Neural-Symbolic Models for Logical Queries on Knowledge Graphs

Zhaocheng Zhu [1 2] **Mikhail Galkin** [1 3] **Zuobai Zhang** [1 2] **Jian Tang** [1 4 5]

Abstract

Answering complex first-order logic (FOL) queries on knowledge graphs is a fundamental task for multi-hop reasoning. Traditional symbolic methods traverse a complete knowledge graph to extract the answers, which provides good interpretation for each step. Recent neural methods learn geometric embeddings for complex queries. These methods can generalize to incomplete knowledge graphs, but their reasoning process is hard to interpret. In this paper, we propose Graph Neural Network Query Executor (GNN-QE), a neural-symbolic model that enjoys the advantages of both worlds. GNN-QE decomposes a complex FOL query into relation projections and logical operations over fuzzy sets, which provides interpretability for intermediate variables. To reason about the missing links, GNN-QE adapts a graph neural network from knowledge graph completion to execute the relation projections, and models the logical operations with product fuzzy logic. Experiments on 3 datasets show that GNN-QE significantly improves over previous state-of-the-art models in answering FOL queries. Meanwhile, GNN-QE can predict the number of answers without explicit supervision, and provide visualizations for intermediate variables.[1]

1. Introduction

Knowledge graphs (KGs) encapsulate knowledge about the world in a collection of relational edges between entities, and are widely adopted by many domains (Miller, 1998; Vrandečić & Krötzsch, 2014; Himmelstein et al., 2017; Szklarczyk et al., 2019). Reasoning on knowledge graphs has attracted much attention in artificial intelligence, since it can be used to infer new knowledge or answer queries based on existing knowledge. One particular reasoning task we are interested in is answering complex First-Order Logic (FOL) queries on knowledge graphs, which involves logic operations like existential quantifier ($\exists$), conjunction ($\wedge$), disjunction ($\vee$) and negation ($\neg$). For example, the question "*Which universities do the Turing Award winners of deep learning work in?*" can be represented as a FOL query, as showed in Fig. 1.

Traditionally, the problem of reasoning is handled by symbolic approaches, such as logic programming (Lloyd, 2012), fuzzy logic (Klir & Yuan, 1995) or probabilistic reasoning (Pearl, 2014). In the same vein, several algorithms (Dalvi & Suciu, 2007; Schmidt et al., 2010; Zou et al., 2011) have been developed for searching the answers to complex queries on graph databases. These methods traverse a graph and extract all possible assignments for intermediate variables, which provides good interpretation for each step. Besides, symbolic methods are guaranteed to produce the correct answer if all facts are given (Stuart & Peter, 2016). However, many real-world knowledge graphs are known to be incomplete (Nickel et al., 2015), which limits the usage of symbolic methods on knowledge graphs.

Recently, neural methods, such as embedding methods (Bordes et al., 2013; Trouillon et al., 2016; Sun et al., 2018) and graph neural networks (GNNs) (Schlichtkrull et al., 2018; Vashishth et al., 2019; Teru et al., 2020; Zhu et al., 2021), have achieved significant progress in knowledge graph completion. Based on the success of these neural methods, many works have been proposed to solve FOL queries on incomplete graphs by learning an embedding for each FOL query (Hamilton et al., 2018; Ren et al., 2019; Ren & Leskovec, 2020; Chen et al., 2021; Zhang et al., 2021b). Typically, these methods translate the logic operations into neural logic operators in the embedding space. Nevertheless, it is hard to interpret what set of entities an intermediate embedding encodes, leaving the reasoning process unknown to users. The only interpretable method is CQD-Beam (Arakelyan et al., 2021), which applies beam search to a pretrained embedding model in the entity space. However, the complexity of exhaustive search prevents CQD-Beam from being trained directly on complex queries.

[1]Mila - Québec AI Institute [2]Université de Montréal [3]McGill University [4]HEC Montréal [5]CIFAR AI Chair. Correspondence to: Zhaocheng Zhu <zhaocheng.zhu@umontreal.ca>, Jian Tang <jian.tang@hec.ca>.

Proceedings of the 39^{th} International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.

[1]Code is available at https://github.com/DeepGraphLearning/GNN-QE

In this paper, we marry the advantages from both neural and symbolic approaches, and propose Graph Neural Network Query Executor (GNN-QE), a neural-symbolic method for answering FOL queries on incomplete knowledge graphs. Following symbolic methods that output a set of assignments for each intermediate variable, we decompose a complex FOL query into an expression over fuzzy sets (i.e., a continuous relaxation of sets), which attains interpretability for intermediate variables. Each basic operation in the expression is either a relation projection or a logic operation (e.g., conjunction, disjunction and negation). We design the relation projection to be a GNN that predicts the fuzzy set of tail entities given a fuzzy set of head entities and a relation. The logic operations are transformed to the product fuzzy logic operations over fuzzy sets, which satisfy logic laws and enable differentiation of logic operations. We also propose traversal dropout to regularize the model, and batch expression execution to speed up training and inference.

We evaluate our method on 3 standard datasets for FOL queries. Experiments show that GNN-QE achieves new state-of-the-art performance on all datasets, with an average relative gain of 22.3% on existential positive first-order (EPFO) queries and 95.1% on negation queries (Sec. 5.2). By disentangling the contribution of knowledge graph completion and complex query framework, we find that GNN-QE achieves one of the best generalization performances from knowledge graph completion to EPFO queries among different methods. Additionally, the symbolic formulation of our method enables us to predict the number of answers without explicit supervision (Sec. 5.3), and visualize intermediate variables (Sec. 5.4 & App. E). The visualization provided by GNN-QE may help us better understand the reasoning process taken by the model, leading to more interpretable multi-hop reasoning.

2. Related Work

Knowledge Graph Completion Recent years have witnessed a significant progress in reasoning about missing links on a knowledge graph. Notably, embedding methods (Bordes et al., 2013; Yang et al., 2015; Trouillon et al., 2016; Sun et al., 2018; Amin et al., 2020) learn a low-dimensional vector for each entity and relation, which preserves the structure of the knowledge graph. Reinforcement learning methods (Xiong et al., 2017; Das et al., 2018; Hildebrandt et al., 2020; Zhang et al., 2021a) train an agent to collect necessary paths for predicting the link between entities. Rule learning methods (Yang et al., 2017; Sadeghian et al., 2019; Qu et al., 2021) first extract interpretable logic rules from the knowledge graph, and then use the rules to predict the links. Another stream of works adopts graph neural networks (GNNs) to learn the entity representations (Schlichtkrull et al., 2018; Vashishth et al., 2019), or the pairwise representations (Teru et al., 2020; Zhu et al., 2021) for knowledge graph completion. Our method adapts a GNN from knowledge graph completion (Zhu et al., 2021) to implement the relation projection on knowledge graphs. However, GNN-QE is designed to answer complex logical queries, a more challenging task than KG completion.

Complex Logical Query Complex logical query extends knowledge graph completion to predict answer entities for queries with conjunction, disjunction or negation operators. Guu et al. (2015) proposes compositional training for embedding methods to predict answers for path queries. GQE (Hamilton et al., 2018) learns a geometric intersection operator to answer conjunctive queries ($\wedge$) in the embedding space, which is later extended by Query2Box (Ren et al., 2019) to EPFO queries ($\exists$, $\wedge$, $\vee$) and BetaE (Ren & Leskovec, 2020) to FOL queries ($\exists$, $\wedge$, $\vee$, $\neg$). FuzzQE (Chen et al., 2021) improves embedding methods with t-norm fuzzy logic, which satisfies the axiomatic system of classical logic. Some recent works utilize advanced geometric embeddings to achieve desired properties for operators, e.g., hyperboloid embeddings in HypE (Choudhary et al., 2021) and cone embeddings in ConE (Zhang et al., 2021b). Generally, all these methods compute an embedding for the query, and decode the answers with nearest neighbor search or dot product. However, the interpretability of embedding methods is usually compromised, i.e., there is no simple way to understand intermediate reasoning results.

Some other works combine neural methods with symbolic algorithms to solve the complex query answering problem. EmQL (Sun et al., 2020) ensembles an embedding model and a count-min sketch, and is able to find logically entailed answers. CQD (Arakelyan et al., 2021) extends a pretrained knowledge graph embedding model to infer answers for complex queries, with CQD-CO based on continuous optimization and CQD-Beam based on beam search. Our method shares a similar spirit with CQD-Beam in the sense that both models wrap a knowledge graph completion model with symbolic algorithms. However, CQD-Beam cannot be directly trained on complex query due to the complexity incurred by exhaustive search. By contrast, GNN-QE is trained directly on complex queries without pretrained embedding models.

3. Preliminary

In this section, we introduce the background knowledge of FOL queries on knowledge graphs and fuzzy sets.

3.1. First-Order Logic Queries on Knowledge Graphs

Given a set of entities $\mathcal{V}$ and a set of relations $\mathcal{R}$, a knowledge graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ is a collection of triplets $\mathcal{E} = \{(h_i, r_i, t_i)\} \subseteq \mathcal{V} \times \mathcal{R} \times \mathcal{V}$, where each triplet is a fact

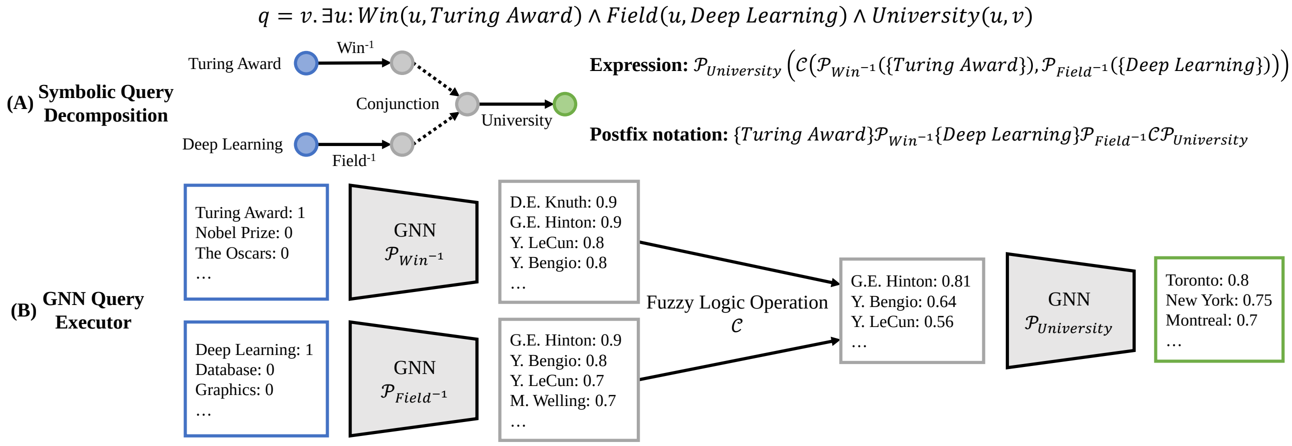


Figure 1: Overview of GNN-QE. **(A)** GNN-QE decomposes a FOL query into an expression of relation projections ($\mathcal{P}$) and logic operations ($\mathcal{C}, \mathcal{D}, \mathcal{N}$). We convert the query into an expression execution problem, where we use the postfix notation to efficiently batch multiple expressions. **(B)** The expression is executed with relation projection learned by GNNs and fuzzy logic operations. All the input, intermediate and output variables are fuzzy sets of entities. Best viewed in color.

from head entity h_i to tail entity t_i with the relation type r_i.

A FOL query on a knowledge graph is a formula composed of constants (denoted with English terms), variables (denoted with a, b, c), relation symbols (denoted with $R(a, b)$) and logic symbols ($\exists$, $\wedge$, $\vee$, $\neg$). In the context of knowledge graphs, each constant or variable is an entity in $\mathcal{V}$. A variable is bounded if it is quantified in the expression, and free otherwise. Each relation symbol $R(a, b)$ is a binary function that indicates whether there is a relation R between a pair of constants or variables. For logic symbols, we consider queries that contain conjunction ($\wedge$), disjunction ($\vee$), negation ($\neg$) and existential quantification ($\exists$)[2]. Fig. 1 illustrates the FOL query for the natural language question "*Which universities do the Turing Award winners of deep learning work in?*". Given a FOL query, the goal is to find answers to the free variables, such that the formula is true.

3.2. Fuzzy Sets and Fuzzy Logic Operations

Fuzzy sets (Klir & Yuan, 1995) are a continuous relaxation of sets whose elements have degrees of membership. A fuzzy set $\mathcal{A} = (\mathcal{U}, x)$ contains a universal set $\mathcal{U}$ and a membership function $x : \mathcal{U} \to [0, 1]$. For each $u \in \mathcal{U}$, the value of $x(u)$ defines the degree of membership (i.e., probability) for u in $\mathcal{A}$. Similar to Boolean logic, fuzzy logic defines three logic operations, AND, OR and NOT, over the real-valued degree of membership. There are several alternative definitions for these operations, such as product fuzzy logic, Gödel fuzzy logic and Łukasiewicz fuzzy logic.

In this paper, fuzzy sets are used to represent the assignments of variables in FOL queries, where the universe $\mathcal{U}$ is always the set of entities $\mathcal{V}$ in the knowledge graph. Since the universe is a finite set, we represent the membership function x as a vector $\boldsymbol{x}$. We use x_u to denote the degree of membership for element u. For simplicity, we abbreviate a fuzzy set $\mathcal{A} = (\mathcal{U}, x)$ as $\boldsymbol{x}$ throughout the paper.

4. Proposed Method

Here we present our model, Graph Neural Network Query Executor (GNN-QE). The high-level idea of GNN-QE is to first decompose a FOL query into an expression of 4 basic operations (relation projection, conjunction, disjunction and negation) over fuzzy sets, then parameterize the relation projection with a GNN adapted from KG completion, and instantiate the logic operations with product fuzzy logic operations. Besides, we introduce traversal dropout to prevent the GNN from converging to a trivial solution, and batched expression execution for speeding up training and inference.

4.1. Symbolic Query Decomposition

Given a FOL query, the first step is to convert it into an expression of basic operations, so that we can retrieve answers by executing the expression. Previous works define basic operations as either relation projections and logic operations over *embeddings* (Ren et al., 2019; Ren & Leskovec, 2020; Chen et al., 2021; Zhang et al., 2021b), or a score function over triplets (Arakelyan et al., 2021). To achieve better interpretability for intermediate variables, we explicitly define 4 basic operations over *fuzzy sets of entities* as follows

- **Relation Projection**: $\mathcal{P}_q(\boldsymbol{x})$ computes the fuzzy set of *tail* entities that are reachable by the input fuzzy set of *head* entities through relation q. $\mathcal{P}_{q^{-1}}(\boldsymbol{x})$ computes the fuzzy set of *head* entities that can reach the input fuzzy

[2]Note universal quantification ($\forall$) is excluded, since none of the entities connects to all entities in a real-world knowledge graph.

set of *tail* entities through relation q.

