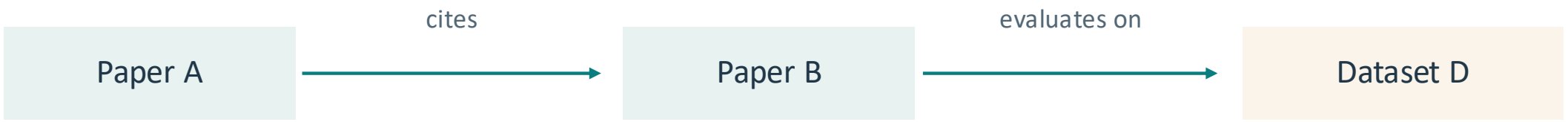
Integrating Graphs and Large Language Models

Architectures, evidence retrieval, and evaluation



Graph structure represents relations.

Language models interpret attributes, queries, and retrieved evidence.

01 / TASKS AND REPRESENTATIONS

Graph–language integration depends on the prediction task

TEXT-ATTRIBUTED GRAPH

Paper A

Node attribute

“We predict traffic speeds from nearby sensors.”

NODE CLASSIFICATION

Input: paper features and citation edges.
Output: a class label for the target paper.

GRAPH QUESTION ANSWERING

Input: a question and a graph.
Output: an answer supported by relevant relations.

A text-attributed graph contains nodes, edges, and textual attributes.
The task determines how language and topology interact.

01 / TASKS AND REPRESENTATIONS

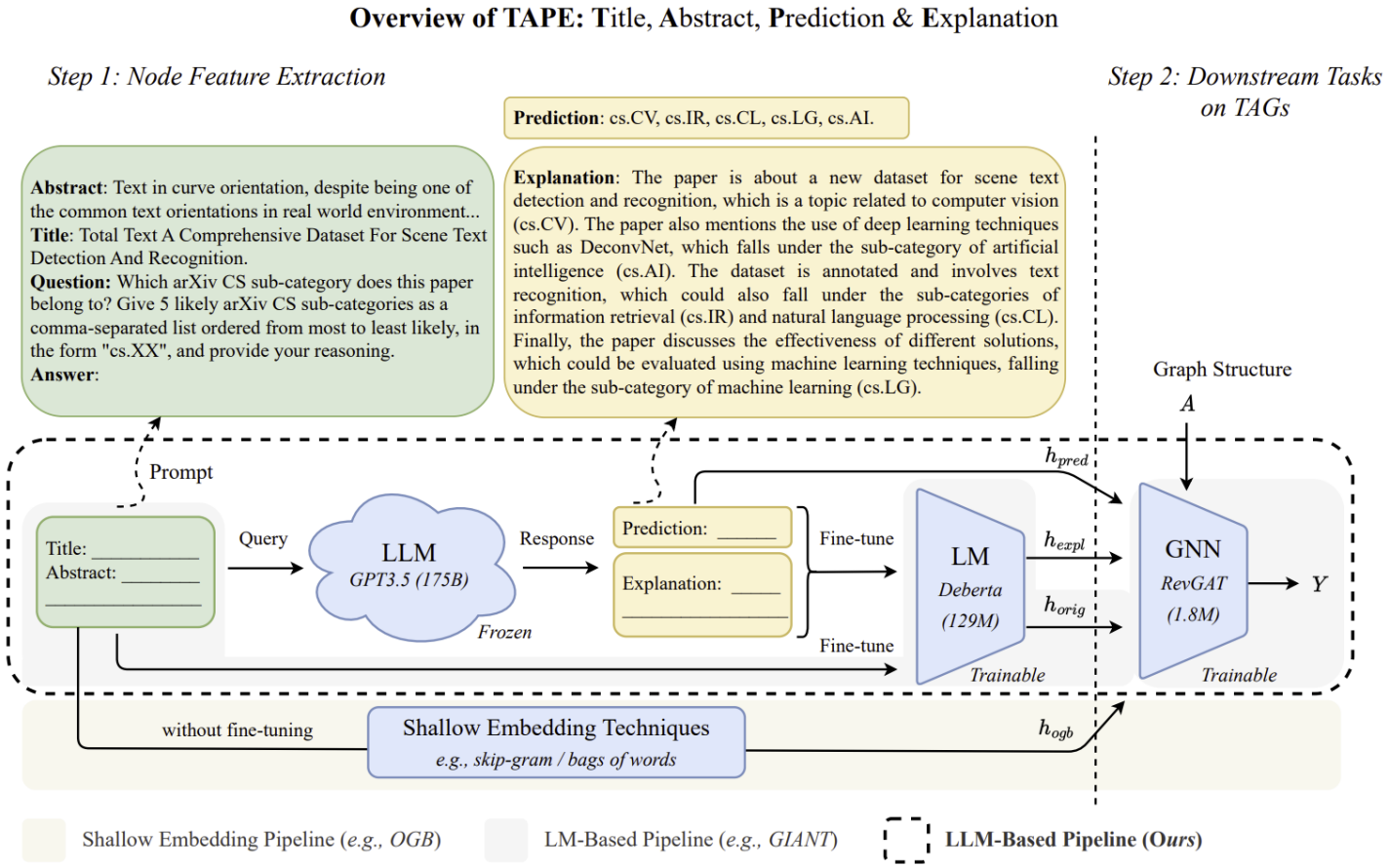
Three interfaces connect graph structure and language

