ICML International Conference On Machine Learning

Forty-Third International Conference on Machine Learning

COEX Convention & Exhibition Center
Jul 6, 2026 - Jul 11, 2026

ICML 2026 Main Conference

Graph-R1: Towards Agentic GraphRAG Framework via End-to-end Reinforcement Learning

Haoran Luo[1,2], Haihong E[1]*, Guanting Chen[1], Qika Lin[3], Yikai Guo[4], Fangzhi Xu[5], Zemin Kuang[6], Meina Song[1], Xiaobao Wu[7], Yifan Zhu[1], Luu Anh Tuan[2,8]

[1] Beijing University of Posts and Telecommunications [2] Nanyang Technological University
[3] National University of Singapore [4] Beijing Institute of Computer Technology and Application
[5] Xi'an Jiaotong University [6] Beijing Anzhen Hospital, Capital Medical University
[7] Shanghai Jiao Tong University [8] VinUniversity
haoran.luo@ieee.org, ehaihong@bupt.edu.cn, anhtuan.luu@ntu.edu.sg

CONTENTS

Graph-R1: Towards Agentic GraphRAG Framework via End-to-end Reinforcement Learning

ICML
International Conference
On Machine Learning

1

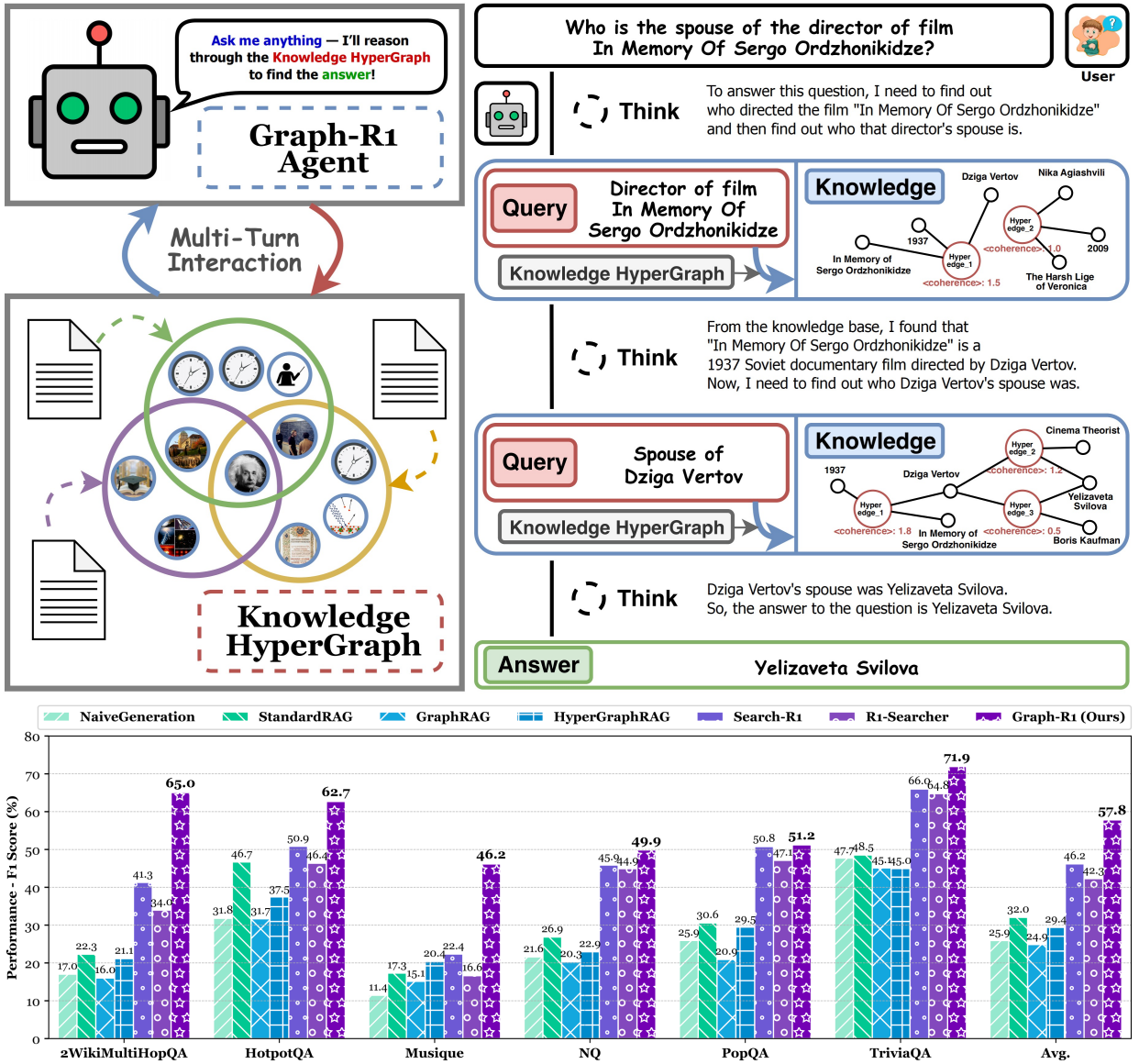
PART

Background of Study

ICML
International Conference
On Machine Learning

Graph-R1: When GraphRAG meets Agentic RL

Retrieval-Augmented Generation (RAG) mitigates hallucination in LLMs by incorporating external knowledge, but relies on chunk-based retrieval that lacks structural semantics. GraphRAG methods improve RAG by modeling knowledge as entity-relation graphs, but still face challenges in high construction cost, fixed one-time retrieval, and reliance on long-context reasoning and prompt design. To address these challenges, we propose Graph-R1, the first agentic GraphRAG framework via end-to-end reinforcement learning (RL). It introduces lightweight knowledge hypergraph construction, models retrieval as a multi-turn agent-environment interaction, and optimizes the agent process via an end-to-end reward mechanism. Experiments on standard RAG datasets show that Graph-R1 outperforms traditional GraphRAG and RL-enhanced RAG methods in reasoning accuracy, retrieval efficiency, and generation quality.