- **Conjunction**: $\mathcal{C}(\boldsymbol{x}, \boldsymbol{y})$ computes the logical conjunction for each element in $\boldsymbol{x}$ and $\boldsymbol{y}$.
- **Disjunction**: $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$ computes the logical disjunction for each element in $\boldsymbol{x}$ and $\boldsymbol{y}$.
- **Negation**: $\mathcal{N}(\boldsymbol{x})$ computes the logical negation for each element in $\boldsymbol{x}$.

where $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^{\mathcal{V}}$ are two vector representations of fuzzy sets. We then decompose a FOL query into an expression of the above operations. For the example in Fig. 1, the corresponding expression is

$$\mathcal{P}_{\textit{University}}\left(\mathcal{C}\left(\mathcal{P}_{\textit{Win}^{-1}}(\{\textit{Turing Award}\}), \mathcal{P}_{\textit{Field}^{-1}}(\{\textit{Deep Learning}\})\right)\right) \quad (1)$$

where $\{\textit{Turing Award}\}$ and $\{\textit{Deep Learning}\}$ denote singleton sets of *Turing Award* and *Deep Learning*, respectively.

4.2. Neural Relation Projection

In order to solve complex queries on incomplete knowledge graphs, we learn a neural model to perform the relation projection $\boldsymbol{y} = \mathcal{P}_q(\boldsymbol{x})$. Specifically, the neural relation projection model should predict the fuzzy set of tail entities $\boldsymbol{y}$ given the fuzzy set of head entities $\boldsymbol{x}$ and a relation q in the presence of missing links. This is in contrast to the common GNNs (Schlichtkrull et al., 2018; Vashishth et al., 2019) and embedding methods (Bordes et al., 2013; Sun et al., 2018) for knowledge graph completion, which operate on individual entities x and y. While it is possible to apply such GNNs or embedding methods for relation projection, it takes at least $O(|\mathcal{V}|^2 d)$ time to compute them for every $x \in \boldsymbol{x}$ and $y \in \boldsymbol{y}$, which is not scalable.

Recently, Zhu et al. (2021) introduced a new GNN framework for knowledge graph completion, which can predict the set of tail entities $\boldsymbol{y}$ given an entity x and a relation q in $O(|\mathcal{V}|d^2 + |\mathcal{E}|d)$ time. Inspired by such a framework, we propose a scalable GNN solution for relation projection.

Graph Neural Networks. Our goal is to design a GNN model that predicts a fuzzy set of tail entities given a fuzzy set of head entities and a relation. A special case of the input is a singleton set, where we need to model the probability $p_q(y|x)$ for every $y \in \boldsymbol{y}$. Such a problem can be solved by GNNs in a single-source fashion (You et al., 2021; Zhu et al., 2021). For example, the recent work NBFNet (Zhu et al., 2021) derives a GNN framework based on the generalized Bellman-Ford algorithm for single-source problems on graphs. Given a head entity u and a projection relation q, we use the following iteration to compute a representation $\boldsymbol{h}_v$ for each entity $v \in \mathcal{V}$ w.r.t. the source entity u:

$$\boldsymbol{h}_v^{(0)} \leftarrow \textsc{Indicator}(u, v, q) \quad (2)$$

$$\boldsymbol{h}_v^{(t)} \leftarrow \textsc{Aggregate}(\{\textsc{Message}(\boldsymbol{h}_z^{(t-1)}, (z, r, v)) \\ |(z, r, v) \in \mathcal{E}(v)\}) \quad (3)$$

where the Indicator function initializes a relation embedding $\boldsymbol{q}$ on entity v if v equals to u and a zero embedding otherwise, and $\mathcal{E}(v)$ is the set of edges going into v. The Message and Aggregate functions can be instantiated with any neural function from popular GNNs. To apply the above framework to a fuzzy set $\boldsymbol{x}$ of head entities, we propose to replace Eqn. 2 with the following initialization

$$\boldsymbol{h}_v^{(0)} \leftarrow x_v \boldsymbol{q} \quad (4)$$

where x_v is the probability of entity v in $\boldsymbol{x}$. Intuitively, this GNN model initializes an embedding $\boldsymbol{q}$ for the projection relation q on all entities, where the scale of the initialization on an entity depends on its probability in the fuzzy set. The original Indicator function can be viewed as a special case of Eqn. 4, with the fuzzy set being a singleton set.

For the Aggregate and the Message functions, we follow the design in NBFNet (Zhu et al., 2021) and parameterize the Message function as

$$\textsc{Message}(\boldsymbol{h}_z^{(t-1)}, (z, r, v)) = \boldsymbol{h}_z^{(t-1)} \odot (\boldsymbol{W}_r \boldsymbol{q} + \boldsymbol{b}_r) \quad (5)$$

where $\boldsymbol{W}_r^{(t)}$ and $\boldsymbol{b}_r^{(t)}$ are the weight matrix and bias vector for relation r in the t-th iteration respectively, and $\odot$ is the element-wise multiplication operator. The Aggregate function is parameterized as the principal neighborhood aggregation (PNA) (Corso et al., 2020). Our GNN has the same time complexity as NBFNet, and therefore takes $O(|\mathcal{V}|d^2 + |\mathcal{E}|d)$ time for each message passing iteration. Note it is possible to parameterize the framework with other GNN models, such as RGCN (Schlichtkrull et al., 2018) or CompGCN (Vashishth et al., 2019). See Sec. 5.5 for experiments with different GNN models.

To apply the GNN framework for relation projection, we propagate the representations with Eqn. 3 for T layers. Then we take the representations in the last layer, and pass them into a multi-layer perceptron (MLP) f followed by a sigmoid function σ to predict the fuzzy set of tail entities.

$$\mathcal{P}_q(\boldsymbol{x}) = \sigma(f(\boldsymbol{h}^{(T)})) \quad (6)$$

4.3. Fuzzy Logic Operations

The logic operations (i.e., $\mathcal{C}(\boldsymbol{x}, \boldsymbol{y})$, $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$, $\mathcal{N}(\boldsymbol{x})$) glue multiple relation projection results and generate the input fuzzy set for the next relation projection. Ideally, they should satisfy certain logic laws, such as commutativity, associativity and non-contradiction. Most previous works (Hamilton et al., 2018; Ren et al., 2019; Ren & Leskovec, 2020; Zhang et al., 2021b) propose dedicated geometric operations to learn these logic operations in the embedding space. Nevertheless, these neural operators are not guaranteed to satisfy most logic laws, which may introduce additional error when they are chained together.

Here we model the conjunction, disjunction and negation with product fuzzy logic operations. Given two fuzzy sets $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^{\mathcal{V}}$, the operations are defined as follows

$$\mathcal{C}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x} \odot \boldsymbol{y} \tag{7}$$

$$\mathcal{D}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x} + \boldsymbol{y} - \boldsymbol{x} \odot \boldsymbol{y} \tag{8}$$

$$\mathcal{N}(\boldsymbol{x}) = \mathbf{1} - \boldsymbol{x} \tag{9}$$

where $\odot$ is the element-wise multiplication and $\mathbf{1}$ is a vector of all ones (i.e., the universe). Compared to geometric operations in previous works, such fuzzy logic operations satisfy many logic laws, e.g., De Morgan's laws $\mathcal{N}(\mathcal{C}(\boldsymbol{x}, \boldsymbol{y})) = \mathcal{D}(\mathcal{N}(\boldsymbol{x}), \mathcal{N}(\boldsymbol{y}))$, $\mathcal{N}(\mathcal{D}(\boldsymbol{x}, \boldsymbol{y})) = \mathcal{C}(\mathcal{N}(\boldsymbol{x}), \mathcal{N}(\boldsymbol{y}))$. Note FuzzQE (Chen et al., 2021) also adopts fuzzy logic operations and satisfies logic laws. However, FuzzQE applies fuzzy logic operations to *embeddings*. By contrast, our GNN-QE applies fuzzy logic operations to *fuzzy sets of entities*, which provides better interpretability (See Sec. 5.4).

4.4. Learning

Following previous works (Ren et al., 2019; Ren & Leskovec, 2020; Zhang et al., 2021b), we train our model to minimize the binary cross entropy loss.

$$\begin{aligned}\mathcal{L} = &-\frac{1}{|\mathcal{A}_Q|}\sum_{a \in \mathcal{A}_Q} \log p(a|Q) \\ &-\frac{1}{|\mathcal{V}\backslash\mathcal{A}_Q|}\sum_{a' \in \mathcal{V}\backslash\mathcal{A}_Q} \log(1 - p(a'|Q))\end{aligned} \tag{10}$$

where $\mathcal{A}_Q$ is the set of answers to the complex query Q and $p(a|Q)$ is the probability of entity a in the final output fuzzy set. Since GNN-QE always outputs the probability for all entities (Eqn. 6), we do not perform negative sampling and compute the loss with all negative answers.

Traversal Dropout. One challenge in training GNN-QE is to let the model generalize to incomplete KGs at test time. This is because all the training queries are generated by assuming the training graph is complete (Ren & Leskovec, 2020). In other words, all the training queries can be perfectly solved by a simple relation traversal model on the training graph, without modeling any missing link. GNN models can easily discover this mode, which does not generalize to incomplete knowledge graphs at test time.

To solve this issue, we introduce traversal dropout to create an incomplete KG at training time. Specifically, we first run a relation traversal model to extract all the edges corresponding to the query. We then randomly mask out the traversed edges in each relation projection with probability p. Intuitively, the probability p trades off between a simple relation traversal model and a full reasoning model. If p is small, the GNN model may converge to a trivial relation traversal model, otherwise it is forced to encode non-trivial reasoning features. Since some of the edges in the test queries may be present in the KG, it is not always optimal to use a large p to discourage a relation traversal model. In practice, we treat p as a hyperparameter, and tune it based on the performance on the validation set. See Sec. 5.5 for experiments with different values of p.

Batched Expression Execution[3]. Modern machine learning relies on batch processing on GPUs to accelerate the computation of neural (or even symbolic) models. However, it is challenging to batch the expressions of FOL queries, since different query structures require different recursive computation steps. Previous works (Hamilton et al., 2018; Ren et al., 2019; Ren & Leskovec, 2020) divide a batch based on the query structure of each sample, and only batch the computation of samples that have the same structure. However, such an implementation needs to enumerate every query structure, and is not scalable when the vocabulary of query structures grows large.

To solve this issue, we need to find a way to execute the expressions without recursion. This can be achieved by converting the expressions into postfix notation. The postfix notation, a.k.a. reverse Polish notation (Lukasiewicz, 1951), writes operators *after* their operands in an expression. For example, the postfix expression of Eqn. 1 is

$$\{\textit{Turing Award}\}\mathcal{P}_{\textit{Win}^{-1}}\{\textit{Deep Learning}\}\mathcal{P}_{\textit{Field}^{-1}}\mathcal{C}\mathcal{P}_{\textit{University}} \tag{11}$$

The advantage of postfix expressions is that they are unambiguous without parentheses, and therefore can be executed easily without recursion. To execute a postfix expression, we allocate a stack and scan the expression from left to right. When we encounter an operand, we push it into the stack. When we encounter an operator, we pop the corresponding number of operands from the stack, apply the operation and push the result into the stack. Such an algorithm can be easily batched for the same operator even in samples of different query types. Examples and pseudo code for batched expression execution are provided in App. C.

5. Experiments

In this section, we evaluate GNN-QE by answering FOL queries on 3 standard datasets. Our experiments demonstrate that: (1) GNN-QE outperforms existing methods on both EPFO queries and queries with negation. (2) GNN-QE can predict the number of answers out-of-the-box without any explicit supervision. (3) We can visualize the intermediate variables of GNN-QE and interpret its reasoning process.

[3]Expression execution is formally known as expression evaluation in computer science. In this paper, we use the term "expression execution" to avoid ambiguity in machine learning contexts.

5.1. Experiment Setup

We evaluate our method on FB15k (Bordes et al., 2013), FB15k-237 (Toutanova & Chen, 2015) and NELL995 (Xiong et al., 2017) knowledge graphs. To make a fair comparison with baselines, we use the standard train, validation and test FOL queries generated by the BetaE paper (Ren & Leskovec, 2020), which consist of 9 EPFO query types and 5 query types with negation. We follow previous works (Ren & Leskovec, 2020; Chen et al., 2021; Zhang et al., 2021b) and train our model with 10 query types (*1p/2p/3p/2i/3i/2in/3in/inp/pni/pin*). The model is evaluated on 10 training query types, plus 4 query types (*ip/pi/2u/up*) that have never been seen during training. A full list of query types and their statistics is provided in App. A.

Evaluation Protocol. Following the evaluation protocol in (Ren et al., 2019), we separate the answers to each query into two sets: easy answers and hard answers. For test (validation) queries, easy answers are the entities that can be reached on the validation (train) graph via a symbolic relation traverse model. Hard answers are those that can only be reached with predicted links. In other words, the model must perform reasoning to get the hard answers. We compute the ranking of each hard answer against all non-answer entities. The performance is measured by mean reciprocal rank (MRR) and HITS at K (H@K) metrics.

Implementation Details. Our work is implemented based on the open-source codebase of GNNs for KG completion[4]. Following (Zhu et al., 2021), we augment each triplet with a flipped one of its inverse relation, so that the GNN can propagate information in both directions. The neural relation projection model is set to a 4-layer GNN model. We train the model with the self-adversarial negative sampling (Sun et al., 2018). Note we only instantiate 1 GNN model and share it across all neural relation projections in the query. For query types that contain multiple relation projections in a chain (*2p/3p/inp/pni/pin*), we observe very noisy gradients for the relation projections early in the chain. Therefore, we zero out the gradients of those relation projections, and only update the GNN with gradients from the last relation projections close to the loss. Our model is trained with Adam optimizer (Kingma & Ba, 2014) on 4 Tesla V100 GPUs. Hyperparameters of GNN-QE are given in App. B.

Baselines. We compare GNN-QE against both embedding methods and neural-symbolic methods. The embedding methods include GQE (Hamilton et al., 2018), Q2B (Ren et al., 2019), BetaE (Ren & Leskovec, 2020), FuzzQE (Chen et al., 2021) and ConE (Zhang et al., 2021b). The neural-symbolic methods include CQD-CO (Arakelyan et al., 2021) and CQD-Beam (Arakelyan et al., 2021). For CQD-CO and CQD-Beam, we obtain their performance using the codebase[5] provided by the original authors.

