A Notations and mathematical proofs

A.1 Notations

Table 3: Notations.

Notation	Description
$\mathcal{K}$	The knowledge base
$\mathcal{E}$	The entity set
$\mathcal{R}$	The relation set
$\mathcal{O}, \mathcal{H}$	The set of observed and unobserved facts
$\boldsymbol{x}, \boldsymbol{y}$	The assignments of $\mathcal{O}$ and $\mathcal{H}$, respectively
$\{F_q, W_q\}_{q=1}^m$	The set of logic rules and attached weights
$\mathcal{I}_q^-, \mathcal{I}_q^+$	The index set of premise atoms and conclusion atoms of rule F_q, respectively
$A, B, ...$	Variables in logic rules
$\{G_q^{(j)}, j \in t_q\}$	All ground formulas created by the q_{th} logic rule
$\Phi_q(\boldsymbol{y}, \boldsymbol{x})$	The sum of potentials of all ground formulas of F_q
θ	The embedding parameters

A.2 Derivation of rule weight gradient

Given

$$P_{\boldsymbol{w}}^*(\boldsymbol{y} \mid \boldsymbol{x}) = \prod_{i=1}^n P^*\left(y_i \mid \mathrm{MB}\left(y_i\right), \boldsymbol{x}\right) = \prod_{i=1}^n \frac{\exp\left[-f_{\boldsymbol{w}}^i\left(y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x}\right)\right]}{Z_i(\boldsymbol{W}, y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x})},$$

$$Z_i(\boldsymbol{W}, y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x}) = \int_{y_i} \exp\left[-f_{\boldsymbol{w}}^i\left(y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x}\right)\right], \quad f_{\boldsymbol{w}}^i = \sum_{q=1}^m W_q \sum_{j=1}^{n_q} \mathbf{1}_{\{y_i \to G_q^{(j)}\}} d(G_q^{(j)}),$$

we have

$$\frac{\partial \log P^*(\boldsymbol{y} \mid \boldsymbol{x})}{\partial W_q} = \sum_{i=1}^n \frac{\partial \log P^*(y_i \mid \mathrm{MB}\left(y_i\right), \boldsymbol{x})}{\partial W_q}. \tag{9}$$

The partial derivative in the left side of Eq. (9) is a summation of n terms, each term represents the partial derivatives of the pseudo-log-likelihood for each y_i, conditioned on its Markov blankets. Each term can be further simplified as follows:

$$\begin{aligned}
&\frac{\partial \log P^*(y_i \mid \mathrm{MB}\left(y_i\right), \boldsymbol{x})}{\partial W_q} \\
=&\frac{\partial \left\{-f_{\boldsymbol{w}}^i\left(y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x}\right) - \log Z_i(\boldsymbol{W}, y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x})\right\}}{\partial W_q} \\
=&\frac{\partial \left\{-f_{\boldsymbol{w}}^i\left(y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x}\right) - \log \int_{y_i} \exp\left[-f_{\boldsymbol{w}}^i\left(y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x}\right)\right]\right\}}{\partial W_q}.
\end{aligned}$$

Here, we can easily get

$$\frac{\partial f_{\boldsymbol{w}}^i\left(y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x}\right)}{\partial W_q} = \sum_j \mathbf{1}_{\{y_i \to G_q^{(j)}\}} d(G_q^{(j)}). \tag{10}$$

To make the writing concise, we replace the right term of Eq. (10) with the following notation:

$$\Psi_{q,MB(i)} = \sum_j \mathbf{1}_{\{y_i \to G_q^{(j)}\}} d(G_q^{(j)}).$$

In this way, we can deduce that:

$$\begin{aligned}&\frac{\partial \log P^*(y_i \mid \mathrm{MB}(y_i), \boldsymbol{x})}{\partial W_q} \\ &= -\Psi_{q,MB(i)} - \frac{1}{Z_i(\boldsymbol{W}, y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x})} \frac{\partial \int_{y_i} \exp\left[-f^i_{\boldsymbol{w}}\left(y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x}\right)\right]}{\partial W_q}.\end{aligned} \tag{11}$$

