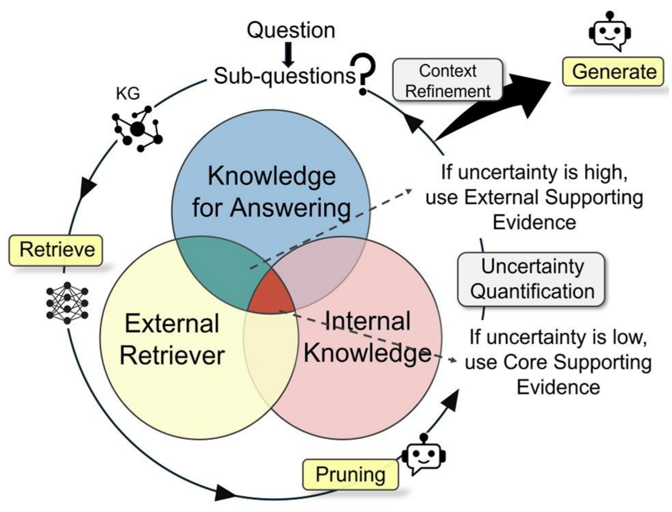
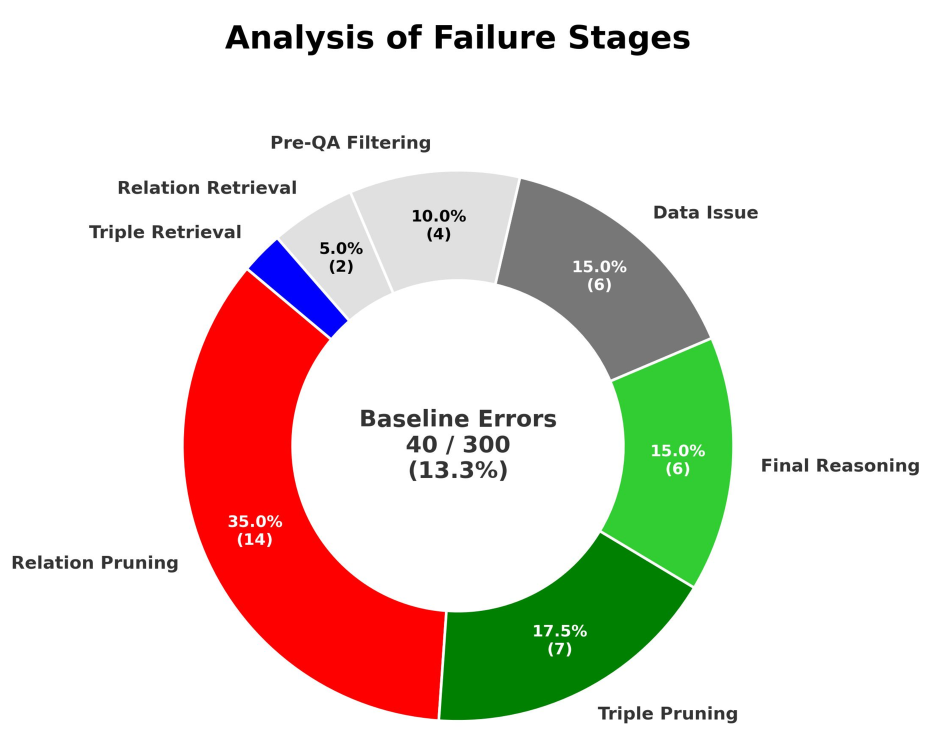
Bridging the Gap in ProgRAG: A Robust Hybrid Strategy for KGQA

Jeonghyun Kim, Dayoon Lee | Advisor: Prof. Hyunjoon Kim
School of Data Science, Hanyang University