5.2. Complex Query Answering

Tab. 1 shows the MRR results of different models for answering FOL queries. GQE, Q2B, CQD-CO and CQD-Beam do not support queries with negation, so the corresponding entries are empty. We observe that GNN-QE achieves the best result for both EPFO queries and queries with negation on all 3 datasets. Notably, GNN-QE achieves an average relative gain of 22.3% in avg_p and 95.1% in avg_n compared to previous best model ConE. We attribute this gain to the advantage of fuzzy sets over geometric embeddings. Fuzzy sets can easily model intermediate variables with many possible assignments, while it is hard to embed a large number of entities in a low-dimensional vector. Such an advantage is especially useful for negation operations, since the output of a negation operation usually contains nearly $|\mathcal{V}|$ entities.

Intuitively, the performance of complex query models should benefit from better KG completion performance, i.e., *1p* queries. Here we disentangle the contribution of KG completion and complex query framework in answering EPFO queries. Fig. 2 plots the performance of EPFO queries w.r.t. the performance of KG completion on all datasets. Methods on the top-left corner of each plot show a better generalization from KG completion to EPFO queries, which implies their complex query frameworks are better. These include GQE, BetaE, FuzzQE, ConE and GNN-QE. By contrast, CQD-CO and CQD-Beam generalize worse than other methods, because they rely on a pretrained embedding model and cannot be trained for complex queries.

5.3. Answer Set Cardinality Prediction

One advantage of GNN-QE is that it can predict the cardinality of the answer set (i.e., the number of answers) without explicit supervision. Specifically, the cardinality of a fuzzy set is computed as the sum of entity probabilities exceeding a certain threshold. We use 0.5 for the threshold as it is a natural choice for our binary classification loss (Eqn. 10). Tab. 2 shows the mean absolute percentage error (MAPE) between our model prediction and the ground truth. Note none of existing methods can predict the number of answers without explicit supervision. Ren & Leskovec (2020) and Zhang et al. (2021b) observe that the uncertainty of Q2B, BetaE and ConE are positively correlated with the number of answers. We follow their setting and report the Spearman's rank correlation between our model prediction and the ground truth. As showed in Tab. 3, GNN-QE outperforms existing methods by a large margin on all query types.

[4] https://github.com/DeepGraphLearning/NBFNet

[5] https://github.com/pminervini/KGReasoning

Table 1: Test MRR results (%) on answering FOL queries. avg_p is the average MRR on EPFO queries ($\wedge$, $\vee$). avg_n is the average MRR on queries with negation. Results of GQE and Q2B are taken from (Ren & Leskovec, 2020). Results of BetaE, FuzzQE and ConE are taken from their original papers. Results of other metrics can be found in App. D.

Model	avg_p	avg_n	1p	2p	3p	2i	3i	pi	ip	2u	up	2in	3in	inp	pin	pni
FB15k																
GQE	28.0	-	54.6	15.3	10.8	39.7	51.4	27.6	19.1	22.1	11.6	-	-	-	-	-
Q2B	38.0	-	68.0	21.0	14.2	55.1	66.5	39.4	26.1	35.1	16.7	-	-	-	-	-
BetaE	41.6	11.8	65.1	25.7	24.7	55.8	66.5	43.9	28.1	40.1	25.2	14.3	14.7	11.5	6.5	12.4
CQD-CO	46.9	-	**89.2**	25.3	13.4	74.4	78.3	44.1	33.2	41.8	21.9	-	-	-	-	-
CQD-Beam	58.2	-	**89.2**	54.3	28.6	74.4	78.3	58.2	67.7	42.4	30.9	-	-	-	-	-
ConE	49.8	14.8	73.3	33.8	29.2	64.4	73.7	50.9	35.7	55.7	31.4	17.9	18.7	12.5	9.8	15.1
GNN-QE	**72.8**	**38.6**	88.5	**69.3**	**58.7**	**79.7**	**83.5**	**69.9**	**70.4**	**74.1**	**61.0**	**44.7**	**41.7**	**42.0**	**30.1**	**34.3**
FB15k-237																
GQE	16.3	-	35.0	7.2	5.3	23.3	34.6	16.5	10.7	8.2	5.7	-	-	-	-	-
Q2B	20.1	-	40.6	9.4	6.8	29.5	42.3	21.2	12.6	11.3	7.6	-	-	-	-	-
BetaE	20.9	5.5	39.0	10.9	10.0	28.8	42.5	22.4	12.6	12.4	9.7	5.1	7.9	7.4	3.5	3.4
CQD-CO	21.8	-	**46.7**	9.5	6.3	31.2	40.6	23.6	16.0	14.5	8.2	-	-	-	-	-
CQD-Beam	22.3	-	**46.7**	11.6	8.0	31.2	40.6	21.2	18.7	14.6	8.4	-	-	-	-	-
FuzzQE	24.0	7.8	42.8	12.9	10.3	33.3	46.9	26.9	17.8	14.6	10.3	8.5	11.6	7.8	5.2	5.8
ConE	23.4	5.9	41.8	12.8	11.0	32.6	47.3	25.5	14.0	14.5	10.8	5.4	8.6	7.8	4.0	3.6
GNN-QE	**26.8**	**10.2**	42.8	**14.7**	**11.8**	**38.3**	**54.1**	**31.1**	**18.9**	**16.2**	**13.4**	**10.0**	**16.8**	**9.3**	**7.2**	**7.8**
NELL995																
GQE	18.6	-	32.8	11.9	9.6	27.5	35.2	18.4	14.4	8.5	8.8	-	-	-	-	-
Q2B	22.9	-	42.2	14.0	11.2	33.3	44.5	22.4	16.8	11.3	10.3	-	-	-	-	-
BetaE	24.6	5.9	53.0	13.0	11.4	37.6	47.5	24.1	14.3	12.2	8.5	5.1	7.8	10.0	3.1	3.5
CQD-CO	**28.8**	-	**60.4**	17.8	12.7	39.3	46.6	30.1	22.0	17.3	**13.2**	-	-	-	-	-
CQD-Beam	28.6	-	**60.4**	**20.6**	11.6	39.3	46.6	25.4	**23.9**	**17.5**	12.2	-	-	-	-	-
FuzzQE	27.0	7.8	47.4	17.2	14.6	39.5	49.2	26.2	20.6	15.3	12.6	7.8	9.8	11.1	4.9	5.5
ConE	27.2	6.4	53.1	16.1	13.9	40.0	50.8	26.3	17.5	15.3	11.3	5.7	8.1	10.8	3.5	3.9
GNN-QE	**28.9**	**9.7**	53.3	18.9	**14.9**	**42.4**	**52.5**	**30.8**	18.9	15.9	12.6	**9.9**	**14.6**	**11.4**	**6.3**	**6.3**

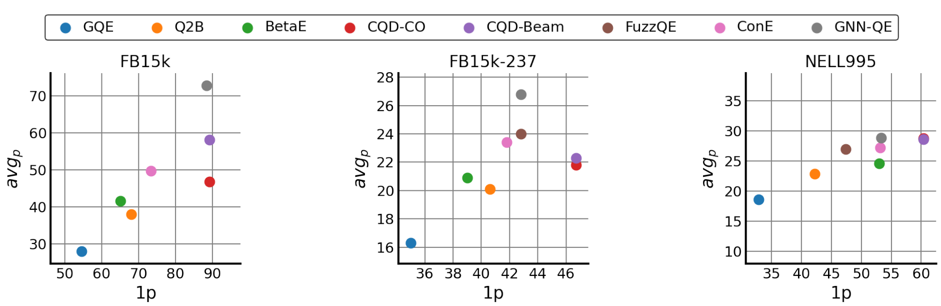


Figure 2: MRR results on EPFO queries w.r.t. MRR results on knowledge graph completion (1p queries). Methods on the top left boundary of each plot generalize better from knowledge graph completion to EPFO queries. Best viewed in color.

5.4. Intermediate Variables Visualization

Another advantage of GNN-QE is that we can interpret its reasoning process by investigating the intermediate variables. As the intermediate fuzzy sets may contain hundreds of entities, we consider two kinds of visualization to qualitatively analyze the precision and the recall of our model. The first one examines the entities with the top probabilities in each fuzzy set, and checks if they are an easy entity (i.e., those can be traversed on the training graph), a hard entity (i.e., those require reasoning) or a false positive one. For each fuzzy set, we visualize the top-3 easy entities and top-6 hard entities that have a minimum probability of 0.1. The second one draws a random ground truth assignment for each variable, such that the assignments form a valid grounding of the query and lead to a hard answer. We report the filtered ranking for each entity in the grounding.

Table 2: MAPE (%) of the number of answers predicted by GNN-QE. avg is the average on all query types.

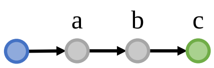
Dataset	avg	1p	2p	3p	2i	3i	pi	ip	2u	up	2in	3in	inp	pin	pni
FB15k	37.1	34.4	29.7	34.7	39.1	57.3	47.8	34.6	13.5	26.5	31.4	50.3	50.3	39.4	29.8
FB15k-237	38.9	40.9	23.6	27.4	34.8	53.4	39.9	60.0	27.8	20.3	40.3	52.6	49.6	44.8	29.0
NELL995	44.0	61.9	38.2	47.1	56.6	72.3	49.5	45.8	19.9	36.2	30.0	47.0	42.3	39.8	29.4

Table 3: Spearman's rank correlation between the model prediction and the number of ground truth answers on FB15k-237. avg is the average correlation on all 12 query types in the table. Results of baselines are taken from (Zhang et al., 2021b). Results on FB15k and NELL can be found in Tab. 9 in Appendix.

Model	avg	1p	2p	3p	2i	3i	pi	ip	2in	3in	inp	pin	pni
Q2B	-	0.184	0.226	0.269	0.347	0.436	0.361	0.199	-	-	-	-	-
BetaE	0.540	0.396	0.503	0.569	0.598	0.516	0.540	0.439	0.685	0.579	0.511	0.468	0.671
ConE	0.738	0.70	0.71	0.74	0.82	0.72	0.70	0.62	0.90	0.83	0.66	0.57	0.88
GNN-QE	**0.940**	**0.948**	**0.951**	**0.895**	**0.992**	**0.970**	**0.911**	**0.937**	**0.981**	**0.968**	**0.864**	**0.880**	**0.987**

Table 4: Visualization of a *3p* query from FB15k-237 test set. More visualizations can be found in App. E.

What team plays the sport that is in the Olympics that Greece participated in?



3p

Query $q = ?c : \exists a, b : \text{ParticipateCountry}(a, \text{Greece}) \wedge \text{OlympicSports}(a, b) \wedge \text{TeamSports}(c, b)$

Variable	Top Predictions (≥ 0.1) Easy	Hard		Random Ground Truth	Filtered Ranking
a	1936 Summer Olympics 1980 Winter Olympics 2002 Winter Olympics	2010 Winter Olympics ✓ 2012 Summer Olympics ✗ 1920 Summer Olympics ✗	1988 Summer Olympics ✗ 1928 Summer Olympics ✗ 1992 Summer Olympics ✗	2010 Winter Olympics (hard)	1
b	soccer track and field water polo	luge ✓ ice hockey ✓ short track speed skating ✓	tennis ✗ - -	ice hockey (hard)	1
c	Sacramento Kings Utah Jazz Seattle SuperSonics	Algeria soccer team ✓ Cincinnati Reds ✓ Washington Nationals ✓	Chile soccer team ✓ Cardiff City ✓ Blackburn Rovers ✓	Florida Panthers (hard)	433

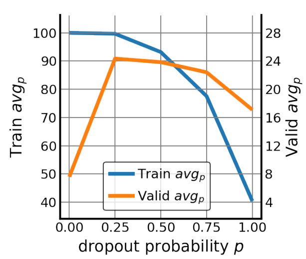


Figure 3: Average MRR on EPFO queries (%) of train / validation sets w.r.t. traversal dropout probability p. The best validation performance is achieved with $p = 0.25$.

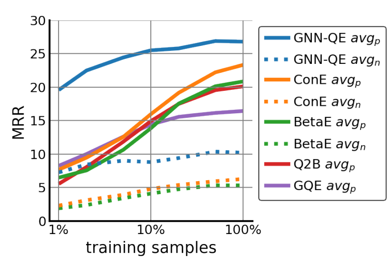


Figure 4: Test MRR results w.r.t. number of training samples. GNN-QE is not only better than embeddings, but also less sensitive to the number of training samples.

Tab. 4 shows the visualization of GNN-QE on a *3p* query from FB15k-237 test set. Among the top hard entities, GNN-QE correctly predicts most of the intermediate entities, which indicates our method has a good precision for this sample. For the random ground truth assignments, GNN-QE recalls the first two hops (2010 Winter Olympics & ice hockey) perfectly, but fails for the last hop. Such analysis would be beneficial to identify the steps where error occurs.

5.5. Ablation Study

To provide a more comprehensive understanding of GNN-QE, we conduct three ablation studies on FB15k-237.

Traversal Dropout Probability p. Fig. 3 shows the average MRR on EPFO queries of train and validation sets w.r.t. different probability p. The model can achieve a perfect training MRR of 1 when $p = 0$, which suggests that the model is able to learn the behavior of a relation traversal model. However, a relation traversal model cannot solve queries on incomplete graphs, which is revealed by its low performance on the validation set. With a non-zero probability p, traversal dropout makes the training problem more difficult, and enforces the model to learn a reasoning model that predicts the dropped link from its surrounding graph structure. However, it is not optimal to learn a fully reasoning model with $p = 1$, since it cannot perform relation traversal and some links in the validation queries can be perfectly solved by a relation traversal model.

Performance w.r.t. Number of Training Samples. Fig. 4 plots the MRR curves of different query types in GNN-QE and BetaE under different number of training samples. It is observed that the performance of GNN-QE is not only better than BetaE, but also less sensitive to the number of training samples. Even with 1% training samples (i.e., only 8,233 training queries for FB15k-237), GNN-QE achieves a comparative avg_p and better avg_n compared with BetaE trained with the full dataset. We conjecture the reason is that BetaE needs to learn a separate embedding for each entity, while our neural-symbolic method only learns relation embeddings (Eqn. 5) for relation projection, which requires less samples to converge.

