Mitigating Degree Bias: Hybrid Anchors Sampling and Budgeted Paths Pruning for Efficient Rule Mining in KG Completion

Anonymous ACL submission

Abstract

Knowledge graph completion (KGC) predicts missing links in multi-relational knowledge graphs. Existing rule-based KGC methods often sample paths around high-degree entities, resulting in biased rules, weak long-tailed coverage, and high computational cost. Inspired by the modern cognitive framework of Resource-Rational Analysis, which formalizes human cognition as the optimal use of limited computational resources, we propose Hybrid Anchors Sampling and Budgeted Paths Pruning (HASBPP), a lightweight soft rule mining framework. HASBPP simulates resource-rational intelligence via three modules: Hybrid Degree-aware Anchor Sampling (HDAS) acts as optimal information sampling to prevent hub bias; Layer-Budgeted Path Pruning (LBPP) mimics bounded mental simulation to avoid combinatorial explosion; and Stability-Preserving Quality Calibration (SPQC) functions as multi-attribute utility evaluation to comprehensively rank and filter candidate paths. Experiments on WN18RR, UMLS, and YAGO3-10 show that HASBPP produces high-quality soft rules, achieves strong KGC performance, and runs significantly faster than existing baselines.

1 Introduction

Knowledge graphs (KGs) store real-world facts in the form of multi-relational graphs, and have become the backbone of many language-related information systems, including entity disambiguation, question answering, and retrieval-augmented generation (Yang et al., 2022; Kotov, 2016; Mulang et al., 2020). However, even large-scale KGs such as Freebase and YAGO are highly incomplete, with over 70% of possible facts missing (Zamini et al., 2022). This incompleteness fundamentally limits the utility of downstream applications, especially for long-tail entities and sparse relations that rarely appear in existing data (Li et al., 2024a; Shen et al., 2022). Knowledge graph completion (KGC) aims to automatically infer such missing facts, and the choice of inference paradigm has important practical implications: in many high-stakes scenarios, interpretability, robustness, and coverage are as critical as raw predictive accuracy.

Existing KGC approaches fall roughly into two families: embedding-based methods and rule-based methods. Embedding-based models, such as TransE (Bordes et al., 2013) and RotatE (Sun et al., 2019), learn continuous vector representations of entities and relations, enabling smooth generalization and efficient scoring. However, these models are inherently black-box, making their predictions difficult to explain, and they tend to struggle with rare relations where training signals are scarce (Amador-Dominguez et al., 2023; Rossi et al., 2021). In contrast, rule-based reasoning methods extract explicit logical rules, usually in the form of Horn clauses, from the graph. Such rules are human-readable, support transparent inference, and can be more robust to noise when expressed as *soft rules* equipped with confidence scores (Cheng et al., 2021; Zhang et al., 2024).

Despite these advantages, existing rule mining approaches lack the resource-rational efficiency of human cognition. According to the modern cognitive framework of Resource-Rational Analysis (Lieder and Griffiths, 2020), humans overcome combinatorial explosions in memory retrieval through optimal information sampling, bounded mental simulation, and multi-attribute utility evaluation. In contrast, current KGC methods face three practical bottlenecks. First, anchor degree bias arises because naive sampling strategies select starting points uniformly at random or gravitate toward popular entities. Due to the scale-free nature of most KGs, random sampling inevitably gravitates toward high-degree hubs, causing the discovered rules to over-represent frequent relations and neglect long-tail entities (Peng et al., 2025; Li et al., 2024b). Second, path search space ex-

plosion occurs when unconstrained breadth-first search (BFS) or random walks are used to collect candidate paths. With an average out-degree d, the number of explored paths grows as $\mathcal{O}(d^L)$ where L is the path length, quickly becoming computationally intractable (Sun et al., 2024). Third, rule quality assessment is often limited to simple statistics such as support and confidence. These metrics can be inflated by trivial loops or overly generic patterns, and they fail to capture other important dimensions such as relational diversity, path length appropriateness, and transition coherence (Meilicke et al., 2024; Li et al., 2025).

To overcome these limitations, drawing inspiration from Resource-Rational Analysis (Lieder and Griffiths, 2020), we introduce Hybrid Anchors Sampling and Budgeted Paths Pruning (HASBPP), a lightweight soft rule mining framework. HASBPP integrates three modules to simulate resource-rational cognition under bounded computation. First, reflecting optimal information sampling, Hybrid Degree-aware Anchor Sampling (HDAS) limits hub dominance by combining random, degree-stratified, and relation-aware sampling to acquire a balanced set of initial seed facts. Second, analogous to bounded mental simulation, Layer-Budgeted Path Pruning (LBPP) performs a breadth-first search with a hard per-layer width limit and a precomputed reverse index. This reduces the per-anchor computational complexity from exponential $\mathcal{O}(d^L)$ to linear $\mathcal{O}(L \cdot B \cdot d)$. Third, functioning as multi-attribute utility evaluation, Stability-Preserving Quality Calibration (SPQC) evaluates candidate paths based on length appropriateness, relation importance, relation diversity, and sequence coherence, dynamically discarding noisy correlations to yield robust soft rules.

