ÉCOLE POLYTECHNIQUE

INSTITUT POLYTECHNIQUE DE PARIS

T.EN TECHNIP ENERGIES

Agentic GraphRAG with Adaptive Graph Traversal in Tendering EPC

Intern: SENG Hok

Supervisor: JULIEN Cadart

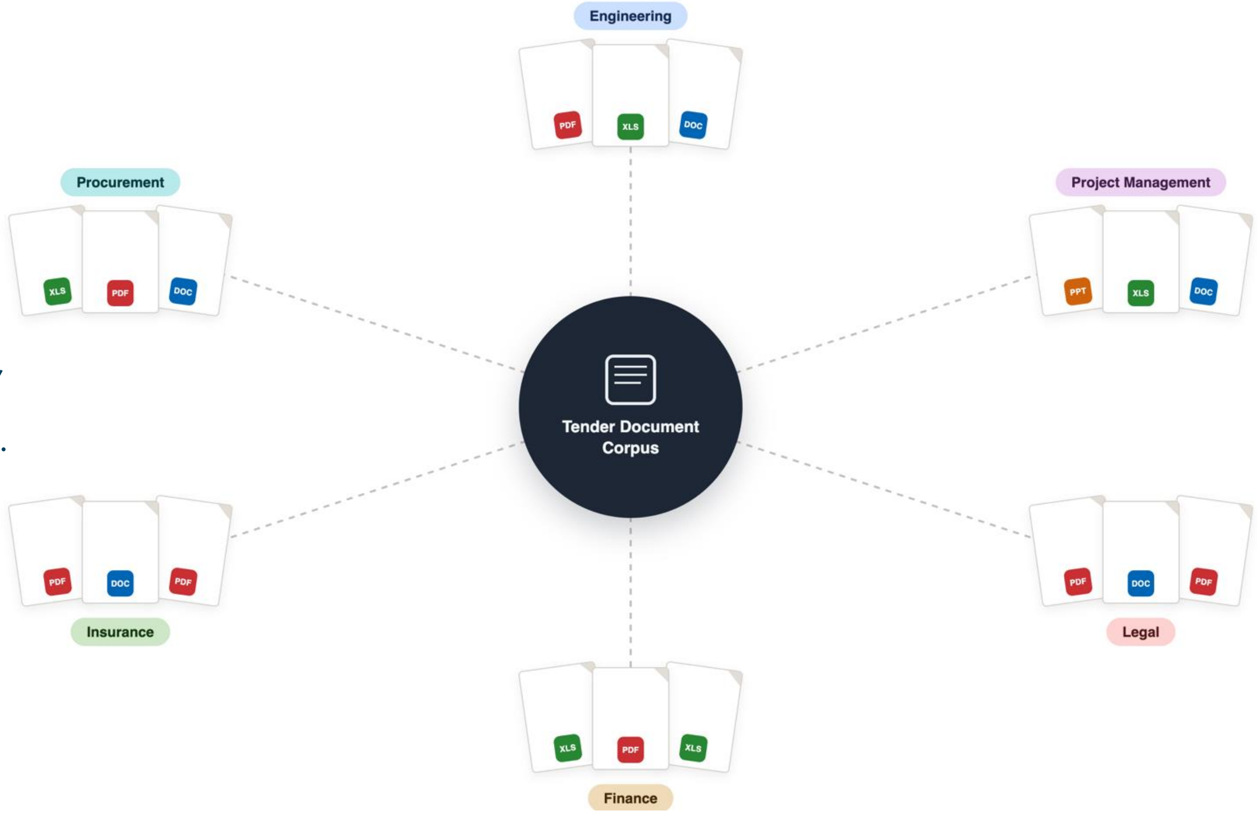
30 July 2026

Table of Contents

Context

Industrial Context: A Multidisciplinary Tender Corpus

- Multidisciplinary corpus: engineering, procurement, project management, legal, finance, insurance, and safety
- 40 specialised lower-level domains, including rotating equipment, piping, document control, and risk management
- Heterogeneous formats: PDF, Word, Excel , PowerPoint, Docx, etc.



