Bayesian Inference with Complex Knowledge Graph Evidence

Armin Toroghi[1], Scott Sanner[1,2]

[1]Department of Mechanical and Industrial Engineering, University of Toronto
[2]Vector Institute of Artificial Intelligence, Toronto
armin.toroghi@mail.utoronto.ca, ssanner@mie.utoronto.ca

Abstract

Knowledge Graphs (KGs) provide a widely used format for representing entities and their relationships and have found use in diverse applications including question answering and recommendation. A majority of current research on KG inference has focused on reasoning with atomic facts (triples) and has disregarded the possibility of making complex evidential observations involving logical operators (negation, conjunction, disjunction) and quantifiers (existential, universal). Further, while the application of complex evidence has been explored in KG-based query answering (KGQA) research, in many practical online settings, observations are made sequentially. For example, in KGQA, additional context may be incrementally suggested to narrow down the answer. Or in interactive recommendation, user critiques may be expressed sequentially in order to narrow down a set of preferred items. Both settings are indicative of information filtering or tracking tasks that are reminiscent of belief tracking in Bayesian inference. In fact, in this paper, we precisely cast the problem of belief tracking over unknown KG entities given incremental complex KG evidence as a Bayesian filtering problem. Specifically, we leverage Knowledge-based Model Construction (KBMC) over the logical KG evidence to instantiate a Markov Random Field (MRF) likelihood representation to perform closed-form Bayesian inference with complex KG evidence (BIKG). We experimentally evaluate BIKG in incremental KGQA and interactive recommendation tasks demonstrating that it outperforms non-incremental methodologies and leads to better incorporation of conjunctive evidence vs. existing complex KGQA methods like CQD that leverage fuzzy T-norm operators. Overall, this work demonstrates a novel, efficient, and unified perspective of logic, KGs, and online inference through the lens of closed-form BIKG.

Introduction

Effective storage and utilization of information from different sources is a central problem in AI. In recent years, the use of Knowledge Graphs (KG) for storing and representing relational and interconnected domain knowledge has garnered substantial interest in various domains ranging from healthcare (Rastogi and Zaki 2020) to recommendation (Guo et al. 2020). Despite the considerable advantages offered by KGs, leveraging them as knowledge bases

for AI agents presents significant challenges due to their inherent incompleteness and noisiness (Pujara, Augustine, and Getoor 2017). These challenges have motivated extensive research about different methodologies for performing reasoning and inference over KGs. To date, most existing works have devoted their focus to inference using atomic level observations, i.e. observations that involve KG triples, for tasks such as KG completion (Bordes et al. 2013), or entity classification (Steenwinckel et al. 2022). However, more complex tasks such as query answering (QA) (Ojokoh and Adebisi 2018) or interactive recommendation (He, Parra, and Verbert 2016) that can potentially leverage KG content may generally require more complex combinations of logical operators (conjunction ($\wedge$), disjunction ($\vee$), negation ($\neg$)) and quantifiers (existential ($\exists$), universal ($\forall$)) over atomic KG triples.

Concrete examples of such complex evidence involving logical operators can be observed in the task of KG-based query answering (KGQA) (Zhang et al. 2018; Jain 2016) with queries such as *"which diseases cause fever ($\wedge$) are transmitted by an ($\exists$) insect endemic to New Zealand ($\vee$) Australia"*. Or in interactive movie recommendation (Sun and Zhang 2018), a user may ask for a *"Korean ($\wedge$) action movie ($\wedge$) featuring a ($\exists$) female actress practicing a ($\exists$) martial art"*. Although works in KGQA have proposed frameworks to answer complex KG-based queries involving logical operators (Arakelyan et al. 2020; Ren, Hu, and Leskovec 2020), their methodologies only operate in a single-shot manner, leaving open the question of how to reason over KGs in a sequentially interactive setting. Note that in both above examples, additional evidence could be provided in an incremental setting, e.g., information about the presence of other symptoms for narrowing down the answer in the KGQA example, or user critiques regarding different genres of recommended movies to narrow down the set of preferred items in the interactive recommendation example.

The sequential inference required in such applications has a high resemblance to the belief tracking problem for which Bayesian Inference methods have been extensively leveraged (Roumeliotis and Bekey 2000; Henderson 2015). Inspired by this similarity, we cast the problem of reasoning over unobserved KG entities with incremental complex evidence as a Bayesian filtering problem, and propose a novel framework for Bayesian Inference with Complex Knowledge Graph Evidence (BIKG) which is based on
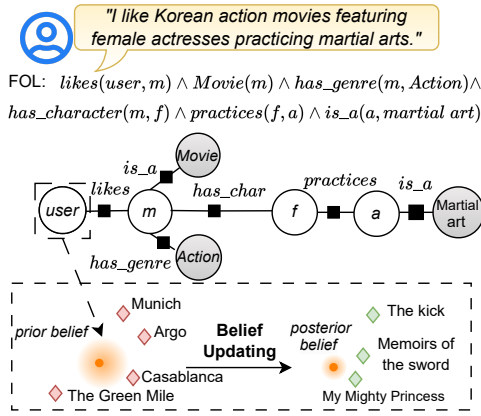
Figure 1: A schematic illustration of inference using BIKG framework for a movie recommender AI chatbot.

Knowledge-based Model Construction (KBMC) (Wellman, Breese, and Goldman 1992). In particular, we construct a Markov Random Field (MRF) regarding the complex logical KG evidence instantiated with tensor-factorized likelihood representations. We then propose an efficient solution to perform Bayesian updating over an unknown entity in light of this incremental evidence. Finally, due to the limitations raised by applying existential quantification for multipath reasoning, we introduce the *Marginal* quantifier to the BIKG framework and provide a closed-form solution for its application. To make this framework concrete, Figure 1 provides a schematic view of the graphical model constructed to update the user belief in the aforementioned example.

We experimentally evaluate the reasoning capability of BIKG in two diverse tasks: incremental KGQA and sequential recommendation, involving five KGs in total. We compare BIKG with different variants of CQD (Arakelyan et al. 2020), a strong methodology for complex KGQA that utilizes T-norm and T-conorm operators from fuzzy logic. The results demonstrate that BIKG outperforms CQD for incorporating complex evidence in incremental reasoning tasks.