The resulting rules are treated as soft rules, each associated with standard statistical metrics (support, coverage, confidence, PCA confidence), and can be used directly for interpretable link prediction.

We evaluate HASBPP on three widely used KG benchmarks: WN18RR, UMLS, and YAGO3-10. Experimental results demonstrate that HASBPP consistently outperforms strong baselines including symbolic (AMIE), neural (Neural-LP, Ruleformer), neuro-symbolic (NCRL), and LLM-based (ChatRule) methods in terms of rule quality, achieving state-of-the-art support, coverage, confidence, and PCA confidence across datasets. Furthermore, when applying the mined rules to the downstream task of KG completion, HASBPP obtains the best MRR and Hits@K on WN18RR and UMLS, while running orders of magnitude faster than LLM-based alternatives. Ablation studies confirm the necessity of each of the three components, and efficiency analysis shows that HASBPP produces a compact rule set with substantially reduced runtime.

In summary, our main contributions are as follows. First, we propose an efficient soft rule mining framework (HASBPP) that explicitly mitigates anchor degree bias, search space explosion, and single-metric quality assessment through three complementary modules. Second, we conduct extensive experiments on three diverse datasets, demonstrating that HASBPP yields more reliable, diverse, and efficient rules compared to state-of-the-art baselines, and achieves strong downstream link prediction performance. Third, our framework features a lightweight design that does not require neural network training or large language model queries, making it easily deployable for real-world knowledge-intensive NLP applications.

2 Related Work

2.1 Rule Mining for Knowledge Graphs

Early rule mining systems, such as AMIE (Galárraga et al., 2013), follow an inductive logic programming (ILP) paradigm and enumerate closed Horn rules over the entire graph. While they offer solid interpretability, the exhaustive search quickly becomes infeasible on large KGs, and the discovered rules tend to favor frequent relations. To improve scalability, subsequent methods introduced neural guidance into the rule learning process. Neural-LP (Yang et al., 2017) casts rule learning as a differentiable operation over relation matrices, enabling end-to-end optimization of rule confidence, but the reliance on dense tensor representations limits its scalability to graphs with many entities. AnyBURL (Meilicke et al., 2024) adopts an anytime bottom-up sampling strategy, which efficiently discovers rules on large-scale KGs by avoiding exhaustive search, yet its random sampling strategy can still suffer from hub dominance.

More recently, Transformer-based architectures and contrastive learning have been explored for rule mining. Ruleformer (Xu et al., 2022) models relation paths as sequences and applies self-attention to generate context-aware candidate rules, improving semantic coherence. NCRL (Cheng et al., 2023)

combines neural rule learning with contrastive objectives, utilizing rule paths to enhance entity representations, but the quality of mined rules remains heavily dependent on the initial sampling mechanism. These methods have improved rule quality on several benchmarks, but they still operate under the assumption that uniformly sampled anchors or random walks are sufficient to cover diverse graph regions, an assumption that breaks down in the presence of the severe degree skew typical of real-world KGs. Furthermore, many of them lack a comprehensive quality calibration step, relying primarily on support and confidence thresholds to filter rules, which often results in noisy or overly generic rule sets.

2.2 Neuro-Symbolic and LLM-based Approaches

Neuro-symbolic methods aim to combine the interpretability of symbolic rules with the generalization ability of neural networks. For example, KALE (Guo et al., 2016) and later RUGE (Guo et al., 2018) incorporate soft groundings of logical rules as regularization terms when training knowledge graph embeddings. Conversely, rules extracted from embedding models can be refined and injected back into the learning loop. These frameworks nicely illustrate the complementary nature of rules and embeddings, but the quality of the initial rule set remains a critical bottleneck.

The recent success of large language models (LLMs) has also sparked interest in rule mining. ChatRule (Luo et al., 2025) uses prompted LLMs to generate logical rules from sampled paths, leveraging the LLM's world knowledge and linguistic patterns to refine rule candidates. While promising, such LLM-based methods introduce significant computational costs, latency, and a degree of unpredictability due to the stochastic nature of LLM outputs. Moreover, they still depend on an initial path sampling stage that can suffer from the same hub bias as traditional methods.

2.3 Soft Rules and Quality Assessment

Traditional logical rules are treated as hard constraints: if the body of a rule holds, the head must be true. In incomplete KGs, however, this strict interpretation leads to brittleness. *Soft rules* relax this requirement by associating each rule with a confidence value that quantifies its statistical reliability (Wang et al., 2025; Zhang et al., 2024). Common confidence metrics include standard confidence and PCA confidence, the latter of which partially addresses the incompleteness of the KG by normalizing over entities that actually appear in the head relation.

