: Explaining graph neural networks with logical rules. The field has organized itself around several complementary directions. Rule Extraction and Logic-Based Explanation Generation focuses on distilling symbolic rules from trained GNNs, either at a global level (capturing overall model behavior) or locally (explaining individual predictions). Concept-Based and Neuron-Level Interpretability examines what internal representations learn, often linking hidden activations to human-understandable concepts. Neural-Symbolic Integration and Rule-Guided Learning explores hybrid architectures that combine neural learning with symbolic reasoning, allowing logic to guide training or inference. Expressivity and Theoretical Foundations investigates the formal capabilities of GNNs in capturing logical structures, while Self-Explainable and Transparent GNN Architectures designs models that are interpretable by construction. Explanation Evaluation and Validation Frameworks addresses how to rigorously assess the quality and faithfulness of explanations, and

Domain-Specific Applications and Hybrid Approaches tailors these techniques to specialized settings such as knowledge graphs or temporal reasoning.

Within Rule Extraction, a particularly active line of work seeks to produce global logical summaries of GNN decision boundaries. LogicXGNN[0] exemplifies this direction by extracting interpretable rules that describe the model's overall behavior across the input space. Nearby efforts like Global Logic Explainability[10] and Extracting Logic Rules[11] similarly aim to distill symbolic patterns from trained networks, though they may differ in the granularity or formalism of the extracted rules. In contrast, works such as Global Concept Interpretability[3] shift focus from rule extraction to identifying high-level concepts encoded in neuron activations, offering a complementary lens on what the model has learned. The central tension across these branches is balancing expressiveness—capturing complex, nuanced patterns—with human readability, as overly detailed rules can become as opaque as the original neural network. LogicXGNN[0] sits squarely in the global rule extraction cluster, emphasizing symbolic summaries that remain interpretable while faithfully representing the GNN's learned logic.

Related Works in Same Category

The following **3 sibling papers** share the same taxonomy leaf node with the original paper:

1. Global explainability of gnns via logic combination of learned concepts

Authors: Azzolin, Steve, Longa, Antonio, Steve Azzolin, et al. (14 authors total) | **Year/Venue**: 2022 | **URL**: View paper

Abstract

While instance-level explanation of GNN is a well-studied problem with plenty of approaches being developed, providing a global explanation for the behaviour of a GNN is much less explored, despite its potential in interpretability and debugging. Existing solutions either simply list local explanations for a given class, or generate a synthetic prototypical graph with maximal score for a given class, completely missing any combinatorial aspect that the GNN could have learned. In this work, we pr...

Relationship Analysis

Both papers belong to the Global Rule Extraction from GNN Behavior category, focusing on extracting logical rules that characterize entire GNN model behavior across classes. They overlap in their goal of generating global, logic-based explanations using Boolean combinations of learned concepts to explain GNN predictions. The key difference is that LogicXGNN explicitly models message-passing structure through predicates designed to capture receptive field patterns and introduces data-grounded fidelity evaluation, while GLGExplainer clusters local explanations into graphical concepts and combines them into logic formulas without the explicit structural grounding mechanism.

---

2. Extracting Interpretable Logic Rules from Graph Neural Networks

Authors: Geng, Chuqin, Zhao, Ziyu, Chuqin Geng, et al. (14 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Graph neural networks (GNNs) operate over both input feature spaces and combinatorial graph structures, making it challenging to understand the rationale behind their predictions. As GNNs gain widespread popularity and demonstrate success across various domains, such as drug discovery, studying their interpretability has become a critical task. To address this, many explainability methods have been proposed, with recent efforts shifting from instance-specific explanations to global concept-based...

⚠ Similarity Notice

This paper appears to be highly similar to the original paper; both share the title 'LogicXGNN' and describe a framework for extracting interpretable logic rules from GNNs with nearly identical technical approaches, including predicate identification, rule extraction, and grounding mechanisms. The abstracts and core contributions are essentially the same, suggesting this may be a variant or near-duplicate. Please manually verify.