Task Definition

We formalize the GraphRAG pipeline into three stages:

(a) Knowledge Graph Construction. Given a knowledge collection $K = \{d_1, d_2, \ldots, d_N\}$, the goal is to extract facts f_d from each semantic unit $d \in K$ into a unified graph $\mathcal{G}_K$:

$$\mathcal{G}_K \sim \sum_{d \in K} \pi_{\text{ext}}(f_d \mid d), \tag{1}$$

where π_{ext} denotes an LLM-based extractor that parses each d into a set of relation-entity pairs $f_d = \{(r_i, \mathcal{V}_{r_i})\}$, with r_i as the relation and $\mathcal{V}_{r_i} = \{v_1, \ldots, v_n\}$ the entities.

(b) Graph Retrieval. Graph retrieval is formulated as a two-step process over $\mathcal{G}_K$: (1) retrieving candidate reasoning paths and (2) pruning irrelevant ones. Conditioned on a query q, the model first retrieves a candidate set $\mathcal{X}_q = \{x_1, \ldots, x_m\}$ and then selects a relevant subset $\mathcal{Z}_q \subseteq \mathcal{X}_q$. The overall objective is to maximize the joint likelihood:

$$\max_{\theta} \ \mathbb{E}_{\mathcal{Z}_q \sim P(\mathcal{Z}_q \mid q, \mathcal{G}_K)} \left[\prod_{t=1}^{T_x} P_\theta(x_t \mid x_{<t}, q, \mathcal{G}_K) \cdot \prod_{t=1}^{T_z} P_\theta(z_t \mid z_{<t}, \mathcal{X}_q, q) \right], \tag{2}$$

where T_x & T_z are numbers of retrieved & selected paths.

(c) Answer Generation. Given a query q and selected paths $\mathcal{Z}_q$, answer generation produces a natural language answer y grounded in graph-based evidence, formulated as:

$$P(y \mid q, \mathcal{G}_K) = \sum_{\mathcal{Z}_q \subseteq \mathcal{X}_q} P(y \mid q, \mathcal{Z}_q) \cdot P(\mathcal{Z}_q \mid q, \mathcal{G}_K), \tag{3}$$

where $P(y \mid q, \mathcal{Z}_q)$ is generation likelihood and $P(\mathcal{Z}_q \mid q, \mathcal{G}_K)$ is retrieval-pruning distribution.