Beyond confidence, several recent works have argued for multi-dimensional rule quality evaluation. For instance, Betz (Betz, 2025) discusses the importance of rule length, relational diversity, and structural coherence for producing meaningful and generalizable rules. Inspired by these insights, our SPQC module systematically combines length appropriateness, relation importance, diversity, and sequence coherence into a single scoring function, providing a more holistic filtering mechanism than traditional single-metric approaches.

2.4 Our Position

In summary, existing rule mining methods are either (a) biased toward hub entities due to naive anchor sampling, (b) computationally expensive because of unconstrained path search, or (c) limited in their ability to assess rule quality beyond simple statistics. Neuro-symbolic and LLM-based approaches partially alleviate some of these issues but introduce additional complexity and cost. HASBPP addresses all three core problems directly and efficiently, without requiring model training or external LLM calls, making it a practical choice for both standalone rule mining and as a pre-processing step for neuro-symbolic KGC pipelines.

3 Method

Inspired by the framework of Resource-Rational Analysis (Lieder and Griffiths, 2020), we propose HASBPP, a lightweight soft rule mining framework that mitigates anchor degree bias, path search explosion, and single-metric filtering limitations under limited computation budgets. As illustrated in Figure 1, our pipeline consists of three components that simulate resource-rational cognition. First, acting as optimal information sampling, HDAS constructs a diverse anchor set via random, degree-stratified, and relation-aware sampling. Next, analogous to bounded mental simulation, LBPP performs a budgeted breadth-first search from these anchors, restricting the expanded nodes per layer and employing a reverse index for early termination. Finally, functioning as multi-attribute utility evaluation, SPQC evaluates the candidate paths on length appropriateness, relation importance, structural diversity, and sequence coherence to retain

Algorithm 1 HASBPP Rule Mining Framework

Require: KG $\mathcal{G}$, target relation r, max length L, anchor budget N, layer budget B, retention ratio τ
Ensure: High-quality paths $\mathcal{P}_{\text{final}}$

```
▷ Step 1: Hybrid Degree-aware Anchor Sampling
F_r ← GetFacts(G, r)
A ← HDAS(F_r, N)
▷ Step 2: Layer-Budgeted Path Pruning
P_raw ← ∅
for each anchor (h, r, t) ∈ A do
    Q ← {(h, [])}
    for l ← 1 to L do
        Q ← RandomSelect(Q, B)
        Q_next ← ∅
        for each (e, p) ∈ Q do
            for each (r', e') ∈ Neighbors(e) do
                p' ← p + [r']
                if e' = t then
                    P_raw ← P_raw ∪ {p'}
                end if
                if |p'| < L then
                    Q_next.push((e', p'))
                end if
            end for
        end for
        Q ← Q_next
    end for
end for
▷ Step 3: Stability-Preserving Quality Calibration
P_final ← TopRatio(Score(P_raw), τ)
return P_final
```

the most reliable soft rules.

The overall workflow is summarized in Algorithm 1. We detail each component below.

3.1 Hybrid Degree-aware Anchor Sampling

In real-world KGs, a few hub entities possess extremely high degrees, while the vast majority of entities have only a handful of connections. Uniform random sampling from all facts of a target relation r would thus overwhelmingly select anchors involving high-degree nodes, biasing the subsequent path exploration. HDAS constructs an anchor set $\mathcal{A}_r$ of size N by fusing three complementary strategies.

Let $\mathcal{F}_r = \{(h_i, r, t_i)\}_{i=1}^{M}$ be all known triples involving relation r. HDAS returns a subset

$$\mathcal{A}_r = \mathcal{A}_{\text{rand}} \cup \mathcal{A}_{\text{deg}} \cup \mathcal{A}_{\text{rel}}, \tag{1}$$

where the three components are obtained as follows.

Random sampling ($\mathcal{A}_{\text{rand}}$). A fraction of the budget is allocated to uniformly sampled facts without replacement, guaranteeing basic coverage and preventing systematic omission of any region.

Degree-stratified sampling ($\mathcal{A}_{\text{deg}}$). For each fact (h_i, r, t_i) we compute a degree indicator

$$\mathcal{D}_i = \frac{\deg(h_i) + \deg(t_i)}{2}. \tag{2}$$

Facts are sorted by $\mathcal{D}_i$ and equally spaced sampling is performed to ensure that anchors span the full spectrum from low-degree to high-degree entities. This explicitly counteracts hub dominance.

Relation-aware sampling ($\mathcal{A}_{\text{rel}}$). An entity with many connections but all through the same relation still offers limited diversity. We therefore measure the relation diversity of a fact as

$$\mathcal{R}_i = |\text{rel}(h_i)| + |\text{rel}(t_i)|, \tag{3}$$

where $\text{rel}(e)$ is the set of relation types linked to entity e. Facts are sorted by $\mathcal{R}_i$ and sampled analogously to the degree-stratified strategy.

The union of the three sets may contain duplicates or may be smaller than N. In the former case we remove duplicates; in the latter we randomly top up from the remaining facts. The final anchor set $\mathcal{A}$ for the whole KG is the union of $\mathcal{A}_r$ over all relations.