---

3. GraphTrail: Translating GNN Predictions into Human-Interpretable Logical Rules

Authors: Burouj Armgaan, Manthan Dalmia, Sourav Medya, Sayan Ranu | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Instance-level explanation of graph neural networks (G NN s) is a well-studied area. These explainers, however, only explain an instance (e.g., a graph) and fail to uncover the combinatorial reasoning learned by a G NN from the training data towards making its predictions. In this work, we introduce G RAPH T RAIL , the first end-to-end , post-hoc, global G NN explainer that translates the functioning of a black-box G NN model to a boolean formula over the (sub)graph-level concepts without relyin...

Relationship Analysis

Both papers belong to the Global Rule Extraction from GNN Behavior category, focusing on extracting logical rules that characterize entire GNN model behavior across classes. They overlap in their goal of generating interpretable logical formulas (DNF/boolean formulas) over graph-level concepts to explain GNN predictions globally. The key difference is that LogicXGNN emphasizes grounding predicates in the GNN's message-passing structure with data-grounded fidelity evaluation, while GraphTrail focuses on mining discriminative subgraph concepts using Shapley values and symbolic regression without explicit message-passing alignment.

Contributions Analysis

Overall novelty summary. The paper proposes LogicXGNN, a framework for extracting global logical rules from trained GNN models to explain predictions. It resides in the 'Global Rule Extraction from GNN Behavior' leaf, which contains four papers including the original work. This leaf sits within the broader 'Rule Extraction and Logic-Based Explanation Generation' branch, indicating a moderately populated research direction. The taxonomy shows this is an active area with multiple complementary approaches, though not as crowded as some domain-specific application categories.

The taxonomy reveals several neighboring research directions. Adjacent leaves include 'Logic Formula and Symbolic Representation Extraction' (3 papers) and 'Path-Based and Subgraph Rule Explanation' (3 papers), both focused on symbolic explanation but with different structural emphases. The broader taxonomy also shows parallel branches in 'Concept-Based and Neuron-Level Interpretability' (4 papers) and 'Explanation Evaluation and Validation Frameworks' (6 papers). LogicXGNN bridges rule extraction with evaluation concerns by introducing data-grounded fidelity, connecting to the validation framework branch while remaining rooted in symbolic rule generation.

Among 30 candidates examined across three contributions, no clearly refuting prior work was identified. The data-grounded fidelity metric examined 10 candidates with 0 refutable, suggesting this evaluation approach may be relatively novel within the limited search scope. Similarly, the LogicXGNN framework and reliable predicates design each examined 10 candidates without finding overlapping prior work. However, this analysis reflects top-K semantic search results, not an exhaustive literature review, so the absence of refutation indicates novelty within the examined candidate set rather than absolute originality across all published work.

Based on the limited search scope of 30 semantically similar papers, the work appears to introduce distinct contributions in both methodology and evaluation. The taxonomy position in a moderately populated leaf suggests the paper addresses an established problem with fresh techniques. The lack of refuting candidates across all three contributions, while not definitive proof of novelty, indicates the specific combination of data-grounded evaluation and message-passing-aware predicates may represent a meaningful advance within the examined literature.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Data-grounded fidelity metric for evaluating rule-based GNN explanations

Description: The authors propose a new evaluation metric called data-grounded fidelity (FidD) that assesses rule-based explanations directly on the final subgraph explanations presented to end users, rather than in an intermediate concept space. This metric is complemented by utility metrics including coverage and validity to provide a more realistic assessment of explanation quality.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Bagel: A benchmark for assessing graph neural network explanations

URL: View paper

Brief Assessment

BAGEL[64] proposes a general benchmark with multiple evaluation regimes including faithfulness for diverse GNN explanation approaches, but does not specifically address data-grounded fidelity for rule-based explanations evaluated on final subgraph forms rather than intermediate concept spaces.

---

2. Evaluating link prediction explanations for graph neural networks

URL: View paper

Brief Assessment

Link Prediction Explanations[66] focuses on link prediction tasks and provides metrics for that specific context, while the original paper addresses graph classification with rule-based explanations evaluated on final subgraph forms. The candidate does not demonstrate prior work on data-grounded fidelity for rule-based graph classification explanations.

---

3. Evaluating attribution for graph neural networks

URL: View paper

Brief Assessment

Attribution Evaluation[65] focuses on evaluating attribution methods (feature importance) for GNN predictions using metrics like accuracy, faithfulness, stability, and consistency. The original paper introduces data-grounded fidelity (FidD) specifically for rule-based explanations evaluated on final subgraph forms. These are distinct evaluation contexts with different scopes and methodologies.