Three Main Challenges

However, current GraphRAG methods still face three key challenges:

(i) High cost and semantic loss in knowledge construction process.
Compared to standard RAG, GraphRAG methods convert natural language knowledge into graph structures using LLMs, which results in high cost and often causes semantic loss relative to the original content.

(ii) Fixed retrieval process with only one-time interaction in graph retrieval process.
Although existing GraphRAG methods design various retrieval strategies to improve efficiency, most of them aim to gather sufficient knowledge in a single fixed retrieval, which limits performance in complex queries.

(iii) Dependence on large LLMs for long-context analysis and prompt quality in answer generation process.
Generation based on retrieved graph-structured knowledge often requires strong long-context reasoning ability, making the output quality highly dependent on the LLM's parameter size and prompt design.

2

PART

Contributions

Contributions

To address these challenges, we propose Graph-R1, the first agentic GraphRAG framework enhanced by end-to-end reinforcement learning (RL), inspired by DeepSeek-R1.

First, we propose a lightweight knowledge hypergraph construction method to establish a standard agent environment for the query action space.

Moreover, we model the retrieval process as a multi-turn agentic interaction process, enabling LLMs to repeatedly perform the reasoning loop of “think-retrieve-rethink-generate” within the knowledge hypergraph environment.

Furthermore, we design an end-to-end reward mechanism that integrates generation quality, retrieval relevance, and structural reliability of graph paths into a unified optimization objective. Using RL, the agent learns a generalizable graph reasoning strategy and achieves tighter alignment between structured knowledge and language.