GNN Parameterization. Tab. 5 shows the MRR results of GNN-QE w.r.t. different GNN parameterizations. We consider three parameterizations for the MESSAGE and AGGREGATE functions in Eqn. 3, namely RGCN (Schlichtkrull et al., 2018), CompGCN (Vashishth et al., 2019) and NBFNet (Zhu et al., 2021). It is observed that all three parameterizations outperform BetaE with significant improvement on avg_n, which suggests the advantages of fuzzy sets in modeling negation queries. Besides, GNN-QE benefits from stronger GNN models (NBFNet $>$ CompGCN $>$ RGCN). The performance of GNN-QE might be further improved with better GNN models.

Table 5: Test MRR results (%) w.r.t. GNN models. GNN-QE benefits from better GNN models.

Model	**avg$_p$**	**avg$_n$**
BetaE	20.9	5.5
GNN-QE (RGCN)	20.9	7.3
GNN-QE (CompGCN)	22.5	7.3
GNN-QE (NBFNet)	**26.8**	**10.2**

6. Conclusion

In this paper, we present a novel neural-symbolic model, namely Graph Neural Network Query Executor (GNN-QE), for answering complex FOL queries on incomplete knowledge graphs. Our method decomposes complex queries into an expression of basic operations over fuzzy sets, and executes the expression with a learned GNN relation projection model and fuzzy logic operations. GNN-QE not only significantly outperforms previous state-of-the-art models on 3 datasets, but also provides interpretability for intermediate variables. Besides, GNN-QE can predict the number of answers without explicit supervision. Future works include combining GNN-QE with a parser to answer logical queries in the natural language form, and scaling up GNN-QE to large-scale knowledge graphs with millions of entities.

Acknowledgements

This project is supported by the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant, the Canada CIFAR AI Chair Program, collaboration grants between Microsoft Research and Mila, Samsung Electronics Co., Ltd., Amazon Faculty Research Award, Tencent AI Lab Rhino-Bird Gift Fund and a NRC Collaborative R&D Project (AI4D-CORE-06). This project was also partially funded by IVADO Fundamental Research Project grant PRF-2019-3583139727. The computation resource of this project is supported by Calcul Québec[6] and Compute Canada[7].

We would like to thank Meng Qu for discussion on research ideas, and Pasquale Minervini for discussion on the CQD model. We also appreciate all anonymous reviewers for their constructive suggestions on this paper.

References

Amin, S., Varanasi, S., Dunfield, K. A., and Neumann, G. Lowfer: Low-rank bilinear pooling for link prediction. In *International Conference on Machine Learning*, pp. 257–268. PMLR, 2020.

Arakelyan, E., Daza, D., Minervini, P., and Cochez, M. Complex query answering with neural link predictors. In *International Conference on Learning Representations*, 2021.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26, 2013.

Chen, X., Hu, Z., and Sun, Y. Fuzzy logic based logical query answering on knowledge graph. In *International Conference on Machine Learning*. PMLR, 2021.

Choudhary, N., Rao, N., Katariya, S., Subbian, K., and Reddy, C. K. Self-supervised hyperboloid representations from logical queries over knowledge graphs. In *Proceedings of the Web Conference 2021*, pp. 1373–1384, 2021.

Corso, G., Cavalleri, L., Beaini, D., Liò, P., and Veličković, P. Principal neighbourhood aggregation for graph nets. volume 33, 2020.

Dalvi, N. and Suciu, D. Efficient query evaluation on probabilistic databases. *The VLDB Journal*, 16(4):523–544, 2007.

Das, R., Dhuliawala, S., Zaheer, M., Vilnis, L., Durugkar, I., Krishnamurthy, A., Smola, A., and McCallum, A. Go

[6] https://www.calculquebec.ca/

[7] https://www.computecanada.ca/

for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. In *International Conference on Learning Representations*, 2018.

Guu, K., Miller, J., and Liang, P. Traversing knowledge graphs in vector space. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 318–327, 2015.

Hamilton, W., Bajaj, P., Zitnik, M., Jurafsky, D., and Leskovec, J. Embedding logical queries on knowledge graphs. *Advances in Neural Information Processing Systems*, 31, 2018.

Hildebrandt, M., Serna, J. A. Q., Ma, Y., Ringsquandl, M., Joblin, M., and Tresp, V. Reasoning on knowledge graphs with debate dynamics. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 4123–4131, 2020.

Himmelstein, D. S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S. L., Hadley, D., Green, A., Khankhanian, P., and Baranzini, S. E. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *Elife*, 6: e26726, 2017.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Klir, G. and Yuan, B. *Fuzzy sets and fuzzy logic*, volume 4. Prentice hall New Jersey, 1995.

Lloyd, J. W. *Foundations of logic programming*. Springer Science & Business Media, 2012.

Lukasiewicz, J. Aristotle's syllogistic from the standpoint of modern formal logic. 1951.

Miller, G. A. *WordNet: An electronic lexical database*. MIT press, 1998.

Nickel, M., Murphy, K., Tresp, V., and Gabrilovich, E. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, 2015.

Pearl, J. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

Qu, M., Chen, J., Xhonneux, L.-P., Bengio, Y., and Tang, J. Rnnlogic: Learning logic rules for reasoning on knowledge graphs. In *International Conference on Learning Representations*, 2021.

Ren, H. and Leskovec, J. Beta embeddings for multi-hop logical reasoning in knowledge graphs. *Advances in Neural Information Processing Systems*, 33, 2020.

Ren, H., Hu, W., and Leskovec, J. Query2box: Reasoning over knowledge graphs in vector space using box embeddings. In *International Conference on Learning Representations*, 2019.

Sadeghian, A., Armandpour, M., Ding, P., and Wang, D. Z. Drum: End-to-end differentiable rule mining on knowledge graphs. volume 32, pp. 15347–15357, 2019.

Schlichtkrull, M., Kipf, T. N., Bloem, P., Van Den Berg, R., Titov, I., and Welling, M. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pp. 593–607. Springer, 2018.

Schmidt, M., Meier, M., and Lausen, G. Foundations of sparql query optimization. In *Proceedings of the 13th International Conference on Database Theory*, pp. 4–33, 2010.

Stuart, R. and Peter, N. Artificial intelligence-a modern approach 3rd ed, 2016.

Sun, H., Arnold, A., Bedrax Weiss, T., Pereira, F., and Cohen, W. W. Faithful embeddings for knowledge base queries. *Advances in Neural Information Processing Systems*, 33, 2020.

Sun, Z., Deng, Z.-H., Nie, J.-Y., and Tang, J. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*, 2018.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., et al. String v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, 47(D1):D607–D613, 2019.

Teru, K., Denis, E., and Hamilton, W. Inductive relation prediction by subgraph reasoning. In *International Conference on Machine Learning*, pp. 9448–9457. PMLR, 2020.

Toutanova, K. and Chen, D. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd workshop on continuous vector space models and their compositionality*, pp. 57–66, 2015.

Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., and Bouchard, G. Complex embeddings for simple link prediction. In *International conference on machine learning*, pp. 2071–2080. PMLR, 2016.

Vashishth, S., Sanyal, S., Nitin, V., and Talukdar, P. Composition-based multi-relational graph convolutional networks. In *International Conference on Learning Representations*, 2019.

Vrandečić, D. and Krötzsch, M. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57 (10):78–85, 2014.

Xiong, W., Hoang, T., and Wang, W. Y. Deeppath: A reinforcement learning method for knowledge graph reasoning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, Copenhagen, Denmark, September 2017. ACL.

Yang, B., Yih, W.-t., He, X., Gao, J., and Deng, L. Embedding entities and relations for learning and inference in knowledge bases. *International Conference on Learning Representations*, 2015.

Yang, F., Yang, Z., and Cohen, W. W. Differentiable learning of logical rules for knowledge base reasoning. In *Advances in Neural Information Processing Systems*, pp. 2316–2325, 2017.

You, J., Gomes-Selman, J., Ying, R., and Leskovec, J. Identity-aware graph neural networks. *arXiv preprint arXiv:2101.10320*, 2021.

Zhang, D., Yuan, Z., Liu, H., Lin, X., and Xiong, H. Learning to walk with dual agents for knowledge graph reasoning. *arXiv preprint arXiv:2112.12876*, 2021a.

Zhang, Z., Wang, J., Chen, J., Ji, S., and Wu, F. Cone: Cone embeddings for multi-hop reasoning over knowledge graphs. *Advances in Neural Information Processing Systems*, 34, 2021b.

Zhu, Z., Zhang, Z., Xhonneux, L.-P., and Tang, J. Neural bellman-ford networks: A general graph neural network framework for link prediction. *arXiv preprint arXiv:2106.06935*, 2021.

Zou, L., Mo, J., Chen, L., Özsu, M. T., and Zhao, D. gstore: answering sparql queries via subgraph matching. *Proceedings of the VLDB Endowment*, 4(8):482–493, 2011.

A. Dataset Statistics

We use the complex query datasets generated by (Ren & Leskovec, 2020). There is a total number of 14 query types, as showed in Fig. 5. Statistics of all query types is summarized in Tab. 6.

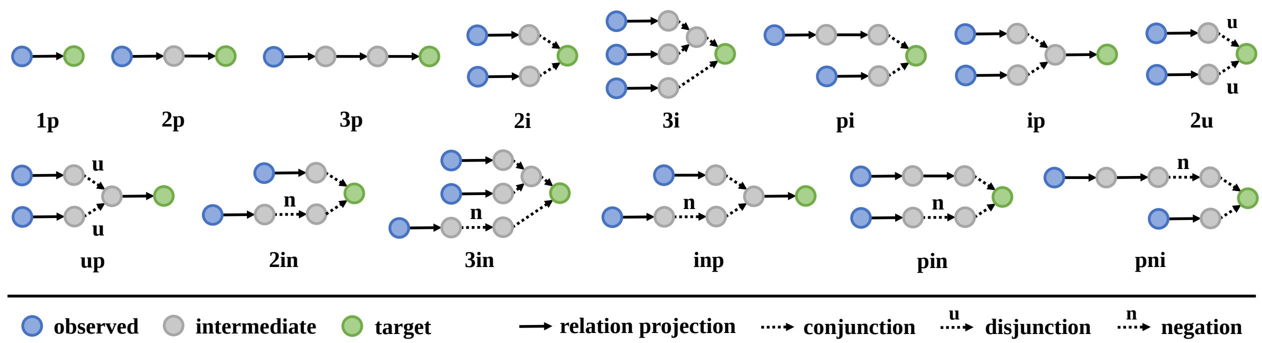


Figure 5: Types of complex FOL queries used in training and inference.

Table 6: Statistics of different query types used in the benchmark datasets.

Split	Query Type	FB15k	FB15k-237	NELL995
Train	1p/2p/3p/2i/3i 2in/3in/inp/pin/pni	273,710 27,371	149,689 14,968	107,982 10,798
Valid	1p Others	59,078 8,000	20,094 5,000	16,910 4,000
Test	1p Others	66,990 8,000	22,804 5,000	17,021 4,000

B. Hyperparameters

Tab. 7 lists the hyperparameter configurations of GNN-QE on different datasets.

Table 7: Hyperparameters of GNN-QE on different datasets. All the hyperparameters are selected by the performance on the validation set.

Hyperparameter		FB15k	FB15k-237	NELL-995
GNN	#layer hidden dim.	4 32	4 32	4 32
MLP	#layer hidden dim.	2 64	2 64	2 64
Traversal Dropout	probability	0.25	0.25	0.25
Learning	batch size sample weight optimizer learning rate iterations (#batch) adv. temperature	192 uniform across queries Adam 5e-3 10,000 0.2	192 uniform across queries Adam 5e-3 10,000 0.2	32 uniform across answers Adam 5e-3 30,000 0.2

C. Batched Expression Execution

Alg. 1 shows the pseudo code for converting expression to postfix notation. The idea is to recursively parse the expression from outside to inside, and construct the postfix notation from inside to outside. We preprocess all query samples in training and evaluation with Alg. 1.

Alg. 2 illustrates the steps of batch execution over postfix expressions. For clarity, we describe the algorithm as one for loop over samples in the pseudo code, while samples that fall into the same case (Line 8, 10, 13, 16 & 19) are executed in parallel. Since the GNN in relation projection takes ($O(|\mathcal{V}|d^2 + |\mathcal{E}|d)$ time (see App. C of (Zhu et al., 2021) for proofs), i.e., much more time than fuzzy logic operations ($O(|\mathcal{V}|)$ time), we synchronize different samples before neural relation projection (Line 20) to maximize the utilization of GPU. Fig. 6 shows the procedure of Alg. 2 over a batch of two queries.

The overall time complexity of our batched execution is $O(t(|\mathcal{V}|d^2 + |\mathcal{E}|d))$, where t is the maximal number of projections in a single query in the batch. Compared to existing implementation (Hamilton et al., 2018; Ren et al., 2019; Ren & Leskovec, 2020) that scales *linearly* w.r.t. the number of query types, batched expression execution scales *independently* w.r.t. the number of query types, and can be applied to arbitrary large number of query types without scalability issues.