Background and Related Work

Knowledge Graphs

A Knowledge Graph (KG) is a graph-structured knowledge base that stores interlinked information about objects, events, persons, abstract concepts etc. (Paulheim 2017). A KG is represented as a directed graph $K = (\mathcal{E}, \mathcal{R})$ where $\mathcal{E}$ is the set of entities (nodes) and $\mathcal{R}$ the set of relations (edges) between entities in $\mathcal{E}$. It can also be viewed as a set of triple facts (h, r, t) which indicate existence of the relationship r, from entity h to entity t. Despite the colossal amount of information that real-world KGs contain, they are severely incomplete and noisy (Pujara, Augustine, and Getoor 2017). To overcome this issue, several methods have been proposed for Knowledge Graph Completion (KGC) that aim to infer plausible relations based on the observed KG information. Works in this area include variations of neural networks for learning embeddings for KG entities and

relations and estimating the plausibility of a triplet (Trouillon et al. 2016; Xiao et al. 2015), symbolic methods that try to extract logical rules governing KG triples (Chen, Wang, and Goldberg 2016; Galárraga et al. 2013), and neurosymbolic methods aiming to combine and preserve advantages of both approaches (Guo et al. 2016, 2018).

Knowledge Graph Embeddings

Knowledge Graph Embedding (KGE) is a major line of work in KGC research. KGE methodologies assign embeddings to the KG nodes and relations in a specific representation space and leverage a *scoring function* that maps the learned representations for a candidate triple to a *plausibility score*. This score is an estimate for the plausibility of existence of that triple in the KG. The main objective of KGE methods is to train embeddings in a way to increase (decrease) the plausibility score of correct (incorrect) triples.

In this work, we use the SimplE KGE method (Kazemi and Poole 2018) because of its efficient computation and simple multiplicative scoring function that aligns with the Bayesian formulation by introducing a tensor factorized likelihood. In order to account for directionality of KG relations, SimplE assigns two embedding vectors $\mathbf{z}_e \in \mathbb{R}^N$ and $\mathbf{z}_{e_{\text{inv}}} \in \mathbb{R}^N$ to each entity e and two vectors $\mathbf{z}_r \in \mathbb{R}^N$ and $\mathbf{z}_{r_{\text{inv}}} \in \mathbb{R}^N$ to each relation r, and for a triple (h, r, t), defines its scoring function as $\Phi(h, r, t) = \frac{1}{2}(\langle \mathbf{z}_h, \mathbf{z}_r, \mathbf{z}_{t_{\text{inv}}} \rangle + \langle \mathbf{z}_t, \mathbf{z}_{r_{\text{inv}}}, \mathbf{z}_{h_{\text{inv}}} \rangle)$. Here, $\langle \mathbf{v}, \mathbf{w}, \mathbf{x} \rangle = (\mathbf{v} \odot \mathbf{w}) \cdot \mathbf{x}$ where $\odot$ and $\cdot$ denote Hadamard and dot product respectively. The model parameters are trained by forming a batch of KG triples and negative samples and solving the following problem:

$$\min_{\theta} \sum_{m=1}^{M} \log(1 + \exp(-y_m . \Phi(h, r, t)_m) + \lambda \|\theta\|_2^2, \quad (1)$$

in which $y \in \{+1, -1\}$ denotes the triple label, θ stands for the model parameters, λ is the regularization hyperparameter, and m is the iterator of batch samples.

Complex Question Answering on KGs

A majority of methods proposed for KGC have focused on answering atomic queries over a KG by predicting missing links. In recent years, a group of works have tried to extend this setting to answering complex FOL queries, i.e. atomic queries that are combined by FOL operators and quantifiers. Due to the incompleteness of KGs, most complex queries cannot be answered by graph traversal algorithms, and thus, query answering methods are required that possess the ability to account for intermediate unobserved triples that are highly plausible. Inspired by the generalization ability provided by KGE methods, a majority of works seek answers by embedding queries and sets of entities to geometric shapes (e.g., boxes or cones) (Hamilton et al. 2018; Zhang et al. 2021; Ren, Hu, and Leskovec 2020; Choudhary et al. 2021) or probability distributions (Ren and Leskovec 2020). These methods require a large set of training queries, substantial training computation, and are unable to answer Out Of Distribution (OOD) queries. In response, CQD (Arakelyan et al. 2020) has proposed means to answer complex queries using gradient-based optimization and beam search leveraging
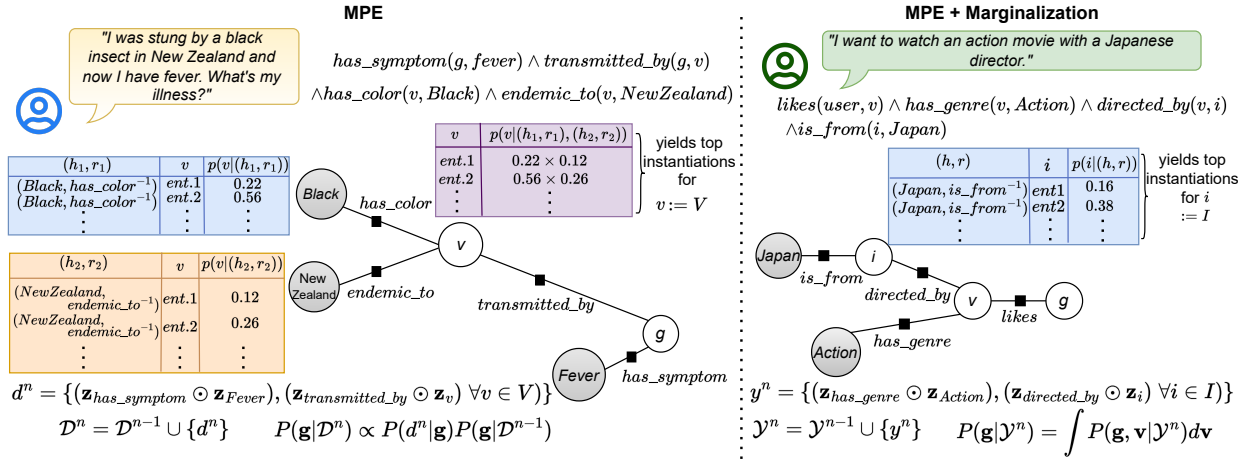
Figure 2: An overview of inference procedure using BIKG for the MPE approach with and without using the Marginal quantifier.

only entity and relation embeddings. All of these methods provide means to reason over KG with complex evidence in a single-shot setting, and a methodology for performing incremental reasoning over KGs remains missing.