---

4. Explainability methods for graph convolutional neural networks

URL: View paper

Brief Assessment

GCN Explainability Methods[69] focuses on adapting CNN explainability methods (gradient-based saliency, CAM, Grad-CAM, EB) to GCNs for node-level explanations. The original paper proposes data-grounded fidelity (FidD) specifically for evaluating rule-based global explanations at the graph level, which is a fundamentally different evaluation context and methodology.

---

5. Refining Fidelity Metrics for Explainable Recommendations

URL: View paper

Brief Assessment

Refining Fidelity Metrics[68] focuses on counterfactual evaluation metrics for recommender systems, not graph neural network explanations. The domains, evaluation frameworks, and technical approaches are fundamentally different.

---

6. Xgexplainer: Robust evaluation-based explanation for graph neural networks

URL: View paper

Brief Assessment

Xgexplainer[61] focuses on evaluation metrics for general GNN explanations, not specifically on rule-based explanations or the concept of data-grounded fidelity that assesses explanations in their final subgraph form rather than intermediate concept spaces.

---

7. Evaluating explainability for graph neural networks

URL: View paper

Brief Assessment

Evaluating Explainability[60] focuses on evaluating GNN explanations using ground-truth explanations on synthetic and real-world graphs, not specifically on rule-based explanations or the concept of data-grounded fidelity for final subgraph forms versus intermediate concept spaces.

---

8. Towards robust fidelity for evaluating explainability of graph neural networks

URL: View paper

Brief Assessment

Robust Fidelity[67] focuses on evaluating instance-level explanation functions (subgraph extraction) and addresses distribution shift in existing fidelity metrics (Fid+, Fid-, Fidδ). The original paper proposes FidD specifically for rule-based explanations evaluated on final subgraph forms, not for instance-level explainers. These are distinct evaluation contexts.

---

9. Is your explanation reliable: Confidence-aware explanation on graph neural networks

URL: View paper

Brief Assessment

Confidence-Aware Explanation[62] focuses on instance-level post-hoc explanations with confidence scoring for reliability assessment, not on evaluating rule-based global explanations or proposing fidelity metrics for final subgraph explanations.

---

10. Gnnexplainer: Generating explanations for graph neural networks

URL: View paper

Brief Assessment

Gnnexplainer[63] focuses on instance-level explanations using mutual information optimization, not on evaluating rule-based explanations with data-grounded fidelity metrics. The candidate addresses a different problem scope (generating explanations) rather than evaluating rule-based methods.

Contribution 2: LogicXGNN framework for generating faithful logical rule-based explanations

Description: The authors introduce LogicXGNN, a novel post-hoc explanation framework that generates logical rules using predicates specifically designed to capture structural patterns from the GNN's message-passing mechanism. This design ensures effective grounding of explanations in observable data, producing both representative subgraphs and generalizable grounding rules.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Symbolic rule-based knowledge graph completion

URL: View paper

Brief Assessment

Symbolic Rule Completion[70] focuses on knowledge graph completion using symbolic rules for relational facts, not on post-hoc explanations for graph neural networks. The domains and technical objectives are fundamentally different.

2. Global Graph Counterfactual Explanation: A Subgraph Mapping Approach

URL: View paper

Brief Assessment

Global Graph Counterfactual[72] focuses on counterfactual explanations through subgraph mapping rules that change predictions, not on generating faithful logical rules that explain the GNN's existing decision-making patterns as LogicXGNN does.

3. Fusing logic rule-based hybrid variable graph neural network approaches to fault diagnosis of industrial processes

URL: View paper

Brief Assessment

Logic Rule Hybrid[5] focuses on fault diagnosis in industrial processes using logic rules for classification, not on post-hoc explanations for graph neural networks or ensuring grounding quality in GNN explanations.

4. Encoding concepts in graph neural networks

URL: View paper

Brief Assessment

Encoding Concepts[73] focuses on making GNNs explainable by design through differentiable concept discovery, whereas the original paper proposes a post-hoc explanation framework. The candidate does not demonstrate prior work on post-hoc logical rule generation with predicates designed to capture message-passing structure.

