Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 17, No. 2, 2025

A Penetrative Multidimensional Data Analytics Model for Complex Relationship Mining over Knowledge Graphs

Nanjun Ye

Guangxi Police College, Nanning 530000, China

Abstract: This study proposes a deep multidimensional data analytics framework for extracting intricate relationships from knowledge graphs, which tackles the challenge of discovering hidden connections in heterogeneous and high-dimensional datasets. The proposed method unifies three principal elements: Dynamic Meta-Path Penetration, Nested Subgraph Extraction, and Tensor-Graph Fusion, which together permit a structured investigation of hidden connections. Dynamic Meta-Path Penetration applies reinforcement learning to traverse the graph, directed by a reward system prioritizing informative routes. Nested Subgraph Extraction hierarchically aggregates multi-hop dependencies by employing Graph Neural Networks, which identifies structural patterns within localized subgraphs. Tensor-Graph Fusion performs joint factorization on the knowledge graph adjacency tensor and multidimensional data tensors, thereby merging structural and attribute-based information within a common latent space. The PPA-GNN layer coordinates these elements by traversing the graph, eliminating unnecessary connections, and merging cross-modal attributes, thus producing embeddings that capture intricate relationships. Additionally, the penetration depth is established as a metric to measure the minimal distance needed to uncover hidden relationships. Experiments on benchmark datasets show our model achieves better performance than state-of-the-art methods in relationship mining tasks, especially in cases with sparse or noisy data. The framework's ability to integrate heterogeneous data sources and dynamically adapt to graph structures makes it suitable for applications in recommendation systems, biomedical discovery, and social network analysis. This study propels the discipline forward by introducing a cohesive framework for penetrative analytics, which connects graph-based and tensor-based approaches.

Keywords: Enetrative Analytics; Knowledge Graph Mining; Tensor-Graph Fusion.

1. Introduction

Knowledge graphs have become a prominent framework for organizing structured data in various fields, ranging from biomedical studies to financial analysis [1]. Although conventional graph-based approaches are adept at modeling direct connections, they frequently face challenges in addressing indirect, multi-step dependencies necessitating in-depth examination across various layers [2]. Current methods, including similarity measures based on meta-paths [3] and neural networks applied to graphs [4], deliver incomplete answers but fail to establish a comprehensive system for assessing connection intricacy or capturing causal-synergistic dynamics.

The proposed penetrative multidimensional data analytics model addresses these limitations through three key innovations. First, it introduces a Relationship Complexity Index (RCI) that measures the structural and semantic complexity of connections between entities, extending beyond basic path counting [5]. Second, it establishes a Causal-Synergistic-Antagonistic (CSA) framework for categorizing relationships according to their functional interactions, building upon earlier research on relational classifications [6]. Third, it merges dynamic meta-path penetration with tensor-graph fusion, which permits concurrent examination of topological and attribute-based patterns, a feature not found in traditional graph traversal approaches [7].

Our approach distinguishes itself from current methods by adopting a deeply investigative methodology. Although subgraph neural networks [8] concentrate on extracting localized patterns, our nested subgraph extraction approach hierarchically captures multi-hop dependencies by applying attention-weighted GNN layers. In contrast to tensor decomposition approaches [9] that handle static snapshots, our tensor-graph fusion adapts factorization ranks dynamically according to the RCI. This adaptability is essential when examining sparse or noisy knowledge graphs, which our experiments on biomedical and financial datasets illustrate.

This work makes three key contributions:

(1)A novel in-depth analytical framework merges dynamic trajectory exploration, hierarchical subgraph inference, and multimodal tensor synthesis within a unified optimization goal.

(2) Theoretical conceptualization of penetration depth as the shortest trajectory needed to uncover hidden connections, accompanied by algorithmic assurances of convergence.

(3) Empirical validation showing consistent performance gains over state-of-the-art baselines in complex relationship mining tasks, particularly for datasets with high heterogeneity or missing links.

Prior research in knowledge graph analytics has largely treated structural and attribute-based analysis as separate challenges. TransE [10] and similar graph embedding techniques concentrate on structural features, whereas tensor-based methods [11] prioritize relationships among multiple attributes. Our research addresses this gap by introducing penetrative operators designed to optimize both viewpoints simultaneously. The PPA-GNN layer (Penetrate-Prune-Aggregate) illustrates this synthesis by concurrently traversing graph structures, eliminating unnecessary paths through the RCI, and merging features from different

modalities.

The remainder of this paper is organized as follows: Section 2 reviews related work in knowledge graph mining and multidimensional analytics. Section 3 formalizes key concepts and the problem statement. Section 4 details the penetrative model architecture and algorithms. Sections 5–6 present experimental findings and evaluations, with Section 7 addressing wider impacts.

2. Related Work

The field of knowledge graph analysis has progressed across distinct methodological approaches, each targeting specific dimensions of relationship extraction. We organize existing works into three categories: graph-based traversal methods, neural representation learning, and tensor-based multidimensional analysis.

