MANTECH Publications
Journal of Data Mining and Knowledge Engineering
Volume 1, Issue 2, July-December 2016

Explainable Reasoning over Knowledge Graphs

Ramchander Singh, Papu S Yadav
Associate Professor, Students
Department of Computer Science
Greenfield University, India
Email: *Ramchander57s@gmail.com, papussy5@yahoo.com*

Abstract

Knowledge graphs (KGs) have become a foundational tool in artificial intelligence, supporting applications in recommendation systems, natural language understanding, and decision-making. While reasoning over knowledge graphs enhances the ability to infer implicit knowledge, many methods remain opaque, leading to limited trust and interpretability. Explainable reasoning over knowledge graphs aims to make the inference process transparent, interpretable, and understandable by humans. This review provides an overview of current methods for explainable reasoning, including rule-based, embedding-based, and hybrid approaches. We discuss evaluation metrics, datasets, and challenges in the field. Moreover, we highlight emerging trends such as neuro-symbolic reasoning and graph attention mechanisms. Through comparative analysis, this paper presents insights into the trade-offs between reasoning performance and explainability, emphasizing the need for methods that balance accuracy with interpretability.

Keywords: *Knowledge Graphs, Explainable AI, Reasoning, Interpretability, Graph Embeddings, Rule-based Reasoning*

INTRODUCTION

Knowledge graphs represent structured knowledge as entities (nodes) and relationships (edges) and provide a flexible way to encode complex information about the world. They have been widely adopted in applications such as search engines, recommendation systems, biomedical informatics,

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

and question-answering systems. Examples include **DBpedia**, **Freebase**, and **Wikidata**, which encode millions of facts across various domains.

Traditional reasoning over KGs focuses on inferring new facts or completing missing relationships, commonly referred to as **knowledge graph completion (KGC)**. While state-of-the-art embedding-based and neural models achieve high accuracy, they are often **black-box** in nature, providing little insight into why a particular prediction is made. In critical applications, such as healthcare or legal domains, such opacity limits trust.

Explainable reasoning bridges this gap by providing understandable explanations for inferred relationships. It combines symbolic and statistical methods to offer transparency while maintaining reasonable predictive performance.

FUNDAMENTALS OF EXPLAINABLE REASONING

Explainable reasoning over knowledge graphs (KGs) combines the **structural representation of knowledge** with methods that make inferences **transparent and understandable** to humans. Before diving into explainability, it is important to understand the basic structure and reasoning operations over knowledge graphs.

1. Knowledge Graph Structure

A **knowledge graph (KG)** is a structured representation of knowledge that encodes entities, their relationships, and associated facts. Formally, a knowledge graph can be defined as:

KG=(E,R,T)KG = (E, R, T)KG=(E,R,T)

Where:

- EEE is the set of **entities**, representing real-world objects or concepts (e.g., people, cities, products).
- RRR is the set of **relations**, representing how entities are connected (e.g., *born_in*, *located_in*, *works_at*).
- T⊆E×R×ET \subseteq E \times R \times ET⊆E×R×E is the set of **triples**, representing factual statements in the form of *(head entity, relation, tail entity)*.

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

Example Triple:

(Barack_Obama,born_in,Honolulu)(Barack_Obama, born_in, Honolulu)(Barack_Obama,born_in,Honolulu)

Here, "Barack_Obama" is the head entity, "born_in" is the relation, and "Honolulu" is the tail entity.

Reasoning Tasks over Knowledge Graphs

Reasoning involves **deriving new knowledge** from the existing graph. The common reasoning tasks include:

1. **Link Prediction**
   - Goal: Infer missing edges between entities.
   - Example: Predict that (Barack_Obama,nationality,USA)(Barack_Obama, nationality, USA)(Barack_Obama,nationality,USA) holds, even if it is missing in the KG.
   - Methods: Embedding-based scoring, path-based inference, or rule-based reasoning.