Features	Node text	→	Language encoder	→	GNN prediction
Joint reasoning	Question + concepts	→	LM + graph updates	→	Candidate score
Retrieved evidence	Question + graph	→	Evidence selection	→	Generated answer

Trace the inputs and the prediction target before comparing model architectures.

02 / LANGUAGE-ENHANCED GRAPH LEARNING

TAPE augments node representations with LLM outputs



Tailored prompt LLM:

Ranked prediction list +
textual explanation as features

Small LM:
Interpret LLM outputs to GNN

He et al. Harnessing Explanations: LLM-to-LM Interpreter for Enhanced Text-Attributed Graph Representation Learning. ICLR, 2024.

02 / LANGUAGE-ENHANCED GRAPH LEARNING

TAPE converts an explanation into a supervised node feature

ILLUSTRATIVE INPUT

"Predict traffic speeds using measurements from neighboring road sensors."

LLM EXPLANATION

"The task predicts a traffic variable and uses spatial dependence among sensors."

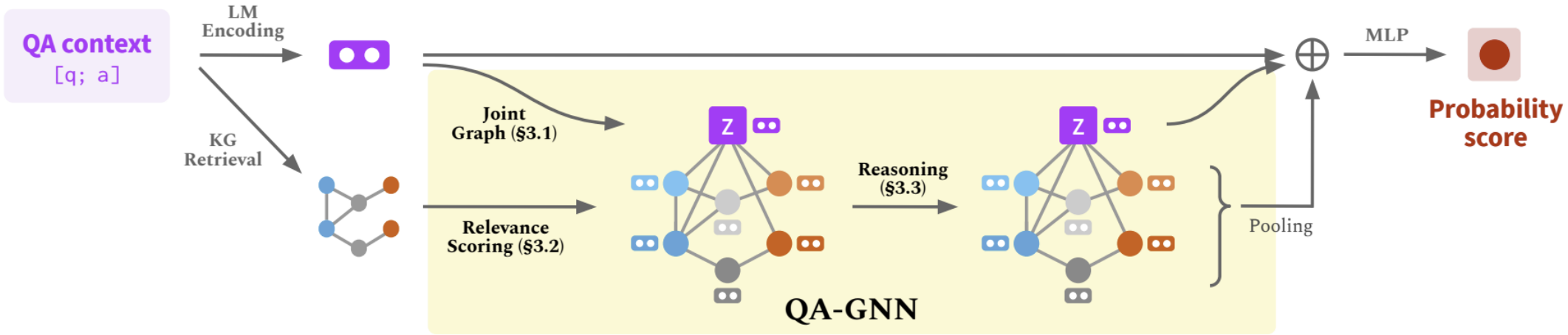
Title + abstract	Explanation text	Predicted classes
Text-based features	Explanation-based features	Prediction features

Each feature set is fed to its own GNN, the paper reports accuracy per feature set and averaged ensemble.

He et al. Harnessing Explanations: LLM-to-LM Interpreter for Enhanced Text-Attributed Graph Representation Learning. ICLR, 2024.