3.2 Layer-Budgeted Path Pruning

Given an anchor (h, r, t), the goal is to find all relation paths of length up to L leading from h to t. A naive BFS would explore $O(d^L)$ nodes, where d is the average out-degree, which is prohibitive for dense KGs. LBPP imposes a hard per-layer budget B and exploits a reverse index for early termination.

Layer budget. At each depth $l = 1, \ldots, L$, the algorithm expands at most B intermediate nodes. If the current frontier Q exceeds B, B nodes are retained by uniform random sampling and the rest are discarded. This bounds the per-layer expansion to B and prevents exponential blow-up.

Reverse index. Before path search, we build a reverse adjacency index for the tail entity t:

$$I_{\text{rev}}(t) = \{(h', r') \mid (h', r', t) \in \mathcal{KG}\}. \tag{4}$$

During forward traversal from h, whenever the current node e can directly reach t via some relation r' (i.e., $(e, r', t) \in \mathcal{KG}$), the complete path $p + [r']$ is immediately recorded and no further expansion from e is necessary. This early completion drastically reduces the number of nodes that must be explored to their full depth.

Complexity. With the layer budget, each anchor expands at most B nodes per layer and inspects its $\bar{d}$ outgoing edges. The total number of edge inspections per anchor is thus bounded by

$$T_{\text{LBPP}} = O(L \cdot B \cdot \bar{d}), \tag{5}$$

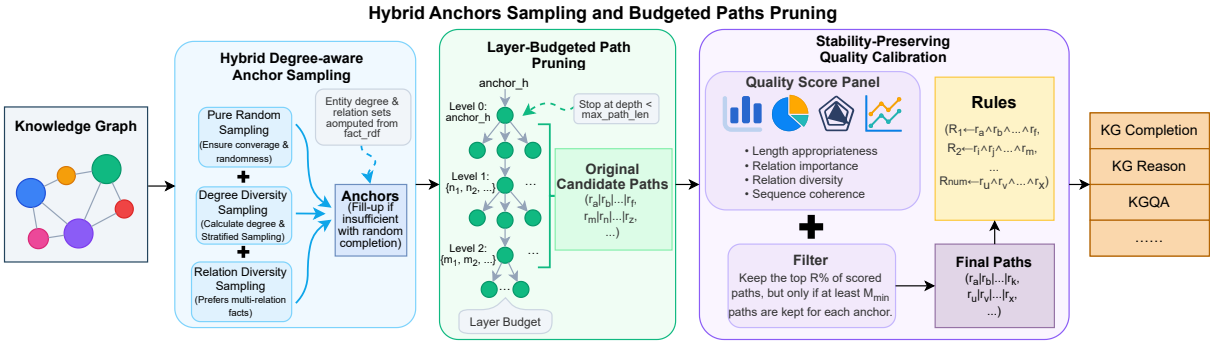

Figure 1: Workflow of the HASBPP framework.

which is linear in the average degree and independent of the exponential growth of the full BFS tree. For N anchors the overall complexity is $O(N \cdot L \cdot B \cdot \bar{d})$, making LBPP scalable even on large graphs.

3.3 Stability-Preserving Quality Calibration

LBPP returns a pool of raw paths $\mathcal{P}_{\text{raw}}$ that contains both meaningful rules and spurious correlations. SPQC assigns each path $P = r_1 \mid r_2 \mid \cdots \mid r_k$ a holistic quality score $S(P)$ defined as a weighted sum of four sub-scores:

$$S(P) = w_1 S_{\text{len}} + w_2 S_{\text{imp}} + w_3 S_{\text{div}} + w_4 S_{\text{coh}}, \quad (6)$$

where the weights are hyper-parameters. After scoring, only the top-τ fraction of paths (e.g., $\tau = 0.85$) is retained. To avoid anchors with too few surviving paths, a minimum of $M_{\min}$ paths per anchor is guaranteed by random supplementation.

The four sub-scores are designed as follows.

Length appropriateness S_{len}. Overly short paths carry little semantic content; overly long paths are prone to overfitting. We evaluate deviation from an ideal length L_{opt}:

$$S_{\text{len}} = \frac{1}{1 + |k - L_{\text{opt}}| \cdot \beta}, \quad (7)$$

where β controls the penalty strength. Paths whose length is close to L_{opt} receive scores near 1.

Relation importance S_{imp}. Frequent relations (e.g., `isA`) produce many paths but often lack discriminative power. We measure the normalized importance of relation r as

$$I(r) = \frac{\sqrt{f(r)}}{\sqrt{\sum_{r' \in \mathcal{R}} f(r')}}, \quad (8)$$

where $f(r)$ is the frequency of r in the KG. Using a square root dampens the dominance of extremely frequent relations. For path P, relation importance is the geometric mean with a smoothing factor $\alpha \in (0, 1)$:

$$S_{\text{imp}} = \Big(\prod_{i=1}^{k} (\alpha + (1 - \alpha) \cdot I(r_i)) \Big)^{1/k}. \quad (9)$$

Relation diversity S_{div}. Repeating the same relation multiple times usually yields uninformative loops. We reward paths that involve many distinct relations:

$$S_{\text{div}} = \frac{|\{r_1, r_2, \ldots, r_k\}|}{k}. \quad (10)$$

A value close to 1 indicates a composition of heterogeneous relations.