1. Research Motivation

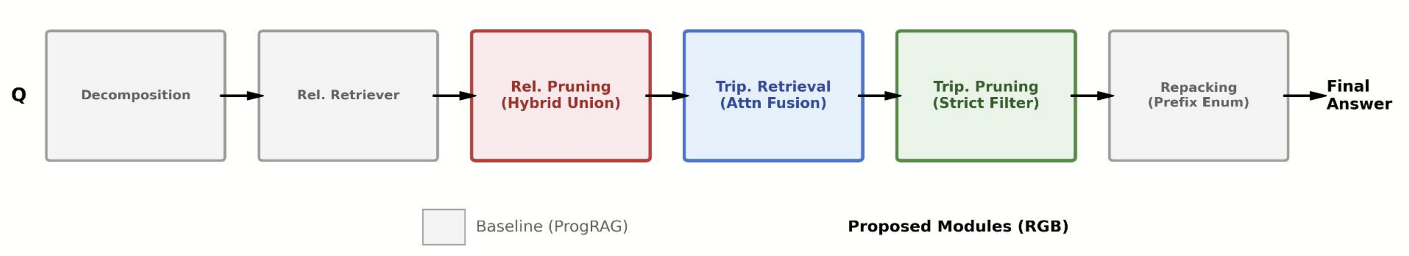


- Knowledge Graph Question Answering(KGQA) systems often suffer from over-retrieval and LLM-induced hallucination.
- ProgRAG addresses these challenges by decomposing complex questions into sub-questions and incrementally expanding partial reasoning paths.
- At each iteration, KG evidence is retrieved using external retrievers and refined through LLM-based pruning.



- However, ProgRAG relies solely on LLM-based pruning, making the system vulnerable to hallucinations.
- Error analysis shows that most failures with a hit score of 0 stem from incorrect pruning decisions.

