

# Graph RAG

## Motivation

在当前大语言模型（LLM）驱动的回答系统中，检索增强生成（RAG）已成为主流方案。而 **GraphRAG** 作为一种结合图结构与文本语义的检索增强方法，具有结构化信息处理和多跳推理的潜力。本次调研主要围绕两个问题展开：

- GraphRAG** 作为向量数据库能否输出结构化数据：传统向量数据库更偏向于“黑箱式”的 Top-K 检索，而 GraphRAG 是否具备结构化输出能力（如实体、关系、社区）将决定其在知识组织和追溯中的适用范围。
- GraphRAG** 的图关系处理流程是否适用于人物关系分析：人物关系往往具有复杂的多跳连接、角色依赖和社群结构，GraphRAG 能否以结构化方式刻画并支持推理，是评估其在情报研判、社交网络分析等场景中的关键。

## GraphRAG survey

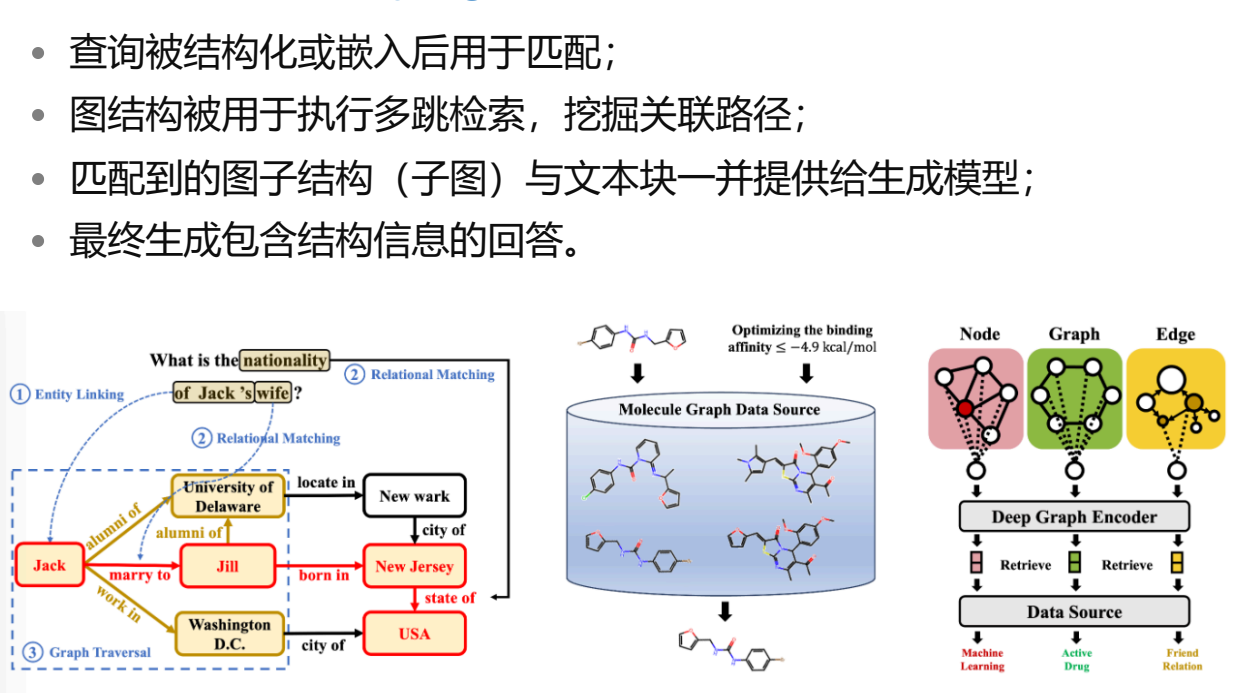


Figure 3: A holistic framework of GraphRAG and representative techniques for its key components.

GraphRAG 在 RAG 框架基础上引入图结构建模，整体机制包括两个核心阶段：

### 1. 索引阶段（Indexing）

- 文本被分块（chunking）成较小的单元（TextUnits）；
- 从文本中提取实体与关系，构建为图结构；
- 文本块与图结构分别进行嵌入，供后续查询使用；
- 图的节点/边与文本单元相链接，实现图-文联合索引。

### 2. 查询阶段（Querying）

- 查询被结构化或嵌入后用于匹配；
- 图结构被用于执行多跳检索，挖掘关联路径；
- 匹配到的图子结构（子图）与文本块一并提供给生成模型；
- 最终生成包含结构信息的回答。