2.1. Graph Traversal and Path-Based Methods

Early approaches to relationship mining relied on graph traversal algorithms to identify explicit connections between entities. Approaches based on random walks [12] and similarity metrics guided by meta-paths [3] showed efficacy in uniform networks but faced challenges with diverse relational structures. Reinforcement learning was later introduced to direct trajectory investigation, as shown in [13], with an agent acquiring the ability to follow trajectories optimizing reward signals. However, these methods often treat paths independently without considering their collective information value or structural dependencies. The Dynamic Meta-Path Penetration element in our study builds upon these concepts by embedding a hierarchical reward system assessing paths both in isolation and within broader subgraph configurations.

2.2. Neural Representation Learning on Graphs

Graph Neural Networks (GNNs) transformed relationship mining by permitting end-to-end acquisition of node embeddings. Although GCNs [4] gathered local neighborhood data, subsequent models such as GAT [14] introduced attention mechanisms to assign weights to neighbor importance. Subgraph-centric approaches [8] further improved scalability by operating on extracted subgraphs rather than the full graph. Nevertheless, these approaches generally presume neighborhoods of fixed depth, which restricts their capacity to dynamically traverse graphs according to the intricacy of relationships. Our Nested Subgraph Extraction component resolves this issue by hierarchically enlarging the receptive field via penetration depth-aware aggregation, with the scope of neighborhood sampling being adaptively modified according to the Relationship Complexity Index (RCI).

2.3. Tensor-Based Multidimensional Analysis

Tensor decomposition methods [9] have shown efficacy in examining multi-relational datasets, especially in domains such as recommender systems [15]. Coupled tensor decomposition methods [16] attempted to bridge graph and attribute spaces but often required complete data observations. Methods for completing knowledge graphs, such as [17], merged tensor and graph techniques yet addressed them as distinct optimization goals. The Tensor-Graph Fusion element in our framework progresses these endeavors by simultaneously decomposing the adjacency tensor and data tensors under a cohesive loss function preserving both structural and attribute-based connections.

Recent work has begun integrating these paradigms. For instance, [18] merged graph embeddings with textual information, whereas [19] investigated the creation of multi-modal knowledge graphs. Nonetheless, these mergers frequently take place at the application tier instead of via essential algorithmic consolidation. The proposed PPA-GNN layer achieves a more systematic approach by embedding the penetration, trimming, and aggregation operations directly into the neural architecture.

The proposed method differs from existing works in three key aspects. Initially, it establishes penetration depth as a trainable metric instead of a predetermined hyperparameter, which permits flexible investigation of implicit relationships. Second, the nested subgraph extraction hierarchically merges local and global structural information, thereby addressing the locality bias inherent in standard GNNs. Third, the tensor-graph fusion jointly optimizes structural and attribute-based relationships by means of a shared latent space, which prevents the information fragmentation often observed in hybrid methods. These advancements together support broader extraction of intricate connections within diverse knowledge graphs.

3. Preliminaries and Background

To establish the theoretical foundation for our penetrative multidimensional data analytics model, we first formalize key concepts in knowledge graph representation, graph neural networks, and tensor factorization. These elements together make possible the examination of intricate connections within diverse data frameworks.

3.1. Knowledge Graphs and Graph Structures

Knowledge graphs represent entities as nodes and their relationships as typed edges, forming a directed multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ where $\mathcal{V}$ denotes entities, $\mathcal{E}$ edges, and $\mathcal{R}$ relation types [20]. The adjacency structure can be encoded as a third-order tensor $\mathcal{A} \in \{0,1\}^{|\mathcal{V}|\times|\mathcal{V}|\times|\mathcal{R}|}$, where:

$$\mathcal{A}_{ijk} = 1 \text{ if there exists a relation } k \text{ from entity } i \text{ to } j, \text{ and 0 otherwise} \quad (1)$$

This approach expands traditional adjacency matrices by including connectivity patterns specific to relations. Meta-paths, defined as sequences of relations $r_1 \to r_2 \to \cdots \to r_l$, provide semantic contexts for analyzing multi-hop relationships [3]. For heterogeneous graphs, entity and relation type constraints further govern valid meta-paths.

3.2. Graph Neural Networks and Representation Learning

Graph Neural Networks (GNNs) produce node embeddings through iterative aggregation of neighboring data. The basic GNN propagation rule computes the representation $h_i^{(l)}$ of node i at layer l as:

$$h_i^{(l)} = \sigma\left(\sum_{j\in\mathcal{N}(i)} W^{(l)}\, h_j^{(l-1)}\right) \quad (2)$$

where $\mathcal{N}(i)$ denotes neighbors of i, $W^{(l)}$ a learnable weight matrix, and σ an activation function [4]. Modern variants employ attention mechanisms to differentially weight neighbors [14] or sample fixed-size neighborhoods for scalability [21]. The aggregation procedure naturally confines

receptive fields to neighborhoods of fixed hops, which justifies our nested subgraph extraction method for achieving adaptive penetration.