Algorithm 1 Convert expression to postfix notation

Input: a query expression
Output: postfix notation of the query expression
function GetPostfix(exp)
  postfix $\leftarrow$ []
  op, *exps_or_vars* $\leftarrow$ GetOutmostOperation(*exp*)
  for *exp_or_var* in *exps_or_vars* **do**
    if *exp_or_var* is an expression **then**
      postfix $\leftarrow$ *postfix* + GetPostfix(*exp_or_var*)
    else
      postfix $\leftarrow$ *postfix* + *exp_or_var*
    end if
  end for
  return *postfix* + *op*
end function

Algorithm 2 Batched expression execution

1: **Input:** a batch of expressions in postfix notation
2: **Output:** a batch of fuzzy sets for answers
3: *stacks* $\leftarrow$ allocate *batch_size* stacks for fuzzy sets
4: **for** $i \leftarrow 0$ to *batch_size* -1 **do**
5:   // parallelized loop
6:   **for** *instruction* in *queries[i]* **do**
7:     **switch** *instruction* **do**
8:       **case** operand
9:         *stacks*[i].push(*instruction*)
10:       **case** conjunction
11:         $\boldsymbol{x}, \boldsymbol{y} \leftarrow$ *stacks*[i].pop(), *stacks*[i].pop()
12:         *stacks*[i].push($\mathcal{C}(\boldsymbol{x}, \boldsymbol{y})$)
13:       **case** disjunction
14:         $\boldsymbol{x}, \boldsymbol{y} \leftarrow$ *stacks*[i].pop(), *stacks*[i].pop()
15:         *stacks*[i].push($\mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$)
16:       **case** negation
17:         $\boldsymbol{x} \leftarrow$ *stacks*[i].pop()
18:         *stacks*[i].push($\mathcal{N}(\boldsymbol{x})$)
19:       **case** projection
20:         wait until all samples are in this case
21:         $\boldsymbol{x} \leftarrow$ *stacks*[i].pop()
22:         *relation* $\leftarrow$ instruction.relation
23:         *stacks*[i].push($\mathcal{P}_{relation}(\boldsymbol{x})$)
24:     **end switch**
25:   **end for**
26: **end for**
27: **return** *stacks*.pop()

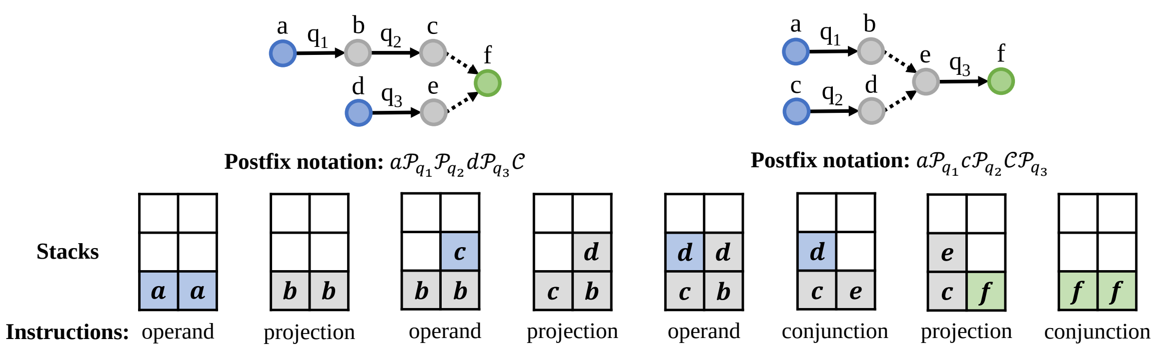


Figure 6: Illustration of batched expression execution (Alg. 2) over a batch of two queries.

D. More Experiment Results

Here we provide additional experiment results.

Tab. 8 shows the H@1 results of different models for answering FOL queries. GNN-QE significantly outperforms existing methods in both EPFO queries and negation queries on all datasets.

Tab. 9 compares the Spearman's rank correlation for answer set cardinality prediction on FB15k and NELL-995. GNN-QE achieves the best rank correlation on all query types.

Table 8: Test H@1 results (%) on answering FOL queries. avg_p is the average H@1 on EPFO queries ($\wedge$, $\vee$). avg_n is the average H@1 on queries with negation. Results of GQE and Q2B are taken from (Ren & Leskovec, 2020).

Model	**avg$_p$**	**avg$_n$**	**1p**	**2p**	**3p**	**2i**	**3i**	**pi**	**ip**	**2u**	**up**	**2in**	**3in**	**inp**	**pin**	**pni**
								FB15k								
GQE	16.6	-	34.2	8.3	5.0	23.8	34.9	15.5	11.2	11.5	5.6	-	-	-	-	-
Q2B	26.8	-	52.0	12.7	7.8	40.5	53.4	26.7	16.7	22.0	9.4	-	-	-	-	-
BetaE	31.3	5.2	52.0	17.0	16.9	43.5	55.3	32.3	19.3	28.1	16.9	6.4	6.7	5.5	2.0	5.3
CQD-CO	39.7	-	85.8	17.8	9.0	67.6	71.7	34.5	24.5	30.9	15.5	-	-	-	-	-
CQD-Beam	51.9	-	85.8	48.6	22.5	67.6	71.7	51.7	62.3	31.7	25.0	-	-	-	-	-
ConE	39.6	7.3	62.4	23.8	20.4	53.6	64.1	39.6	25.6	44.9	21.7	9.4	9.1	6.0	4.3	7.5
GNN-QE	**67.3**	**28.6**	**86.1**	**63.5**	**52.5**	**74.8**	**80.1**	**63.6**	**65.1**	**67.1**	**53.0**	**35.4**	**33.1**	**33.8**	**18.6**	**21.8**
								FB15k-237								
GQE	8.8	-	22.4	2.8	2.1	11.7	20.9	8.4	5.7	3.3	2.1	-	-	-	-	-
Q2B	12.3	-	28.3	4.1	3.0	17.5	29.5	12.3	7.1	5.2	3.3	-	-	-	-	-
BetaE	13.4	2.8	28.9	5.5	4.9	18.3	31.7	14.0	6.7	6.3	4.6	1.5	7.7	3.0	0.9	0.9
CQD-CO	14.7	-	**36.6**	4.7	3.0	20.7	29.6	15.5	9.9	8.6	4.0	-	-	-	-	-
CQD-Beam	15.1	-	**36.6**	6.3	4.3	20.7	29.6	13.5	12.1	8.7	4.3	-	-	-	-	-
ConE	15.6	2.2	31.9	6.9	5.3	21.9	36.6	17.0	7.8	8.0	5.3	1.8	3.7	3.4	1.3	1.0
GNN-QE	**19.1**	**4.3**	32.8	**8.2**	**6.5**	**27.7**	**44.6**	**22.4**	**12.3**	**9.8**	**7.6**	**4.1**	**8.1**	**4.1**	**2.5**	**2.7**
								NELL-995								
GQE	9.9	-	15.4	6.7	5.0	14.3	20.4	10.6	9.0	2.9	5.0	-	-	-	-	-
Q2B	14.1	-	23.8	8.7	6.9	20.3	31.5	14.3	10.7	5.0	6.0	-	-	-	-	-
BetaE	17.8	2.1	43.5	8.1	7.0	27.2	36.5	17.4	9.3	6.9	4.7	1.6	2.2	4.8	0.7	1.2
CQD-CO	21.3	-	**51.2**	11.8	9.0	28.4	36.3	22.4	15.5	9.9	**7.6**	-	-	-	-	-
CQD-Beam	21.0	-	**51.2**	**14.3**	6.3	28.4	36.3	18.1	**17.4**	**10.2**	7.2	-	-	-	-	-
ConE	19.8	2.2	43.6	10.7	9.0	28.6	39.8	19.2	11.4	9.0	6.6	1.4	2.6	5.2	0.8	1.2
GNN-QE	**21.5**	**3.6**	43.5	12.9	**9.9**	**32.5**	**42.4**	**23.5**	12.9	8.8	7.4	**3.2**	**5.9**	**5.4**	**1.6**	**2.0**

Table 9: Spearman's rank correlation between the model prediction and the number of ground truth answers. avg is the average correlation on all 12 query types in the table. Results of baseline methods are taken from (Zhang et al., 2021b).

Model	**avg**	**1p**	**2p**	**3p**	**2i**	**3i**	**pi**	**ip**	**2in**	**3in**	**inp**	**pin**	**pni**
						FB15k							
Q2B	-	0.301	0.219	0.262	0.331	0.270	0.297	0.139	-	-	-	-	-
BetaE	0.494	0.373	0.478	0.472	0.572	0.397	0.519	0.421	0.622	0.548	0.459	0.465	0.608
ConE	0.659	0.60	0.68	0.70	0.68	0.52	0.59	0.56	0.84	0.75	0.61	0.58	0.80
GNN-QE	**0.945**	**0.958**	**0.970**	**0.940**	**0.984**	**0.927**	**0.936**	**0.916**	**0.980**	**0.907**	**0.905**	**0.944**	**0.978**
						NELL995							
Q2B	-	0.154	0.288	0.305	0.380	0.410	0.361	0.345	-	-	-	-	-
BetaE	0.552	0.423	0.552	0.564	0.594	0.610	0.598	0.535	0.711	0.595	0.354	0.447	0.639
ConE	0.688	0.56	0.61	0.60	0.79	0.79	0.74	0.58	0.90	0.79	0.56	0.48	0.85
GNN-QE	**0.891**	**0.913**	**0.851**	**0.780**	**0.974**	**0.935**	**0.825**	**0.737**	**0.994**	**0.980**	**0.882**	**0.848**	**0.976**

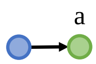
E. More Visualization Results

We provide more visualization for intermediate variables in Tab. 10. For each of the 14 query types, we *randomly* draw 3 query samples from the test set of FB15k-237. Therefore, we can observe both successful and failure cases of our method. For all expressions in the visualization, the operations follow the priority $\neg > \wedge > \vee$.

Note the correctness of some predictions are contradictory to common sense. This is not a failure of our visualization, but is a result of the incomplete knowledge graph. For these contradictory tables, we add a footnote below to illustrate this problem.

Table 10: Visualization of random samples from FB15k-237 test set.

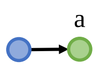
What language does Mattel provide service in?



1p

Query	$q = ?a$: ServiceLanguage(Mattel, a)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	English Spanish -	Chinese ✗ Russia ✗ French ✓	Arabic ✗ - -	French (hard)	3

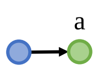
What film did Amy Irving star in?



1p

Query	$q = ?a$: Star(Amy Irving, a)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Never Say Never Again Carrie Who Framed Roger Rabbit	She's Having a Baby ✗ Mickey's Magical Christmas ✗ Terror in the Aisles ✓	Think like a Man ✗ Evolution ✗ I Love You, Man ✗	Terror in the Aisles (hard)	3

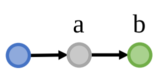
What sport does Nicaragua play?



1p

Query	$q = ?a$: SportCountry(a, $Nicaragua$)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	beach volleyball shooting sport rowing	road running ✗ track and field ✓ judo ✗	taekwondo ✗ boxing ✓ road cycling ✗	boxing (hard)	4

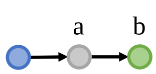
Who was nominated for the award that Kuch Kuch Hota Hai received?



2p

Query	$q = ?b$: Award(Kuch Kuch Hota Hai, a) $\wedge$ Nominated(b, a)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Filmfare Award for Best Actor Filmfare Award for Best Film Filmfare Award for Best Actress	Filmfare Award for Best Director ✗ - -	- - -	Filmfare Award for Best Supporting Actor (easy)	1
b	Girish Karnad Kamal Haasan Dilip Kumar	Vijay Anand ✗ Aamir Khan ✓ K. Balachander ✗	Asit Sen ✗ Sridevi ✗ Aishwarya Rai ✗	Arjun Rampal (hard)	25

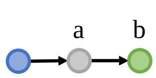
What institution studies the industry that Harland & Wolff specializes in?



2p

Query	$q = ?b$: Industry(Harland & Wolff, a) $\wedge$ FieldOfStudy(b, a)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	civil engineering shipbuilding -	video game ✗ media ✗ airline ✗	publishing ✗ broadcasting ✗ entertainment ✗	civil engineering (easy)	1
b	West Point Yale University Newcastle University	University of Minnesota ✗ Princeton University ✗ Brown University ✗	University of Edinburgh ✗ George Washington University ✗ Rutgers University ✗	University of Utah (hard)	26

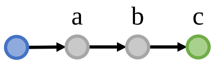
Who is the winner of the award that Lee Grant was nominated for?



2p

Query	$q = ?b$: Nominated(Lee Grant, a) $\wedge$ Winner(a, b)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Emmy Award for Best Supporting Actress (Drama) Emmy Award for Best Lead Actress (Comedy) Academy Award for Best Supporting Actress	Emmy Award for Best Lead Actress (Movie) ✗ Cannes Award for Best Actress ✗ Tony Award for Best Actress (Play) ✗	Tony Award for Best Featured Actress (Play) ✗ SAG Award for Best Female Actor (Movie) ✗ Emmy Award for Best Guest Actress (Drama) ✗	Academy Award for Best Supporting Actress (easy)	1
b	Cate Blanchett Anna Paquin Juliette Binoche	Laura Linney ✗ Jessica Lange ✗ Sally Field ✗	Helen Mirren ✗ Jane Fonda ✗ Meryl Streep ✗	Angelina Jolie (hard)	32

Whose profession is a specialization of Novak Djokovic's profession?

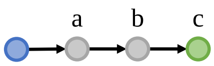


3p

Query $q = ?c : \exists a, b : \text{Profession}(\text{Novak Djokovic}, a) \wedge \text{SpecializationOf}(b, a) \wedge \text{Profession}(c, b)$

Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	athlete - -	actor ✗ screenwriter ✗ voice actor ✗	film director ✗ politician ✗ soccer player ✗	athlete (easy)	1
b	soccer player attacker (soccer) strongman	- - -	- - -	soccer player (easy)	1
c	Jermaine Easter Jonathan Walters Taiwo Atieno	Delroy Facey ✓ Kevin Betsy ✓ Carl Cort ✗	Jung Sung-ryong ✗ Leon Cort ✗ Hameur Bouazza ✓	Roy Carroll (hard)	4

What is the organization that a politician of a WTO member state came from?

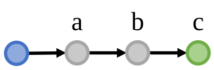


3p

Query $q = ?c : \exists a, b : \text{MemberStates}(\text{WTO}, a) \wedge \text{JurisdicationOfOffice}(b, a) \wedge \text{Organization}(b, c)$

Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Philippines Hong Kong Fiji	Yugoslavia ✗ Guyana ✓ Luxembourg ✓	Greece ✓ Ghana ✓ Tajikistan ✗	Fiji (easy)	1
b	Robert F. Kennedy Thomas Jefferson John F. Kennedy	Mao Zedong ✓ Saddam Hussein ✗ Yasser Arafat ✗	Chiang Kai-shek ✗ Georgy Zhukov ✗ Donald M. Payne ✗	president (easy)	1
c	University of Manitoba Kansas State University Columbia University	West Point ✗ U.S. Naval Academy ✗ Spyglass Media Group ✗	Yale Law School ✗ Harvard Law School ✗ London School of Economics ✗	Manhattan School of Music (hard)	121

What team plays the sport that is in the Olympics that Greece participated in?

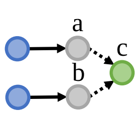


3p

Query $q = ?c : \exists a, b : \text{ParticipateCountry}(a, \text{Greece}) \wedge \text{OlympicSports}(a, b) \wedge \text{TeamSports}(c, b)$

Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	1936 Summer Olympics 1980 Winter Olympics 2002 Winter Olympics	2010 Winter Olympics ✓ 2012 Summer Olympics ✗ 1920 Summer Olympics ✗	1988 Summer Olympics ✗ 1928 Summer Olympics ✗ 1992 Summer Olympics ✗	2010 Winter Olympics (hard)	1
b	soccer track and field water polo	luge ✓ ice hockey ✓ short track speed skating ✓	tennis ✗ - -	ice hockey (hard)	1
c	Sacramento Kings Utah Jazz Seattle SuperSonics	Algeria soccer team ✓ Cincinnati Reds ✓ Washington Nationals ✓	Chile soccer team ✓ Cardiff City ✓ Blackburn Rovers ✓	Florida Panthers (hard)	433

What artist plays both teen pop and pop rock?