2. **Entity Classification**
   - Goal: Assign types or categories to entities.
   - Example: Classifying "Honolulu" as a *City* and "Barack_Obama" as a *Person*.
   - Methods: Graph neural networks (GNNs), type propagation, or logic-based approaches.
3. **Path Reasoning / Multi-Hop Reasoning**
   - Goal: Discover chains of relationships connecting two entities.
   - Example: Inferring that "Barack_Obama is a citizen of USA" via: (Barack_Obama,bornin,Honolulu)→(Honolulu,locatedin,USA)(Barack_Obama, born_in, Honolulu) \rightarrow (Honolulu, located_in, USA)(Barack_Obama,borni n,Honolulu)→(Honolulu,locatedin,USA)
   - Significance: Multi-hop reasoning provides richer explanations for inferred facts.

2. Explainability in Reasoning

Explainable reasoning ensures that the inference process over a KG is **transparent, interpretable, and justifiable** to humans. Without explainability, predictions—no matter how accurate—may be unreliable in sensitive domains like healthcare or finance.

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

Types of Explanations

1. **Rule-Based Explanations**
   - Logical rules or Horn clauses are used to explain predictions.
   - Example:
   - Rule: If *X born_in Y* and *Y located_in Z*, then *X nationality Z.*
   - Explanation: "Barack_Obama is a US citizen because he was born in Honolulu, which is located in USA."
   - Advantages: Fully interpretable and precise; easy to trace reasoning steps.
   - Limitation: May not cover all facts, especially in large, incomplete KGs.

2. **Path-Based Explanations**
   - Reasoning is explained through paths in the graph connecting entities.
   - Example: Predicting (Alice,worksat,CompanyX)(Alice, works_at,
   - CompanyX)(Alice,worksat,CompanyX) via:

Alice→colleague_ofBob→works_atCompanyXAlice \xrightarrow{colleague_of} Bob \xrightarrow{works_at} CompanyXAlicecolleague_ofBobworks_atCompanyX

   - Visualization: Highlighted paths in the KG make reasoning intuitive for humans.

3. **Attention-Based Explanations**
   - Embedding-based or neural models assign **attention weights** to edges or neighboring nodes that contribute to a prediction.
   - Example: A graph neural network might assign high attention to the edge (Honolulu,locatedin,USA)(Honolulu, located_in, USA)(Honolulu,locatedin,USA) when predicting Barack_Obama's nationality.
   - Advantage: Works in large-scale graphs where direct rules may be infeasible.

4. **Counterfactual Explanations**
   - Show **what-if scenarios** to justify a prediction.
   - Example: "Barack_Obama would not be predicted as a US citizen if he were born outside of the USA."

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

- Significance: Provides causal insights and improves user trust.

METHODS FOR EXPLAINABLE REASONING

Reasoning over knowledge graphs can be broadly categorized into **rule-based**, **embedding-based**, and **hybrid methods**. Each approach balances **accuracy, scalability, and explainability** differently. Here, we elaborate on **rule-based reasoning**, which is the most interpretable method.

1. Rule-Based Reasoning

Rule-based reasoning relies on **symbolic logic** to extract patterns from the knowledge graph. It uses **if-then rules** to infer new facts, providing transparent and easily understandable explanations for each prediction.

Formal Representation:

A **Horn rule** in a KG can be expressed as:

r1(x,y)∧r2(y,z) ⟹ r3(x,z)r_1(x, y) \wedge r_2(y, z) \implies r_3(x, z)r1(x,y)∧r2(y,z)⟹r3(x,z)

- Here, r1,r2,r3r_1, r_2, r_3r1,r2,r3 are relations, and x,y,zx, y, zx,y,z are entities.
- Example:
- Rule: *(X born_in Y) ∧ (Y located_in Z) ⇒ (X nationality Z)*
- Explanation: “Barack Obama is predicted to have nationality USA because he was born in Honolulu, which is located in USA.”

