Plan-Answer-Refine-on-Graph: Structured Planning and Self-Refinement for Large Language Model Reasoning on Knowledge Graphs

Yuxin Shi, Han Fu, Zhuo Li, Chenghao Liu, Xiaoxue Ren, Jianling Sun

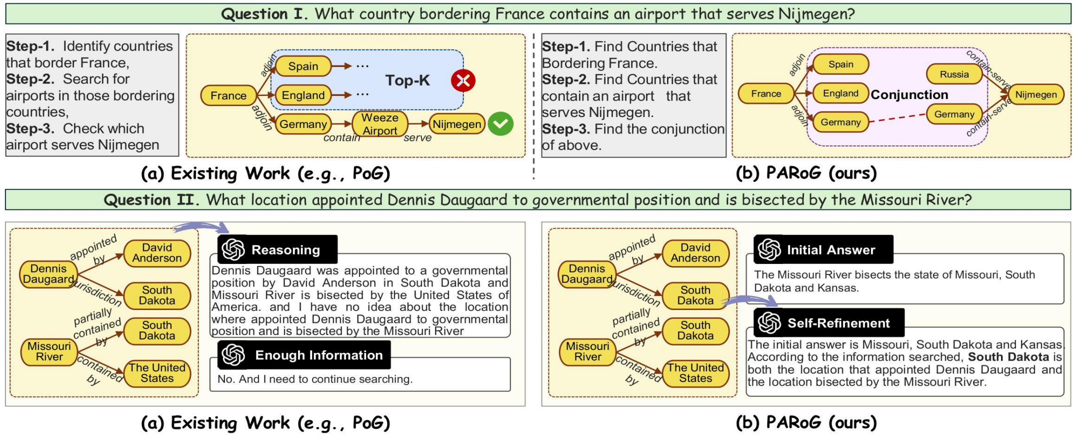
Motivation: Why Existing LLM+KG Methods Still Fail ?

Search Space Truncation Bias

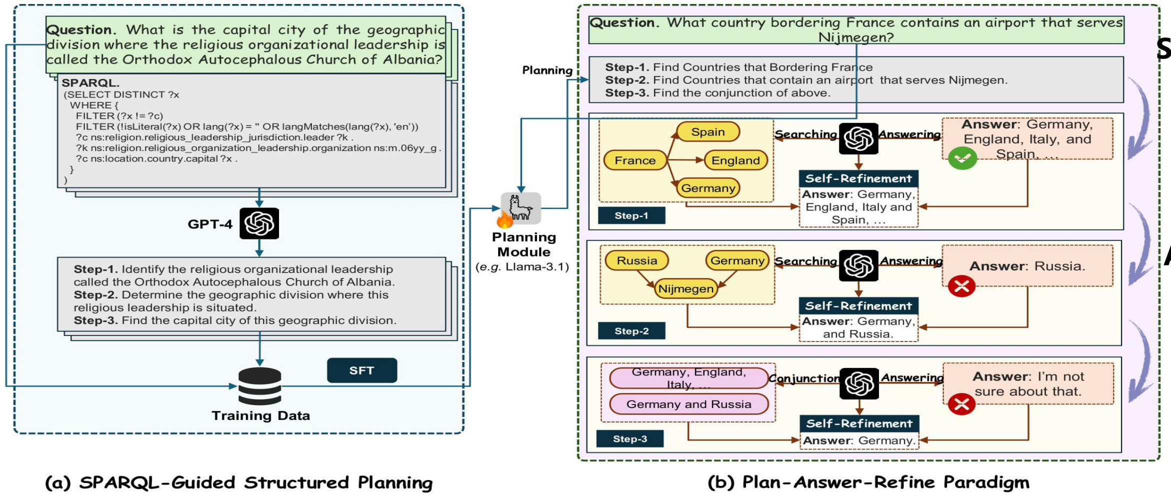
Linear path expansion often prunes correct candidates too early

Error Amplification

Retrieve-and-answer pipelines may over-trust partially relevant triples



Plan-Answer-Refine: Two Components, One Workflow



SPARQL-grounded decomposition ✔

Compositional, operator-aware, and SPARQL-consistent ✔

Answer→KG retrieval→refine ✔

Explicit error override using retrieved triples ✔

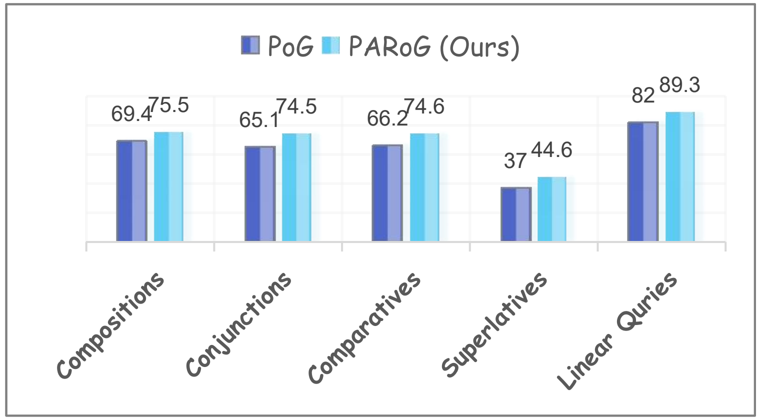
Gains grow with logical complexity

91.2 % +3.9
WebQSP
GPT-4 as Agent

79.3 % +4.3
CWQ
GPT-4 as Agent

The **larger** the logical burden, the **larger** the benefit of better planning and refinement.

Planner 73.3
GPT-3.5 65.2
DeepSeek-R1 68.7
outperforms SOTA model !



Key Findings

1. Biggest gains appear on complex logical query types.
2. Self-refinement corrects 62–77% of initially wrong answers.
3. A SPARQL-supervised 8B planner can beat much larger generic planners.