Bayesian Inference with Complex KG Evidence

Our proposed Bayesian Inference with Complex Knowledge Graph (BIKG) framework is based on constructing a Markov Random Field (MRF) following the structure of the KG subgraph encompassing evidence and target entities. This formalism enables us to adopt a probabilistic perspective on KG entities and relations, framing the KG inference problem as a probabilistic inference task over the created MRF. This idea is founded on KBMC (Wellman, Breese, and Goldman 1992) framework that crafts a Bayesian Network (BN) for performing inference on knowledge bases, but due to limitations of BNs (Koller and Pfeffer 1997), we develop an MRF to formalize the probabilistic inference problem.

In the primary phase of BIKG, we train the SimplE KGE model to obtain embeddings for all KG relations and entities. Next, upon observing complex evidence concerning an entity of interest, we construct the corresponding MRF as shown in Figure 2. In brief, the overall objective of BIKG is to exploit the observed evidence to update the belief distribution maintained over the target entity.

The constructed MRF corresponding to the observations at an increment n may involve three types of nodes: (i) *Target node* (g) which is the entity of interest necessitating inference, (ii) *Evidence nodes* (d^n) corresponding to the observed KG entities, and (iii) *Intermediate variable nodes* $(\mathcal{V})$ that lie between evidence and target nodes and could be partitioned into three subsets: variable nodes directly connected to evidence nodes $(\mathcal{V}_d)$, variable nodes connected to other variable nodes $(\mathcal{V}_v)$, and variable nodes connected to the target node $(\mathcal{V}_g)$ such that $\mathcal{V}_d \cup \mathcal{V}_v \cup \mathcal{V}_g = \mathcal{V}$. We instantiate the observed evidence nodes with $\mathbf{z}_d^n$, the embedding vectors corresponding to their KG entities. Also, the factors between MRF nodes are the tensor-factorized embedding vectors of

the corresponding relations $\mathbf{z}_r^n$.

Following the general setting of Bayesian inference, we maintain a belief distribution over the target variable node $P(\mathbf{g})$ and aim to incrementally update it conditioned on the observed complex evidence to obtain its posterior distribution. Denoting the set of observations up to increment $n-1$ by $\mathcal{D}^{n-1}$ and the new observation made at increment n as d^n, the posterior belief distribution $P(\mathbf{g}|\mathcal{D}^n)$ is found as:

$$P(\mathbf{g}|\mathcal{D}^n) \propto P(d^n|\mathbf{g}, \mathcal{D}^{n-1})P(\mathbf{g}|\mathcal{D}^{n-1})$$
$$\propto P(d^n|\mathbf{g})P(\mathbf{g}|\mathcal{D}^{n-1}), \quad (2)$$

where $\mathcal{D}^n = \mathcal{D}^{n-1} \cup \{d^n\}$. For the sake of generality, we assume having no prior information about the target node and consider a zero-centered Gaussian distribution with a relatively large covariance matrix, $P(\mathbf{g}) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ to represent the high uncertainty in our initial belief which will be reduced after making sequential observations. In order to update this belief, we have to propagate the observations from evidence nodes toward the target node on the MRF, for which we propose two different approaches.

Most Probable Explanation for Existentially Quantified Variables

The unobserved intermediate variable nodes pose a challenge to the propagation of the observed evidence toward the target node. To handle these variables, a possible approach is to instantiate them with the KG entities that are in strongest alignment with the observations. To this end, as indicated in Figure 2, we traverse the MRF from evidence nodes toward the target, and when confronting an intermediate variable node, we use the scoring function of the tensor factorization-based KGE model to obtain the plausibility scores for each candidate entity to serve as its instantiation. These scores are compiled into a table that we call *Conditional Plausibility Table* (CPT) which is analogous to conditional probability tables but its values may not be valid probability distributions. When coming across a conjunction, we require to find the joint plausibilities of the conditioning sides, which due

to the tree structure of the MRF, is obtained by multiplying the CPTs. Since finding the most plausible sequence of intermediate variables requires a combinatorial search within the joint CPTs, we restrain the search space by only keeping the values of the top-k candidates for each variable node resulting in a *beam* of most plausible instantiations. By searching through this beam, we identify the set of intermediate entities that result in the highest plausibility scores, a procedure akin to the Most Probable Explanation (MPE) concept in Bayesian inference. Formally, we aim to find instantiations of variable nodes that are solutions to the following problem:

$$\mathcal{V}^* = \arg\max_{\mathcal{V}} \prod_{v_d \in \mathcal{V}_d, d \in \{d^n\}} \Phi(d, r_{d,v}, v_d) \prod_{v_i, v_j \in \{\mathcal{V}_v \cup \mathcal{V}_g\}} \Phi(v_i, r_{i,j}, v_j),$$
(3)

where $r_{i,j}$ denotes the relation between nodes i and j. This approach resembles the CQD-beam algorithm (Arakelyan et al. 2020) when using Product t-norm for complex query answering modulo a key difference. Since CQD-beam aims to find the answer nodes in a single-shot manner, it continues the beam search to instantiate the target node, but since BIKG aims to incrementally update the target belief distribution, we halt the beam search at $\mathcal{V}_g$ nodes and instantiate the variable nodes with their final top k candidates. In fact, we are only interested in the instantiations of $\mathcal{V}_g$ nodes to use them with evidence nodes directly connected to the target node to update the target belief distribution by performing a Bayesian update following Equation 2.

We address disjunctive evidence by first transforming it to Disjunctive Normal Form (DNF) (Davey and Priestley 2002) and follow the MPE procedure to instantiate the variable nodes for each conjunctive constituent. Since the truth of any conjunctive constituent of a DNF evidence will make the whole statement true, instead of finding a single updated belief, we find a candidate posterior belief for each conjunctive constituent resulting in a set of possible posteriors.