2. Popular Rule-Based Methods

AMIE (Association Rule Mining under Incomplete Evidence)

- **Function:** Automatically extracts **Horn rules** from large KGs even if some facts are missing.
- **Process:**
- Scans triples in the KG.
- Generates candidate rules based on co-occurrence patterns.
- Computes **confidence scores** to select reliable rules.
- **Example Rule Learned:**

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

- (Person, born_in, City) ∧ (City, located_in, Country) ⇒ (Person, nationality, Country)

AnyBURL (Anytime Bottom-Up Rule Learning)

- **Function:** Learns **probabilistic rules** to predict missing triples.
- **Process:**

1. Randomly samples paths in the KG.
2. Constructs candidate rules from observed patterns.
3. Assigns **probabilities** indicating rule reliability

Table 1: Comparison of Rule-Based Methods

Method	Explanation Type	Scalability	Accuracy	Reference
AMIE	Horn rules	Medium	Moderate	[1]
AnyBURL	Probabilistic rules	High	Moderate	[2]

EMBEDDING-BASED REASONING

Embedding-based reasoning over knowledge graphs (KGs) has emerged as a **powerful approach** for link prediction, entity classification, and multi-hop inference. Unlike rule-based reasoning, which relies on symbolic logic, embedding-based methods **represent entities and relations as vectors in a continuous space**, allowing neural models to generalize patterns and infer new knowledge efficiently.

1. Fundamentals of Embedding-Based Reasoning

In embedding-based methods, a knowledge graph KG=(E,R,T)KG = (E, R, T)KG=(E,R,T) is transformed into a **low-dimensional vector space**:

- **Entities:** Each entity e∈Ee \in Ee∈E is represented as a vector e∈Rd\mathbf{e} \in \mathbb{R}^de∈Rd.
- **Relations:** Each relation r∈Rr \in Rr∈R is represented as a vector or transformation r∈Rd\mathbf{r} \in \mathbb{R}^dr∈Rd.

Objective: For each triple (h,r,t)(h, r, t)(h,r,t), the embeddings are learned such that valid triples have **higher plausibility scores** than invalid ones.

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

Example Models

1. **TransE** (Bordes et al., 2013)
   - Models relations as **translations in the vector space**:

h+r≈t\mathbf{h} + \mathbf{r} \approx \mathbf{t}h+r≈t

   - Suitable for **one-to-one relations**, but limited for complex or symmetric relations.
2. **RotatE** (Sun et al., 2019)
   - Models relations as **rotations in the complex vector space**:

t=h∘r\mathbf{t} = \mathbf{h} \circ \mathbf{r}t=h∘r

   - Can capture **asymmetry, inversion, and composition** of relations.
3. **ComplEx** (Trouillon et al., 2016)
   - Embeddings in **complex space**, enabling modeling of **symmetric and antisymmetric relations**.

2. Explainability Mechanisms in Embedding-Based Models

Since embeddings are **opaque vectors**, auxiliary methods are required to make reasoning interpretable. Two widely used approaches are **path-based attention** and **feature attribution**.

Path-Based Attention

- **Concept:** Highlights sequences of relations (paths) connecting entities that contribute to a prediction.

Process:

- Identify multi-hop paths between head and tail entities.
- Assign **attention weights** to each path, representing its contribution to the prediction.
- **Example:** Predicting (Barack_Obama,nationality,USA)(Barack_Obama, nationality, USA)(Barack_Obama,nationality,USA)
- Path 1: (Barack_Obama,bornin,Honolulu)→(Honolulu,locatedin,USA)(Barack_Obama, born_in, Honolulu) → (Honolulu, located_in, USA)(Barack_Obama,borni n,Honolulu)→(Honolulu,locatedin,USA), attention weight = 0.8
- Path 2: (Barack_Obama,presidentof,USA)(Barack_Obama, president_of, USA)(Barack_Obama,presidentof,USA), attention weight = 0.2

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

- **Interpretation:** Path 1 contributes most to the prediction, providing an intuitive explanation.