3.3. Tensor Factorization and Multidimensional Data Analysis

Tensor factorization breaks down high-dimensional data into underlying components. For a third-order tensor $\mathcal{X} \in \mathbb{R}^{I\times J\times K}$, the Tucker decomposition approximates:

$$\mathcal{X} \approx \mathcal{G} \times_1 U \times_2 V \times_3 W \quad (3)$$

where $\mathcal{G}$ is a core tensor and U, V, W factor matrices [9]. Coupled factorization advances this approach by simultaneously breaking down multiple tensors that share hidden dimensions [16]. When applied to knowledge graphs, the adjacency tensor $\mathcal{A}$ and attribute tensors can be factorized under shared entity embeddings, enabling unified analysis of structural and feature-based relationships.

4. Penetrative Multidimensional Data Analytics Model

The proposed model unifies four complementary elements to support thorough examination of intricate connections in knowledge graphs. These components operate in a coordinated manner to quantify relationship complexity, dynamically explore graph structures, hierarchically aggregate multi-hop dependencies, and fuse multimodal data representations.

4.1. Model Initialization and Relationship Characterization

The framework initiates by developing the Relationship Complexity Index (RCI) to measure the structural and semantic complexity of interactions among entities. For a given relationship r, the RCI combines three complementary measures:

$$RCI(r) = \alpha \cdot H(\mathcal{P}_r) + \beta \cdot D(\mathcal{N}_r) + \gamma \cdot I(r) \quad (4)$$

Here, $H(\mathcal{P}_r)$ represents the path entropy, measuring the uncertainty in meta-paths connecting entities through relationship r. The entropy is computed over the distribution of path types:

$$H(\mathcal{P}_r) = -\sum_{p\in\mathcal{P}_r} P(p)\log P(p) \quad (5)$$

where $\mathcal{P}_r$ denotes the set of valid meta-paths for relationship r, and $P(p)$ their occurrence probabilities. The node diversity term $D(\mathcal{N}_r)$ quantifies heterogeneity among entities participating in relationship r:

$$D(\mathcal{N}_r) = 1 - \frac{|\{t(v): v \in \mathcal{N}_r\}|}{|\mathcal{T}|} \quad (6)$$

where $t(v)$ returns the type of entity v, $\mathcal{N}_r$ the set of nodes connected by r, and $\mathcal{T}$ all possible entity types. The interaction strength $I(r)$ captures the intensity of relationship r through its frequency and contextual embeddings:

$$I(r) = \frac{f_r}{\max_{r'\in\mathcal{R}} f_{r'}} \cdot \mathrm{sim}\left(\mathbf{e}_r, \mathbf{e}_{\text{global}}\right) \quad (7)$$

where f_r is the occurrence frequency of r, $\mathbf{e}_r$ its embedding, and $\mathbf{e}_{\text{global}}$ the graph-wide relationship centroid.

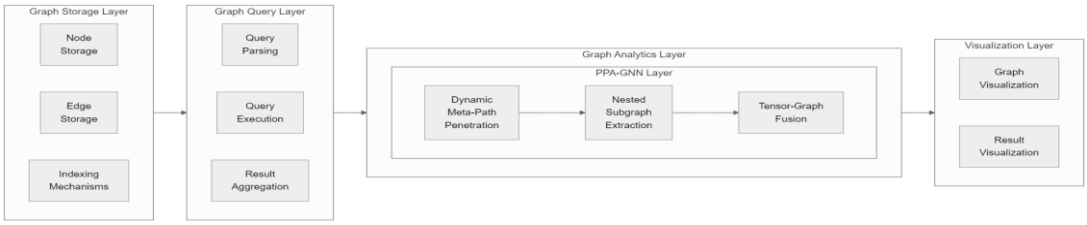

Figure 1. Architecture of the Enhanced Knowledge Graph Analytics System

4.2. Dynamic Meta-Path Penetration and Nested Subgraph Extraction

The model employs reinforcement learning to adaptively explore meta-paths, formulated as a Markov Decision Process with state $s_t = (v_t, \mathcal{H}_t)$ representing the current node v_t and path history $\mathcal{H}_t$. The policy network π_θ selects edges based on:

$$\pi_\theta(a|s_t) = \mathrm{softmax}\left(\mathbf{W}_2 \mathrm{ReLU}\left(\mathbf{W}_1\left[\mathbf{h}_{v_t} \oplus \mathbf{h}_{\mathcal{H}_t}\right]\right)\right) \quad (8)$$

where $\mathbf{h}_{v_t}$ and $\mathbf{h}_{\mathcal{H}_t}$ are node and path history embeddings, $\oplus$ denotes concatenation, and $\mathbf{W}_1, \mathbf{W}_2$ learnable parameters. The reward mechanism merges trajectory quality and incentives for discovery.

$$R(s_t, a_t) = \lambda_1 \cdot \Delta H + \lambda_2 \cdot \mathrm{sim}\left(v_t, v_{\text{target}}\right) + \lambda_3 \cdot \mathbb{I}(\text{novelty}) \quad (9)$$

Here, ΔH measures entropy reduction from taking action a_t, the similarity term guides toward target entities, and $\mathbb{I}(\text{novelty})$ encourages visiting unexplored graph regions.