Marginalization: A Probabilistic Quantifier for Multi-path Reasoning

Using the MPE approach to obtain posterior belief by instantiating variable nodes with top candidates is apt for functional or nearly functional relations, i.e. cases where the number of true candidate tails for the *(head, relation)* pair is limited. This approach provides a probabilistic view of existential quantification for all intermediate variable nodes. However, for unobserved variables that can be correctly instantiated by multiple entities, instantiating them with a limited number of top candidates imposes clear limitations. For instance, in examples provided in Figure 2, the number of *Black insects* that are *endemic_to, New Zealand* is expected to be much lower than *action* movies with *Japanese* directors. We later verify this hypothesis in real-world KGs. In addition, the *likes* relation in the recommendation example is inherently different from other objective relations since users possess various multi-modal tastes and their preference for movies can stem from multiple diverse reasons.

As a refined solution to overcome shortcomings of the existential quantification view of the MPE, we introduce the *Marginal* quantifier into BIKG to allow for *"multi-path"* probabilistic reasoning. Since several true candidates for such an intermediate variable are in agreement with the observed evidence, instead of restricting the density of its probability distribution $P(\mathbf{v})$ to a limited number of candidates, we preserve its entire distribution and eventually marginalize it out from its joint distribution with the target belief.

$$P(\mathbf{g}|\mathcal{D}^{n+1}) = \int P(\mathbf{g}, \mathbf{v}|\mathcal{D}^{n+1}) d\mathbf{v}.$$
(4)

In our framework, we assume the distribution of all variables and factors to be Gaussian and update the target belief distribution using Gaussian Belief Propagation (BP) (Weiss and Freeman 1999). First, unobserved variable nodes corresponding to functional relations are instantiated following the MPE approach. For Marginally quantified variables, we use the set of neighboring evidence and instantiated variable nodes, $\mathcal{Y}$, to update the unobserved variable distribution. Considering instantiations of $y \in \mathcal{Y}$ to be distributed as $\mathbf{y} \sim \mathcal{N}^{-1}(\mathbf{h_y}, \mathbf{J_y})$ and the prior distribution over the unobserved variable node v to be $\mathbf{v} \sim \mathcal{N}^{-1}(\mathbf{h_v}, \mathbf{J_v})$ in which $\mathbf{h} \in \mathbb{R}^N$ and $\mathbf{J} \in \mathbb{R}^{N \times N}$ are the potential vector and precision matrix of the corresponding Gaussian distribution in the canonical form (Bishop and Nasrabadi 2006), we first update $P(\mathbf{v})$ conditioned on this evidence.

$$P(\mathbf{v}|\mathbf{y}) = \mathcal{N}^{-1}(\mathbf{h_v} + \mathbf{h_y}, \mathbf{J_v} + \mathbf{J_y}).$$
(5)

Next, we pass on the updated variable distribution to the target node g with prior $P(\mathbf{g}) = \mathcal{N}^{-1}(\mathbf{h_g}, \mathbf{J_g})$ and form the joint distribution of $\mathbf{g}$ and $\mathbf{v}$ conditioned on the evidence.

$$P(\mathbf{g}, \mathbf{v}|\mathbf{y}) \propto \exp\{-\frac{1}{2}\mathbf{g}^T \mathbf{J_g} \mathbf{g} +$$
$$\mathbf{h_g}^T \mathbf{g} - \frac{1}{2}\mathbf{v}^T \mathbf{J_v} \mathbf{v} + \mathbf{h_v}^T \mathbf{v} - \mathbf{g}^T \mathbf{J_{g,v}} \mathbf{v}\}.$$
(6)

At this point, the challenge is to obtain $\mathbf{J_{g,v}}$. Assuming Gaussian likelihood and considering the scoring function of SimplE, the likelihood factor between $\mathbf{g}$ and $\mathbf{v}$ becomes $\exp\{\langle \mathbf{g}, \mathbf{r_{g,v}}, \mathbf{v}\rangle\}$ in which $\mathbf{r_{g,v}}$ is the embedding vector of the relation between $\mathbf{g}$ and $\mathbf{v}$. Although this term is log-bilinear in $\mathbf{g}$ and $\mathbf{v}$ and would appear to stymie closed-form Gaussian belief propagation, we can rewrite $\langle \mathbf{g}, \mathbf{r_{g,v}}, \mathbf{v}\rangle$ as $\mathbf{g}^T \mathbf{D_{r_{g,v}}} \mathbf{v}$ to obtain $\mathbf{J_{g,v}} = -\mathbf{D_{r_{g,v}}}$, in which $\mathbf{D_{r_{g,v}}}$ is obtained by reshaping $\mathbf{r_{g,v}}$ as a diagonal matrix. Next, in order to obtain $P(\mathbf{g}|\mathbf{y})$, we marginalize the joint distribution found in Equation (6) over $\mathbf{v}$.

$$P(\mathbf{g}|\mathbf{y}) = \int P(\mathbf{g}, \mathbf{v}|\mathbf{y}) d\mathbf{v} \propto \exp\left\{-\frac{1}{2}\mathbf{g}^T \mathbf{J_g} \mathbf{g} + \mathbf{h_g}^T \mathbf{g}\right\}$$
$$\times \int \exp\left\{-\frac{1}{2}\mathbf{v}^T \mathbf{J_v} \mathbf{v} + \mathbf{h_v}^T \mathbf{v} - \mathbf{g}^T \mathbf{J_{g,v}} \mathbf{v}\right\} d\mathbf{v}.$$
(7)

In order to calculate the latter integral, we first write the exponentiated term in matrix form to obtain:

$$P(\mathbf{g}|\mathbf{y}) \propto \exp\left\{-\frac{1}{2}\mathbf{g}^T \mathbf{J_g} \mathbf{g} + \mathbf{h_g}^T \mathbf{g}\right\} \times$$
$$\int \exp\left\{-\frac{1}{2}\begin{bmatrix}\mathbf{g}\\\mathbf{v}\end{bmatrix}^T \begin{bmatrix}\mathbf{0} & \mathbf{J_{g,v}}\\\mathbf{J_{g,v}} & \mathbf{J_v}\end{bmatrix} \begin{bmatrix}\mathbf{g}\\\mathbf{v}\end{bmatrix} + \begin{bmatrix}\mathbf{0}\\\mathbf{h_v}^T\end{bmatrix} \begin{bmatrix}\mathbf{g}\\\mathbf{v}\end{bmatrix}\right\} d\mathbf{v}.$$
(8)