Challenges

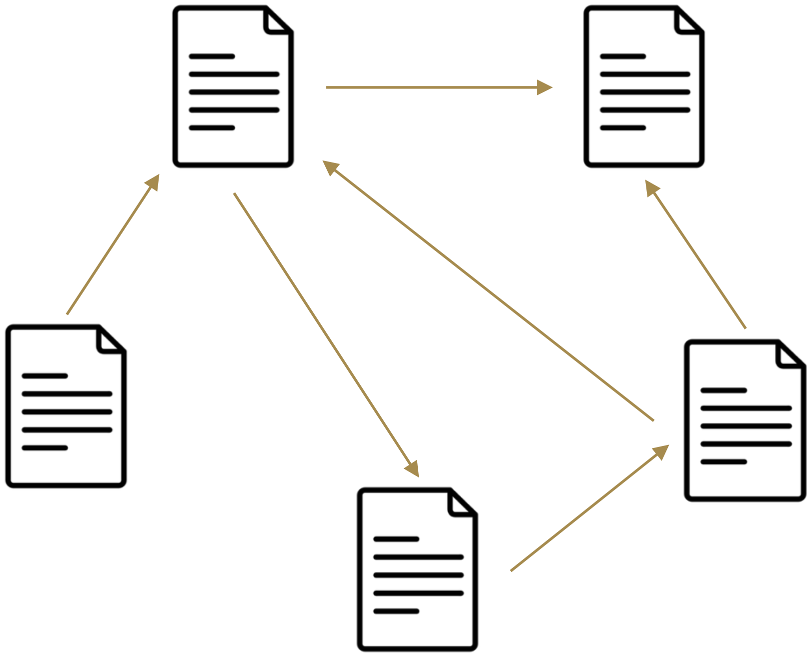
Manual search

↓

Time-consuming cross-checking

↓

Risk of missing critical information



Link through domain sharing, references, etc

Research Question & Contributions

Can knowledge graph-based retrieval improve the accuracy and completeness of the current RAG systems for large-scale engineering tender documents?

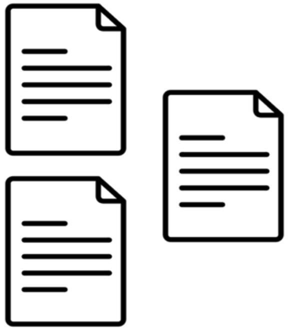
Contributions:

- Domain-aware KG construction
- Development of an adaptive graph retrieval pipeline
- Implementation of Dual-Level RAG and Weighted Hub-aware PPR-RAG including cross-encoder
- Quantitative & Qualitative evaluation against the baseline Hybrid RAG

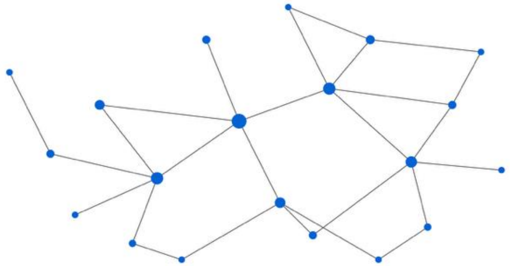
State of the art

From RAG to GraphRAG

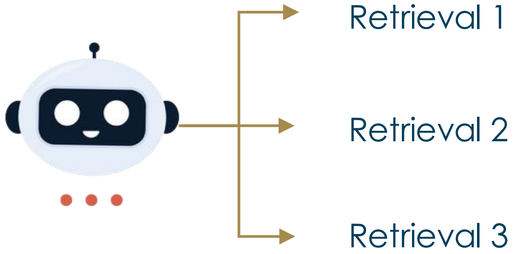
2020 - Lewis et al.



2024 - Edge et al.



2025 — Liu et al.



Main idea

Retrieve chunks using BM25 & dense embeddings

Limitations

- Chunks are ranked independently
- Relationships not explicitly represented

Main idea

- Extract entities and relationships using an ontology
- Use Cypher queries or Graph Algorithm to retrieve connected evidence