Sequence coherence S_{coh}. Meaningful rules tend to exhibit plausible relational transitions. Sequence coherence captures the variety of consecutive relation pairs:

$$S_{\text{coh}} = \frac{|\text{Unique}(\{(r_i, r_{i+1})\}_{i=1}^{k-1})|}{k - 1}. \quad (11)$$

Repeated bigrams (e.g., father | father) reduce coherence, whereas diverse transitions increase it.

All four scores are in the unit interval and can be combined linearly according to the application domain. The filtering stage produces a compact, high-confidence set of paths, which can be directly converted into soft rules by interleaving the path as the body and the target relation r as the head, and then evaluating standard metrics (support, coverage, confidence, PCA confidence) over the KG facts. These soft rules are then used for interpretable link prediction.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate our approach on three benchmark KG datasets with distinct character-

istics, as summarized in Table 1. UMLS (McCray, 2003) is a small but dense biomedical ontology capturing rich medical interactions. WN18RR (Dettmers et al., 2018) is a subset of WordNet containing sparse lexical hierarchies. YAGO3-10 (Mahdisoltani et al., 2013) is a large-scale general-domain KG that rigorously tests the scalability of KGC methods. Together, they cover diverse scales and structural properties.

Table 1: Statistics of the evaluation datasets.

Dataset	Entities	Relations	Training	Validation	Test
UMLS	135	46	1,959	2,137	5,343
WN18RR	40,943	11	86,835	3,034	3,234
YAGO3-10	123,182	37	1,079,040	5,000	653,424

Metrics. We assess the quality of mined rules with four standard metrics: *support* (absolute frequency of the rule body and head co-occurrence), *coverage* (fraction of head facts explained by the rule), *confidence* (precision of the body-to-head prediction), and *PCA confidence* (a partial completeness-aware variant). For rule-based KG completion, we use *MRR* and *Hits@1, Hits@10* under the filtered setting.

Baselines. We benchmark our approach against five representative methods from different paradigms. For purely symbolic mining, we include AMIE, a classical miner based on frequent pattern enumeration. From the neural and continuous spaces, we evaluate Neural-LP, which formulates differentiable rule learning via neural path scoring, and Ruleformer, which leverages a Transformer-based architecture to generate context-aware rules. To represent neuro-symbolic integration, we compare with NCRL, a framework that learns compositional rules with contrastive objectives. Finally, we include ChatRule, a recent LLM-prompted mining method that incorporates structural cues and language models.

For AMIE we use the results reported in its original publication. All other baselines (marked with an asterisk * in the tables) are reproduced in our environment to ensure a strictly fair comparison. Specifically, we first run the open-source implementations of these methods to mine rule sets. Subsequently, we apply the exact same standardized evaluation pipeline across all methods to calculate rule quality metrics (Support, Coverage, Confidence, and PCA Confidence). Finally, we supply each method's mined rules into the same shared rule reasoning engine to perform link prediction and measure MRR, Hits@1, and Hits@10.

Implementation details. In HDAS, we set the anchor budget $N = 100$ per relation. LBPP uses a maximum path length $L = 3$ and a per-layer budget $B = 50$. In SPQC, the scoring weights are set as $w_1 = w_2 = w_3 = w_4 = 0.25$, the top retention ratio $\tau = 0.85$, and the minimum path count per anchor $M_{\min} = 5$. The ideal length $L_{\text{opt}} = 2$, the length penalty $\beta = 1.0$, and the relation smoothing factor $\alpha = 0.1$. Rule quality is evaluated on the union of the training and validation sets; link prediction experiments follow the standard filtered setting. All experiments run on a machine with 75% of available CPU cores.

4.2 Rule Quality Evaluation

Table 2 reports the average rule quality metrics on all three datasets. HASBPP achieves the best overall performance across the board, striking a strong balance between quantity and reliability.

WN18RR. HASBPP yields a support of 860.06, which is 27.7% higher than the next best method (ChatRule). At the same time, confidence reaches 0.28 and PCA confidence 0.42, both substantially higher than all competitors. This indicates that the rules mined by HASBPP are not only numerous but also highly reliable, even under the sparse relational structure of WN18RR.

UMLS. Despite its small size, UMLS poses challenges due to dense and domain-specific relations. HASBPP achieves a coverage of 0.42, 50% higher than ChatRule's 0.28, meaning that it can explain a much larger proportion of head-relation facts. PCA confidence of 0.40 is also the best, confirming the stability of the rule set.