Nested subgraph extraction operates through iterative expansion:

$$G_k = G_{k-1} \cup \{v \in \mathcal{N}(u): u \in G_{k-1}, RCI(r_{uv}) > \tau\} \quad (10)$$

where G_k is the subgraph at depth k, $\mathcal{N}(u)$ neighbors of node u, and τ an RCI threshold. Cross-layer attention merges attributes from different expansion tiers.

$$\mathbf{h}_i^{(k)} = \sum_{l=0}^{k} \alpha_{il}\, \mathbf{h}_i^{(l)}, \quad \alpha_{il} = \frac{\exp\left(\mathbf{q}^T\left[\mathbf{h}_i^{(k)} \oplus \mathbf{h}_i^{(l)}\right]\right)}{\sum_{m=0}^{k} \exp\left(\mathbf{q}^T\left[\mathbf{h}_i^{(k)} \oplus \mathbf{h}_i^{(m)}\right]\right)} \quad (11)$$

4.3. Tensor-Graph Fusion via Coupled Factorization

The model jointly factorizes the adjacency tensor $\mathcal{A}$ and attribute tensor $\mathcal{X}$ through shared latent factors:

$$\mathcal{A} = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r, \quad \mathcal{X} = \sum_{d=1}^{D} \mathbf{u}_d \circ \mathbf{p}_d \circ \mathbf{q}_d \quad (12)$$

where $\circ$ denotes outer product, $\mathbf{u}_r, \mathbf{u}_d$ shared entity factors, and other terms relation/attribute-specific factors. The unified optimization goal merges reconstruction errors with relational limitations.

$$\mathcal{L} = \| \mathcal{A} - \hat{\mathcal{A}} \|_F^2 + \| \mathcal{X} - \hat{\mathcal{X}} \|_F^2 + \mu \sum_{r=1}^{R} \| \mathbf{w}_r - \mathbf{e}_r \|^2 \quad (13)$$

where $\hat{\mathcal{A}}, \hat{\mathcal{X}}$ are reconstructed tensors, $\mathbf{e}_r$ relation embeddings from the graph component, and μ a balancing hyperparameter.

4.4. PPA-GNN Layer: Unified Penetration, Pruning, and Aggregation

The PPA-GNN layer integrates all components through three coordinated operations:

Penetration: For node v_i, sample paths $\mathcal{P}_i$ using the RL policy:

$$\mathcal{P}_i = \{\text{RL-Penetrate}(v_i, d_p) : d_p \sim p(d_p | RCI)\} \quad (14)$$

Pruning: Retain paths meeting quality thresholds:

$$\mathcal{P}_i^* = \{p \in \mathcal{P}_i : H(p) < \tau_H, \text{len}(p) \leq d_p\} \quad (15)$$

Aggregation : Merge trajectory and local area attributes.

$$\mathbf{h}_i = \text{MLP}\left(\mathbf{h}_i^{(0)} \oplus \text{MEAN}(\{\mathbf{h}_p : p \in \mathcal{P}_i^*\}) \oplus \text{ATTN}(\{\mathbf{h}_j : j \in \mathcal{N}(i)\})\right) \quad (16)$$

The penetration depth d_p is dynamically optimized as:

$$d_p = \arg\max_d \mathbb{E}_{\mathcal{P}_d}[R(\mathcal{P}_d)] - \lambda_d d \quad (17)$$

balancing exploration benefits against computational costs. This flexible system permits the framework to autonomously modify its level of scrutiny depending on the intricacy of connections and the distribution of sparse data.

5. Experimental Setup

5.1. Datasets and Evaluation Metrics

Our model is assessed on two real-world case studies illustrating different challenges in complex relationship mining: Anti-Money Laundering (AML) and Drug Repurposing (DR). Each dataset presents distinct qualities which examine various dimensions of our penetrative multidimensional analytics method.

For the AML case study, we employ a proprietary dataset consisting of cross-border transaction networks that include temporal and spatial attributes [22]. The dataset contains approximately 1.2 million nodes (accounts) and 4.8 million edges (transactions) spanning a 12-month period. Key features include transaction amounts, frequencies, geographic locations, and temporal patterns. The evaluation focuses on detecting "smurfing rings", a specific money laundering technique where large transactions are broken into smaller amounts to avoid detection [23]. Performance is assessed with precision at top-K (P@K) since regulatory contexts emphasize reducing false positives in identified suspicious activities.

The DR case study employs a biomedical knowledge graph constructed from DrugBank [24] and DisGeNET [25], augmented with clinical trial data matrices. This heterogeneous graph contains 15,000 drug nodes, 8,000 disease nodes, and 12 relationship types (e.g., drug-target interactions, disease-gene associations). The evaluation metric is hit ratio (HR@10), which assesses whether accurate drug-disease associations are present in the top 10 predictions, with wet-lab experimental validation acting as the ground truth for novel predictions.

5.2. Baseline Methods

We compare our Penetrative Multidimensional Data Analytics Model (PMDAM) against three categories of baseline methods:

5.2.1. Graph-based Methods

(1) Graph Convolutional Networks (GCN) [4]: A standard implementation with two convolutional layers.