Main idea

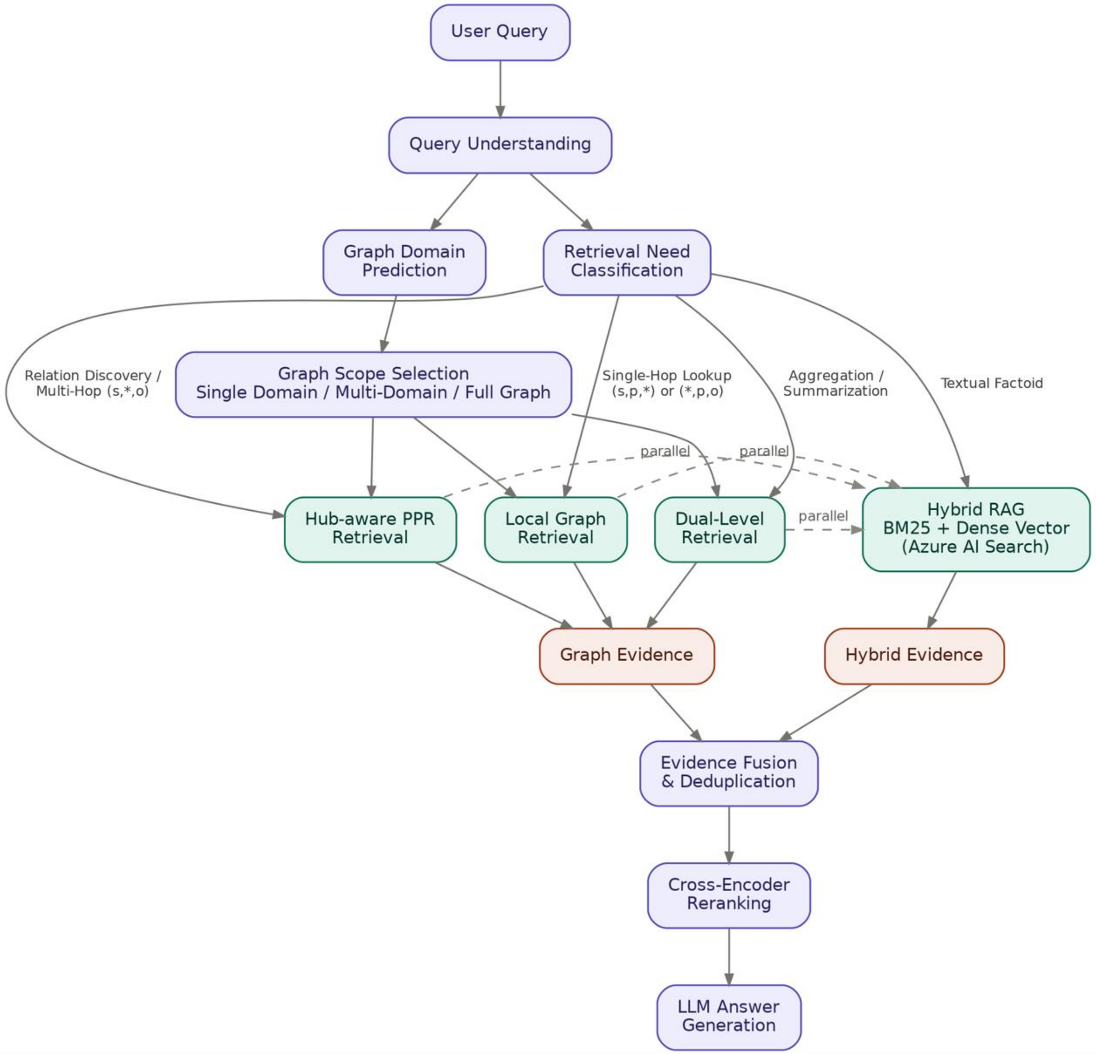
- Select a retrieval strategy based on the query
- Avoid using expensive graph traversal for every question