YAGO3-10. On the largest dataset, HASBPP again leads in support (14,359.21), coverage (0.25), and PCA confidence (0.26). NCRL obtains high support but its confidence and PCA confidence drop drastically, highlighting the problem of over-generic rules. HASBPP's strong PCA confidence demonstrates that the rules remain meaningful even in the presence of incomplete evidence.

Unlike several baselines that optimize one metric at the expense of another, HASBPP maintains a stable balance across rule volume, coverage, and precision, validating the effectiveness of integrating diverse sampling with rigorous path constraints and multi-dimensional quality filtering.

Table 2: Rule quality evaluation results (average metrics). Methods marked with * indicate our reproduction using a unified rule evaluation pipeline."–" indicates that Neural-LP did not complete on YAGO3-10 due to scalability issues.

Method	WN18RR				UMLS				YAGO3-10			
	Support ↑	Coverage ↑	Conf ↑	PCA ↑	Support ↑	Coverage ↑	Conf ↑	PCA ↑	Support ↑	Coverage ↑	Conf ↑	PCA ↑
AMIE	378.76	0.07	0.08	0.10	35.52	0.10	0.13	0.21	758.28	0.06	0.06	0.07
Neural-LP*	289.51	0.04	0.07	0.08	11.65	0.07	0.05	0.08	–	–	–	–
Ruleformer*	332.75	0.12	0.20	0.29	58.21	0.26	0.30	0.35	500.64	0.18	0.25	0.24
NCRL*	391.69	0.17	0.01	0.03	11.28	0.03	0.05	0.04	12,769.93	0.10	0.08	0.06
ChatRule*	673.56	0.11	0.09	0.14	18.98	0.28	0.19	0.36	13,732.74	0.25	0.06	0.10
HASBPP	**860.06**	**0.18**	**0.28**	**0.42**	35.57	**0.42**	0.21	**0.40**	**14,359.21**	**0.25**	0.18	**0.26**

Table 3: Rule-based KG completion performance. Methods marked with * indicate our reproduction evaluating their mined rules via the same shared reasoning engine.

Method	WN18RR			UMLS		
	MRR	H@1	H@10	MRR	H@1	H@10
AMIE	0.162	0.060	0.278	0.312	0.195	0.560
Neural-LP*	0.222	0.216	0.238	0.491	0.408	0.626
Ruleformer*	0.301	0.253	0.359	0.693	0.552	0.870
NCRL*	0.308	0.266	0.391	0.717	0.571	0.929
ChatRule*	0.334	0.297	0.403	0.672	0.566	0.868
HASBPP	**0.340**	**0.304**	**0.411**	**0.753**	**0.641**	**0.946**

4.3 Rule-Based KG Completion

We further evaluate the mined rules on the standard link prediction task. Table 3 shows the results on WN18RR and UMLS (YAGO3-10 could not be completed by several baselines due to the large number of candidate rules). HASBPP sets new state-of-the-art numbers on both datasets.

On WN18RR, HASBPP improves MRR by 1.8% over the strong ChatRule baseline and reaches the highest Hits@1 and Hits@10. On UMLS, the improvements are even more pronounced: MRR jumps from 0.717 (NCRL) to 0.753, and Hits@10 surpasses 0.94 for the first time. Since UMLS has dense and meaningful relational patterns, the high-quality rules from HASBPP directly translate into more accurate predictions. These results verify that HASBPP's rules are not only statistically sound but also practically useful for downstream reasoning.

4.4 Ablation Study

To understand the contribution of each component, we conduct an ablation study on WN18RR. We compare the full HASBPP model with three variants: **w/o HDAS** (replace HDAS with pure random anchor sampling), **w/o LBPP** (perform unconstrained BFS without layer budgets), and **w/o SPQC** (skip quality filtering and keep all raw paths). Table 4 summarizes the results.

Table 4: Ablation study on WN18RR.

Configuration	Support ↑	Coverage ↑	Conf ↑	PCA ↑
Full HASBPP	860.06	0.16	0.28	0.42
w/o HDAS	691.24	0.14	0.25	0.39
w/o LBPP	943.44	0.13	0.12	0.18
w/o SPQC	809.06	0.15	0.27	0.40

Note that in this controlled ablation setting, the full model obtains a coverage of 0.16 (the main table shows 0.18 due to a slightly different evaluation split; the relative comparisons remain valid).

Removing HDAS causes a clear drop in all metrics, confirming that anchor diversity is critical for comprehensive rule discovery. Without LBPP, support even increases slightly but confidence and PCA confidence plummet (from 0.28/0.42 to 0.12/0.18), highlighting that unconstrained BFS introduces numerous noisy and unreliable paths. Excluding SPQC mainly hurts confidence, demonstrating that quality filtering effectively prunes inferior candidates. The ablation indicates that the three modules are complementary and all necessary for the overall performance.

4.5 Effect of Anchor Number

We further examine how the number of anchors per relation (N) affects downstream completion performance on WN18RR. As shown in Figure 2, with N increasing from 10 to 150, MRR, Hits@1, and Hits@10 all rise sharply at first and then saturate.