The partial derivative and the integration in Eq. (11) can be swapped using Lebesgue's dominated convergence theorem, the Eq. (11) thus becomes:

$$\begin{aligned}&\frac{\partial \log P^*(y_i \mid \mathrm{MB}(y_i), \boldsymbol{x})}{\partial W_q} \\ &= -\Psi_{q,MB(i)} - \frac{1}{Z_i(\boldsymbol{W}, y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x})} \int_{y_i} \frac{\partial \exp\left[-f^i_{\boldsymbol{w}}\left(y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x}\right)\right]}{\partial W_q} \\ &= -\Psi_{q,MB(i)} + \int_{y_i} \frac{\exp\left[-f^i_{\boldsymbol{w}}\left(y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x}\right)\right]}{Z_i(\boldsymbol{W}, y_i \cup \boldsymbol{y}_{\setminus i}, \boldsymbol{x})} \Psi_{q,MB(i)} \\ &= -\Psi_{q,MB(i)} + \int_{y_i} P^*\left(y_i \mid \mathrm{MB}(y_i), \boldsymbol{x}\right) \Psi_{q,MB(i)} \\ &= -\Psi_{q,MB(i)} + \mathbb{E}_{y_i \mid \mathrm{MB}}\left[\Psi_{q,MB(i)}\right].\end{aligned}$$

Therefore, the partial derivative of pseudo-log-likelihood with respect to rule weight W_q is computed by:

$$\frac{\partial \log P^*(\boldsymbol{y} \mid \boldsymbol{x})}{\partial W_q} = \sum_{i=1}^{n} \left\{ \mathbb{E}_{y_i \mid \mathrm{MB}} \left[\sum_j \mathbf{1}_{\{y_i \to G_q^{(j)}\}} d(G_q^{(j)}) \right] - \sum_j \mathbf{1}_{\{y_i \to G_q^{(j)}\}} d(G_q^{(j)}) \right\}.$$

A.3 Calculation of number of ground formulas for Kinship datasets

We present a detailed calculation of the number of ground formulas considered by PSL in Kinship datasets as follows.

Given

- a first-order logical rule F_q containing $|\mathcal{I}_q^-|$ premise atoms, and
- a knowledge base containing $|\mathcal{E}|$ number of entities,

the number of variables in F_q is $|\mathcal{I}_q^-| + 1$.

PSL grounds each rule by substituting the variables with all possible entities. The number of ground formulas created by this logic rule F_q on the knowledge base is:

$$|\mathcal{E}|^{|\mathcal{I}_q^-|+1}.$$

Thus the overall ground formulas created by the rule set $\{F_q\}_{q=1}^m$ is:

$$\sum_{q=1}^{m} |\mathcal{E}|^{|\mathcal{I}_q^-|+1}.$$

Given the statistics of Kinship datasets in Table 6, rules statistics are shared across different sizes of Kinship datasets, each dataset contains 12 rules that contain two variables and 9 rules that contain 3 variables. The number of ground formulas considered by PSL is thus computed by:

$$12 \times |\mathcal{E}|^2 + 9 \times |\mathcal{E}|^3. \tag{12}$$

By applying the Eq. (12), we can get the ground formula number for each size of the Kinship dataset, as presented in Table 4:

Table 4: Number of ground formulas of Kinship datasets created by classical grounding method.

Kinship Size	S1	S2	S3	S4	S5
Number of ground formulas	1,373,601	10,853,976	35,798,376	74,671,320	172,162,935

B Experimental details

B.1 Dataset statistics

We list the statistics of the real-world knowledge graph datasets in Table 5 and the synthetic Kinship dataset in Table 6. We present detailed descriptions for each dataset below.

CodeX. The CodeX dataset, recently proposed for knowledge graph completion tasks, is a comprehensive collection extracted from both Wikidata and Wikipedia. This challenging dataset comes in three versions: small (S), medium (M), and large (L), allowing for comprehensive evaluation.

YAGO3-10. YAGO3-10 is a subset of YAGO3 (Suchanek et al., 2007), a large knowledge base completion dataset, with the majority of triples describing attributes of persons, including their citizenship, gender, and profession.

WN18. WordNet 18 (WN18) dataset is one of the most commonly used subsets of WordNet.