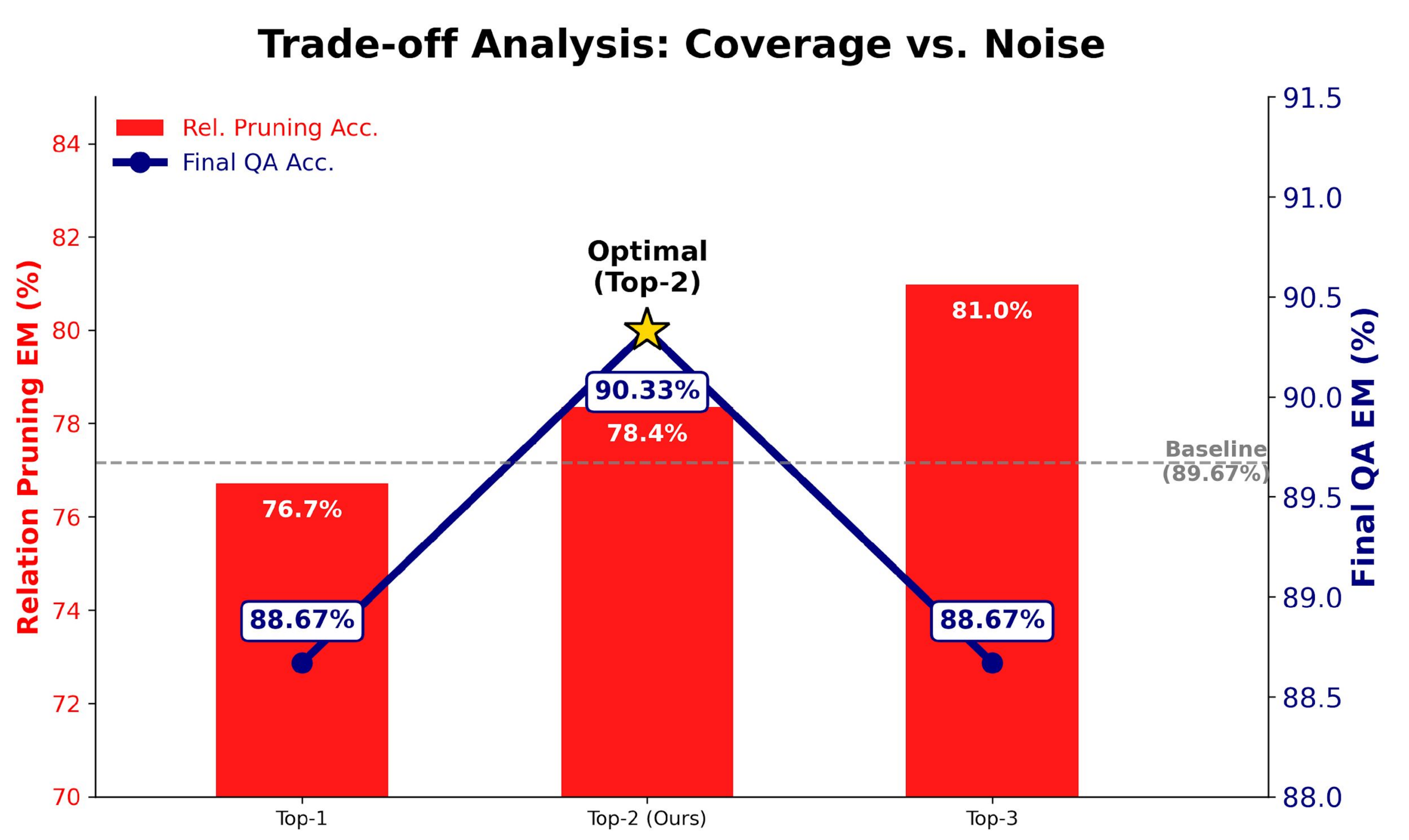
2. Proposed Method Overview



- Integrated ProgRAG (Gray) with Task-Adaptive Modules (RGB).
  - Red: Hybrid Union Strategy in Relation Pruning
  - Blue: Attention-Fusion Layer in Triple Retrieval
  - Green: Prompt Engineering in Triple Pruning & Final Reasoning

3. Methods

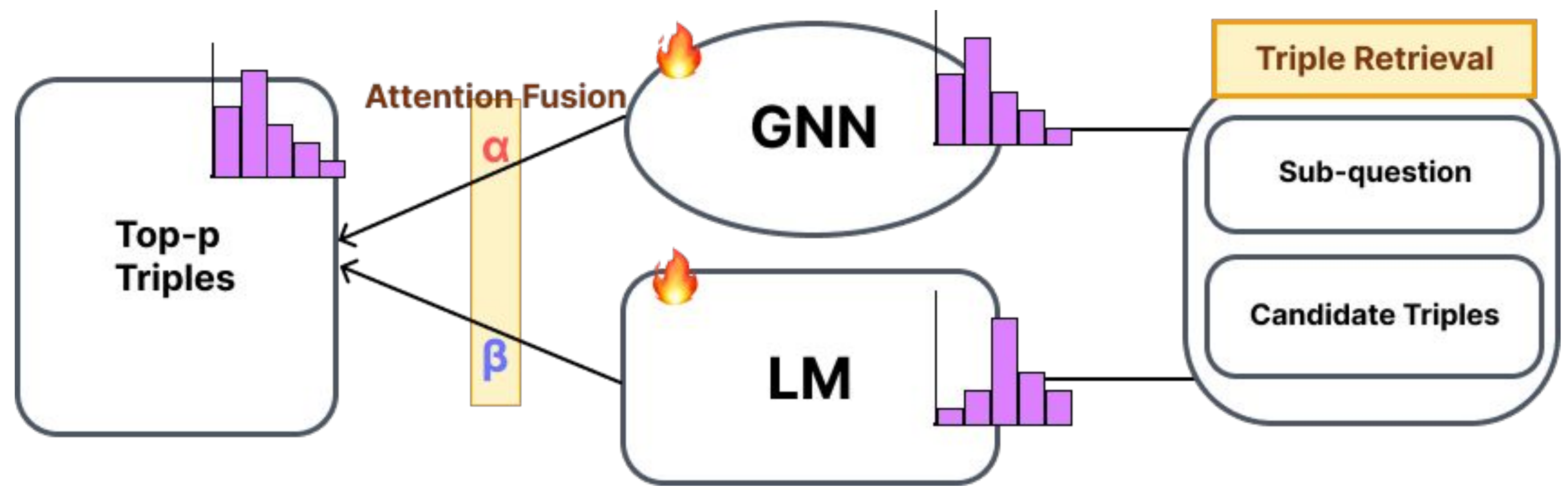
3. 1. Hybrid Union Strategy



- Why Top-2 ?
  - Optimal Point (Top-2): "Golden Cross" balancing Recall & Noise.
  - Maximize Coverage: Captures valid answers efficiently.
  - Minimize Noise: Filters out over-retrieved, irrelevant relations (Top-3+).