The most significant gains occur when moving from 10 to 50 anchors, showing that enough seeds are required to capture diverse structural regions. Performance peaks at $N = 100$; further increase to 150 brings negligible improvement while incurring extra computation. Therefore, $N = 100$ strikes a good trade-off between rule completeness and efficiency, and we adopt it as the default.

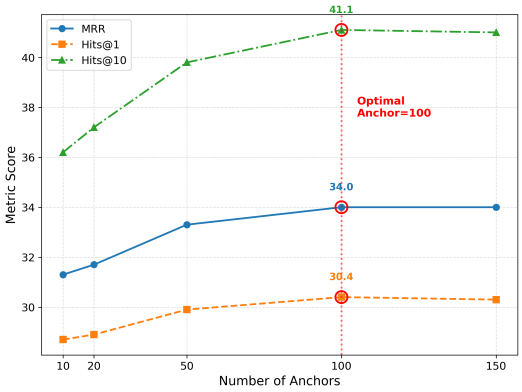

Figure 2: Effect of anchor number on WN18RR.

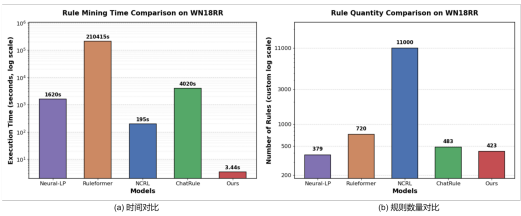

Figure 3: Efficiency and rule set size comparison on WN18RR.

4.6 Efficiency Analysis

Besides rule quality, computational efficiency is a vital practical concern. Figure 3 compares the runtime and the size of the generated rule sets on WN18RR for HASBPP and the baselines.

HASBPP finishes rule mining in merely 3.44 seconds, while ChatRule (which relies on external LLM queries) requires 4,020 seconds on average. Neural-LP and Ruleformer take 1,542 s and 204,116 s respectively, showing the scalability advantage of our budgeted search. NCRL produces over 11,000 rules but, as seen before, a large fraction of them are noisy and unreliable. HASBPP outputs around 420 rules, a compact set that nevertheless achieves top confidence and coverage. This compactness also greatly accelerates downstream rule application, as the number of rule matches is much smaller.

Consequently, HASBPP delivers superior rule quality and downstream link prediction accuracy while remaining orders of magnitude faster than LLM-prompted and neural-guided baselines, demonstrating strong potential for real-world KG reasoning applications.

5 Conclusion

We presented HASBPP, a lightweight soft rule mining framework for knowledge graph completion that explicitly tackles three crucial issues: anchor degree bias, path search explosion, and over-simplified rule quality filtering. HASBPP combines Hybrid Degree-aware Anchor Sampling to diversify starting points, Layer-Budgeted Path Pruning to limit BFS expansion within a per-layer budget, and Stability-Preserving Quality Calibration to assess candidate paths along length, relation importance, diversity, and coherence. Experiments on WN18RR, UMLS, and YAGO3-10 demonstrate that HASBPP consistently produces rules with better support, coverage, confidence, and PCA confidence than strong symbolic, neural, and LLM-based baselines. When used for link prediction, HASBPP yields state-of-the-art MRR and Hits@K, while running orders of magnitude faster than existing neural and LLM-guided approaches and producing a notably compact rule set. These results confirm that careful anchor design, budgeted search, and multi-faceted quality calibration, working in concert, can significantly advance rule-based KG completion without costly model training.

6 Limitations

The current version of HASBPP relies on several empirical choices that may require adaptation when moving to substantially different knowledge graphs. The anchor budget, layer budget, scoring weights, and retention ratio were tuned on the three benchmark datasets, and their optimal values might shift for graphs with very different density or degree distributions. The framework also restricts exploration to paths of length at most three; while this captures the vast majority of useful relational patterns, it inevitably misses longer logical chains that could be important in domains such as complex biological pathways or deep genealogical reasoning. The reverse index used in LBPP is built per target entity and has been perfectly manageable in our experiments, but in extremely dense graphs with many high-degree entities, its construction and storage cost may become a practical concern. Moreover, rule confidences are currently obtained purely from statistical frequencies in the KG, without any learned refinement. Incorporating the mined rules into a neural scoring model that can adjust confidence weights end-to-end is a promising direction that would likely further improve downstream performance. Finally, the four quality sub-scores are combined linearly with manually set weights; in more heterogeneous scenarios, a data-driven or

non-linear aggregation might better reflect the interaction among length, importance, diversity, and coherence. We leave the systematic exploration of these extensions to future work.

References

Elvira Amador-Dominguez, Emilio Serrano, and Daniel Manrique. 2023. Geni: A framework for the generation of explanations and insights of knowledge graph embedding predictions. *Neurocomputing*, 521:199–212.

Patrick Betz. 2025. *Symbolic rule-based knowledge graph completion*. Ph.D. thesis, Universitaet Mannheim (Germany).