5. Learning rule-induced subgraph representations for inductive relation prediction

URL: View paper

Brief Assessment

Rule-Induced Subgraph[74] focuses on inductive relation prediction in knowledge graphs using subgraph-based GNN methods, not on post-hoc explanation frameworks for GNN interpretability. The candidate addresses a fundamentally different problem domain (knowledge graph completion) rather than explaining GNN predictions through logical rules.

6. GNNBoundary: Towards explaining graph neural networks through the lens of decision boundaries

URL: View paper

Brief Assessment

GNNBoundary[4] focuses on explaining GNN decision boundaries through boundary graph generation, not on logical rule-based post-hoc explanations. The candidate addresses a fundamentally different aspect of GNN explainability.

7. Logical rule-based knowledge graph reasoning: A comprehensive survey

URL: View paper

Brief Assessment

Rule-Based Reasoning Survey[75] focuses on knowledge graph reasoning using logical rules for link prediction and knowledge completion, not on post-hoc explanations for graph neural networks. The survey addresses a fundamentally different problem domain (knowledge graphs) compared to the original paper's focus on explaining GNN predictions.

8. Rule-Guided Graph Neural Networks for Explainable Knowledge Graph Reasoning

URL: View paper

Brief Assessment

Rule-Guided GNN[2] focuses on knowledge graph reasoning with rule-guided training for GNNs, not post-hoc explanation generation for general graph classification tasks. The candidate addresses a different problem domain (KG completion) with different technical objectives (rule-guided training vs. post-hoc explanation).

9. Special issue on feature engineering editorial

URL: View paper

Brief Assessment

Feature Engineering Editorial[71] is a survey paper discussing general feature engineering techniques across various domains (networks, text, time series, multimedia). It does not address post-hoc explanation frameworks for GNNs or logical rule generation for model interpretability, which is the focus of the original paper's contribution.

10. Explainable Deep Learning Models for Detecting Sophisticated Cyber-Enabled Financial Fraud Across Multi-Layered FinTech Infrastructure

URL: View paper

Brief Assessment

Cyber Fraud Detection[76] focuses on explainable deep learning for financial fraud detection in fintech infrastructure, not on generating logical rule-based explanations for graph neural networks. The candidate addresses a completely different application domain (financial fraud) and does not discuss GNN explanation methods or logical rule generation frameworks.

Contribution 3: Reliable predicates preserving GNN message-passing structure

Description: The framework constructs predicates that explicitly model recurring structural patterns induced by the GNN's message-passing computation, using techniques like Weisfeiler-Lehman graph hashing to capture receptive field topologies. This approach addresses unreliable grounding issues in existing methods by ensuring predicates are both structurally grounded and model-faithful.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Learning symbolic models for graph-structured physical mechanism

URL: View paper

Brief Assessment

Symbolic Physical Mechanism[54] focuses on discovering symbolic formulas for graph-structured physical mechanisms by searching message-passing flows in GNNs, not on constructing predicates for GNN explanations. The candidate addresses formula discovery in physics domains, while the original addresses explainability through logical rules.

2. Heterophilic Graph Neural Networks Optimization with Causal Message-passing

URL: View paper

Brief Assessment

Heterophilic Causal[53] focuses on learning causal structures for heterophilic message-passing in GNNs through intervention-based causal inference, not on constructing predicates that capture message-passing structure for explanation purposes as in the original paper.

3. KnowGNN: a knowledge-aware and structure-sensitive model-level explainer for graph neural networks

URL: View paper

Brief Assessment

KnowGNN[55] focuses on knowledge-aware explanations using graphical models, not on constructing predicates that explicitly model message-passing structure through techniques like Weisfeiler-Lehman hashing.

4. Flowx: Towards explainable graph neural networks via message flows

URL: View paper

Brief Assessment

Flowx[51] focuses on explaining GNNs through message flows using Shapley values, not on constructing predicates that model message-passing structure for rule-based explanations. The approaches address different aspects of GNN explainability.

5. A theory of graph comprehension

URL: View paper

Brief Assessment

Graph Comprehension Theory[58] focuses on human cognitive processes for understanding graph visualizations, not on constructing predicates for GNN explanations or capturing message-passing computational structures.