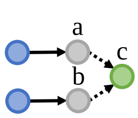


2i

Query $q = ?c : \text{Artist}(\text{teen pop}, c) \wedge \text{Artist}(\text{pop rock}, c)$

Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Jonas Brothers Ashley Tisdale Emma Roberts	Vanessa Hudgens ✓ Jesse McCartney ✓ Emily Osment ✗	SM Town ✗ La Toya Jackson ✗ Justin Bieber ✓	Kevin Jonas (easy)	1
b	Blondie Rupert Holmes Bon Jovi	The Monkees ✓ Ringo Starr & His All-Starr Band ✗ The Who ✓	Queen ✗ Thin Lizzy ✗ Band Aid ✗	Kevin Jonas (hard)	336
c	Ashley Tisdale Emma Roberts Morning Musume	Jonas Brothers ✓ NSYNC ✗ Joe Jonas ✓	Britney Spears ✗ Kevin Jonas ✓ Dr. Luke ✗	Kevin Jonas (hard)	4

Which science fiction film is distributed by Warner Bros.?

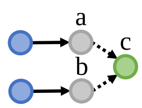


2i

Query $q = ?c : \text{Genre}(c, \text{Science Fiction}) \wedge \text{DistributeFilm}(\text{Warner Bros.}, c)$

Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Men in Black 3 The Fifth Element Frankenweenie	Star Trek 10 ✓ Star Trek 8 ✓ Superman 4 ✓	Spiderman 3 ✓ It ✗ Bleach Movie 3 ✗	Priest (hard)	82
b	The Legend of Tarzan, Lord of the Apes Never Say Never Again Sayonara	The Hangover 2 ✗ Scooby-Doo ✗ Glee: The 3D Concert Movie ✗	Hamlet ✗ The Animatrix ✗ Superman/Batman: Public Enemies ✗	Priest (easy)	1
c	A Clockwork Orange Terminator Salvation Superman: Man Of steel	The Animatrix ✗ Superman 4 ✓ Superman/Batman: Public Enemies ✗	The Matrix 3 ✗ Batman: Gotham Knight ✗ The Matrix 2 ✓	Priest (hard)	28

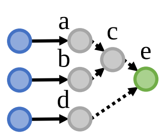
What dog breed is in both San Francisco and Miami?



2i

Query	$q = ?c : \text{DogBreed}(\text{San Francisco}, c) \wedge \text{DogBreed}(\text{Miami}, c)$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	German Shepherd dog labrador retriever Bulldog	Yorkshire Terrier ✓ - -	- - -	Yorkshire Terrier (hard)	1
b	labrador retriever Bulldog Yorkshire Terrier	German Shepherd dog ✓ - -	- - -	Yorkshire Terrier (easy)	1
c	labrador retriever Bulldog Golden Retriever	German Shepherd dog ✓ Yorkshire Terrier ✓ -	- - -	Yorkshire Terrier (hard)	1

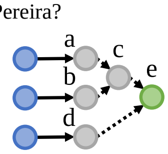
Which songwriter won the same award as BeBeWinans, and was nominated for the same award as Babyface?



3i

Query	$q = ?e : \text{Profession}(e, \text{songwriter}) \wedge \text{WinnerOfSameAward}(e, \text{BeBe Winans}) \wedge \text{NominatedForSameAward}(\text{Babyface}, e)$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Phil Vischer Dr. Seuss Michael Nesmith	Walter Scharf ✗ Paul Francis Webster ✓ DMX ✗	Al Jourgensen ✗ Trey Anastasio ✗ George Duning ✗	L.A. Reid (easy)	1
b	Whitney Houston - -	CeCe Winans ✗ Babyface ✗ David Foster ✗	- - -	L.A. Reid (hard)	6
c	Whitney Houston - -	Babyface ✗ David Foster ✗ CeCe Winans ✗	- - -	L.A. Reid (hard)	6
d	Stephen Schwartz Eric Clapton T-Pain	Barry White ✗ David Banner ✗ Stevie Wonder ✗	Static Major ✗ Whitney Houston ✗ BeBe Winans ✗	L.A. Reid (easy)	1
e	- - -	Whitney Houston ✗ David Foster ✗ -	- - -	L.A. Reid (hard)	3

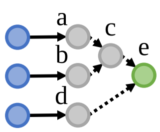
Who was born in Pennsylvania, nominated for the same award as Robert F. Boyle, and nominated for the same award as Hal Pereira?



3i

Query	$q = ?e : \text{PlaceOfBirth}(e, \text{Pennsylvania}) \wedge \text{NominatedForSameAward}(e, \text{Robert F. Boyle}) \wedge \text{NominatedForSameAward}(\text{Hal Pereira}, e)$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	- - -	Bam Mangera ✗ Jim Thorpe ✗ Matthew Fox ✗	Dean Koontz ✗ Clark Gable ✗ Poul Anderson ✗	Frank R. McKelvy (hard)	770
b	Edward G. Boyle Frank R. McKelvy William A. Horning	Peter Lamont ✗ Ken Adam ✗ Henry Grace ✓	Cedric Gibbons ✗ Edwin B. Willis ✗ Edward Carfagno ✗	Frank R. McKelvy (easy)	1
c	- - -	Frank R. McKelvy ✓ Edward G. Boyle ✗ -	- - -	Frank R. McKelvy (hard)	1
d	Frank R. McKelvy Ray Moyer Henry Bumstead	Joseph Kish ✗ Hans Dreier ✗ Richard Day ✗	- - -	Frank R. McKelvy (easy)	1
e	- - -	Frank R. McKelvy ✓ - -	- - -	Frank R. McKelvy (hard)	1

Which team has positions for defender, midfielder and goalkeeper?

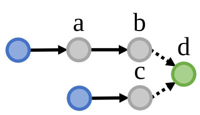


3i

Query	$q = ?e$: Position(e, defender (soccer)) $\wedge$ Position(e, midfielder (soccer)) $\wedge$ Position(e, goalkeeper (soccer))				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Hannover 96 RB Leipzig Bayer 04 Leverkusen	Belarus soccer team ✗ U.S. Lecce ✗ A.C. Cesena ✗	FC Torpedo Moscow ✗ Austria soccer team ✗ 1. FC Union Berlin ✗	Zamalek SC (hard)	149
b	Hannover 96 RB Leipzig FC Bayern Munich	Belarus soccer team ✗ Kazakhstan soccer team ✗ Croatia soccer team ✗	Chile soccer team ✗ FC Torpedo Moscow ✗ FC Energie Cottbus ✗	Zamalek SC (easy)	1
c	Hannover 96 RB Leipzig Bayer 04 Leverkusen	Bulgaria soccer team ✗ Belarus soccer team ✗ Kazakhstan soccer team ✗	Croatia soccer team ✗ Chile soccer team ✗ U.S. Lecce ✗	Zamalek SC (hard)	218
d	RB Leipzig Hannover 96 FC Volga Nizhny Novgorod	D.C. United ✗ F.C. Lorient ✗ Coritiba F.C. ✗	South Africa soccer team ✗ Croatia soccer team ✗ Suwon Samsung Bluewings FC ✗	Zamalek SC (easy)	1
e	Hannover 96 RB Leipzig Bayer 04 Leverkusen	D.C. United ✗ F.C. Lorient ✗ Bulgaria soccer team ✗	Belarus soccer team ✗ Kazakhstan soccer team ✗ U.S. Lecce ✗	Zamalek SC (hard)	241

In reality, every soccer team has positions for defender, midfielder and goalkeeper. The predictions are considered wrong because the corresponding facts are missing in FB15k-237.

What is contained by both the administrative division of Columbia and South Carolina?

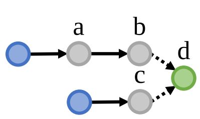


pi

Query	$q = ?d : \exists a$: AdministrativeDivision(Columbia, a) $\wedge$ Contain(a, d) $\wedge$ Contain(South Carolina, d)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	South Carolina - -	Lexington County ✗ - -	- - -	South Carolina (easy)	1
b	University of South Carolina Clemson University Furman University	Florida Keys ✗ Marquette County ✗ Bethesda ✗	Passaic ✗ Fairbanks ✗ Johnstown ✗	Greenville County (hard)	34
c	Clemson University Furman University University of South Carolina	Marquette County ✗ Binghamton ✗ Passaic ✗	Green Bay ✗ Spartanburg County ✓ Rochester ✗	Greenville County (hard)	34
d	Clemson University University of South Carolina Furman University	Marquette County ✗ Florida Keys ✗ -	- - -	Greenville County (hard)	33

Since the *administrative division* of *Columbia* is *South Carolina*, this *pi* query degenerates to a *2p* query.

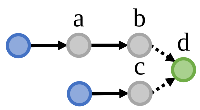
What company was founded by a student from West Point and was also the one that Buzz Aldrin work for?



pi

Query	$q = ?d : \exists a$: Student(West Point, a) $\wedge$ Found(a, d) $\wedge$ Company(Buzz Aldrin, d)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Timothy Leary David Petraeus Edgar Allan Poe	Grover Cleveland ✗ Theodore Roosevelt ✗ James Monroe ✗	DeWitt Clinton ✗ Chris Elliott ✗ Gary Dourdan ✗	Dwight D. Eisenhower (easy)	1
b	CIA NASA -	US Department of Defense ✗ University of California, Berkeley ✗ Columbia University ✗	US Department of Housing and Urban Development ✗ US Department of the Air Force ✗ United Nations ✗	NASA (easy)	1
c	US Department of the Air Force - -	ESPN ✗ Columbia University ✗ Microsoft ✗	Havard University ✗ Oracle ✗ Dartmouth College ✗	NASA (hard)	934
d	- - -	US Department of the Air Force ✗ NASA ✓ -	- - -	NASA (hard)	2

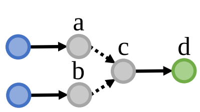
What religion is found in a country that Anugilla exports to and is also believed by Sunny Deol?



pi

Query	$q = ?d : \exists a :$ ExportTo(Anguilla, a) $\wedge$ Religion(a, d) $\wedge$ Religion(Sunny Deol, d)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	United States of America United Kingdom -	- - -	- - -	United Kingdom (easy)	1
b	atheism Hinduism Judaism	Christianity ✓ Protestantism ✗ agnosticism ✗	Lutheranism ✗ Methodism ✓ Anglicanism ✗	Sikhism (hard)	6
c	Hinduism Sikhism -	Hindu ✗ Methodism ✗ Baptists ✗	Presbyterianism ✗ Pentecostalism ✗ Nondenominational Christianity ✗	Sikhism (easy)	1
d	Hinduism - -	Islam ✗ Sikhism ✓ -	- - -	Sikhism (hard)	2

Who died in the place where marriage takes place and C.D. Chivas locates in?

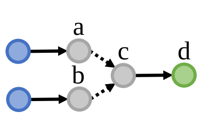


ip

Query	$q = ?d : \exists c :$ LocationOfCeremony(marraige, c) $\wedge$ LocalTeam(c, C.D. Chivas) $\wedge$ PlaceOfDeath(d, c)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Tijuana Jerusalem Puerto Rico	Tehran ✗ Genoa ✓ Tunis ✗	Monterrey ✗ Green Bay ✗ Binghamton ✗	Los Angeles (easy)	1
b	Los Angeles - -	Museo del Prado ✗ Lund ✗ Kharkiv ✗	Rawalpindi ✗ Innsbruck ✗ Katowice ✗	Los Angeles (easy)	1
c	Los Angeles - -	Museo del Prado ✗ Guatemala City ✗ Seville ✗	Bilbao ✗ Rabat ✗ Tunis ✗	Los Angeles (easy)	1
d	Ralph Burns Robert F. Boyle Boris Leven	William Travilla ✗ Fred MacMurray ✗ Chuck Jones ✗	Jerry Wald ✗ Louis Armstrong ✗ Gregory Peck ✗	Ida Lupino (hard)	15

In reality, marriage can take place in any city. The predictions are considered wrong because the corresponding facts are missing in FB15k-237.

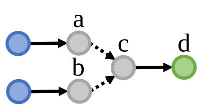
What film is nominated for the award that Freddy Got Fingered won and Peter Hyams was nominated for?



ip

Query	$q = ?d : \exists c :$ Award(Freddy Got Fingered, c) $\wedge$ Nominated(Peter Hyams, c) $\wedge$ NominatedFor(d, c)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Golden Raspberry Award for Worst Screen Couple Golden Raspberry Award for Worst Picture Golden Raspberry Award for Worst Actor	Golden Raspberry Award for Worst Remake ✗ Golden Raspberry Award for Worst Supporting Actor ✗ Golden Raspberry Award for Worst Supporting Actress ✗	Golden Raspberry Award for Worst Actress ✗ - -	Golden Raspberry Award for Worst Director (easy)	1
b	Golden Raspberry Award for Worst Director - -	Golden Raspberry Award for Worst Supporting Actor ✗ Academy Award for Best Animated Short Film ✗ Academy Award for Best Live Action Short Film ✗	Academy Award for Best Story ✗ Academy Award for Best Documentary Feature ✗ Satellite Award for Best Adapted Screenplay ✗	Golden Raspberry Award for Worst Director (easy)	1
c	Golden Raspberry Award for Worst Director - -	Golden Raspberry Award for Worst Screenplay ✗ Golden Raspberry Award for Worst Actor ✗ Golden Raspberry Award for Worst Picture ✗	Golden Raspberry Award for Worst Screen Couple ✗ - -	Golden Raspberry Award for Worst Director (easy)	1
d	Battleship New Year's Eve Dressed to Kill	Swept Away ✗ Showgirls ✗ Ghost Rider: Spirit of Vengeance ✗	Jaws 3-D ✓ Bolero ✗ Gigli ✗	Last Action Hero (hard)	119

What organization has the position that both Miramax and WarnerMedia have?

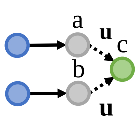


ip

Query	$q = ?d : \exists c : \text{Organization}(c, \text{Miramax}) \wedge \text{Company}(c, \text{WarnerMedia}) \wedge \text{Organization}(c, d)$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	chief executive officer - -	- - -	- - -	chief executive officer (easy)	1
b	president chief operating officer chief executive officer	chief technology officer ✗ treasurer ✗ chief marketing officer ✗	general manager ✗ chief administrative officer ✗ vice president ✗	chief executive officer (easy)	1
c	chief executive officer - -	- - -	- - -	chief executive officer (easy)	1
d	Miramax New Line Cinema Lionsgate	KSA Network ✗ DC Comics ✗ ESPN ✗	LucasArts ✗ GMA Network ✗ Spyglass Media Group ✗	TV5 (hard)	425

In reality, most big companies have positions for CEO, CTO, etc. The predictions are considered wrong because the corresponding facts are missing in FB15k-237.