Existing approaches & research gap

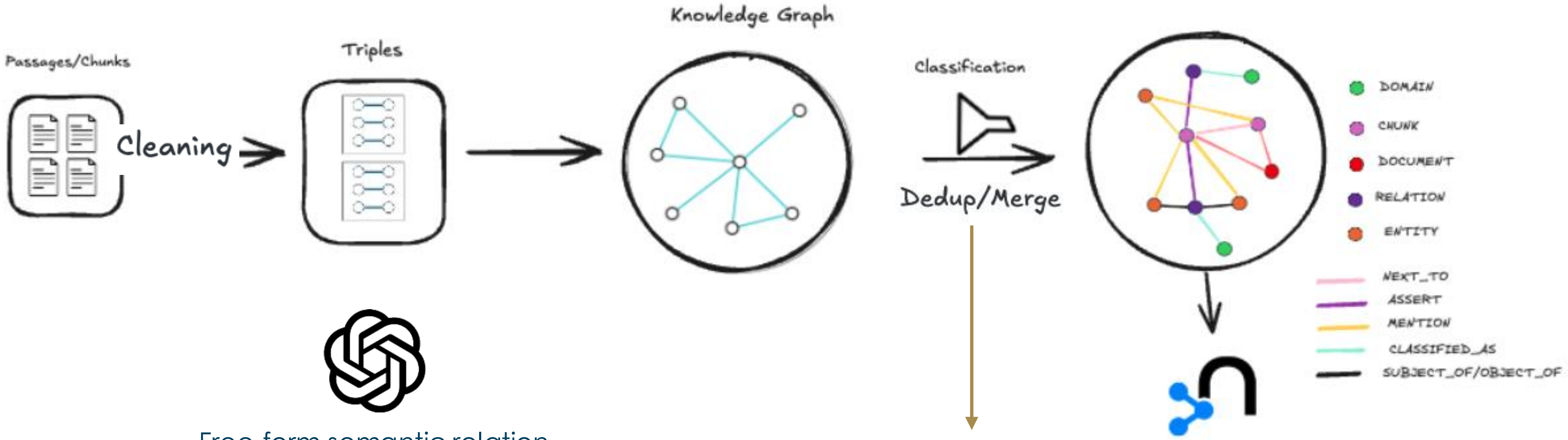
Approach	Main idea	Limitation for this work
Microsoft GraphRAG (2024)	Community detection and hierarchical summaries	High offline indexing cost, fact-level queries, not benchmarked on tendering or any comparable enterprise-specific domain
LightRAG (2024)	Local relation retrieval and global relation retrieval	Fixed set of retrieval modes, not query-adaptive
HippoRAG (2025)	Personalized PageRank from query entities	Free-form extraction, Graph sparsity
PolyG (2025)	Adaptive graph traversal by query type	Assumes a pre-built, Cypher-queryable graph database

Proposed Methodology

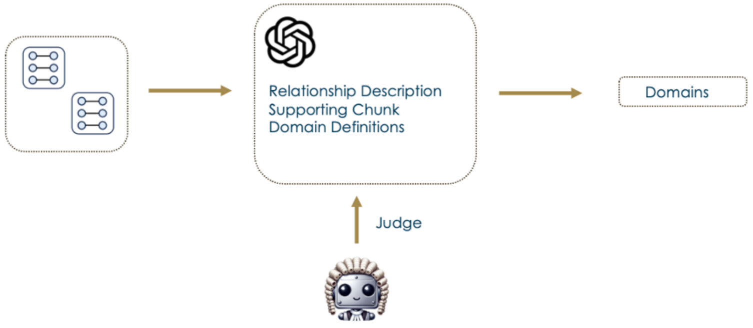
Overall proposed system architecture



Knowledge Graph Construction Pipeline



Free-form semantic relation extraction with constraints on entity type: **Deliverable, risk, requirement, project, system, etc.**