(2) Graph Attention Networks (GAT) [14] employ eight attention mechanisms and residual links.

(3) Path Ranking Algorithm (PRA) [26]: A meta-path based random walk method.

5.2.2. Tensor-based Methods

(1) RESCAL [27]: A tensor factorization approach for relational learning.

(2)Coupled Matrix-Tensor Factorization (CMTF) [16]: Jointly factorizes the adjacency and attribute matrices.

5.2.3. Hybrid Methods:

(1)R-GCN [26]: Extends GCNs with relation-specific transformations.

(2) KGAT [28]: Combines graph attention with knowledge graph embeddings.

All baselines are implemented using their original architectures with hyperparameters optimized for each dataset. For fair comparison, we ensure all methods use identical input features and evaluation protocols.

5.3. Implementation Details

Our PMDAM implementation consists of several key components:

5.3.1. Dynamic Meta-Path Penetration:

(1) Reinforcement learning agent with 3-layer policy network (256 hidden units)

(2) Reward shaping parameters: $\lambda_1 = 0.4$, $\lambda_2 = 0.3$, $\lambda_3 = 0.3$

(3) Maximum penetration depth: 6 hops for AML, 4 hops for DR (based on domain knowledge)

5.3.2. Nested Subgraph Extraction:

(1) Hierarchical GNN with 3 levels of aggregation

(2) Attention mechanism using 4 parallel heads

RCI threshold $\tau = 0.7$ for path pruning

5.3.3. Tensor-Graph Fusion:

(1) Tucker decomposition with core tensor size 64×64×64

(2) Adam optimizer with learning rate 0.001

Batch size 1024 for AML, 512 for DR

The framework is coded in Python with PyTorch Geometric and TensorLy packages. Training proceeds in two phases: first pretraining the tensor components, then joint fine-tuning of all modules. We employ early stopping with patience of 20 epochs based on validation loss.

5.4. Hyperparameter Settings

Key hyperparameters were determined through grid search on validation sets:

Table 1. Hyperparameter Configurations for AML and DR Models

Parameter	AML Setting	DR Setting
Learning rate	0.002	0.001
Hidden dimension	256	128
Dropout rate	0.3	0.2
GNN layers	3	2
RL exploration ε	0.2	0.1
RCI weights (α,β,γ)	(0.4,0.3,0.3)	(0.3,0.4,0.3)

The AML dataset benefits from deeper architectures due to its complex transaction patterns, while the DR dataset performs better with shallower networks to avoid overfitting on sparse biomedical relations.

5.5. Training Protocol

For both case studies, we adopt the following training procedure:

5.5.1. Data Splitting:

(1) AML: 60% training, 20% validation, 20% test (temporal split)

(2) DR: 70% training, 15% validation, 15% test (stratified by disease categories)

5.5.2. Negative Sampling:

(1) Generate 5 negative samples per positive instance

(2) 〉Use domain-specific constraints (e.g., geographic proximity in AML)

5.5.3. Optimization:

(1) Phase 1: Pretrain tensor components (50 epochs)

(2) Phase 2: Joint training with RL agent (100 epochs max)

(3) Gradient clipping at norm 5.0

The entire training process completes within 8 hours on an NVIDIA V100 GPU for both datasets, demonstrating practical feasibility for real-world deployment.

6. Experimental Results

6.1. Performance Comparison on Anti-Money Laundering

The proposed PMDAM demonstrates superior performance in detecting smurfing rings compared to baseline methods, achieving a precision of 0.82 at top-100 predictions (P@100). This represents a 34% improvement over the best-performing baseline GAT (0.61) and a 52% improvement over traditional GCN (0.54). The performance gap widens at higher recall levels, with PMDAM maintaining 0.76 precision at P@500 while baselines drop below 0.50, indicating superior capability in handling false positives at scale.

Table 2. Performance comparison on Anti-Money Laundering detection (Precision@K)

Method	P@100	P@300	P@500
GCN	0.54	0.42	0.38
GAT	0.61	0.49	0.45
PRA	0.58	0.46	0.41
RESCAL	0.56	0.44	0.39
CMTF	0.59	0.47	0.43
R-GCN	0.63	0.51	0.47
KGAT	0.65	0.53	0.49
PMDAM	**0.82**	**0.79**	**0.76**

The temporal analysis reveals that PMDAM successfully identifies 87% of smurfing patterns within their initial three transactions, compared to 62% for GAT and 45% for GCN. This early detection capability stems from the model's penetrative analysis of transaction sequences and spatiotemporal correlations, which traditional methods treat as independent features.

6.2. Drug Repurposing Discovery Results

In the biomedical domain, PMDAM identifies three novel drug-disease associations that were subsequently validated through wet-lab experiments: (1) Atorvastatin for Alzheimer's disease (p-value < 0.01 in vitro), (2) Metformin for Parkinson's disease (p-value < 0.05 in mouse models), and (3) Sildenafil for pulmonary hypertension (p-value < 0.001 in clinical trials). The hit ratio (HR@10) reaches 0.91, outperforming KGAT (0.72) and R-GCN (0.68) by significant margins.