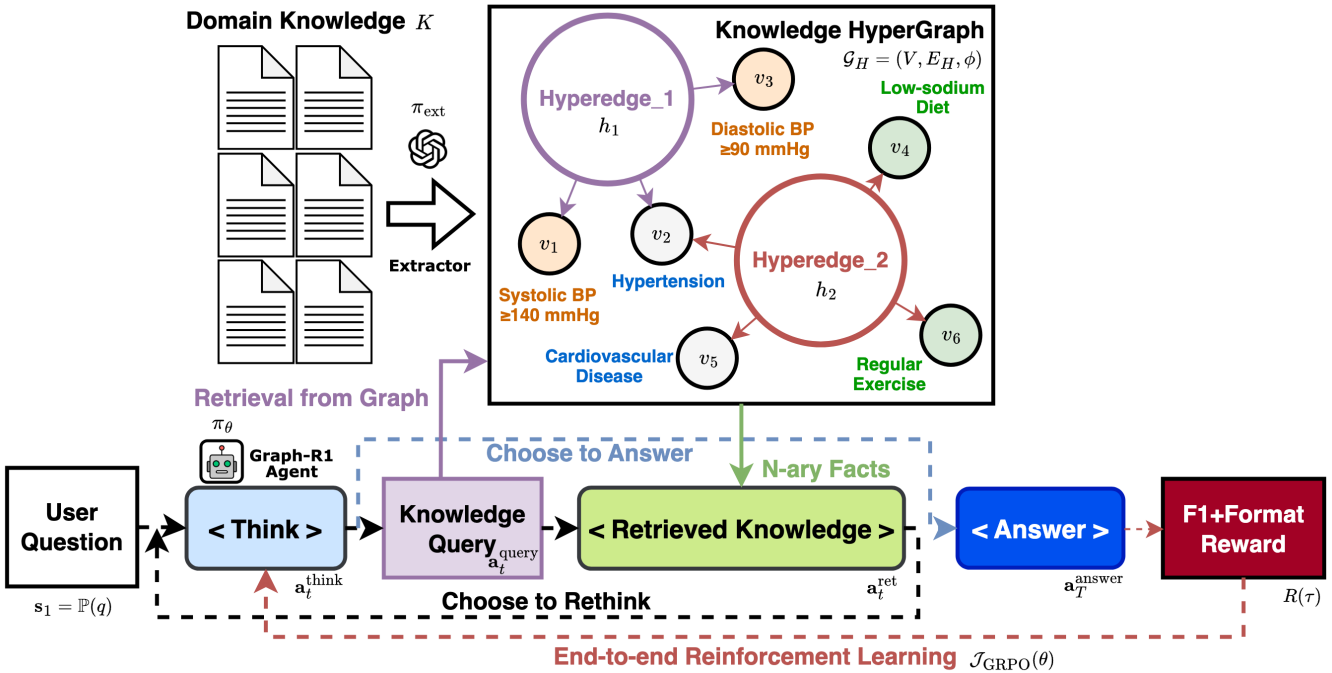
3

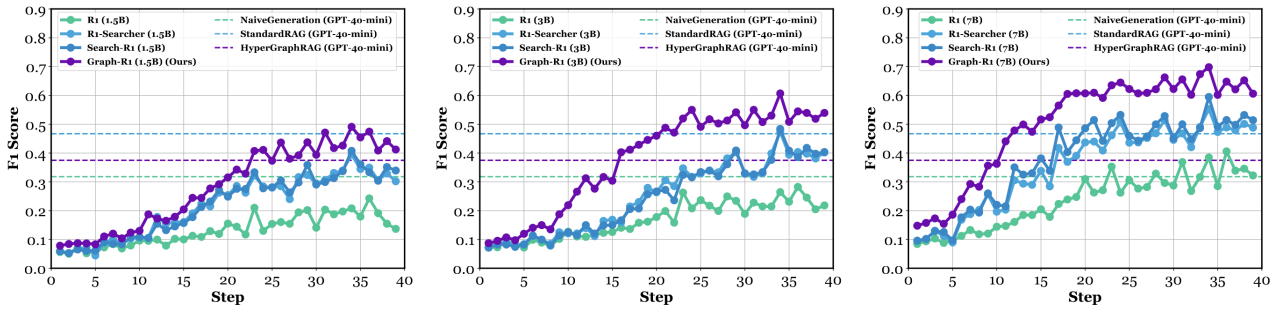
PART

Methodology

ICML
International Conference
On Machine Learning

Method: Graph-R1





(a) Qwen2.5-1.5B-Instruct *(b)* Qwen2.5-3B-Instruct *(c)* Qwen2.5-7B-Instruct

Figure 4. Step-wise F1 score based on Qwen2.5 (1.5B, 3B, 7B), where Graph-R1 outperforms baselines and GPT-4o-mini variants.

1. Lightweight Knowledge Hypergraph Construction

Graph-R1 first converts domain knowledge into a knowledge hypergraph, where each hyperedge represents an n-ary relational fact connecting multiple entities.

This structure preserves richer semantic relations than traditional binary graphs and provides a structured environment for agentic retrieval.

2. Multi-turn Graph Interaction

Given a user question, the Graph-R1 agent follows a Think–Query–Retrieve–Rethink–Answer loop.

Instead of relying on one-shot retrieval, the agent can iteratively query the hypergraph, collect relevant n-ary facts, and decide whether to continue reasoning or generate the final answer.

3. End-to-end Reinforcement Learning Optimization

Graph-R1 optimizes the full reasoning trajectory with end-to-end reinforcement learning.

The reward combines answer correctness and format consistency, encouraging the agent to learn effective graph-based retrieval strategies and produce faithful, well-structured answers.

4

PART

Experimental Results

ICML
International Conference
On Machine Learning

Main Results

Graph-R1 consistently outperforms all baselines across six RAG benchmarks, including 2WikiMultiHopQA, HotpotQA, Musique, NQ, PopQA, and TriviaQA. Compared with traditional RAG and GraphRAG methods, Graph-R1 achieves stronger answer accuracy by combining **graph-structured knowledge** with **end-to-end reinforcement learning**.

The performance improves steadily as the base model scales from **1.5B to 3B and 7B**, showing that larger models can better exploit the synergy between multi-turn graph retrieval and RL-based reasoning.

Most importantly, Graph-R1 achieves the best average F1 score:
40.09 with Qwen2.5-1.5B, 51.26 with Qwen2.5-3B, and 57.82 with Qwen2.5-7B, significantly surpassing Search-R1, R1-Searcher, StandardRAG, and HyperGraphRAG.