6. Explainable GNN-based models over knowledge graphs

URL: View paper

Brief Assessment

Explainable GNN KG[25] focuses on extracting datalog rules from trained GNNs for knowledge graph reasoning, not on constructing predicates that model message-passing structure for explanation purposes. The candidate addresses symbolic rule extraction from GNN models, while the original contribution concerns designing predicates that explicitly capture GNN computational patterns to ensure reliable grounding in explanations.

7. Revelio: Revealing Important Message Flows in Graph Neural Networks

URL: View paper

Brief Assessment

Revelio[57] focuses on quantifying importance of message flows for instance-level explanations, while the original paper constructs predicates that capture recurring structural patterns for global rule-based explanations. These are fundamentally different explanation paradigms.

8. Message-passing selection: Towards interpretable GNNs for graph classification

URL: View paper

Brief Assessment

Message-Passing Selection[56] focuses on selecting critical message aggregation paths for self-explanation during inference, not on constructing predicates that capture message-passing structure for post-hoc rule-based explanations as in the original paper.

9. A logic for reasoning about aggregate-combine graph neural networks

URL: View paper

Brief Assessment

Aggregate-Combine Logic[52] focuses on modal logic formulations and transformations between logical formulas and GNNs, not on constructing predicates for explanation frameworks that explicitly model message-passing patterns using techniques like Weisfeiler-Lehman hashing.