3.2. Attention-Fusion Layer in Triple Retrieval

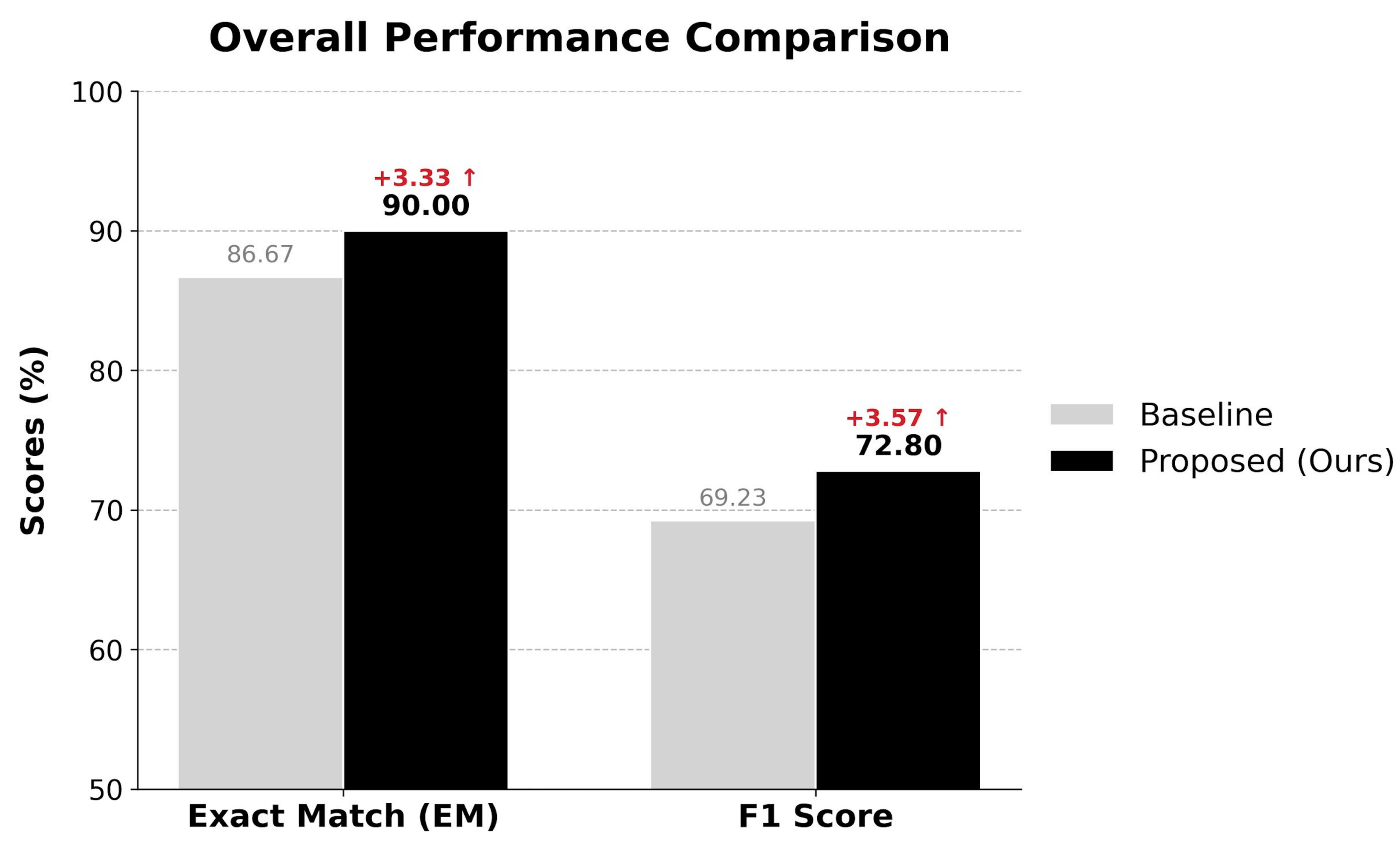
- Triple relevance scoring was modified to dynamically weight two different scores depending on the question embedding