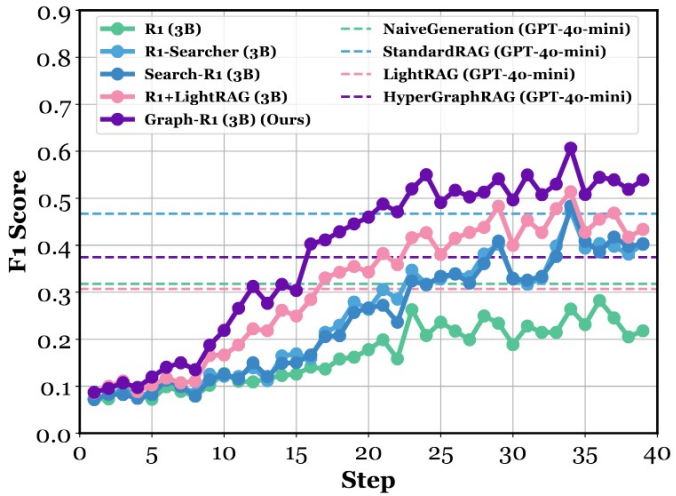
Table 2. Main results with best in **bold**. means prompt engineering, means training, means no knowledge interaction, means chunk-based knowledge, and means graph-based knowledge.

Method	2Wiki.		HotpotQA		Musique		NQ		PopQA		TriviaQA		Avg.			
	F1	G-E	F1	G-E	F1	G-E	F1	G-E	F1	G-E	F1	G-E	EM	F1	R-S	G-E
GPT-4o-mini																
NaiveGeneration	17.03	74.86	31.79	78.48	11.45	76.61	21.59	84.64	25.95	**72.75**	47.73	83.33	11.36	25.92	-	78.45
StandardRAG	**22.31**	73.02	**46.70**	**81.88**	17.31	74.93	**26.85**	84.55	**30.58**	69.42	**48.55**	84.63	**18.10**	**32.05**	52.68	78.07
GraphRAG	16.02	72.81	31.67	77.37	15.14	74.43	20.31	82.36	20.92	65.88	45.13	82.76	12.50	24.87	32.48	75.94
LightRAG	16.59	71.94	30.70	73.42	14.39	73.75	19.09	80.20	20.47	67.76	40.18	81.60	9.77	23.57	47.42	74.78
PathRAG	12.42	67.19	23.12	71.81	11.49	69.94	20.01	81.99	15.65	60.58	37.44	80.94	7.03	20.02	46.71	72.08
HippoRAG2	16.27	68.78	31.78	76.43	12.37	73.05	24.56	**84.65**	21.10	63.31	46.86	83.55	13.80	25.49	36.41	74.96
HyperGraphRAG	21.14	**76.76**	37.46	80.50	**20.40**	**79.29**	22.95	81.22	29.48	70.55	44.95	**85.20**	13.15	29.40	**61.82**	**78.92**
Qwen2.5-1.5B-Instruct																
NaiveGeneration	7.78	49.13	4.27	45.77	2.35	46.63	6.03	46.74	10.06	42.67	8.10	52.92	1.17	6.43	-	47.31
StandardRAG	11.46	55.38	9.93	52.91	3.18	39.46	11.39	59.73	13.08	50.29	17.43	60.52	5.73	11.08	52.84	53.05
SFT	13.26	34.72	13.61	38.93	5.14	28.50	11.56	46.61	15.61	31.35	26.18	46.66	9.83	14.23	-	37.80
R1	26.28	47.48	20.07	44.43	4.84	39.12	16.75	45.95	21.36	44.50	34.78	48.59	14.19	20.68	-	45.01
Search-R1	28.43	60.61	39.99	64.16	4.69	39.32	20.26	59.93	39.63	58.19	44.16	63.01	23.18	29.53	50.45	57.54
R1-Searcher	28.01	58.81	**41.50**	61.54	6.26	38.31	**36.86**	**60.79**	38.37	56.02	42.57	61.24	23.70	32.26	50.68	56.12
Graph-R1 (ours)	**35.13**	**65.73**	40.62	**65.30**	**28.28**	**58.82**	35.62	59.13	**43.55**	**66.46**	**57.36**	**70.83**	**31.90**	**40.09**	**59.35**	**64.38**
Qwen2.5-3B-Instruct																
NaiveGeneration	7.59	55.00	11.16	53.75	3.67	54.00	8.90	57.18	10.89	49.08	10.89	48.16	3.26	8.85	-	52.86
StandardRAG	12.52	60.01	15.41	62.51	2.92	50.40	10.69	65.13	14.70	57.25	21.92	68.43	3.39	13.03	52.69	60.62
SFT	12.40	52.31	16.48	51.35	5.04	51.31	11.23	58.20	16.95	46.42	33.02	59.98	9.64	15.85	-	53.26
R1	28.45	56.92	25.33	55.38	8.07	47.53	21.51	55.11	27.11	48.65	47.91	60.74	19.66	26.40	-	54.06
Search-R1	38.04	54.39	43.84	69.32	7.65	46.43	37.96	52.90	38.67	63.74	47.99	60.37	28.65	35.69	49.99	57.86
R1-Searcher	23.50	55.86	42.44	64.60	12.81	50.07	36.53	63.33	40.18	66.23	54.00	60.52	27.08	34.91	49.98	60.10
Graph-R1 (ours)	**57.56**	**76.45**	**56.75**	**77.46**	**40.51**	**67.84**	**44.75**	**69.92**	**45.65**	**71.27**	**62.31**	**75.01**	**42.45**	**51.26**	**60.19**	**72.99**
Qwen2.5-7B-Instruct																
NaiveGeneration	12.25	66.75	16.58	65.31	4.06	65.47	13.00	69.56	12.82	60.50	24.51	72.65	3.12	13.87	-	66.71
StandardRAG	12.75	60.06	21.10	66.13	4.53	59.84	15.97	70.49	16.10	60.86	24.90	73.71	5.34	15.89	52.67	65.18
SFT	20.28	63.85	27.59	65.65	10.02	63.50	19.02	68.19	27.93	56.31	39.21	70.25	15.57	24.01	-	64.63
R1	30.99	59.19	37.05	60.12	14.53	49.39	28.45	57.63	30.35	53.38	57.33	66.73	25.91	33.12	-	57.74
Search-R1	41.29	70.26	50.85	73.85	22.35	57.68	45.88	67.58	50.76	66.08	65.98	76.15	38.54	46.19	51.60	68.60
R1-Searcher	33.96	69.61	46.36	74.56	16.63	59.05	44.93	68.54	47.12	66.74	64.76	75.95	34.51	42.29	51.26	69.08
Graph-R1 (ours)	**65.04**	**82.42**	**62.69**	**80.03**	**46.17**	**71.42**	**49.87**	**70.97**	**51.22**	**73.43**	**71.93**	**79.11**	**48.57**	**57.82**	**60.40**	**76.23**