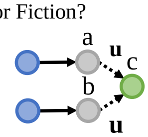
What genre is played by Chick Corea or Keith Jarrett?



2u

Query	$q = ?c : \text{Artist}(c, \text{Chick Corea}) \vee \text{Artist}(c, \text{Keith Jarrett})$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	smooth jazz jazz-funk jazz fusion	free jazz ✗ hard bop ✗ bebop ✓	funk ✗ sophisti-pop ✗ big band ✗	bebop (hard)	3
b	free jazz jazz fusion post-bop	avant-garde jazz ✓ hard bop ✗ bebop ✗	Baroque music ✗ smooth jazz ✓ progressive rock ✗	jazz (easy)	1
c	post-bop classical music jazz	hard bop ✗ bebop ✓ rock music ✗	Baroque music ✗ cool jazz ✗ funk ✗	bebop (hard)	2

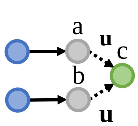
Who is a student of Bucknell University or is nominated for National Book Award for Fiction?



2u

Query	$q = ?c : \text{Student}(\text{Bucknell University}, c) \vee \text{Nominated}(c, \text{National Book Award for Fiction})$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Philip Roth - -	Kate Flannery ✗ Hogan Sheffer ✗ James Buchanan ✗	Pandro S. Berman ✗ Gary Dourdan ✗ Martie Maguire ✗	Philip Roth (easy)	1
b	James Baldwin J. D. Salinger Vladimir Nabokov	Joyce Carol Oates ✗ Michael Chabon ✗ James A. Michener ✗	Larry McMurtry ✗ James Joyce ✗ Fyodor Dostoyevsky ✗	Ursula K. Le Guin (hard)	24
c	Philip Roth James Baldwin J. D. Salinger	Joyce Carol Oates ✗ James A. Michener ✗ Michael Chabon ✗	Andy Warhol ✗ Lloyd Alexander ✗ Noam Chomsky ✗	Ursula K. Le Guin (hard)	40

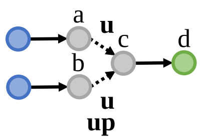
Who is an artist of Warner Bros. Records or is a lead vocalist?



2u

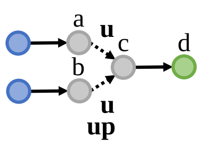
Query	$q = ?c : \text{Artist}(\text{Warner Bros. Records}, c) \vee \text{Role}(c, \text{lead vocalist})$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Sheila E. Linkin Park Metallica	Los Lobos ✗ Gong ✗ Dir En Grey ✗	Gwar ✗ The Smashing Pumpkins ✗ Tony Levin ✗	Randy Newman (hard)	366
b	Bernadette Peters John Gielgud Andy Dick	Rick Springfield ✗ Steve Winwood ✓ Lou Reed ✗	Chris Seefried ✗ David Grisman ✗ Arjen Anthony Lucassen ✗	Eddie Vedder (easy)	1
c	Sheila E. Chaka Khan Metallica	Los Lobos ✗ Glen Ballard ✗ Brad Paisley ✗	Tony Levin ✗ Rick Springfield ✗ Meat Loaf ✗	Randy Newman (hard)	269

What genre shares a derived genre with jazz or is the parent genre of Maroon 5's genre?



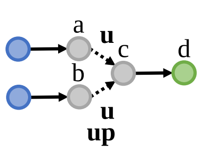
Query	$q = ?d : \exists c : (\text{ParentGenre}(c, \text{jazz}) \vee \text{Artist}(c, \text{Maroon 5})) \wedge \text{ParentGenre}(c, d)$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	drum and bass post-rock technical death metal	Dixieland ✓ bossa nova ✓ sophisti-pop ✗	soul jazz ✓ doo-wop ✗ progressive bluegrass ✗	bluegrass music (easy)	1
b	techno ska punk jazz fusion	power pop ✗ Britpop ✗ synth-pop ✗	rhythm and blues ✗ post-punk ✗ electro ✗	pop music (easy)	1
c	jazz fusion jazz rap techno	jam band ✗ sophisti-pop ✗ doo-wop ✗	power pop ✗ progressive metal ✗ electro ✗	pop music (easy)	1
d	gospel music folk music blues	folk rock ✗ blues rock ✗ hard rock ✓	psychedelic rock ✓ pop rock ✗ glam rock ✓	dance music (hard)	96

Who wins the award that David Kirschner winned or is given at 39th Daytime Emmy Awards?



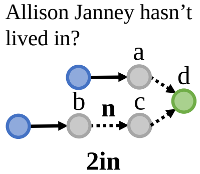
Query	$q = ?d : \exists c : (\text{Award}(\text{David Kirschner}, c) \vee \text{Ceremony}(c, \text{39th Daytime Emmy Awards})) \wedge \text{Award}(d, c)$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Daytime Emmy Award for Outstanding Animated Program - -	Academy Award for Best Animated Short Film ✗ Primetime Emmy Award for Outstanding Comedy Series ✗ Academy Award for Best Adapted Screenplay ✗	Academy Award for Best Story ✗ Academy Award for Best Original Screenplay ✗ Golden Raspberry Award for Worst Director ✗	Daytime Emmy Award for Outstanding Animated Program (easy)	1
b	Daytime Emmy Award for Outstanding Animated Program Daytime Emmy Award for Outstanding Talk Show Host Daytime Emmy Award for Outstanding Younger Actress in Drama	Daytime Emmy Award for Outstanding Drama Series ✓ Writers Guild of America Award for Best Adapted Screenplay ✗ Writers Guild of America Award for Best Original Screenplay ✗	Daytime Emmy Award for Outstanding Talk Show ✗ soap opera ✗ game show ✗	Daytime Emmy Award for Outstanding Game Show (easy)	1
c	Daytime Emmy Award for Outstanding Animated Program Daytime Emmy Award for Outstanding Talk Show Host Daytime Emmy Award for Outstanding Younger Actress in Drama	Daytime Emmy Award for Outstanding Drama Series ✓ Writers Guild of America Award for Best Adapted Screenplay ✗ Writers Guild of America Award for Best Original Screenplay ✗	Daytime Emmy Award for Outstanding Talk Show ✗ Academy Award for Best Animated Short Film ✗ MTV Movie Award for Best On-Screen Duo ✗	Daytime Emmy Award for Outstanding Game Show (easy)	1
d	Bill Melendez Susan Blu Ben Stein	William J. Bell ✓ Stephen Demorest ✓ Andy Heyward ✗	Mary-Ellis Bunim ✓ Garin Wolf ✗ Nancy Williams Watt ✗	Robert Zemeckis (hard)	81

What's the award that a New Zealander or a 76th Academy Awards winner wins?



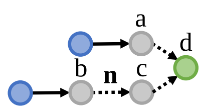
Query	$q = ?d : \exists c : (\text{Nationality}(c, \text{New Zealand}) \vee \text{AwardWinner}(\text{76th Academy Awards}, c)) \wedge \text{Award}(c, d)$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Keith Urban Chris Wood Jemaine Clement	Graeme Revell ✓ Fran Walsh ✓ Jack Thompson ✗	Yoram Globus ✗ Christopher Columbus ✗ David Petraeus ✗	Karl Urban (easy)	1
b	Andrew Stanton Tim Robbins Annie Lennox	Focus Features ✗ Bill Murray ✗ Peter Weir ✗	Brian Helgeland ✗ Scarlett Johansson ✗ Henry Bumstead ✗	Howard Shore (hard)	15
c	Peter Jackson Richard Taylor Keith Urban	Graeme Revell ✓ Jack Thompson ✗ Focus Features ✗	Peter Weir ✗ Bill Murray ✗ Brian Helgeland ✗	Howard Shore (hard)	50
d	Saturn Award for Best Costume BAFTA Award for Best Special Visual Effects MTV Video Music Award for Best Female Video	Academy Award for Best Supporting Actor ✗ BFCA Award for Best Supporting Actor ✗ Independent Spirit Award for Best Male Lead ✗	BFCA Award for Best Cast ✗ Tony Award for Best Musical ✗ BFCA Award for Best Actress ✗	Golden Globe Award for Best Original Score (hard)	26

Which city takes mortgage from US Department of Housing and Urban Development, but Allison Janney hasn't lived in?



Query	$q = ?d : \text{MortgageSource}(d, \text{US Department of HUD}) \wedge \neg\text{PlaceLive}(\text{Allison Janney}, d)$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Tulsa Terre Haute Santa Monica	Meridian ✓ Paducah ✓ Natchez ✓	Albany ✓ Biloxi ✓ Waukesha ✓	Anderson (hard)	1
b	- - -	Boston ✗ Massachusetts ✗ South Carolina ✗	New York City ✗ Ohio ✗ Plainfield ✗	Dayton (hard)	8
d	Lewis County Hidalgo County Delaware County	Meridian ✓ Spokane County ✓ Albany ✓	Odessa ✓ High Point ✓ Paducah ✓	Anderson (hard)	8

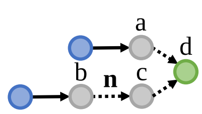
Which school is picked by Golden State Warrior in the draft, but not picked by 2006 NBA Draft?



2in

Query	$q = ?d$: DraftPickSchool(Golden State Warriors, d) $\wedge$ $\neg$DraftPickSchool(2006 NBA Draft, d)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Temple University Stanford University George Washington University	Wake Forest University ✗ University of Texas at Arlington ✗ Centre College ✗	University of Vermont ✗ Walsh School of Foreign Service ✗ The Wharton School ✗	University of Florida (hard)	7
b	Duke University University of Kentucky University of South Carolina	University of Kansas ✗ New Mexico State University ✗ University of Florida ✗	University of North Carolina at Chapel Hill ✗ Boston College ✗ University of Nevada, Reno ✗	Louisiana State University (easy)	1
d	University of San Diego College of the Holy Cross Southern Methodist University	Columbia Business School ✗ The Wharton School ✗ Walsh School of Foreign Service ✗	Harvard Business School ✗ Stanford Graduate School of Business ✗ Stanford Law School ✗	University of Florida (hard)	401

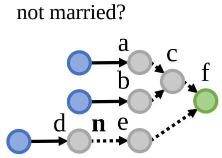
Which musician is not an African American?



2in

Query	$q = ?d$: Profession(d, musician) $\wedge$ $\neg$People(African American, d)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Peter Hyams Adam Levine James Marsters	Al Jourgensen ✗ Walter Scharf ✗ Loudon Wainwright III ✗	Jon Lord ✓ Aimee Mann ✗ Trey Anastasio ✗	Kid Rock (hard)	173
b	Fred Williamson Maya Rudolph Ben Harper	Yolanda Adams ✗ LL Cool J ✗ Flo Rida ✗	Barry White ✗ Fat Joe ✗ Timbaland ✓	Pharrell Williams (easy)	1
d	Sting Guy Pearce Wendy Melvoin	Arjen Anthony Lucassen ✗ Georg Solti ✗ Dmitri Shostakovich ✗	Jerry Duplessis ✗ Franz Schubert ✗ David Gilmour ✓	Kid Rock (hard)	278

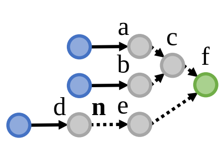
Which American is nominated for SAG Award for Outstanding Performance in a Motion Picture, but is not married?



3in

Query	$q = ?f$: Nominated(f, SAG Award for Outstanding Performance in a Motion Picture) $\wedge$ Nationality(f, USA) $\wedge$ $\neg$MarriageType(f, marriage)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Joey Fatone Judy Greer Dan Hedaya	Alan Tudyk ✓ Peter Stormare ✓ Sam Rockwell ✓	Jerry Stiller ✓ Edward Herrmann ✓ Larry Hagman ✓	Charlize Theron (hard)	1
b	Hannah Arendt Sylvia Plath Vladimir Nabokov	Ad-Rock ✓ George III of Great Britain ✗ Moe Howard ✓	Tim Duncan ✓ Frank Herbert ✓ Lamar Odom ✓	Charlize Theron (easy)	1
c	Thomas Jane Billy Zane Kerry Washington	Paul Dooley ✓ Amy Adams ✓ Jake Gyllenhaal ✓	Mo'Nique ✓ Alan Tudyk ✓ Peter Stormare ✓	Charlize Theron (hard)	1
d	Wong Kar-wai Zhang Yimou Roger Daltrey	Hugh Dancy ✓ Greta Scacchi ✓ Preston Sturges ✓	Jenny McCarthy-Wahlberg ✓ Todd Field ✓ Nathaniel Hawthorne ✓	Joel Silver (easy)	1
f	Leonardo DiCaprio Dwight Yoakam Ludacris	Sam Rockwell ✓ Charlize Theron ✓ Matt Damon ✗	Jake Gyllenhaal ✓ Philip Seymour Hoffman ✓ Heather Graham ✓	Charlize Theron (hard)	24

Which team has positions for both midfielder and goalkeeper, but doesn't play soccer?

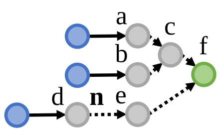


3in

Query	$q = ?f$: Position(f, midfielder (soccer)) $\wedge$ Position(f, goalkeeper (soccer)) $\wedge$ $\neg$Sport(f, soccer)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Hannover 96 RB Leipzig FC Bayern Munich	Bulgaria soccer team ✗ Kazakhstan soccer team ✗ Croatia soccer team ✗	Chile soccer team ✗ FC Torpedo Moscow ✗ FC Energie Cottbus ✗	Wimbledon F.C. (hard)	130
b	RB Leipzig Hannover 96 FC Volga Nizhny Novgorod	D.C. United ✗ F.C. Lorient ✗ Coritiba F.C. ✗	South Africa soccer team ✗ Croatia soccer team ✗ Suwon Samsung Bluewings FC ✗	Wimbledon F.C. (easy)	1
c	Hannover 96 RB Leipzig FC Volga Nizhny Novgorod	D.C. United ✗ F.C. Lorient ✗ Bulgaria soccer team ✗	Kazakhstan soccer team ✗ Coritiba F.C. ✗ FC Torpedo Moscow ✗	Wimbledon F.C. (hard)	202
d	India soccer team Levante UD Singapore soccer team	Croatia soccer team ✓ Tunisia soccer team ✓ Colombia soccer team ✓	1. FSV Mainz 05 ✓ Esporte Clube Bahia ✓ Legia Warsaw ✓	PFC CSKA Moscow (easy)	1
f	Hamilton Academical F.C. Barnet F.C. Lewes F.C.	Saba Qom F.C. ✗ Odense Boldklub ✗ FC Vorskla Poltava ✗	Seongnam FC ✗ AC Arles ✗ FC Vaslui ✗	Wimbledon F.C. (hard)	128

In reality, any team has a midfielder or a goalkeeper plays soccer, and this query has no answer. The failure of generating this query is due to the incompleteness of FB15k-237.