Algorithm 1: BIKG Algorithm

1: **Input:** Evidence at increment n represented in FOL, KG embeddings, $\{P(\mathbf{g}|\mathcal{D}^{n-1})\}$, and quantifier $\in$ $\{MPE, MPE+Marginal\}$.
2: Convert evidence to DNF form.
3: Initialize an empty set of candidate posteriors.
4: **for** each conjunctive constituent **do**
5:     Construct an MRF corresponding to the evidence.
6:     Instantiate observed nodes $d_i \in d^n$ with $\mathbf{z}_{d_i}$.
7:     **for** variable node $v_d \in \mathcal{V}_d$ **do**
8:         Compile CPT with $\Phi(d, r_{d,v}, v_d)$ values.
9:     **end for**
10:     Traverse MRF toward the target node.
11:     **for** variable node $v_i \in \{\mathcal{V}_v \cup \mathcal{V}_g\}$ **do**
12:         Obtain set of neighboring nodes $\mathcal{N}_{v_i}$.
13:         Compile joint CPT with $\prod_{v_j \in \mathcal{N}_{v_i}} \Phi(v_i, r_{i,j}, v_j)$.
14:     **end for**
15:     Solve problem (3) to obtain $\mathcal{V}^* = \mathcal{V}_d^* \cup \mathcal{V}_v^* \cup \mathcal{V}_g^*$.
16:     **for** node $v \in \{\mathcal{V}_d \cup \mathcal{V}_v\}$ **do**
17:         Obtain corresponding v^* from $\mathcal{V}^*$.
18:         Instantiate v with $\mathbf{z}_{v^*}$.
19:     **end for**
20:     **if** quantifier = MPE **then**
21:         **for** node $v_g \in \mathcal{V}_g$ **do**
22:             Obtain corresponding v_g^* from $\mathcal{V}^*$.
23:             Instantiate v_g with $\mathbf{z}_{vg^*}$.
24:         **end for**
25:         Obtain $P(\mathbf{g}|\mathcal{D}^n)$ from (2).
26:     **else if** quantifier = $MPE+Marginal$ **then**
27:         Obtain $P(\mathbf{g}|\mathcal{D}^n)$ from (9) and (10).
28:     **end if**
29:     Add $P(\mathbf{g}|\mathcal{D}^n)$ to set of candidate posteriors.
30: **end for**
31: **Output:** $\{P(\mathbf{g}|\mathcal{D}^n)\}$

Name	#Entities	#Relations	# Avg Legitimate 2p variables
FB15k	14,951	1,345	34.47
FB15k-237	14,505	237	31.85
NELL995	63,361	200	8.46
ML20M + FB	159,606	98	1769.49
LFM-1b + FB	166,830	93	489.69

Table 1: Summary of Dataset Statistics

which the model is given a single piece of evidence represented through a complex FOL query. We extend this task to the sequential setting in which evidence is observed in an incremental manner, i.e. at each session, the model is provided with a new piece of evidence. A clear real-world example of this task is reasoning about patient diagnosis by observing medical test results, which often come sequentially as it is infeasible to take all possible tests at once.

Critiquing with Complex Feedback Critiquing is a common conversational recommendation paradigm in which the model tries to adapt its recommendations according to the user's feedback in an iterative process (Chen and Pu 2012; Toroghi et al. 2023). In each critiquing session, the system recommends a set of items to the user, and the user either accepts them or provides feedback regarding their preference toward a subset of items or item attributes. The system leverages the observed evidence of the user's interests to adjust its recommendations. Existing critiquing methods allow for simple item attributes (Wu et al. 2019; Yang, Shen, and Sanner 2021), but using BIKG, we extend this task to define the new problem of reasoning over user's interests with complex knowledge-based critiques. In this task, in each critiquing session, the user provides complex FOL feedback to the recommender, and the recommender has to adjust its belief over user interests and recommend new items accordingly.

Since the term inside the integral is in the form of a jointly Gaussian distribution, its marginal probability distribution can be obtained using the Schur complement. Thus, we have $P(\mathbf{g}|\mathbf{y}) \sim \mathcal{N}^{-1}(\hat{\mathbf{h}_\mathbf{g}}, \hat{\mathbf{J}_\mathbf{g}})$, where

$$\hat{\mathbf{h}_\mathbf{g}} = \mathbf{h_g} - \mathbf{J_{g,v}}(\mathbf{J_v})^{-1}(\mathbf{h_v}), \quad (9)$$

$$\hat{\mathbf{J}_\mathbf{g}} = \mathbf{J_g} - \mathbf{J_{g,v}}(\mathbf{J_v})^{-1}\mathbf{J_{g,v}}. \quad (10)$$

The BIKG algorithm is summarized in Algorithm 1.

Experiments and Evaluation

We evaluate BIKG[1] on two different tasks involving inference using complex KG-based evidence: sequential complex KGQA and critique-based conversational recommendation.

Task Description

Sequential Query Answering While complex query answering on Knowledge Graphs (KGs) is a well-established field, current works focus on the single-shot paradigm in

Datasets

Sequential Query Answering We evaluate BIKG on three KGs: FB15k (Bordes et al. 2013), FB15k-237 (Toutanova and Chen 2015), and NELL995 (Xiong, Hoang, and Wang 2017). These KGs are established datasets frequently utilized in complex KGQA works. Since queries presented in Ren, Hu, and Leskovec (2020) that are commonly employed in the literature do not support sufficient evidence for all query types to be used in multiple sequences, we extract 3000 queries for each of their query types and divide its evidence nodes into three sessions. For instance, for each 2-intersection query, we extract 6 evidence nodes to use 2 of them per session. In generating queries, we ensure that at least one part of each query is excluded from the training set of KG so that a graph traversal method on the training KG cannot answer any query.