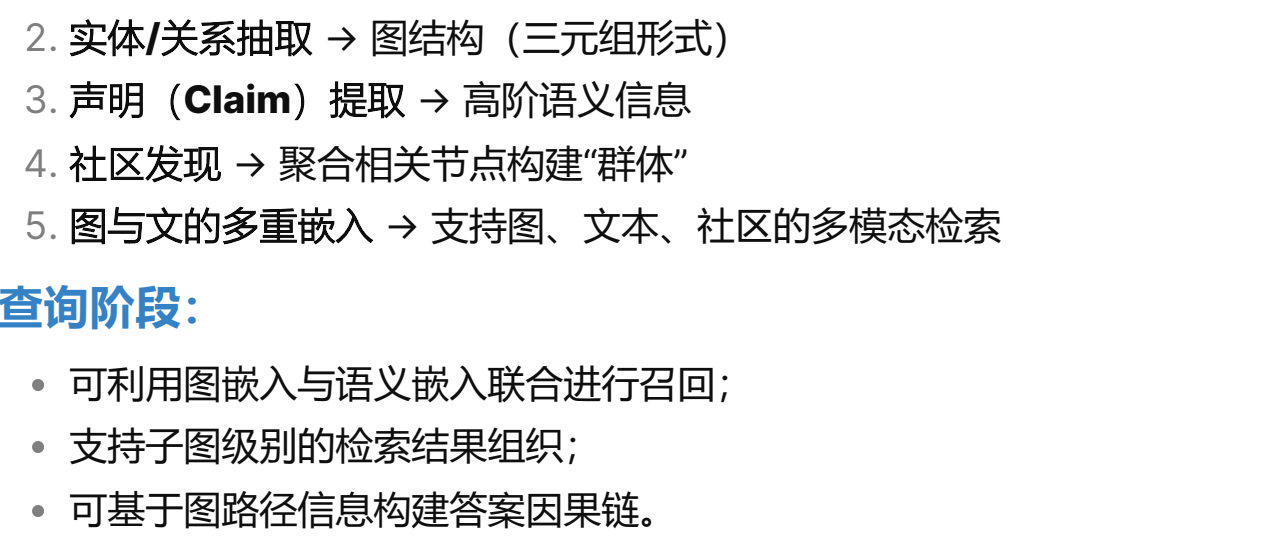
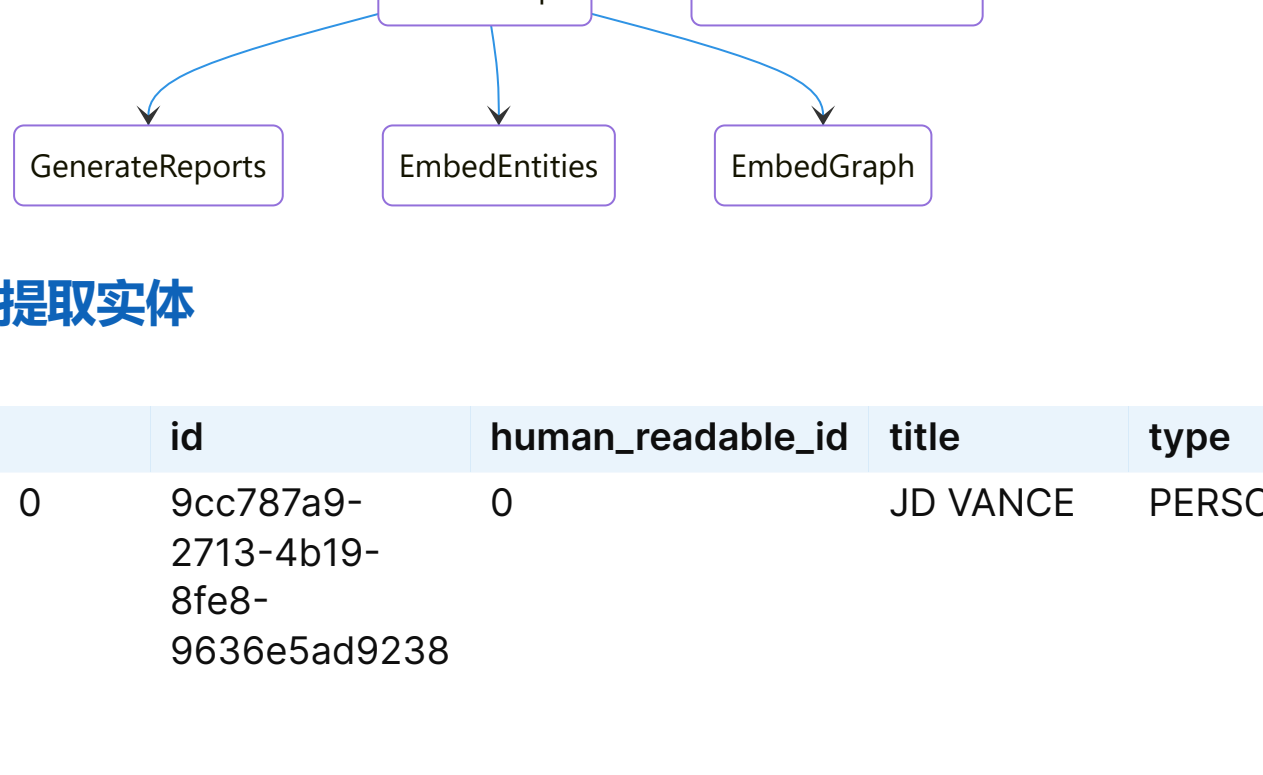