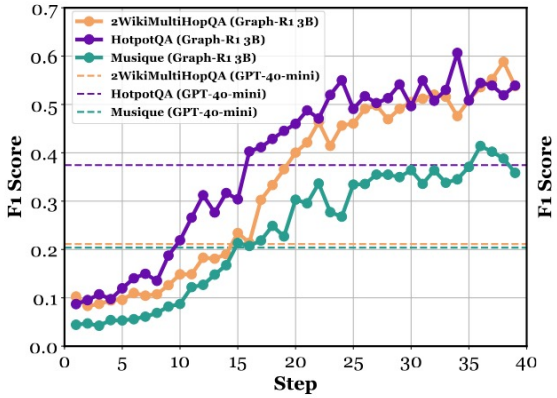
Ablation Study

Method	2Wiki.				HotpotQA				Avg.			
	EM	F1	R-S	G-E	EM	F1	R-S	G-E	EM	F1	R-S	G-E
Qwen2.5-3B-Instruct												
Graph-R1	**50.00**	**57.56**	**55.78**	**76.45**	**50.78**	**56.75**	**54.74**	**77.46**	**50.39**	**57.16**	**55.26**	**76.96**
w/o K.C.	36.33	44.94	53.46	65.83	40.63	47.27	53.23	72.69	38.48	46.11	53.35	69.26
w/o M.I.	21.88	34.34	54.70	66.59	30.86	37.64	53.34	64.75	26.37	35.99	54.02	65.67
w/o R.L.	0.78	8.91	10.16	47.14	5.47	12.56	14.44	58.60	3.13	10.74	12.30	52.87
Qwen2.5-7B-Instruct												
Graph-R1	**55.47**	**65.04**	**55.24**	**82.42**	**57.03**	**62.69**	**56.27**	**80.03**	**56.25**	**63.87**	**55.76**	**81.23**
w/o K.C.	44.14	51.81	54.10	75.90	49.22	55.93	54.14	76.78	46.68	53.87	54.12	76.34
w/o M.I.	37.50	44.78	54.54	69.98	40.63	47.04	54.58	69.63	39.07	45.91	54.56	69.81
w/o R.L.	0.00	18.25	54.63	75.81	3.12	17.33	53.80	78.92	1.56	17.79	54.22	77.37