Critiquing with Complex Evidence Given the absence of pre-existing datasets for sequential critiquing with complex evidence, we extract KG-based queries for items from two frequently utilized recommendation datasets: Movielens 20M (Harper and Konstan 2015) and LFM-1b (Schedl
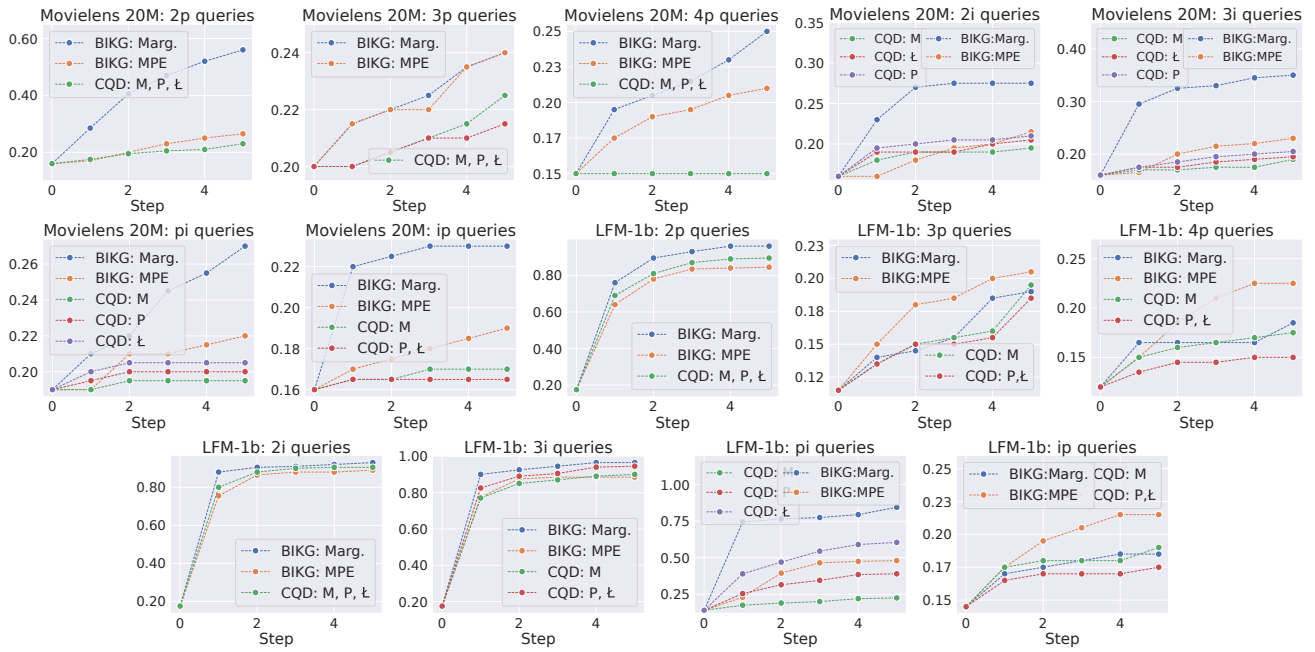
---

[1]https://github.com/atoroghi/BIKG

Figure 3: Comparison of improvement in hit rate@10 during the critiquing sessions for various chain types. CQD-{M, Ł, P} refer to variants of CQD using Minimum, Łukasiewicz, and Product T-norms, and *Marg.* refers to the Marginal quantifier.

2016). To do so, we first use entity matching results provided in Zhao et al. (2019) to match items from each dataset to the corresponding entities from Freebase KG (Bollacker et al. 2008) and extract a number of related triples up to two hops away from each item. Next, by adding a node to the KG for each user in the dataset, we form a user-item KG (Yu et al. 2013). Finally, taking the user-item pairs from the validation and test sets, we extract queries for each item following the same procedure as query extraction for complex query answering and extract 1000 queries for each type. Note that since critiquing queries must contain both user and item nodes, the simplest query type in this task is 2p. A summary of the statistics for all five KGs is provided in Table 1.

Evaluation

For the sequential query answering task, we perform Bayesian inference following the MPE approach and at each increment, select the closest entity to the center of the posterior target distribution as the answer. In critiquing experiments we also use the marginal quantifier for item node to obtain the posterior user belief distribution. For disjunctive queries, we average the scores obtained from all candidate posteriors. To validate our hypothesis regarding the inadequacy of the existential quantifier for the recommendation task, we compare the number of KG entities that are valid instantiations for the variable node in the 2p queries for all datasets and report them in Table 1. This number is substantially higher for the recommendation KGs, and instantiating the variable node corresponding to the item with a few top candidates leads to the oversight of other legitimate entities.

We use CQD (Arakelyan et al. 2020), one of the strongest complex KGQA models that uses fuzzy T-norm and T-

conorm operators as the comparison baseline and use hit rate at 10 as the evaluation metric. The performance of CQD depends on the used T-norm, so we evaluate it with three common T-norms: Minimum, Łukasiewicz, and Product. We use the CQD-beam variant of CQD due to its stronger performance and closer workflow to BIKG. To ensure a fair comparison, since CQD is tailored to answer queries by considering all pieces of evidence in single-shot, we accumulate observations and at each session, present evidence from previous queries to it too. Contrarily, as a Bayesian methodology, BIKG is expected to encompass a historical record of past observations within its belief and thus, it is only provided observations from the most current session.

Results

Sequential Query Answering Sequential query answering results are provided in Table 2, indicating that BIKG outperforms variations of CQD on all datasets, most query types, and most sessions. The superiority of BIKG is more evident in queries including disjunction (*2u* and *up*) and projection (*2p* and *3p*). CQD excels in infrequent cases, often confined to one dataset and BIKG performs at least on par with it on other datasets. These results indicate the limitations of T-norms as a statistically plausible method to integrate multiple evidence pieces into the target beliefs, while the Bayesian approach effectively overcomes this deficiency.