Figure 5: Visualizing representative retrievers used in GraphRAG.

🔍 与传统**RAG**比较优势：传统 RAG 通常基于文本块的相似度检索，忽略实体之间的结构信息。GraphRAG 的引入让检索具备“语义 + 图结构”双重约束，尤其在多跳推理、实体对齐、因果链追踪等任务中具有天然优势。

📊 微博数据集，社会影响力局部性：建模转发行为 | [AMiner --- Social Influence Locality for Modeling Retweeting Behaviors](#) | [AMiner](#)

## Microsoft GraphRAG



## 三、Microsoft GraphRAG 的索引与查询机制及其改进点

微软开源的 GraphRAG 版本在传统 GraphRAG 基础上进一步增强了结构提取和模块化能力，主要流程包括：

### 索引阶段的关键模块：

- 文本分块（**Chunking**）→ 文本单元（TextUnits）
- 实体/关系抽取 → 图结构（三元组形式）
- 声明（**Claim**）提取 → 高阶语义信息
- 社区发现 → 聚合相关节点构建“群体”
- 图与文的多重嵌入 → 支持图、文本、社区的多模态检索

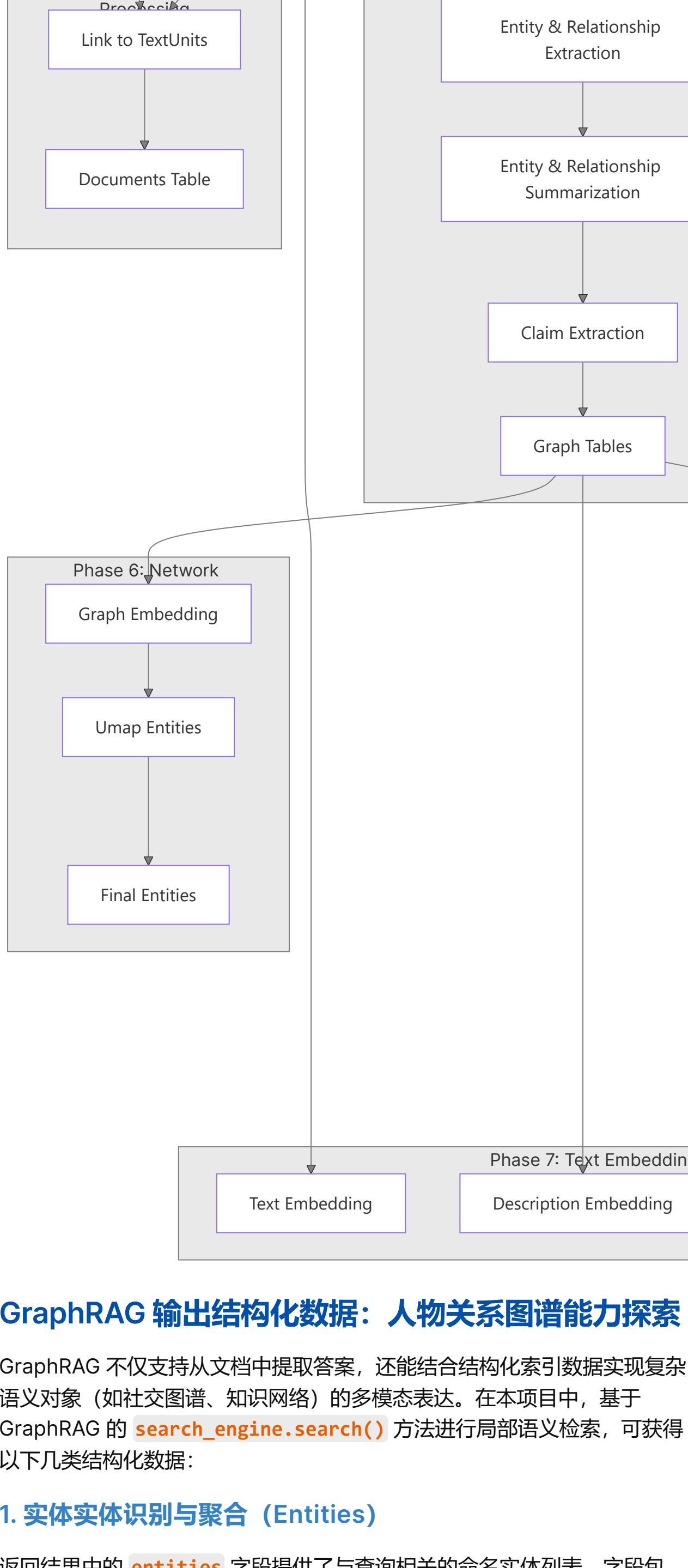
### 查询阶段：

- 可利用图嵌入与语义嵌入联合进行召回；
- 支持子图级别的检索结果组织；
- 可基于图路径信息构建答案因果链。

### VS 区别与优势：

- Microsoft GraphRAG 明确输出结构化结果（表格、社区、UMAP 结构）；
- 支持多阶段数据流（如 Text → Graph → Community → Summary）；
- 更适合用于复杂知识工程、情报挖掘和多跳问答任务。

## Index Workflow



## 提取实体

	id	human_readable_id	title	type
0	9cc787a9-2713-4b19-8fe8-9636e5ad9238	0	JD VANCE	PERSC
1	b5b92777-b01b-445d-a240-83b86ea5b865	1	DONALD TRUMP	PERSC
2	38863ac9-8440-4299-bf39-c96037bc58a8	2	MIKE PENCE	PERSC
3	be99bf04-955c-4787-a549-f518c68f3b8d	3	HILLBILLY ELEGY	EVENT
4	dc8cfd3a-f305-49d6-a829-4861fbc37946	4	MAGA MOVEMENT	ORGANIZATION

	id	human_readable_id	source	target
0	118ff68d-e9c6-4dfb-e853c-f658e03ee46f	0	JD VANCE	DONALD TRUMP
1	3ba9bd73-a908-4426-b8f1-c95263a983ca	1	JD VANCE	MAGA MOVEMENT
2	700d6acb-09eb-4cc8-845b-9ba8d7ee78fe	2	JD VANCE	HILLBILLY ELEGY
3	a482a176-d94f-4f4c-bf05-38579b8d13a0	3	JD VANCE	MIDDLETOWN OHIO