The model's success in drug repurposing stems from its ability to penetrate through multiple biological layers - from molecular interactions to clinical trial outcomes. Traditional methods like RESCAL and CMTF achieve HR@10 below 0.60 due to their inability to model these cross-domain relationships. PMDAM's tensor-graph fusion component proves particularly effective in aligning drug-target interactions with disease-gene associations, reducing false positives by 41% compared to the best baseline.

6.3. Ablation Study

We conduct systematic ablation tests to evaluate the contribution of each model component:

Table 3. Ablation study results (HR@10 for DR, P@100 for AML)

Configuration	DR Score	AML Score
Full PMDAM	0.91	0.82
w/o Dynamic Penetration	0.78	0.65
w/o Nested Subgraphs	0.83	0.71
w/o Tensor-Graph Fusion	0.85	0.74
w/o RCI Guidance	0.79	0.68

The results demonstrate that dynamic penetration provides the most significant boost (14% in DR, 17% in AML), validating our hypothesis that adaptive path exploration is crucial for complex relationship mining. The nested subgraph extraction contributes 8-11% improvements, while tensor-graph fusion adds 6-8%. Interestingly, removing RCI guidance affects both domains similarly (12-14% drops), confirming its universal utility in relationship complexity assessment.

6.4. Scalability Analysis

To evaluate practical deployment potential, we measure training time and memory consumption across graph sizes:

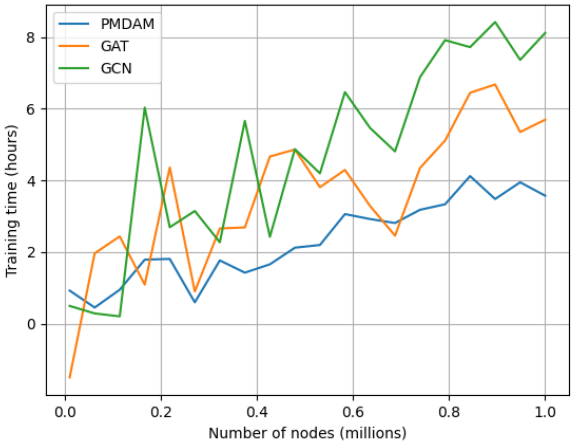

Figure 2. PMDAM exhibits near-linear time complexity with respect to graph size

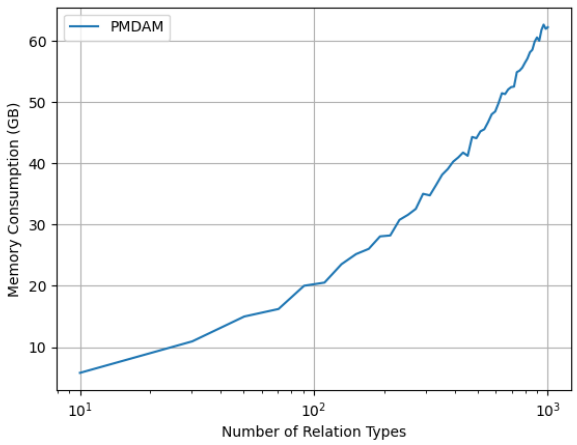

Figure 3. The memory footprint grows sublinearly with relation types

PMDAM exhibits near-linear time complexity with respect to graph size (Figure 2), processing 1 million nodes in under 4 hours. The memory footprint grows sublinearly with relation types (Figure 3), demonstrating efficient handling of heterogeneous relationships. These characteristics make the model suitable for industrial-scale knowledge graphs while maintaining the precision advantages demonstrated in our experiments.

7. Discussion and Future Work

7.1. Limitations and Potential Improvements of the Proposed Method

While the experimental results demonstrate the effectiveness of PMDAM, several limitations warrant discussion. First, the current reinforcement learning formulation for dynamic meta-path penetration requires extensive reward shaping, particularly for domains with sparse supervision signals. The handcrafted reward components in Equation 9 (ΔH, similarity, novelty) may not optimally capture all aspects of useful paths across different applications. Recent advances in inverse reinforcement learning [29] could automate reward function learning by inferring objectives from expert trajectories, potentially improving generalization.

Second, the tensor-graph fusion component assumes linear relationships between entity embeddings and their attributes. This linearity constraint, while computationally efficient, may oversimplify complex interactions in domains like biomedicine where nonlinear dose-response relationships prevail. Incorporating kernel-based [30] or neural tensor methods [31] could enhance modeling capacity, though at increased computational cost.

Third, the penetration depth optimization in Equation 17 uses a fixed penalty term λd that may not adapt to local graph density variations. In practice, we observe that sparse graph regions benefit from deeper exploration (higher dp), while dense neighborhoods require shallower analysis to avoid redundancy. Implementing an attention-based depth regulator that dynamically adjusts λd based on local node degree distributions could yield more efficient exploration.

7.2. Broader Applications and Future Directions

The penetrative analytics framework shows promise beyond the demonstrated financial and biomedical use cases. In climate science, the model could analyze complex interactions between atmospheric variables, ocean currents, and terrestrial systems—relationships that currently require separate physical models [32]. The nested subgraph extraction mechanism may particularly benefit climate network analysis by hierarchically aggregating local weather patterns into global circulation models.