10. L2XGNN: learning to explain graph neural networks

URL: View paper

Brief Assessment

L2XGNN[59] focuses on learning to select explanatory subgraphs used exclusively in message-passing operations, not on constructing predicates that model message-passing structure for rule-based explanations. The technical approaches differ fundamentally in their explanation paradigms.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] LogicXGNN: Grounded Logical Rules for Explaining Graph Neural Networks View paper
- [1] Graph neural networks and logic View paper
- [2] Rule-Guided Graph Neural Networks for Explainable Knowledge Graph Reasoning View paper
- [3] Global concept-based interpretability for graph neural networks via neuron analysis View paper
- [4] GNNBoundary: Towards explaining graph neural networks through the lens of decision boundaries View paper
- [5] Fusing logic rule-based hybrid variable graph neural network approaches to fault diagnosis of industrial processes View paper
- [6] Interpretable Hierarchical Concept Reasoning through Attention-Guided Graph Learning View paper
- [7] Back to the future: Towards explainable temporal reasoning with large language models View paper
- [8] Rule Based Learning with Dynamic (Graph) Neural Networks View paper
- [9] Graph neural networks in neural-symbolic computing View paper
- [10] Global explainability of gnns via logic combination of learned concepts View paper
- [11] Extracting Interpretable Logic Rules from Graph Neural Networks View paper
- [12] Sound Logical Explanations for Mean Aggregation Graph Neural Networks View paper
- [13] Beyond Topological Self-Explainable GNNs: A Formal Explainability Perspective View paper
- [14] Smart Decision Orchestration for Consumer Electronics Management Using Dynamic Neuro-Symbolic AI Fusion View paper
- [15] CORTEX: A Cost-Sensitive Rule and Tree Extraction Method View paper
- [16] An interpretable model for sepsis prediction using multi-objective rule extraction View paper
- [17] Neural Probabilistic Logic Learning for Knowledge Graph Reasoning View paper
- [18] From latent to lucid: Transforming knowledge graph embeddings into interpretable structures View paper
- [19] A comprehensive survey on self-interpretable neural networks View paper
- [20] Generating explanations for conceptual validation of graph neural networks: An investigation of symbolic predicates learned on relevance-ranked sub-graphs View paper
- [21] EDGE: evaluation framework for logical vs. subgraph explanations for node classifiers on knowledge graphs View paper
- [22] Explanation of Deep Learning Models via Logic Rules Enhanced by Embeddings Analysis, and Probabilistic Models View paper
- [23] Logical Expressiveness of Graph Neural Networks on Knowledge Graphs View paper
- [24] Neural-symbolic models for logical queries on knowledge graphs View paper
- [25] Explainable GNN-based models over knowledge graphs View paper
- [26] A New Concept for Explaining Graph Neural Networks. View paper
- [27] Global explanation supervision for Graph Neural Networks View paper
- [28] Faithfully explainable recommendation via neural logic reasoning View paper
- [29] eXpath: Explaining Knowledge Graph Link Prediction with Ontological Closed Path Rules View paper
- [30] Explaining GNN-based Recommendations in Logic View paper
- [31] ARM-Explainer--Explaining and improving graph neural network predictions for the maximum clique problem using node features and association rule mining View paper
- [32] SAFRAN: An interpretable, rule-based link prediction method outperforming embedding models View paper
- [33] Inductive knowledge graph completion with GNNs and rules: An analysis View paper
- [34] Learning reasoning strategies in end-to-end differentiable proving View paper
- [35] Contrastive Graph Representations for Logical Formulas Embedding View paper
- [36] Fuzzy Rule-Based Explainer Systems for Deep Neural Networks: From Local Explainability to Global Understanding View paper
- [37] A Framework for Extracting Rules from Explanation Subgraphs for Gnn-Based Knowledge Graph Link Prediction View paper
- [38] GnnXemplar: Exemplars to Explanations - Natural Language Rules for Global GNN Interpretability View paper
- [39] Learning Signal Temporal Logic through Neural Network for Interpretable Classification View paper
- [40] Understanding expressivity of gnn in rule learning View paper
- [41] Generating Graph-Based Rules for Enhancing Logical Reasoning View paper
- [42] FireGNN: Neuro-Symbolic Graph Neural Networks with Trainable Fuzzy Rules for Interpretable Medical Image Classification View paper
- [43] Ensemble interpretation: a unified framework for explanation methods View paper
- [44] Logical Characterizations of GNNs with Mean Aggregation View paper
- [45] GraphTrail: Translating GNN Predictions into Human-Interpretable Logical Rules View paper
- [46] GraphMind: Theorem Selection and Conclusion Generation Framework with Dynamic GNN for LLM Reasoning View paper
- [47] A Hybrid Transformer and GNN Framework for Interpretable Fair Value Classification in Accounting View paper
- [48] Assessing Natural Language Explanations of Relational Graph Neural Networks View paper
- [49] DT+GNN: A Fully Explainable Graph Neural Network using Decision Trees View paper
- [50] On GNN explainability with activation rules View paper
- [51] Flowx: Towards explainable graph neural networks via message flows View paper
- [52] A logic for reasoning about aggregate-combine graph neural networks View paper
- [53] Heterophilic Graph Neural Networks Optimization with Causal Message-passing View paper
- [54] Learning symbolic models for graph-structured physical mechanism View paper
- [55] KnowGNN: a knowledge-aware and structure-sensitive model-level explainer for graph neural networks View paper
- [56] Message-passing selection: Towards interpretable GNNs for graph classification View paper
- [57] Revelio: Revealing Important Message Flows in Graph Neural Networks View paper
- [58] A theory of graph comprehension View paper
- [59] L2XGNN: learning to explain graph neural networks View paper
- [60] Evaluating explainability for graph neural networks View paper
- [61] Xgexplainer: Robust evaluation-based explanation for graph neural networks View paper
- [62] Is your explanation reliable: Confidence-aware explanation on graph neural networks View paper
- [63] Gnnexplainer: Generating explanations for graph neural networks View paper
- [64] Bagel: A benchmark for assessing graph neural network explanations View paper
- [65] Evaluating attribution for graph neural networks View paper
- [66] Evaluating link prediction explanations for graph neural networks View paper
- [67] Towards robust fidelity for evaluating explainability of graph neural networks View paper

- [68] Refining Fidelity Metrics for Explainable Recommendations View paper
- [69] Explainability methods for graph convolutional neural networks View paper
- [70] Symbolic rule-based knowledge graph completion View paper
- [71] Special issue on feature engineering editorial View paper
- [72] Global Graph Counterfactual Explanation: A Subgraph Mapping Approach View paper
- [73] Encoding concepts in graph neural networks View paper
- [74] Learning rule-induced subgraph representations for inductive relation prediction View paper
- [75] Logical rule-based knowledge graph reasoning: A comprehensive survey View paper
- [76] Explainable Deep Learning Models for Detecting Sophisticated Cyber-Enabled Financial Fraud Across Multi-Layered FinTech Infrastructure View paper