Online Entity Linking

Query → Entity Extraction → Fuzzy matching + Semantic Search

↓

Reciprocal Rank Fusion

↓

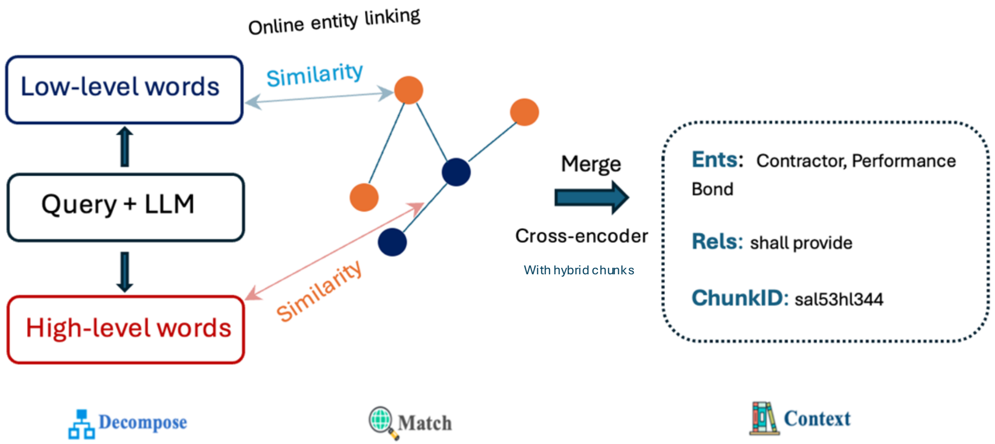
LLM disambiguation & deduplication

↓

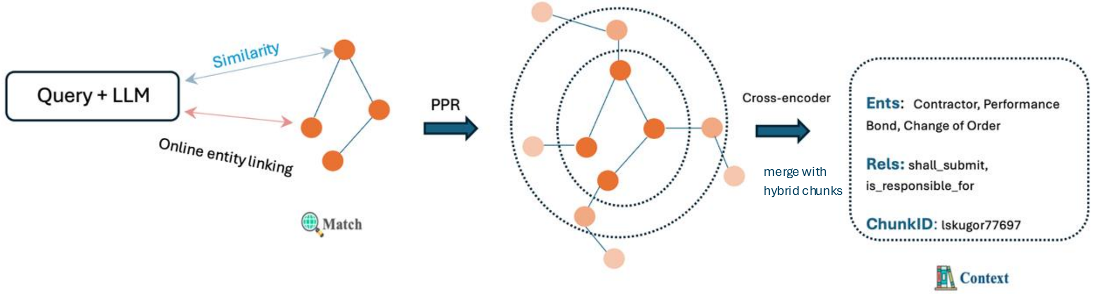
Graph Seed Entities

Graph-Based Retrieval Strategies

Dual-level RAG



Weighted Hub-aware PPR-RAG



Formulation:

$$\widetilde{w}_{ij} = \begin{cases} 0.85, & \text{if} \quad \deg(i) \leq \theta_{\text{hub}} \ \wedge \ \deg(j) \leq \theta_{\text{hub}}, \\ \dfrac{\log(\theta_{\text{hub}})}{\log(d_{\text{hub}}(i,j))}, & \text{otherwise,} \end{cases}$$

$$r_j^{(t+1)} = (1-\alpha)p_j + \alpha \sum_{i \in V} r_i^{(t)} P_{ij} + \alpha S^{(t)} p_j,$$

$$S^{(t)} = \sum_{i:\deg(i)=0} r_i^{(t)}$$

$$P_{ij} = \frac{\widetilde{w}_{ij}}{\sum_{k \in \mathcal{N}(i)} \widetilde{w}_{ik}}.$$

Where:

- P_ij: transition matrix
- p: personalization vector
- S_t: sink mass at time t
- r_t: relevance score at time t

Experimental Results & Analysis

Experimental Setup

200 questions

Question Type	Description	Method	Proportion
Factoid Queries, Single-hop Queries	Queries seeking precise information within a single project, such as contractual terms, bid amounts, or project entities.	Hybrid RAG	40%
Summarization Queries	Queries requiring the aggregation and summarization of information across multiple documents or engineering domains.	Dual-Level RAG	30%
Relational Queries	Queries involving multi-entity or multi-hop reasoning over the knowledge graph to infer implicit relationships.	PPR-RAG	30%

TABLE 9 – Statistical information of the dataset.

Statistic	Value
Total Documents	84
Total Chunks	8,572
Total Tokens	8,397,964
Total Entities	8,965
Total Relations	6,766

Evaluation method

LLM-as-a-judge +

ragas

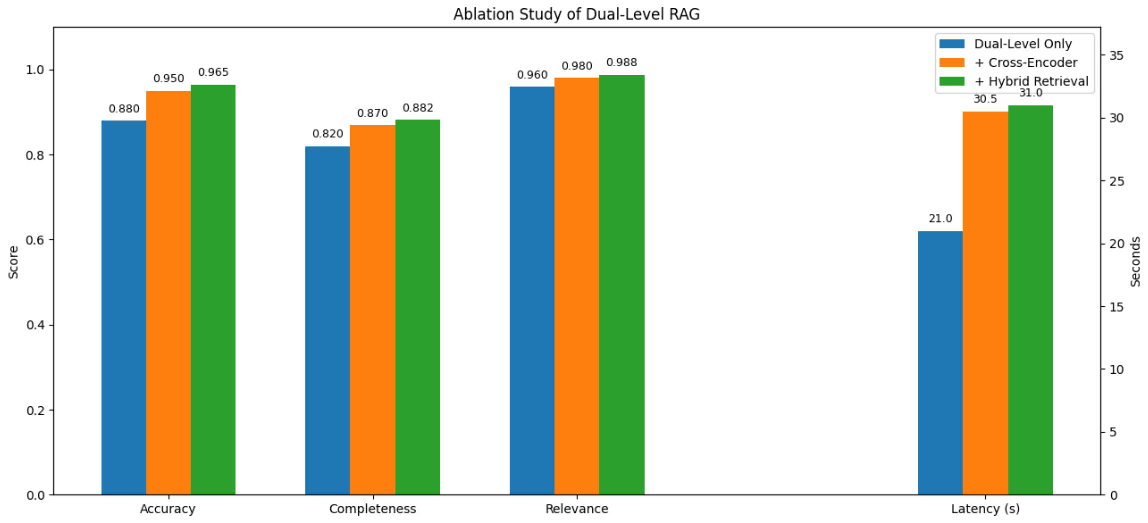
accuracy
completeness
relevance
Latency
context precision