Feature Attribution

- **Concept:** Assigns **importance scores** to entities, relations, or embedding dimensions using gradient-based or perturbation methods.
- **Techniques:**
- **Gradient-based:** Compute gradients of the scoring function with respect to entity/relation embeddings.
- **Perturbation-based:** Mask certain embeddings and observe change in prediction score.
- **Outcome:** Identifies which parts of the graph or which embedding features are most influential for the reasoning process.

3. Workflow of Embedding-Based Explainable Reasoning

Step 1: Embedding Learning

- Train embeddings for entities and relations using loss functions such as **margin ranking loss** or **cross-entropy loss**.

Step 2: Triple Scoring

- Compute a **plausibility score** for candidate triples using the model-specific scoring function.

Step 3: Explanation Extraction

- Apply path attention or feature attribution to highlight influential paths or features.

Step 4: Human-Readable Explanation

- Transform attention weights or feature importance into **symbolic paths or descriptive statements**.

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

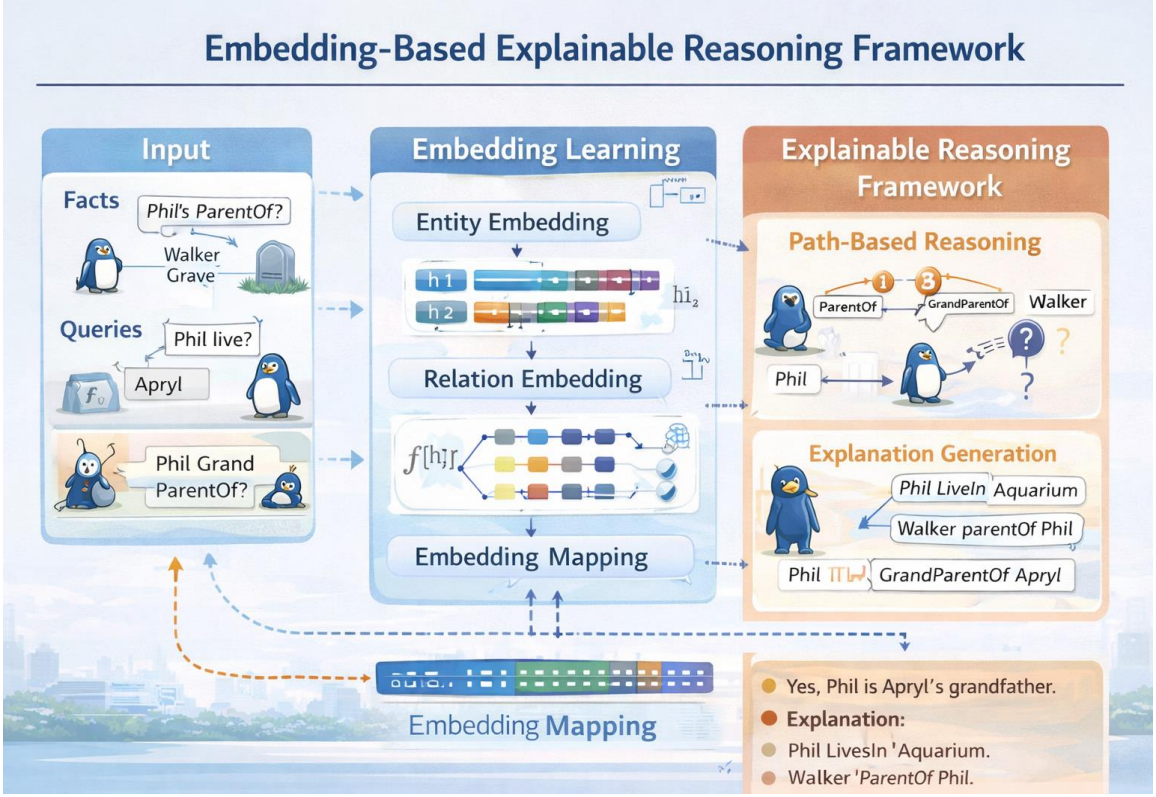


Figure 1: Embedding-Based Explainable Reasoning Framework

HYBRID METHODS

Hybrid methods for explainable reasoning over knowledge graphs (KGs) aim to **combine the interpretability of symbolic approaches** with the **scalability and predictive power of neural embeddings**. By integrating **rule-based logic** with **embedding-based or neural models**, hybrid methods provide a balance between **accuracy** and **explainability**. These approaches are particularly useful for **large, complex KGs** where neither purely symbolic nor purely neural methods are sufficient.