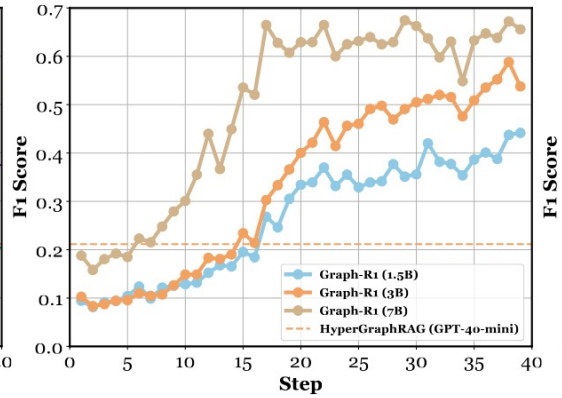
(a) Ablation Study



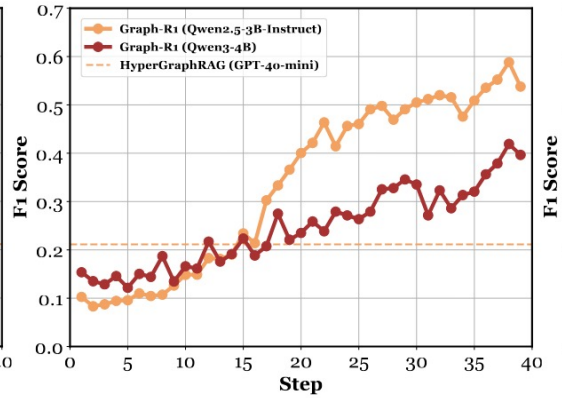
(b) Representations



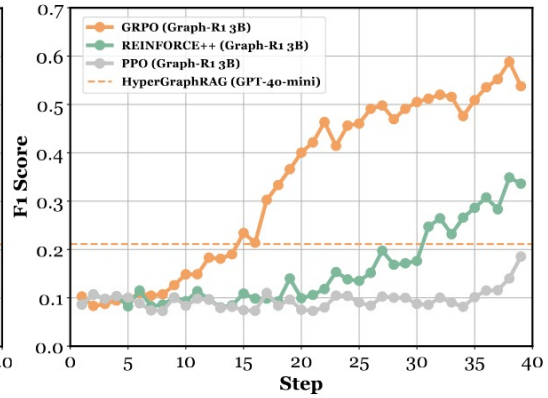
(c) Datasets



(d) Parameters



(e) Qwen3



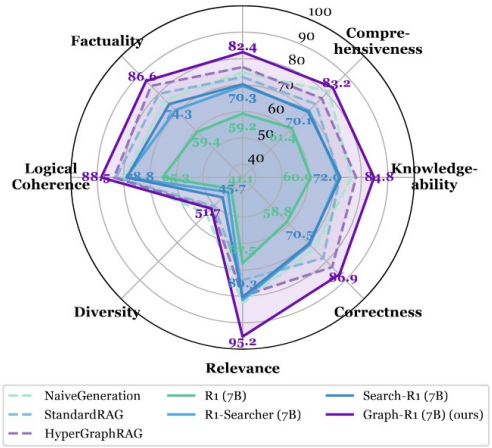
(f) Algorithms

Figure 5. (a) Ablation study of Graph-R1. (b–f) Performance comparison across different kinds of knowledge representations, RAG datasets, model parameters, Qwen versions, and RL algorithms.