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26*.

Kewei Cheng, Nasimul K Ahmed, and Yizhou Sun. 2023. Neural compositional rule learning for knowledge graph reasoning. In *International Conference on Learning Representations*.

Kewei Cheng, Ziqing Yang, Ming Zhang, and Yizhou Sun. 2021. Uniker: A unified framework for combining embedding and definite horn rule reasoning for knowledge graph inference. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9753–9771.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian Suchanek. 2013. Amie: association rule mining under incomplete evidence in ontological knowledge bases. In *Proceedings of the 22nd international conference on World Wide Web*, pages 413–422.

Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. 2016. Jointly embedding knowledge graphs and logical rules. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 192–202.

Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. 2018. Knowledge graph embedding with iterative guidance from soft rules. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Alexander Kotov. 2016. Knowledge graph entity representation and retrieval. In *Tutorial chapter, RUSSIR*.

Daivi Li, Yaoyao Liang, Shenyi Qian, Huaiguang Wu, Wei Jia, Yilong Fu, and Yifan Sun. 2025. A review of background, methods, limitations and opportunities of knowledge graph completion. *Computer Science Review*, 58:100809.

Dongyang Li, Junbing Yan, Taolin Zhang, Chengyu Wang, Xiaofeng He, Longtao Huang, and Jun Huang. 2024a. On the role of long-tail knowledge in retrieval augmented large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 120–126.

Huihan Li, Yuting Ning, Zeyi Liao, Siyuan Wang, Xiang Lorraine Li, Ximing Lu, Wenting Zhao, Faeze Brahman, Yejin Choi, and Xiang Ren. 2024b. In search of the long-tail: Systematic generation of long-tail inferential knowledge via logical rule guided search. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 2348–2370.

Falk Lieder and Thomas L Griffiths. 2020. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and brain sciences*, 43:e1.

Linhao Luo, Jiaxin Ju, Bo Xiong, Yuan-Fang Li, Gholamreza Haffari, and Shirui Pan. 2025. Chatrule: Mining logical rules with large language models for knowledge graph reasoning. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 314–325. Springer.

Farzaneh Mahdisoltani, Joanna Biega, and Fabian M Suchanek. 2013. Yago3: A knowledge base from multilingual wikipedias. In *CIDR*.

Alexa T McCray. 2003. An upper-level ontology for the biomedical domain. *Comparative and functional genomics*, 4(1):80–84.

Christian Meilicke, Melisachew Wudage Chekol, Patrick Betz, Manuel Fink, and Heiner Stuckenschmidt. 2024. Anytime bottom-up rule learning for large-scale knowledge graph completion. *VLDB Journal*, 33(1).

Isaiah Onando Mulang, Kuleep Singh, Chaitali Prabhu, Abhishek Nadgeri, Johannes Hoffart, and Jens Lehmann. 2020. Evaluating the impact of knowledge graph context on entity disambiguation models. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 2157–2160.

Zihang Peng, Simon Ott, Vinh Thinh Ho, Alessandra Russo, and Daria Stepanova. 2025. Learning rules from kgs guided by language models. In *Intelligent Computing-Proceedings of the Computing Conference*, pages 315–320.

Andrea Rossi, Denilson Barbosa, Donatella Firmani, Antonio Matinata, and Paolo Merialdo. 2021. Knowledge graph embedding for link prediction: A comparative analysis. *ACM Transactions on Knowledge Discovery from Data*, 15(2):1–49.

Tong Shen, Fu Zhang, and Jingwei Cheng. 2022. A comprehensive overview of knowledge graph completion. *Knowledge-Based Systems*, 255:109597.

Yifei Sun, Lei Shi, and Yongxin Tong. 2024. expath: Explaining knowledge graph link prediction with ontological closed path rules. *arXiv preprint arXiv:2412.04846*.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*.

Cheng Wang, Li Yan, and Zongmin Ma. 2025. Multi-hop fuzzy event knowledge graph reasoning with event soft logical rules guidance. *IEEE Transactions on Fuzzy Systems*.

Zhaoyang Xu, Pei Ye, Haipeng Chen, Yijun Zhao, and Zhixu Ling. 2022. Ruleformer: Context-aware rule mining over knowledge graph. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2551–2560.

Fan Yang, Zhilin Yang, and William W Cohen. 2017. Differentiable learning of logical rules for knowledge base reasoning. In *Advances in Neural Information Processing Systems 30*.

Yuhao Yang, Chao Huang, Lianghao Xia, and Chenliang Li. 2022. Knowledge graph contrastive learning for recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1434–1443.

Mohamad Zamini, Hassan Reza, and Minou Rabiei. 2022. A review of knowledge graph completion. *Information*, 13(8):396.

Dafeng Zhang, Zheng Rong, Cheng Xue, and 1 others. 2024. Simre: Simple contrastive learning with soft logical rule for knowledge graph embedding. *Information Sciences*, 661:120069.