1. Neuro-Symbolic Knowledge Graph Reasoning

Concept:

Neuro-symbolic approaches merge **neural network representations** with **symbolic logical rules**. The goal is to maintain **human-readable reasoning** while benefiting from the **generalization ability of embeddings**.

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

Workflow:

Embedding Learning

- Entities and relations are represented as vectors using neural methods like TransE, RotatE, or GNN embeddings.

Rule Integration

- Logical rules (e.g., Horn clauses) are incorporated to guide neural reasoning.
- Example:
- If X born_in Y AND Y located_in Z THEN X nationality Z
- Rule provides a prior constraint or regularization for the embeddings.

Inference

- The model predicts missing triples using both the learned embeddings and the symbolic rules.
- Neural embeddings handle **complex patterns** while rules provide **interpretable justifications**.

Explanation Generation

- Explanations can be derived from the rules used during inference or from attention weights on paths influenced by rules.

Example Scenario:

- Predict: (Alice, works_at, CompanyX)
- Rule: (X colleague_of Y) ∧ (Y works_at Z) ⇒ (X works_at Z)
- Embedding model identifies candidate companies with high plausibility score.
- Explanation: "Alice is predicted to work at CompanyX because her colleague Bob is associated with CompanyX. This reasoning is consistent with the colleague-based rule."

GNN-BASED HYBRID REASONING

Concept:

Graph Neural Networks (GNNs) process **graph-structured data** by aggregating information from

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

neighboring nodes. When combined with **attention mechanisms** and **symbolic reasoning modules**, GNNs can provide both **accurate predictions** and **path-level explanations**.

Workflow:

Graph Construction

- Build a KG where nodes are entities and edges are relations.

Node & Relation Embeddings

- Initial embeddings are assigned to nodes and relations.

GNN Message Passing

- Each node aggregates features from its neighbors.
- Attention weights determine which neighbors are more influential for a prediction.

Rule Regularization (Optional)

- Integrate symbolic rules as constraints during training to guide message passing.

Prediction & Explanation

- Predicted triples are generated.
- Explanation: Highlight high-attention neighbors or paths consistent with symbolic rules.

Example:

- Task: Predict (Barack_Obama, nationality, USA)
- GNN aggregates neighborhood information:
- Node Honolulu receives high attention because it is connected to USA.
- Node Hawaii contributes via path (Honolulu → Hawaii → USA).
- Explanation: "Prediction is mainly influenced by Barack Obama's birthplace and its location in the USA."

DATASETS FOR EXPLAINABLE REASONING

Commonly used datasets include:

Dataset	Domain	Size	Features
FB15k-237	General Knowledge	14,541 entities, 237 relations	Cleaned version of Freebase

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

Dataset	Domain	Size	Features
WN18RR	Lexical / WordNet	40,943 entities, 11 relations	Avoids inverse relation leakage
NELL-995	Web-extracted Knowledge	75,492 entities, 200+ relations	Semi-structured knowledge

These datasets are widely used for benchmarking explainable reasoning models.