Critiquing with Complex Evidence Multi-step critiquing results are shown in Figure 3 which depict that BIKG excels CQD in capturing user preferences by reasoning over complex evidence observed during critiquing sessions. While employing BIKG with the MPE approach often outperforms CQD, using the Marginal quantifier leads to a more consid-

FB15k

Method	1p			2p			3p			2i			2u	
	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2
CQD-M	**0.654**	0.685	0.686	**0.592**	0.59	0.592	0.478	0.481	0.481	0.717	**0.852**	**0.883**	0.737	0.716
CQD-Ł	**0.654**	**0.696**	**0.718**	**0.592**	0.593	0.59	**0.482**	0.486	0.484	**0.745**	0.842	0.875	0.61	0.593
CQD-P	**0.654**	0.693	0.713	**0.592**	0.59	0.593	**0.482**	0.486	0.484	**0.745**	0.848	0.878	0.741	0.728
BIKG	**0.654**	**0.696**	**0.718**	**0.592**	0.624	0.649	**0.482**	0.503	0.515	**0.745**	0.842	0.875	**0.744**	**0.773**

Method	2u	3i			pi			ip			up		
	S#3	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2	S#3
CQD-M	0.717	0.935	0.969	0.945	0.526	0.111	0.554	0.35	0.363	0.326	0.386	0.379	0.368
CQD-Ł	0.639	**0.966**	**0.986**	0.95	**0.588**	0.304	**0.717**	0.35	0.388	0.406	0.387	0.379	0.392
CQD-P	0.728	0.963	0.99	0.952	0.574	0.179	0.694	0.35	0.388	0.38	0.399	0.313	0.41
BIKG	**0.783**	**0.966**	**0.986**	**0.996**	**0.588**	**0.662**	**0.717**	**0.455**	**0.445**	**0.465**	**0.419**	**0.447**	**0.448**

FB15k-237

Method	1p			2p			3p			2i			2u	
	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2
CQD-M	**0.522**	**0.587**	**0.604**	0.472	0.468	0.46	0.414	0.415	0.412	0.555	0.682	0.719	0.688	0.656
CQD-Ł	**0.522**	0.572	0.595	**0.479**	0.473	0.479	**0.416**	0.419	0.415	**0.661**	**0.744**	**0.775**	0.541	0.513
CQD-P	**0.522**	0.548	0.603	0.447	0.444	0.457	**0.416**	0.419	0.415	0.649	0.741	0.774	0.691	0.675
BIKG	**0.522**	0.572	0.595	0.474	**0.501**	**0.513**	**0.416**	**0.448**	**0.462**	**0.661**	**0.744**	**0.775**	**0.696**	**0.729**

Method	2u	3i			pi			ip			up		
	S#3	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2	S#3
CQD-M	0.658	0.851	0.898	0.857	0.369	0.097	0.407	0.145	0.142	0.15	0.22	0.225	0.238
CQD-Ł	0.545	0.9	**0.943**	0.896	**0.41**	0.137	**0.473**	0.146	0.143	0.159	0.219	0.228	0.249
CQD-P	0.665	**0.903**	0.942	0.89	0.393	0.11	0.459	0.146	0.143	0.153	0.232	0.224	0.258
BIKG	**0.738**	0.9	**0.943**	**0.953**	**0.41**	**0.457**	**0.473**	**0.176**	**0.174**	**0.171**	**0.275**	**0.282**	**0.291**

NELL995

Method	1p			2p			3p			2i			2u	
	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2
CQD-M	**0.607**	0.654	0.645	0.408	0.397	0.398	0.34	0.359	0.358	0.774	0.781	0.762	0.816	0.81
CQD-Ł	**0.607**	**0.673**	**0.71**	0.417	0.411	0.408	**0.341**	0.359	0.358	**0.83**	**0.879**	**0.893**	0.687	0.653
CQD-P	**0.607**	0.621	0.693	**0.422**	0.41	0.408	**0.341**	0.359	0.358	0.8	0.845	0.858	0.823	0.812
BIKG	**0.607**	**0.673**	**0.71**	**0.422**	**0.451**	**0.472**	**0.341**	**0.379**	**0.394**	**0.83**	**0.879**	**0.893**	**0.824**	**0.838**

Method	2u	3i			pi			ip			up		
	S#3	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2	S#3	S#1	S#2	S#3
CQD-M	0.8	0.876	0.871	0.869	0.513	0.251	0.498	0.296	0.275	0.277	0.323	0.316	0.322
CQD-Ł	0.671	**0.989**	**0.996**	0.943	**0.567**	0.434	**0.638**	0.294	0.277	0.33	0.324	0.312	0.336
CQD-P	0.805	0.958	0.953	0.915	0.538	0.278	0.598	0.294	0.277	0.318	0.324	0.312	0.335
BIKG	**0.843**	**0.989**	**0.996**	**0.996**	**0.567**	**0.622**	**0.638**	**0.338**	**0.361**	**0.372**	**0.359**	**0.393**	**0.402**

Table 2: Results of sequential complex query answering task. CQD-{M, Ł, P} refer to variants of CQD using Minimum, Łukasiewicz, and Product T-norms. Each S#$n \in \{1, 2, 3\}$ column refers to the nth query answering session. Letters {p, i, u} in the query type names refer to projection, intersection, and union respectively. The reported evaluation metric is hit rate@10.

erable improvement, consistent with our hypothesis. These results highlight the significance of the muti-path reasoning capability of BIKG offered by the Marginal quantifier.

Conclusion

We proposed BIKG, a Bayesian framework for incremental reasoning over KGs with complex logical evidence. BIKG takes a probabilistic perspective on KG inference by con-structing an MRF corresponding to observations, inferring unobserved variables using the MPE methodology, and performing closed-form multi-path reasoning by incorporating the Marginal quantifier. The promising performance of BIKG on sequential KGQA and interactive recommendation tasks demonstrates its efficacy for performing incremental reasoning, paving the way for versatile future extensions to other logical systems and applications in different domains.

References

Arakelyan, E.; Daza, D.; Minervini, P.; and Cochez, M. 2020. Complex query answering with neural link predictors. *arXiv preprint arXiv:2011.03459*.

Bishop, C. M.; and Nasrabadi, N. M. 2006. *Pattern recognition and machine learning*, volume 4. Springer.

Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; and Taylor, J. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, 1247–1250.