Baseline

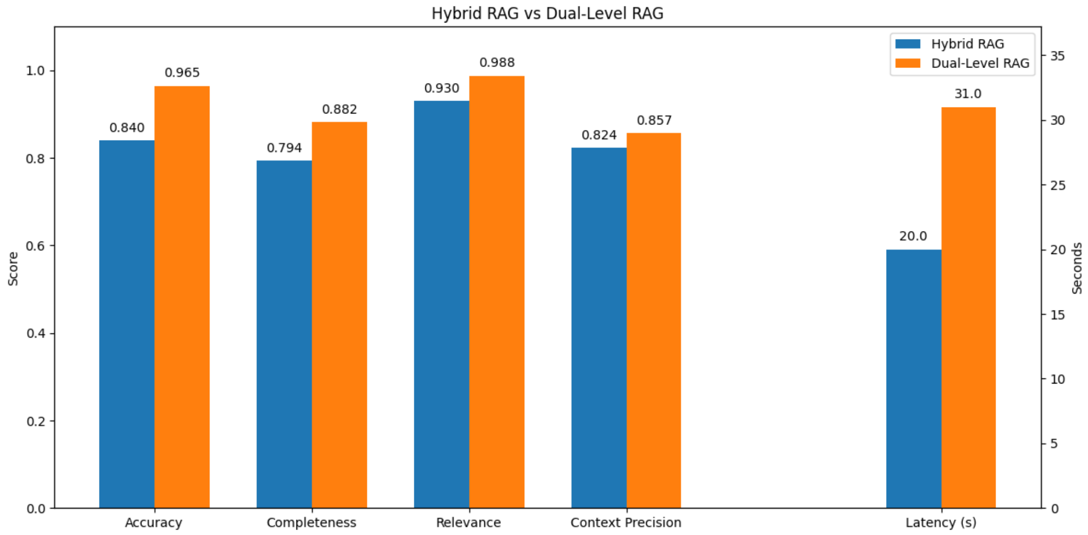
Hybrid RAG

Dual-level RAG result

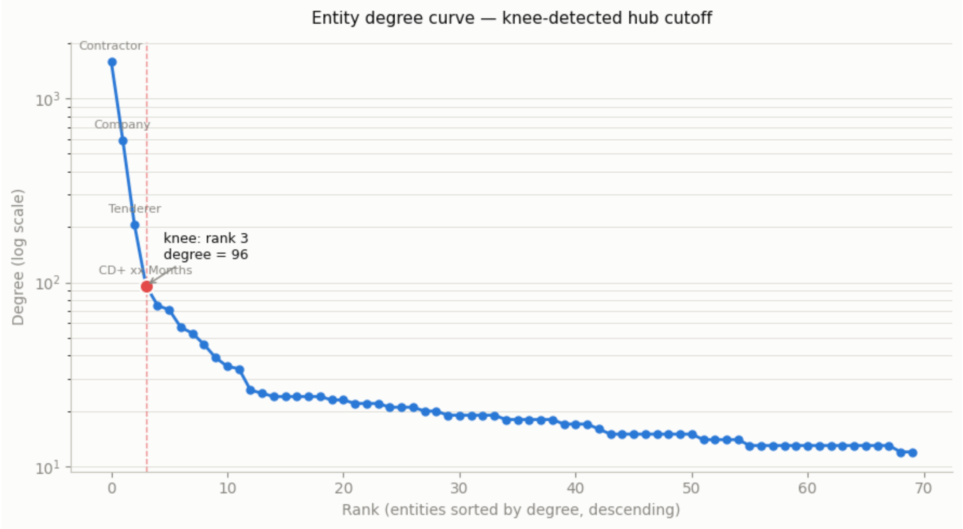
Ablation Study