EVALUATION METRICS

Evaluation of explainable reasoning involves both **prediction quality** and **explanation quality**:

Prediction Metrics:

- Mean Reciprocal Rank (MRR)
- Hits@k

Explanation Metrics:

- Fidelity: How well explanations align with model predictions
- Plausibility: Human evaluation of explanation quality
- Completeness: Fraction of reasoning process captured in explanation

CHALLENGES

Explainable reasoning faces several challenges:

- **Scalability:** Large KGs make rule extraction or path analysis computationally expensive
- **Incomplete Knowledge:** Missing entities or relations reduce the quality of explanations
- **Trade-off Between Accuracy and Explainability:** Highly interpretable models often achieve lower predictive performance
- **Evaluation of Explanations:** No standardized metrics; human evaluation is time-consuming

EMERGING TRENDS

1. Neuro-Symbolic Reasoning

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

Neuro-symbolic approaches combine neural embeddings with logical reasoning. They enable multi-hop inference with explanations that can be traced to symbolic rules

.

2. Graph Attention Networks (GATs)

Attention mechanisms in graph neural networks provide interpretable scores indicating which neighbors or paths influence predictions.

3. Counterfactual Explanations

By simulating hypothetical scenarios in a KG, models can explain why a fact holds or would change, improving human trust.

APPLICATIONS

- **Healthcare:** Explainable KGs can suggest drug interactions while justifying predictions with known biochemical pathways
- **Finance:** Predicting fraud while tracing transaction paths and related entities
- **Recommendation Systems:** Explaining why a particular item is recommended using paths in user-item KGs
- **Question Answering:** Transparent reasoning helps users understand the steps leading to an answer

DISCUSSION

The current state of explainable reasoning shows promising results, yet challenges remain:

- Hybrid methods offer the best trade-off but require careful design
- Path-based explanations and attention weights improve interpretability but are limited by KG sparsity
- Standard benchmarks for explanation evaluation are needed

Future research should focus on scalable neuro-symbolic models, counterfactual reasoning, and human-centric evaluation of explanations.

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved

CONCLUSION

Explainable reasoning over knowledge graphs is critical for increasing transparency, trust, and adoption of AI systems. While rule-based methods provide clear interpretability, embedding-based models offer better scalability and predictive performance. Hybrid methods, integrating symbolic and neural reasoning, are emerging as the most effective approach. This paper has reviewed key methods, datasets, evaluation metrics, and challenges, highlighting opportunities for future research. Improved explainable reasoning will benefit applications in healthcare, finance, and beyond, enabling AI systems that are not only accurate but also understandable and trustworthy.

REFERENCES

1. Galárraga, L., Teflioudi, C., Hose, K., & Suchanek, F. (2013). AMIE: Association Rule Mining under Incomplete Evidence in Ontological Knowledge Bases. *WWW*.
2. Meilicke, C., Chekol, M. W., Stuckenschmidt, H., & Paulheim, H. (2019). Anytime Bottom-Up Rule Learning for Knowledge Graph Completion. *ISWC*.
3. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating Embeddings for Modeling Multi-relational Data. *NIPS*.
4. Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). A Review of Relational Machine Learning for Knowledge Graphs. *Proceedings of the IEEE*.
5. Xie, R., Liu, Z., Sun, M., & Li, Y. (2016). Representation Learning of Knowledge Graphs with Entity Descriptions. *AAAI*.
6. Dettmers, T., Minervini, P., Stenetorp, P., & Riedel, S. (2018). Convolutional 2D Knowledge Graph Embeddings. *AAAI*.
7. Hu, W., Fey, M., Zitnik, M., Dong, Y., Ren, H., Liu, B., Catasta, M., & Leskovec, J. (2020). Open Graph Benchmark: Datasets for Machine Learning on Graphs. *NeurIPS*.
8. Wang, Z., Zhang, J., Feng, J., & Chen, Z. (2014). Knowledge Graph Embedding by Translating on Hyperplanes. *AAAI*.
9. Sun, Z., Deng, Z.-H., Nie, J.-Y., & Tang, J. (2019). Rotate: Knowledge Graph Embedding by Relational Rotation in Complex Space. *ICLR*.
10. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*.

Page 105-118 © MANTECH PUBLICATIONS 2016. All Rights Reserved