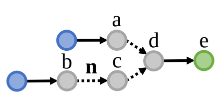
Which film has a DVD version and is released in New Zealand, but doesn't have an English version?



3in

Query	$q = ?f$: ReleaseMedium(f, DVD) $\wedge$ ReleaseRegion(f, New Zealand) $\wedge$ $\neg$Language(f, English)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	The Verdict The Interpreter The Love Guru	The Matrix 3 ✓ Get Him to the Greek ✓ Shattered Glass ✓	A View to a Kill ✓ Airplane! ✓ Superman/Batman: Public Enemies ✗	Detective Dee and the Mystery of the Phantom Flame (hard)	219
b	The Tree of Life Total Recall Prometheus	Snow White and the Huntsman ✓ Submarine ✗ Tyrannosaur ✗	This Is England ✗ The Host ✗ Argo ✓	Detective Dee and the Mystery of the Phantom Flame (easy)	1
c	The Expendables 2 Soul Surfer National Security	The Butterfly Effect ✓ On the Road ✗ Monsters, Inc. ✗	Pineapple Express ✓ Snow White and the Huntsman ✓ Hyde Park on Hudson ✗	Detective Dee and the Mystery of the Phantom Flame (hard)	98
d	Zero Dark Thirty Slaughterhouse-Five The Butterfly Effect	Going My Way ✓ She's Having a Baby ✓ Julius Caesar ✓	The Wizard of Oz ✓ The Country Girl ✓ The Right Stuff ✓	Dangerous Liaisons (easy)	1
f	A Separation The Secret in Their Eyes Rust and Bone	The Good, the Bad the Weird ✗ Submarine ✗ The Orphanage ✗	The Help ✗ The Believer ✗ The City of Lost Children ✗	Detective Dee and the Mystery of the Phantom Flame (hard)	44

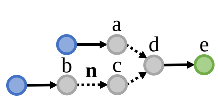
Who plays the instrument that is played by the same musician as flute, but is not in Blondie?



inp

Query	$q = ?e : \exists d$: SameMusician(d, flute) $\wedge$ $\neg$Group(d, Blondie) $\wedge$ Play(e, d)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	zither bassoon timbales	saxophone ✗ clarinet ✗ clapping ✗	solo ✗ Hammond organ ✗ oboe ✗	electronic keyboard (hard)	16
b	lead vocalist drum kit percussion instrument	guitar ✓ lead guitarist ✗ bass guitar ✓	soprano ✗ bass ✗ programming ✗	lead vocalist (easy)	1
d	celesta cornet lute	solo ✗ clapping ✗ clarinet ✗	saxophone ✗ oboe ✗ bagpipes ✗	electronic keyboard (hard)	16
e	Adele Beck John Cale	Jamie Cullum ✓ Devin Townsend ✓ Sun Ra ✓	D'Angelo ✓ Billy Preston ✓ George Duke ✓	Tricky Stewart (hard)	58

Who studies the field that is studied by McGill University, but is not spoken by Nico?

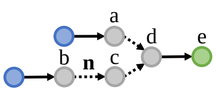


inp

Query	$q = ?e : \exists d$: FieldOfStudy(McGill University, d) $\wedge$ $\neg$Language(Nico, d) $\wedge$ FieldOfStudy(e, d)				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	electrical engineering communication anthropology	computer science ✗ chemical engineering ✗ economics ✗	physics ✓ philosophy ✓ political science ✗	psychology (easy)	1
b	French Spanish -	English ✓ Italian ✗ Chinese ✗	German ✗ Portuguese ✗ Czech ✗	English (hard)	1
d	communication law architecture	chemical engineering ✗ economics ✗ computer science ✗	physics ✓ philosophy ✓ political science ✗	psychology (easy)	1
e	computer engineering mechanical engineering Latin	classics ✗ fine art ✗ literature ✗	music ✗ television ✗ finance ✗	photography (hard)	43

In reality, McGill University studies all the fields that our model predicts for variable a and d. The predictions are considered wrong because the corresponding facts are missing in FB15k-237.

Where is the county that takes mortgage from US Department of Housing and Urban Development, but is not contained by US?



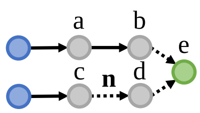
inp

Query $q = ?e : \exists d : \text{MortageSource}(d, \text{US Department of HUD}) \wedge \neg\text{Contain}(\text{USA}, d) \wedge \text{County}(e, d)$

Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Tulsa Terre Haute Santa Monica	Meridian ✓ Paducah ✓ Natchez ✓	Albany ✓ Biloxi ✓ Waukesha ✓	High Point (hard)	1
b	Yuma Fairbanks Santa Clara University	Green Bay ✗ Mansfield ✗ Parkersburg ✓	Valdosta ✓ Beaumont ✗ Muskogee ✗	Pontiac (hard)	25
d	St. Louis County Monroe County Genesee County	Spokane County ✓ Krista Allen ✗ Jeff Gordon ✗	Johnson County ✓ Joan Bennett ✗ Heather O'Reilly ✗	High Point (hard)	223
e	San Jose Evanston Williamsburg	New Haven ✓ Charlottesville ✗ Orlando ✓	Falls Church ✓ Oswego ✓ Iowa City ✓	High Point (hard)	620

In reality, US Department of HUD only provides mortage to US cities, and this query has no answer. The failure of generating this query is due to the incompleteness of FB15k-237.

What genre is the parent genre of the genre that is played by The Prodigy, but is not parent genre of drum and bass?

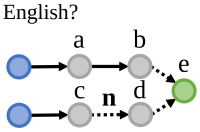


pin

Query $q = ?e : \exists a : \text{Artist}(a, \text{The Prodigy}) \wedge \text{ParentGenre}(a, e) \wedge \neg\text{ParentGenre}(\text{drum and bass}, e)$

Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	industrial rock alternative hip hop dance-punk	techno ✓ breakbeat ✓ intelligent dance music ✗	trance ✗ electronic body music ✗ noise music ✗	post-punk (hard)	21
b	jazz hip hop music hard rock	synth-pop ✓ pop music ✗ soul music ✗	psychedelic rock ✗ pop rock ✗ house music ✓	Krautrock (hard)	13
c	dub music jazz ambient music	electronica ✗ big beat ✗ dark ambient ✗	dance-punk ✗ glitch ✗ trance ✗	electronic dance music (easy)	1
e	hard rock heavy metal rock music	synth-pop ✓ pop rock ✗ pop music ✗	blues ✗ folk music ✗ experimental rock ✗	Krautrock (hard)	28

What film is the actor that is nominated as SAG Award for Outstanding Performance by an Ensemble nominated for, but is not in English?



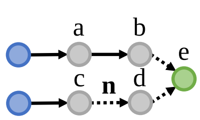
pin

Query $q = ?e : \exists a : \text{Nominated}(a, \text{SAG Award for Outstanding Performance by an Ensemble}) \wedge \text{NominatedFor}(a, e) \wedge \neg\text{Language}(e, \text{English})$

Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Hugh Laurie Jennifer Morrison Fred Thompson	Jude Ciccolella ✓ Jonny Lee Miller ✓ Chris Noth ✓	Aidan Gillen ✓ Kate Walsh ✗ Maura Tierney ✓	Taraji P. Henson (hard)	607
b	X-Men Origins: Wolverine Confessions of a Dangerous Mind Iron Man 2	The Sessions ✗ Syriana ✗ The Green Mile ✗	A Serious Man ✗ The Departed ✗ The Descendants ✗	The Curious Case of Benjamin Button (hard)	95
c	Sesame Street Dancing with the Stars Buffy the Vampire Slayer	/m/06r4f ✓ Tom and Jerry ✓ Ben 10 ✓	Scarlett ✗ Saturday Night's Main Event ✓ Angel ✓	Entourage (easy)	1
e	Transformers: Dark of the Moon Three Kings X-Men Origins: Wolverine	Syriana ✗ The Sessions ✗ The Green Mile ✗	The Departed ✗ The Descendants ✗ Silver Linings Playbook ✓	The Curious Case of Benjamin Button (hard)	92

In reality, all films that win SAG Award have English versions. The failure of generating this query is due to the incompleteness of FB15k-237.

Who was a student of a university in United States, but did not film Malcolm X?

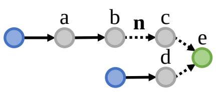


pin

Query $q = ?e : \exists a : \text{Country}(a, \text{USA}) \wedge \text{Student}(a, e) \wedge \neg\text{Film}(e, \text{Malcolm X})$

Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	University of San Francisco Xavier University Adelphi University	The Cannon Group ✗ /m/05f260 ✗ Interplay Entertainment ✗	MicroProse ✗ EuropaCorp ✗ Origin Systems ✓	Los Angeles City College (easy)	1
b	Ann Roth Bruce Berman Roger Corman	Lewis Cass ✓ John Irving ✓ Colin Powell ✗	Milton Friedman ✗ Robert Gates ✗ Gore Vidal ✓	Jerry Goldsmith (hard)	699
c	Nelson Mandela Angela Bassett Christopher Plummer	Frankie Faison ✗ Jim Brown ✗ Anthony LaPaglia ✗	John Leguizamo ✗ Martin Lawrence ✗ Miguel Sandoval ✗	Delroy Lindo (easy)	1
e	Ann Roth James Buchanan Ron Paul	Lewis Cass ✓ Colin Powell ✗ John Irving ✓	J. J. Abrams ✗ Ted Kennedy ✗ Milton Friedman ✗	Jerry Goldsmith (hard)	316

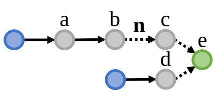
What British is not the gender that is a risk factor of multiple myeloma?



pni

Query	$q = ?e : \exists a : \text{RiskFactor}(a, \text{multiple myeloma}) \land \neg\text{Gender}(e, a) \land \text{People}(\text{British}, e)$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	male organism obesity old age	veganism ✗ vegetarianism ✗ tobacco use disorder ✗	white people ✗ diabetes ✗ alcoholism ✗	male organism (easy)	1
b	Erich Ludendorff Ryan Kavanaugh Stuart Craig	Augustine of Hippo ✓ Ken Adam ✓ R. Madhavan ✓	Peter Stormare ✓ George Bernard Shaw ✓ Fred Tatasciore ✓	Lou Scheimer (easy)	1
d	Eddie Izzard David Niven Julie Christie	Isaac Newton ✗ Charles, Prince of Wales ✗ Anthony B. Richmond ✗	Tom Hardy ✗ Elizabeth II ✗ Joely Richardson ✗	Shirley Ann Russell (hard)	21
e	Siobhan Finneran Natasha Richardson Emma Thompson	Joely Richardson ✗ Elizabeth II ✗ Claire Forlani ✗	Shirley Ann Russell ✓ Emily Mortimer ✗ Hayley Mills ✗	Shirley Ann Russell (hard)	13

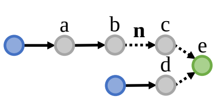
Who believed in Catholicism and did not die from the disease that has the symptom of dyspnea?



pni

Query	$q = ?e : \exists a : \text{SymptomOf}(\text{dyspnea}, a) \land \neg\text{CauseOfDeath}(e, a) \land \text{Religion}(e, \text{Catholicism})$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	colorectal cancer heart failure tuberculosis	bladder cancer ✗ malaria ✗ liver failure ✗	hepatitis ✗ leukemia ✗ -	tuberculosis (easy)	1
b	Jessica Tandy George Washington June Allyson	Truman Capote ✗ Telly Savalas ✗ Stan Winston ✗	Oliver Cromwell ✗ David Watkin ✗ Jorge Luis Borges ✗	Denholm Elliott (easy)	1
d	Robert Zemeckis Lara Flynn Boyle Frank Capra	Ted Kennedy ✓ John Dingell, Jr. ✓ Linda Cardellini ✗	Marcello Mastroianni ✗ Cedric Gibbons ✗ Dante Spinotti ✗	Nicole Kidman (hard)	2804
e	Ann Curry Mario Lopez Leonardo DiCaprio	Linda Cardellini ✗ Carson Daly ✓ Hayden Panettiere ✗	Joe Jonas ✗ Luke Wilson ✗ Minnie Driver ✗	Nicole Kidman (hard)	2149

What team plays ice hockey but is not located in a city that marriage takes place?



pni

Query	$q = ?e : \exists a : \text{LocationOfCeremony}(\text{marriage}, a) \land \neg\text{LocalTeam}(a, e) \land \text{TeamSports}(e, \text{ice hockey})$				
Variable	**Top Predictions** (≥ 0.1) **Easy**	**Hard**		**Random Ground Truth**	**Filtered Ranking**
a	Tijuana Jerusalem Puerto Rico	Tehran ✗ Genoa ✓ Tunis ✗	Monterrey ✗ Green Bay ✗ Binghamton ✗	Moscow (easy)	1
b	C.F. Os Belenenses Williams Grand Prix Engineering FC Torpedo Moscow	Esteghlal F.C. ✗ Club de Fútbol Monterrey ✗ U.C. Sampdoria ✓	Green Bay Packers ✗ GNK Dinamo Zagreb ✗ APOEL F.C. ✗	Spartak Moscow (easy)	1
d	Montreal Canadiens Los Angeles Kings Chicago Blackhawks	Buffalo Sabres ✓ Rockford IceHogs ✓ Manchester Monarchs ✗	Milwaukee Admirals ✗ Portland Pirates ✓ Houston Aeros ✗	Carolina Hurricanes (hard)	6
e	San Jose Sharks New York Rangers Salavat Yulaev Ufa	Manchester Monarchs ✗ Houston Aeros ✗ Rockford IceHogs ✓	Milwaukee Admirals ✗ Portland Pirates ✓ Winnipeg Jets ✗	Carolina Hurricanes (hard)	15

In reality, marriage can take place in any city, and this query has no answer. The failure of generating this query is due to the incompleteness of FB15k-237.