03 / JOINT LANGUAGE–GRAPH REASONING

QA-GNN jointly encodes language context and graph concepts

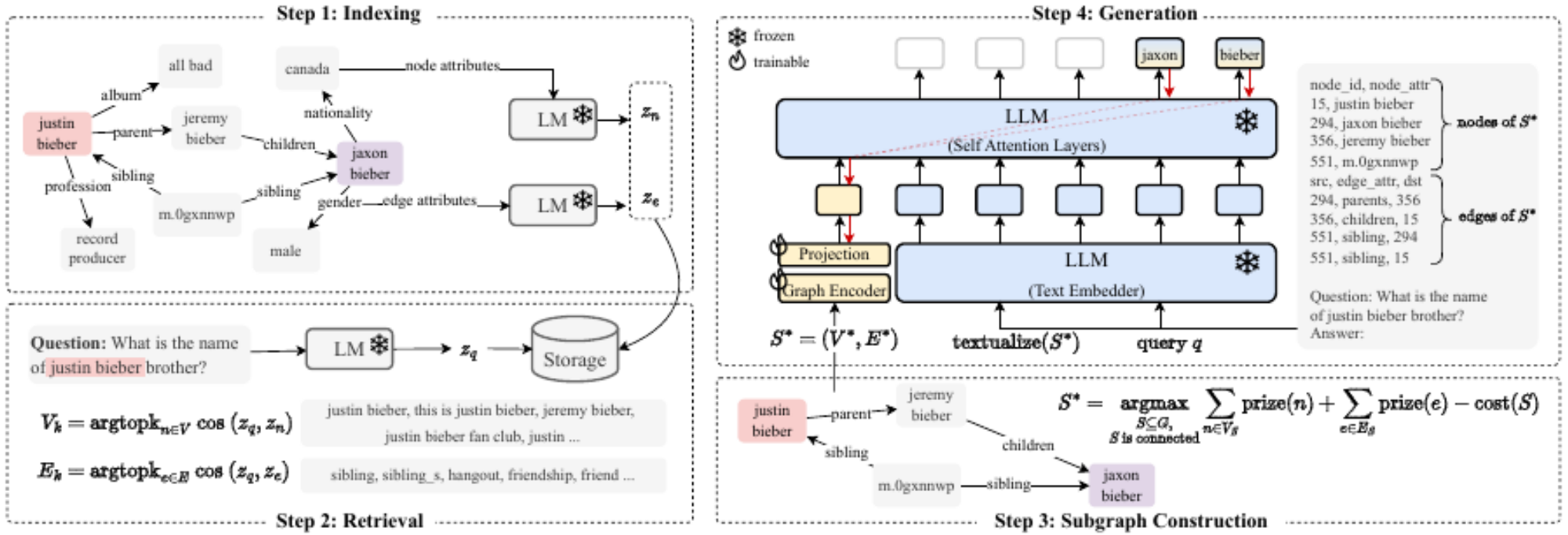


Relevance-aware graph attention updates concept and context representations.
A shared classifier scores each answer candidate.

Yasunaga et al. QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering. NAACL, 2021.