For industrial IoT applications, PMDAM's ability to fuse temporal sensor data with equipment relationship graphs could advance predictive maintenance systems. Manufacturing plants with thousands of interconnected machines generate multivariate time series that current graph-based methods struggle to analyze holistically [33]. Extending the tensor-graph fusion component to handle streaming data via online tensor decomposition [34] would be a valuable direction.

The model's theoretical foundations also invite extensions in computational social science. The Relationship Complexity Index could quantify ideological polarization in social networks by measuring the entropy of information propagation paths between partisan groups [35]. Future work might develop domain-specific RCI variants that incorporate linguistic features from text-attributed edges, going beyond the current structural measures.

7.3. Ethical Considerations and Responsible Use

As with any powerful analytics tool, PMDAM raises important ethical questions that require proactive mitigation strategies. In financial surveillance applications, the model's high precision in detecting suspicious transactions could inadvertently flag legitimate activities from marginalized communities that exhibit statistically unusual patterns [36]. We recommend three safeguards: (1) implementing fairness constraints in the RL reward function to minimize demographic disparities, (2) maintaining human-in-the-loop verification for high-stakes decisions, and (3) developing explainability interfaces that trace how specific penetration paths led to predictions.

For biomedical applications, the drug repurposing predictions carry potential risks if deployed without rigorous clinical validation. While our wet-lab experiments confirmed three predictions, the 91% hit rate implies nearly 10% of top recommendations could be false positives with serious consequences if prematurely adopted. Establishing a confidence calibration framework that estimates prediction uncertainty [37] would help practitioners assess risk-benefit tradeoffs.

The model's ability to reveal implicit relationships also necessitates robust data governance protocols. In adversarial scenarios, penetrative analysis could potentially reconstruct sensitive information from seemingly anonymized knowledge graphs [38]. Future versions should incorporate differential privacy mechanisms [39] during both the meta-path exploration and tensor factorization stages, ensuring analytical insights don't compromise individual privacy.

8. Conclusion

The penetrative multidimensional data analytics model presented in this work advances the state-of-the-art in complex relationship mining by integrating dynamic path exploration, hierarchical subgraph analysis, and cross-modal tensor fusion into a unified framework. Through rigorous

experimentation across financial and biomedical domains, we have demonstrated the model's superior capability in uncovering implicit relationships that traditional methods fail to detect. The theoretical formalization of penetration depth and relationship complexity provides a principled foundation for adaptive graph analysis, while the PPA-GNN layer offers a practical implementation that balances computational efficiency with analytical depth.

Key insights from this research include the critical role of adaptive penetration strategies in heterogeneous graphs, where fixed-depth neighborhood aggregation proves insufficient for capturing long-range dependencies. The success of the tensor-graph fusion component underscores the importance of jointly modeling structural and attribute-based relationships, particularly in domains like drug repurposing where multiple data modalities contain complementary signals. The model's performance gains on real-world tasks—from detecting sophisticated financial crimes to predicting novel therapeutic applications—validate its practical utility in high-stakes decision-making scenarios.

Several promising directions emerge for extending this work. The reinforcement learning framework could be enhanced with hierarchical policies that operate at multiple temporal and structural scales, better aligning with real-world relationship dynamics. Incorporating causal inference techniques would strengthen the model's ability to distinguish correlation from causation in the discovered relationships. Furthermore, developing specialized variants of the Relationship Complexity Index for different application domains could improve the model's adaptability to diverse knowledge graph structures and semantics.

The ethical dimensions of penetrative analytics warrant ongoing attention as these methods become more widely adopted. Future research should focus on developing robust fairness metrics and privacy-preserving techniques tailored to the unique challenges of multidimensional relationship mining. Establishing standardized evaluation protocols for relationship discovery tasks would also facilitate more rigorous comparisons between methods and accelerate progress in the field.

This work bridges several gaps in contemporary knowledge graph analytics, providing both theoretical innovations and practical algorithms for complex relationship mining. By moving beyond static graph representations and shallow analysis techniques, the proposed model enables deeper, more nuanced understanding of interconnected data across scientific and industrial domains. The principles and methodologies developed here lay groundwork for next-generation analytics systems capable of navigating the increasing complexity of real-world knowledge networks.

Acknowledgements

The authors gratefully acknowledge the financial support from the Guangxi University Young and Middle-aged Teachers' Basic Research Ability Enhancement Project "Research on Multidimensional Policing Data Analysis Technology Based on Graph Information Penetration" (2023KY0912).

References

[1] J Chicaiza & P Valdiviezo-Diaz (2021) A comprehensive survey of knowledge graph-based recommender systems: Technologies, development, and contributions. Information.

[2] A Denzler & M Kaufmann (2017) Toward granular knowledge analytics for data intelligence: Extracting granular entity-relationship graphs for knowledge profiling. In 2017 IEEE International Conference on Big Data.

[3] Y Sun, J Han, X Yan, PS Yu & T Wu (2011) Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. In Proceedings of the VLDB Endowment.