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

Chen, L.; and Pu, P. 2012. Critiquing-based recommenders: survey and emerging trends. *User Modeling and User-Adapted Interaction*, 22: 125–150.

Chen, Y.; Wang, D. Z.; and Goldberg, S. 2016. ScaLeKB: scalable learning and inference over large knowledge bases. *The VLDB Journal*, 25(6): 893–918.

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

Davey, B. A.; and Priestley, H. A. 2002. *Introduction to lattices and order*. Cambridge university press.

Galárraga, L. A.; Teflioudi, C.; Hose, K.; and Suchanek, F. 2013. AMIE: association rule mining under incomplete evidence in ontological knowledge bases. In *Proceedings of the 22nd international conference on World Wide Web*, 413–422.

Guo, Q.; Zhuang, F.; Qin, C.; Zhu, H.; Xie, X.; Xiong, H.; and He, Q. 2020. A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering*, 34(8): 3549–3568.

Guo, S.; Wang, Q.; Wang, L.; Wang, B.; and Guo, L. 2016. Jointly embedding knowledge graphs and logical rules. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, 192–202.

Guo, S.; Wang, Q.; Wang, L.; Wang, B.; and Guo, L. 2018. Knowledge graph embedding with iterative guidance from soft rules. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Hamilton, W.; Bajaj, P.; Zitnik, M.; Jurafsky, D.; and Leskovec, J. 2018. Embedding logical queries on knowledge graphs. *Advances in neural information processing systems*, 31.

Harper, F. M.; and Konstan, J. A. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4): 1–19.

He, C.; Parra, D.; and Verbert, K. 2016. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, 56: 9–27.

Henderson, M. 2015. Machine learning for dialog state tracking: A review.

Jain, S. 2016. Question answering over knowledge base using factual memory networks. In *Proceedings of the NAACL student research workshop*, 109–115.

Kazemi, S. M.; and Poole, D. 2018. Simple embedding for link prediction in knowledge graphs. *Advances in neural information processing systems*, 31.

Koller, D.; and Pfeffer, A. 1997. Learning probabilities for noisy first-order rules. In *IJCAI*, 1316–1323.

Ojokoh, B.; and Adebisi, E. 2018. A review of question answering systems. *Journal of Web Engineering*, 717–758.

Paulheim, H. 2017. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, 8(3): 489–508.

Pujara, J.; Augustine, E.; and Getoor, L. 2017. Sparsity and noise: Where knowledge graph embeddings fall short. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, 1751–1756.

Rastogi, N.; and Zaki, M. J. 2020. Personal Health Knowledge Graphs for Patients. In *Workshop–Personal Health Knowledge Graphs (PHKG2020)*.

Ren, H.; Hu, W.; and Leskovec, J. 2020. Query2box: Reasoning over knowledge graphs in vector space using box embeddings. *arXiv preprint arXiv:2002.05969*.

Ren, H.; and Leskovec, J. 2020. Beta embeddings for multi-hop logical reasoning in knowledge graphs. *Advances in Neural Information Processing Systems*, 33: 19716–19726.

Roumeliotis, S. I.; and Bekey, G. A. 2000. Bayesian estimation and Kalman filtering: A unified framework for mobile robot localization. In *Proceedings 2000 ICRA. Millennium conference. IEEE international conference on robotics and automation. Symposia proceedings (Cat. No. 00CH37065)*, volume 3, 2985–2992. IEEE.

Schedl, M. 2016. The lfm-1b dataset for music retrieval and recommendation. In *Proceedings of the 2016 ACM on international conference on multimedia retrieval*, 103–110.

Steenwinckel, B.; Vandewiele, G.; Weyns, M.; Agozzino, T.; Turck, F. D.; and Ongenae, F. 2022. INK: knowledge graph embeddings for node classification. *Data Mining and Knowledge Discovery*, 36(2): 620–667.

Sun, Y.; and Zhang, Y. 2018. Conversational recommender system. In *The 41st international acm sigir conference on research & development in information retrieval*, 235–244.

Toroghi, A.; Floto, G.; Tang, Z.; and Sanner, S. 2023. Bayesian Knowledge-Driven Critiquing with Indirect Evidence. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '23, 1838–1842. New York, NY, USA: Association for Computing Machinery. ISBN 9781450394086.

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

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

Weiss, Y.; and Freeman, W. 1999. Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Advances in neural information processing systems*, 12.

Wellman, M. P.; Breese, J. S.; and Goldman, R. P. 1992. From knowledge bases to decision models. *The Knowledge Engineering Review*, 7(1): 35–53.

Wu, G.; Luo, K.; Sanner, S.; and Soh, H. 2019. Deep language-based critiquing for recommender systems. In *Proceedings of the 13th ACM Conference on Recommender Systems*, 137–145.

Xiao, H.; Huang, M.; Hao, Y.; and Zhu, X. 2015. TransA: An adaptive approach for knowledge graph embedding. *arXiv preprint arXiv:1509.05490*.

Xiong, W.; Hoang, T.; and Wang, W. Y. 2017. Deeppath: A reinforcement learning method for knowledge graph reasoning. *arXiv preprint arXiv:1707.06690*.

Yang, H.; Shen, T.; and Sanner, S. 2021. Bayesian Critiquing with Keyphrase Activation Vectors for VAE-based Recommender Systems. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2111–2115.

Yu, X.; Ren, X.; Gu, Q.; Sun, Y.; and Han, J. 2013. Collaborative filtering with entity similarity regularization in heterogeneous information networks. *IJCAI HINA*, 27.

Zhang, Y.; Dai, H.; Kozareva, Z.; Smola, A.; and Song, L. 2018. Variational reasoning for question answering with knowledge graph. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Zhang, Z.; Wang, J.; Chen, J.; Ji, S.; and Wu, F. 2021. Cone: Cone embeddings for multi-hop reasoning over knowledge graphs. *Advances in Neural Information Processing Systems*, 34: 19172–19183.

Zhao, W. X.; He, G.; Yang, K.; Dou, H.; Huang, J.; Ouyang, S.; and Wen, J.-R. 2019. Kb4rec: A data set for linking knowledge bases with recommender systems. *Data Intelligence*, 1(2): 121–136.