04 / RETRIEVAL-AUGMENTED GENERATION

G-Retriever combines subgraph retrieval and graph prompt tuning

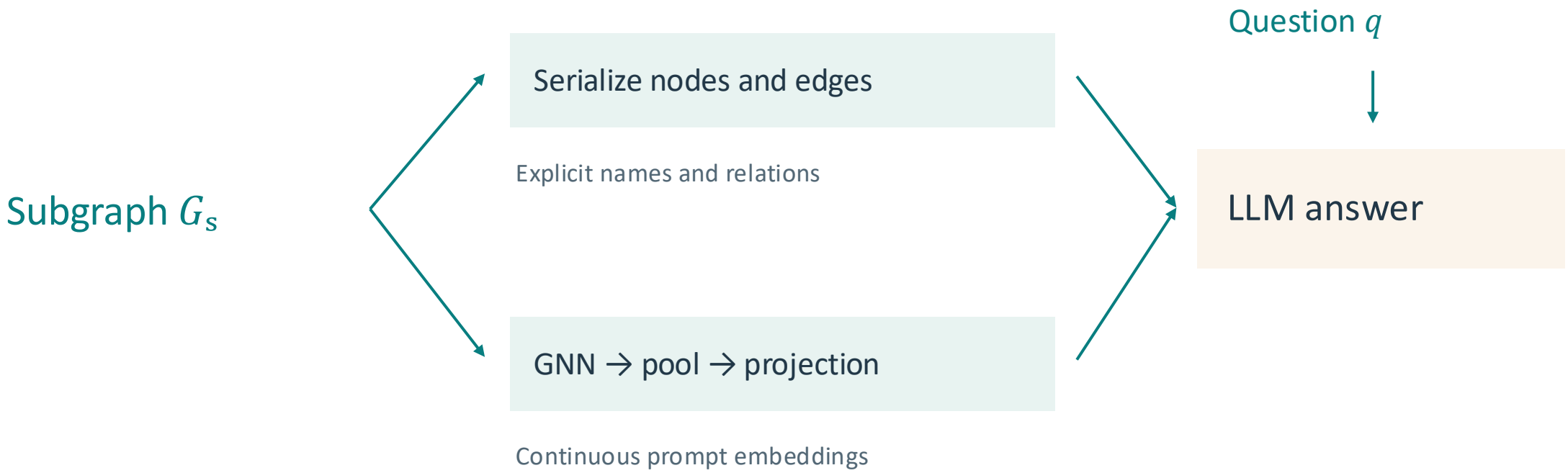


Two evidence channels condition the LLM:
serialized nodes and edges, and a projected GNN representation of the subgraph.

He et al. G-Retriever: Retrieval-Augmented Generation for Textual Graph Understanding and Question Answering. NeurIPS, 2024.

04 / RETRIEVAL-AUGMENTED GENERATION

G-Retriever supplies both text and a learned graph prompt



Frozen-LLM configuration:
answer supervision trains the graph encoder and projection through the language-model loss.

04 / RETRIEVAL-AUGMENTED GENERATION

Multi-hop evidence preserves the roles of intermediate entities

Query: Which dataset was used by a paper cited by A?



Relevant evidence

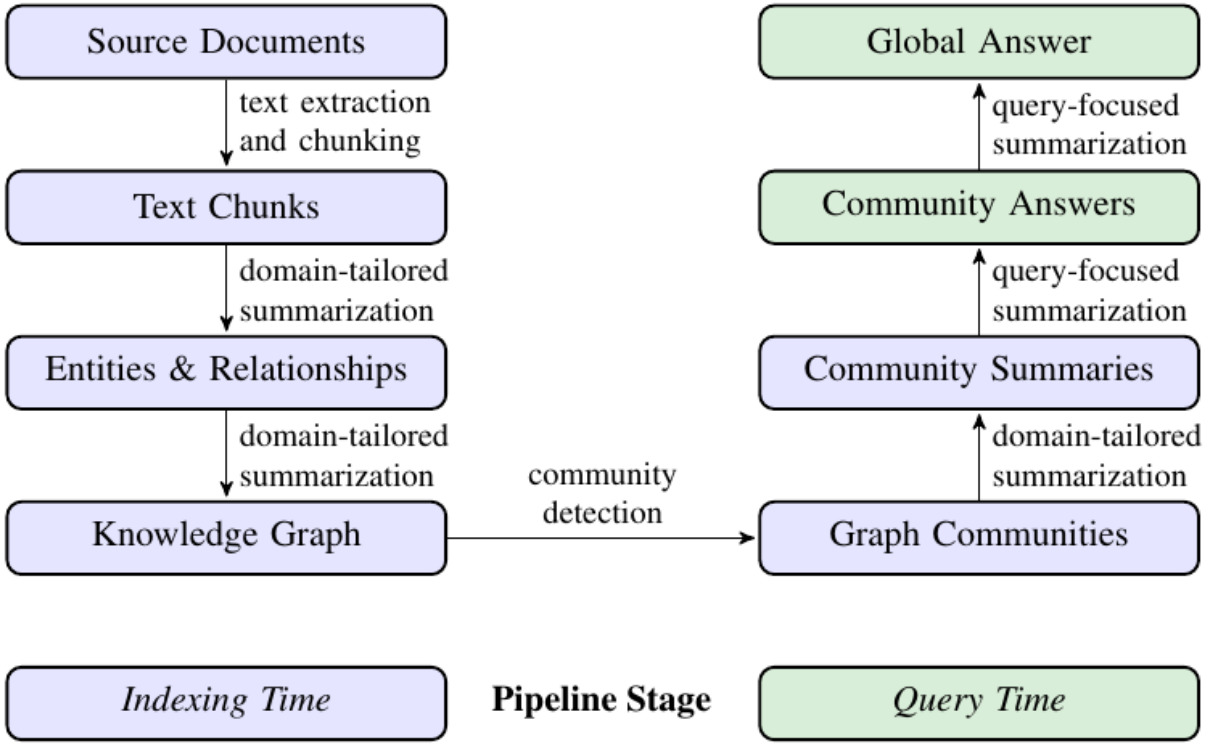
(A, cites, B) and (B, evaluates on, D)

Researcher C

The authorship branch supplies context for a different query.