WN18RR. WN18RR is a modified version of WN18 designed to be more challenging for knowledge graph reasoning algorithms by removing reverse relations in the knowledge graph.

Kinship. A synthetic dataset, widely used (Zhang et al., 2020; Fang et al., 2023) for evaluating the statistical relational learning ability and the scalability of reasoning algorithms. We use five different sizes of the dataset for evaluating its run time efficiency and parameter scalability, namely Kinship-S1/S2/S3/S4/S5, respectively.

Table 5: Statistics of real-world knowledge base datasets.

Dataset	**#Ent**	**#Rel**	**#Train/Valid/Test**	**#Rules**
CodeX-s	2,034	42	32,888/1,827/1,828	35
CodeX-m	17,050	51	185,584/10,310/10,311	52
CodeX-l	77,951	69	551,193/30,622/30,622	57
YAGO3-10	123,182	37	1,079,040/5,000/5,000	22
WN18	40,943	18	141,442/ 5,000/ 5,000	140
WN18RR	40,943	11	86,835/ 3,034/ 3,134	51

Table 6: Statistics for Kinship datasets of varied sizes (S1-S5).

	S1	S2	S3	S4	S5
Number of rules containing 1 premise atom	12	12	12	12	12
Number of rules containing 2 premise atoms	9	9	9	9	9
Number of predicates	15	15	15	15	15
Number of entities	52	106	158	202	267

B.2 Probabilistic logic reasoning on Kinship Dataset

We assess performance on the Kinship dataset across five different sizes. Due to the full confidence of rules, we only perform inference in this experiment and do not need to update weights. We include

Table 7: Comparative evaluation of reasoning performance on the Kinship dataset.

Algorithms	Ground iteration	AUC-ROC S1	S2	S3	S4	S5
PSL	-	.976±.011	.980±.005	.991±.003	.982±.005	.972±.004
ExpressGNN	-	.957±.002	.921±.001	.959±.004	.940±.001	.989±.004
DiffLogic-RotatE	1	.841±.005	.895±.001	.922±.001	.901±.001	.903±.000
	2	.931±.005	.994±.001	.998±.001	.985±.001	.993±.001
	3	.937±.005	.987±.001	.995±.001	.978±.001	.989±.001
DiffLogic-MLP	1	.567±.099	.537±.041	.507±.024	.503±.018	.504±.014
	2	.956±.032	.997±.002	.999±.003	.999±.001	**.999±.000**
	3	**.982±.014**	**.997±.001**	**.999±.001**	**.999±.000**	**.999±.000**

DiffLogic using two different embedding models, i.e., RotatE and MLP, and evaluate their reasoning performance using RGIG with varied iterations (i.e., 1, 2, 3) for grounding. We include PSL and ExpressGNN as baselines, but we exclude pLogicNet due to its inability to utilize handcrafted rules. Given that the Kinship dataset lacks a validation set, we run each model ten times and report the AUC-ROC statistics from the final epoch of each run on the test set. The results are presented in Table 7, with the best results shown in bold.

B.3 Comparing inference time on Kinship

We evaluate the inference time on the Kinship dataset across five different sizes. We include models in Appendix B.2 for this experiment. For two DiffLogic variants, we only evaluate their inference time when using 3 iterations of RGIG for grounding. All the runtime experiments are conducted in the same machine with configurations as in Table 8. All of these models are implemented in Python, thereby ensuring a fair comparison. The inference time results are displayed in Table 9, with the best results shown in bold.

Table 8: Machine configuration.

Component	Specification
GPU	NVIDIA GeForce RTX 3090
CPU	Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz

Table 9: Comparison of runtime of inference on Kinship.

Algorithms	Grounding iteration	Runtime S1	S2	S3	S4	S5
PSL	-	∼3.6min	∼7.9min	∼12.9min	∼13.5min	∼32min
ExpressGNN	-	∼18.4min	∼19.1min	∼18.9min	∼19.4min	∼20.2min
DiffLogic-RotatE	3	37s	∼1.5min	∼3.2min	∼3.6min	∼4min
DiffLogic-MLP	3	**21.8s**	**41.5s**	**45s**	**54.4s**	**∼1.2min**