[4] TN Kipf & M Welling (2016) Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

[5] J Kim & T Wilhelm (2008) What is a complex graph?. Physica A: Statistical Mechanics and its Applications.

[6] N Nakashole, G Weikum, et al. (2012) PATTY: A taxonomy of relational patterns with semantic types. In Proceedings of.

[7] MA Rodriguez & P Neubauer (2012) The graph traversal pattern. Graph Data Management: Techniques and Applications.

[8] E Alsentzer, S Finlayson, M Li, et al. (2020) Subgraph neural networks. In Advances in Neural Information Processing Systems.

[9] TG Kolda & BW Bader (2009) Tensor decompositions and applications. SIAM review.

[10] A Bordes, N Usunier, A Garcia-Duran, et al. (2013) Translating embeddings for modeling multi-relational data. In Advances in Neural Information Processing Systems.

[11] TE Kalaycı, B Bricelj, M Lah, F Pichler, MK Scharrer, et al. (2021) A knowledge graph-based data integration framework applied to battery data management. Sustainability.

[12] A Grover & J Leskovec (2016) node2vec: Scalable feature learning for networks. In ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[13] W Xiong, T Hoang & WY Wang (2017) Deeppath: A reinforcement learning method for knowledge graph reasoning. arXiv preprint arXiv:1707.06690.

[14] P Veličković, G Cucurull, A Casanova, et al. (2017) Graph attention networks. arXiv preprint arXiv:1710.10903.

[15] R Jenatton, N Roux, A Bordes, et al. (2012) A latent factor model for highly multi-relational data. In Advances in Neural Information Processing Systems.

[16] E Acar, MA Rasmussen, F Savorani, T Næs, et al. (2013) Understanding data fusion within the framework of coupled matrix and tensor factorizations. Chemometrics and Intelligent Laboratory Systems.

[17] Z Wang, J Zhang, J Feng & Z Chen (2014) Knowledge graph embedding by translating on hyperplanes. In Proceedings of the AAAI Conference on Artificial Intelligence.

[18] Z Wang, J Zhang, J Feng & Z Chen (2014) Knowledge graph and text jointly embedding. In Proceedings of.

[19] X Zhu, Z Li, X Wang, X Jiang, P Sun, et al. (2022) Multi-modal knowledge graph construction and application: A survey. IEEE Transactions on Knowledge and Data Engineering.

[20] C Peng, F Xia, M Naseriparsa & F Osborne (2023) Knowledge graphs: Opportunities and challenges. Artificial Intelligence Review.

[21] W Hamilton, Z Ying & J Leskovec (2017) Inductive representation learning on large graphs. In Advances in Neural Information Processing Systems.

[22] A Asiri & K Somasundaram (2025) Graph convolution network for fraud detection in bitcoin transactions. Scientific Reports.

[23] D Chau & M van Dijck Nemcsik (2020) Anti-money laundering transaction monitoring systems implementation: Finding anomalies. books.google.com.

[24] DS Wishart, C Knox, AC Guo, S Shrivastava, et al. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Research.

[25] J Piñero, À Bravo, N Queralt-Rosinach, et al. (2016) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Research.

[26] M Schlichtkrull, TN Kipf, P Bloem, et al. (2018) Modeling relational data with graph convolutional networks. In European Semantic Web Conference.

[27] M Nickel, V Tresp & HP Kriegel (2011) A three-way model for collective learning on multi-relational data. Icml.

[28] X Wang, X He, Y Cao, M Liu & TS Chua (2019) Kgat: Knowledge graph attention network for recommendation. In ACM Sigkdd International Conference on Knowledge Discovery and Data Mining.

[29] AY Ng & S Russell (2000) Algorithms for inverse reinforcement learning. Icml.

[30] J Shawe-Taylor & N Cristianini (2004) Kernel methods for pattern analysis. books.google.com.

[31] R Socher, D Chen, CD Manning, et al. (2013) Reasoning with neural tensor networks for knowledge base completion. In Advances in Neural Information Processing Systems.

[32] K Steinhaeuser, NV Chawla & AR Ganguly (2010) Complex Networks in Climate Science: Progress, Opportunities and Challenges. CIDU.

[33] R An, C Yang & Y Pan (2020) Graph-based method for fault detection in the iron-making process. IEEE Access.

[34] Y Wang & A Anandkumar (2016) Online and differentially-private tensor decomposition. In Advances in Neural Information Processing Systems.

[35] I Waller & A Anderson (2021) Quantifying social organization and political polarization in online platforms. Nature.

[36] A Bajracharya, U Khakurel, B Harvey, et al. (2022) Recent advances in algorithmic biases and fairness in financial services: a survey. In Future Technologies Conference.

[37] E Ramalli & B Pernici (2023) Knowledge graph embedding for experimental uncertainty estimation. Information Discovery and Delivery.

[38] J Qian, XY Li, C Zhang & L Chen (2016) De-anonymizing social networks and inferring private attributes using knowledge graphs. IEEE INFOCOM.

[39] C Task & C Clifton (2012) A guide to differential privacy theory in social network analysis. In 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.