Answer: Paper B, cited by A, evaluated its method on Dataset D.

Global GraphRAG synthesizes evidence from communities



OFFLINE INDEX CONSTRUCTION

Extract entities and relations. Detect hierarchical communities and summarize their evidence.

QUERY-TIME SYNTHESIS

Generate partial answers from community reports, then combine the relevant contributions.

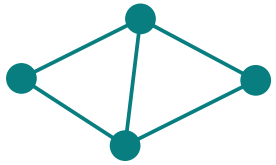
Edge et al. From Local to Global: A Graph RAG Approach to Query-Focused Summarization. arXiv:2404.16130, 2024 (revised 2025).

04 / RETRIEVAL-AUGMENTED GENERATION

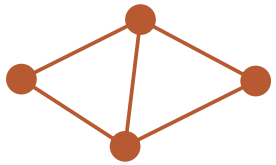
Community reports support questions spanning many papers

Query: What research themes recur across this collection?

COMMUNITY



Traffic prediction



Network optimization

REPORT → QUERY-SPECIFIC ANSWER

Forecasting combines measurements across connected roads.

Allocation methods account for constraints across a network.

SYNTHESIS

Recurring themes: prediction on networks and resource allocation

Illustrative corpus. Each report summarizes evidence from one community before query-specific synthesis.

05 / COMPARISON AND RESEARCH STATUS

Four architectures differ in evidence and prediction mechanism

METHOD	PREDICTION TARGET	COMPUTATIONAL ROLE OF THE GRAPH
TAPE	Node class	GNN propagation over language-derived features
QA-GNN	Answer-choice score	Joint context–concept message passing
G-Retriever	Generated graph answer	Connected evidence + learned graph prompt
GraphRAG	Corpus-level summary	Community-based indexing and report synthesis

Select a baseline with the same task, training labels, language backbone, and permitted evidence.

05 / COMPARISON AND RESEARCH STATUS

A controlled comparison tests what graph structure adds

Hold fixed — Questions, corpus, LLM, token budget, and evaluation split

A	Passage retrieval	Text chunks selected from the corpus
B	Graph retrieval	Selected subgraph serialized as text
C	Graph + soft prompt	Same subgraph, plus the learned embedding

A→B tests the evidence-selection pipeline. B→C tests the graph prompt, using the same selected facts.

05 / RESEARCH STATUS, NOVEMBER 2026

Recent evaluations identify task-dependent benefits of graph access

LLM interaction with node data

The strongest overall node-classification performance with code generation --
for **long text** and **high-degree nodes.**

Retrieval for complex questions

GraphRAG-Bench found competitive passage retrieval for **simple facts** and clearer graph-based gains for **combining related evidence**.

Research priorities	Robustness to missing or incorrect edges, faithful use of retrieved evidence, and quality under fixed token and latency budgets.

Finkelshtein et al. Actions Speak Louder than Prompts: A Large-Scale Study of LLMs for Graph Inference. ICLR, 2026.

Xiang et al. When to use Graphs in RAG: A Comprehensive Analysis for Graph Retrieval-Augmented Generation. arXiv:2506.05690v3, 2026.

06 / SYNTHESIS

Evaluate the graph contribution at each stage of prediction

Node classification	Compare text-only and graph-enhanced predictors with matched labels, splits, and feature encoders.
Graph question answering	Measure required-evidence retrieval and answer correctness, including relation direction and entity attribution.
Global summarization	Assess coverage across communities and factual support in the original source passages.

Report prediction quality together with graph-construction cost, retrieval cost, and query-time latency.

REFERENCES / CLICKABLE PAPER TITLES

References for the architectures and evaluation studies

He et al. (ICLR 2024)

Harnessing Explanations: LLM-to-LM Interpreter for Enhanced Text-Attributed Graph Representation Learning

Yasunaga et al. (NAACL 2021)

QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering

He et al. (NeurIPS 2024)

G-Retriever: Retrieval-Augmented Generation for Textual Graph Understanding and Question Answering

Edge et al. (arXiv 2024, revised 2025)

From Local to Global: A Graph RAG Approach to Query-Focused Summarization

Finkelshtein et al. (ICLR 2026)

Actions Speak Louder than Prompts: A Large-Scale Study of LLMs for Graph Inference

Xiang et al. (arXiv v3, 2026)

When to use Graphs in RAG: A Comprehensive Analysis for Graph Retrieval-Augmented Generation

Full references also accompany the relevant architecture slides and speaker notes.