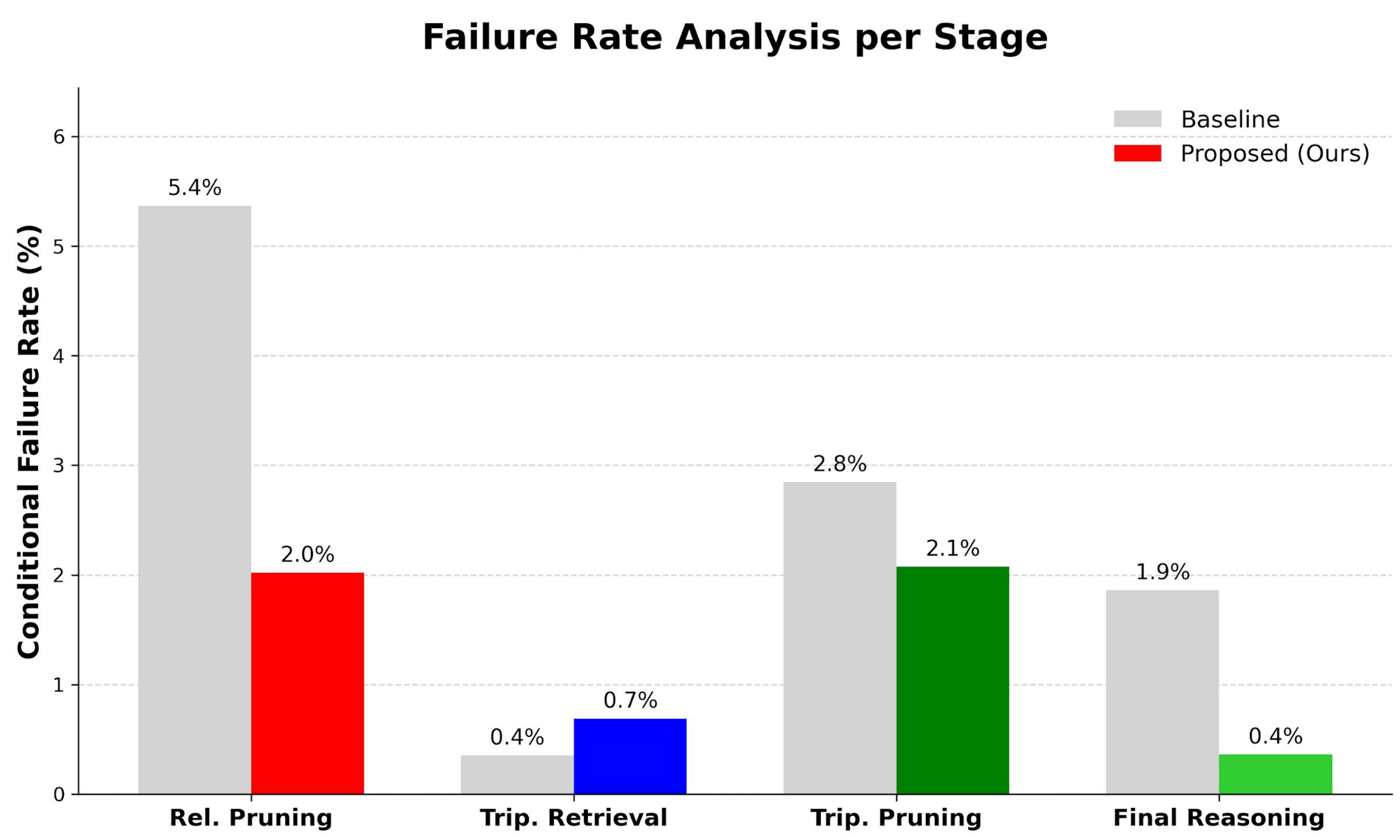
3.3. Prompt Engineering

- 3-Shot Guidance: Demonstrates Abstinence (["None"]) for irrelevant paths.
- Triple Pruning: Precision Protocol → Filters noise via Hard Constraints.
- Final Reasoning: Recall Protocol → Aggregates answers via Hybrid Fallback.

4. Experimental Results



- High Accuracy: 90.00% EM, 72.80 F1
- Substantial Gain: +3.33%p over Baseline.
- Data Scope: Evaluated on randomly sampled subset (N=300).



- Rel. Pruning: Drastic error reduction (5.4% → 2.0%).
- Trip. Retrieval: Stable defense against Hard Samples.
- Trip.Pruning & Final Reasoning: Strict Filtering → Near-Zero (0.4%) Error.

5. Conclusion

- Contribution: Proposed Task-Adaptive Reasoning, achieving +3.33%p gain.
- Robustness: Significantly reduced noise and hallucinations across all failure stages.
- Future Work: Extend evaluation from current subset (N=300) to full WebQSP dataset.