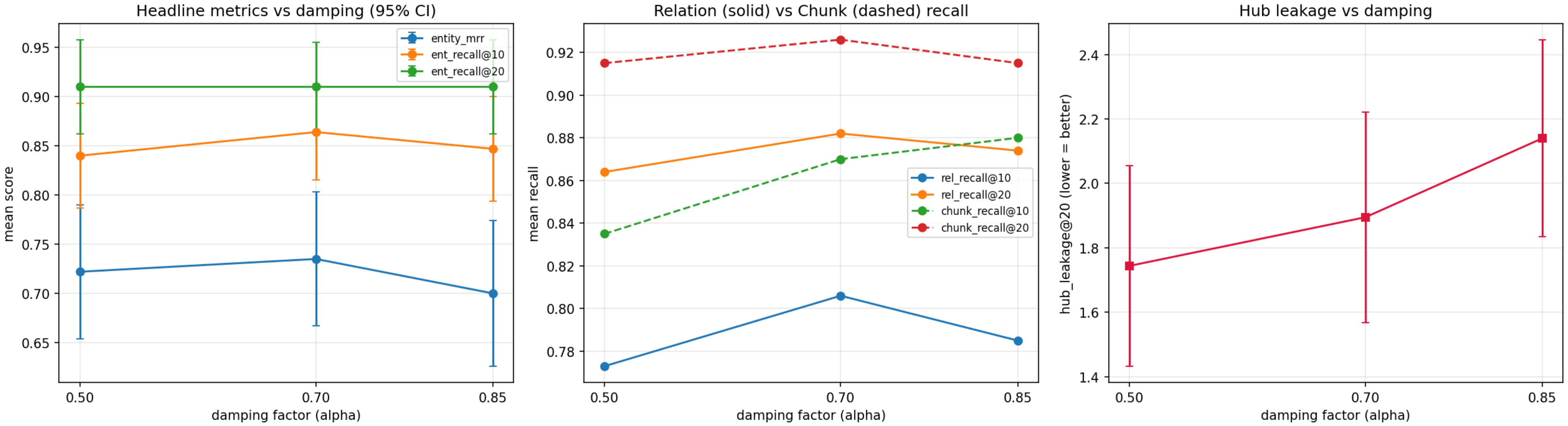


Global comparison



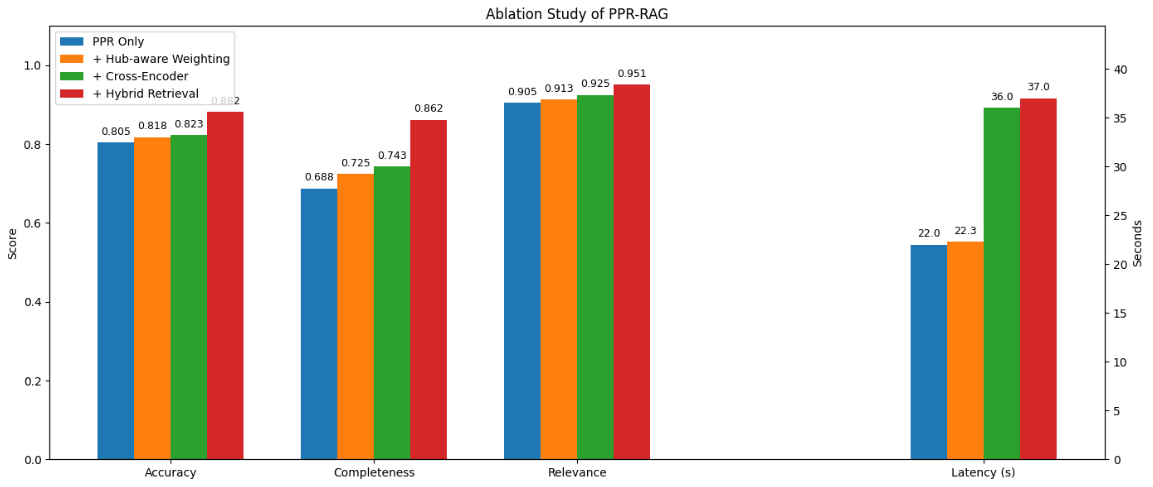
Tuning damping factor for PPR & Hub entity finding