4	0cfb2b8f-f8f9-4a2b-b5de-d828e8df525d	4	JD VANCE	YALE LAW SCHOOL

	id	source	target	description	weight
0	112	VANCE	U.S. SENATE	Vance serves as a senator in the U.S. Senate, ...	8.0
1	97	VANCE	HOWIE BEIGELMAN	Vance has worked closely with Howie Beigelman ...	7.0
2	114	VANCE	HALF ISLAND TV	Half Island TV published commentary on Vance's...	4.0
3	35	J.D. VANCE	DONALD TRUMP	J.D. Vance and Donald Trump are prominent figu...	27.0
4	50	DONALD TRUMP	J.D. VANCE	Donald Trump and J.D. Vance share a complex po...	17.0
...	...	...	...	...	...
68	41	J.D. VANCE	UKRAINE	J.D. Vance has strongly opposed providing war ...	7.0
69	42	J.D. VANCE	TAIWAN	J.D. Vance's stance on Taiwan aligns with Dona...	6.0
70	60	J.D. VANCE	MIDDLETOWN	J.D. Vance was born and raised in Middletown, ...	25.0
71	74	J.D. VANCE	KENTUCKY	J.D. Vance's grandparents migrated from Kentuc...	7.0
72	75	J.D. VANCE	OUR OHIO RENEWAL	J.D. Vance founded Our Ohio Renewal to address...	6.0

例如，对于“Vance的社交关系”查询，系统输出了其与 Donald Trump、Howie Beigelman、Yoram Hazony、Mitt Romney 等人物的多条明确互动关系，涉及政治联盟、宗教合作、思想传承等维度。

构建和索引图数据库是 GraphRAG 的基础，其中图数据库的质量直接影响 GraphRAG 的性能。图数据来源主要包括两类：开放知识图谱和自建知识图谱。

实体别名

## Prompt Tuning

```
graphrag init --root ./ragtest
```

```
graphrag index --root ./ragtest
```

```
graphrag query \
--root ./ragtest \
--method local \
--query "Vance的社交人物关系"
```

```
python -m graphrag prompt-tune \
--root ./ragtest \
--config ./ragtest/settings.yaml \
--language Chinese \
--domain "Social Relationship Analysis"
```