Comparison Analysis

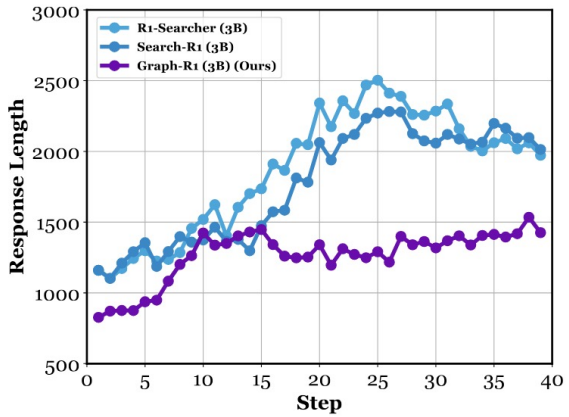
Method	Knowledge Construction				Retrieval & Generation		
	TP1KT	CP1MT	#Node	#Edge	TPQ	CP1KQ	F1
NaiveGeneration	0 s	0 $	-	-	3.7 s	0.16 $	17.0
StandardRAG	0 s	0 $	-	-	4.1 s	1.35 $	22.3
GraphRAG	8.04 s	3.35 $	7,771	4,863	7.4 s	3.97 $	16.0
LightRAG	6.84 s	4.07 $	59,197	24,596	12.2 s	8.11 $	16.6
PathRAG	6.84 s	4.07 $	59,197	24,596	15.8 s	8.28 $	12.4
HippoRAG2	3.25 s	1.26 $	11,819	40,654	8.8 s	7.68 $	16.3
HyperGraphRAG	6.76 s	4.14 $	173,575	114,426	9.6 s	8.76 $	21.1
Graph-R1 (7B) (ours)	5.69 s	2.81 $	120,499	98,073	7.0 s	0 $	65.0

(a)

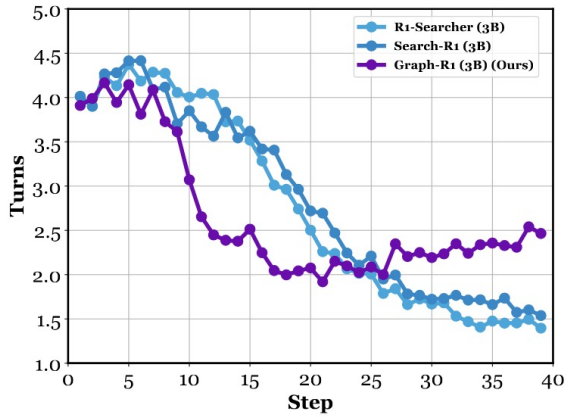


(b)

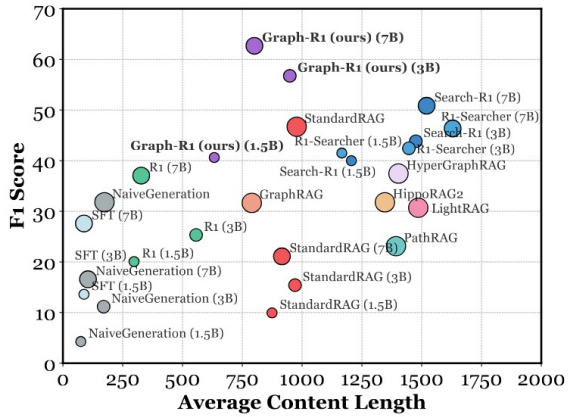
Figure 6. (a) Time & Cost Comparisons on 2Wiki. (b) Generation Evaluations.



(a) Response Length



(b) Turns of Interaction



(c) Efficiency Comparison

Figure 7. Step-wise response length & turns of interaction, and efficiency comparison on HotpotQA.

Cost Efficiency
Graph-R1 builds the hypergraph with lower time and cost than most GraphRAG baselines.

Retrieval Efficiency
Graph-R1 achieves higher F1 with fewer interaction turns and moderate response length.

Generation Quality
Graph-R1 shows better factuality, relevance, correctness, and logical coherence, indicating stronger graph-grounded reasoning.

Forty-Third International Conference on Machine Learning

COEX Convention & Exhibition Center
Jul 6, 2026 - Jul 11, 2026

ICML 2026 Main Conference

Thanks for Listening

Code & Data https://github.com/LHRLAB/Graph-R1

Contact haoran.luo@ieee.org