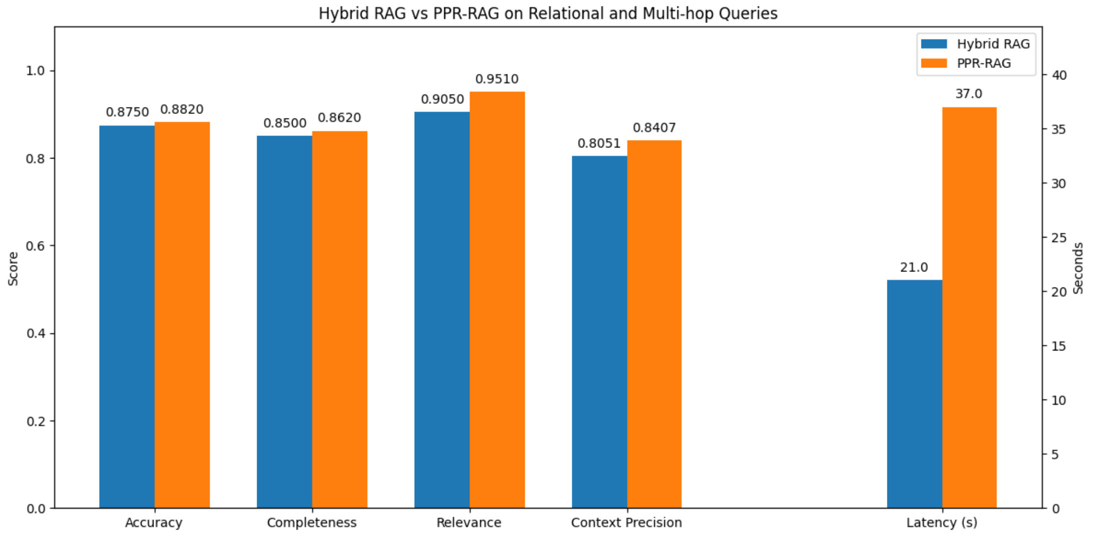


Weighted Hub-aware PPR-RAG result

Ablation Study



Global comparison

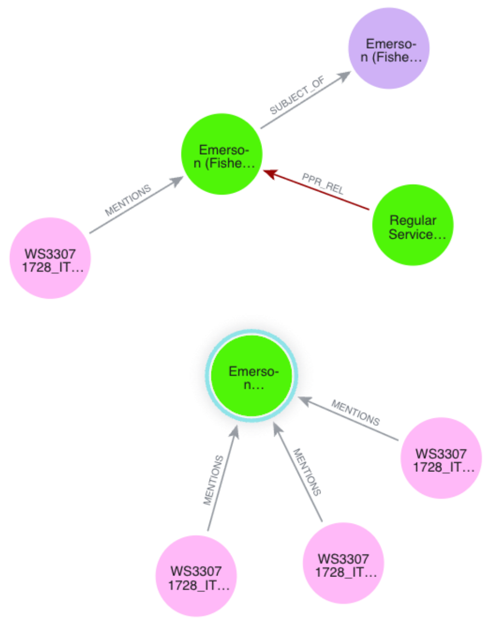


Limitations & Failure Analysis

Graph fragmentation

3,921 connected components
75.2% of components contain five nodes or fewer
Most entities are isolated or belong to very small local neighbourhoods , so PPR cannot propagate across disconnected components

Imperfect entity resolution



Conclusion & Future work

Conclusion

- Developed an **end-to-end domain-aware GraphRAG framework** for engineering tender documents
- **Dual-Level RAG clearly improved summarisation and aggregation** over Hybrid RAG
- **PPR-RAG does not give significant gains** for relational and multi-hop queries
- Graph retrieval and Hybrid RAG are **complementary rather than competing approaches**
- Main limitation: performance strongly depends on **predefined ontology**, **connectivity, entity resolution & entity linking**

Main takeaway
Knowledge graph constructed improve information aggregation, but reliable multi-hop retrieval requires a cleaner and better-connected graph.

Future work

- **Ontology refinement**
  - Clearer entity categories, relationship categories, and domain definitions
- **Do entity resolution**
  - Better acronym, alias, and terminology normalisation & definition
  - Merge semantically equivalent entities
- **Evaluation improvement**
  - Gold-standard question–answer benchmark
- **Cross-project generalisation**
  - Evaluate on other EPC tenders and industrial document project