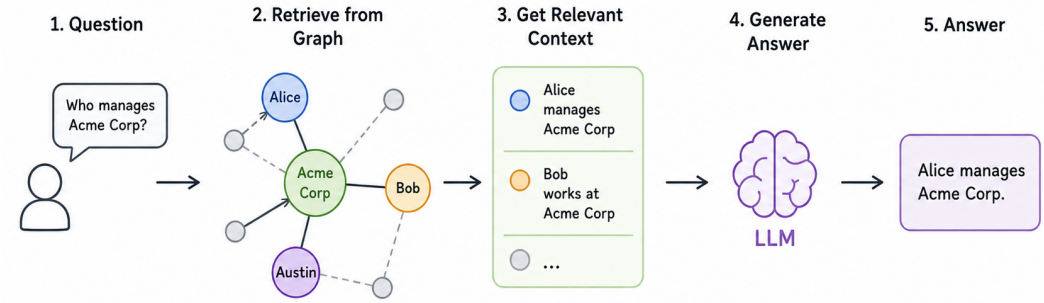
# How Graph Retrieval Design Shapes LLM Reasoning

Valentin SIX, Gaël DE CHALENDAR, Evan DUFRAISSE

May 16, 2026

# What is GraphRAG



# Why decompose questions?

**Complex Question:**

*"Which city is the birthplace of the author of the novel "1984"?"*

**Subquestions:**

1. Who is the author of the novel "1984"?
2. Where was this author born?

# Why decompose questions?

**Complex Question:**

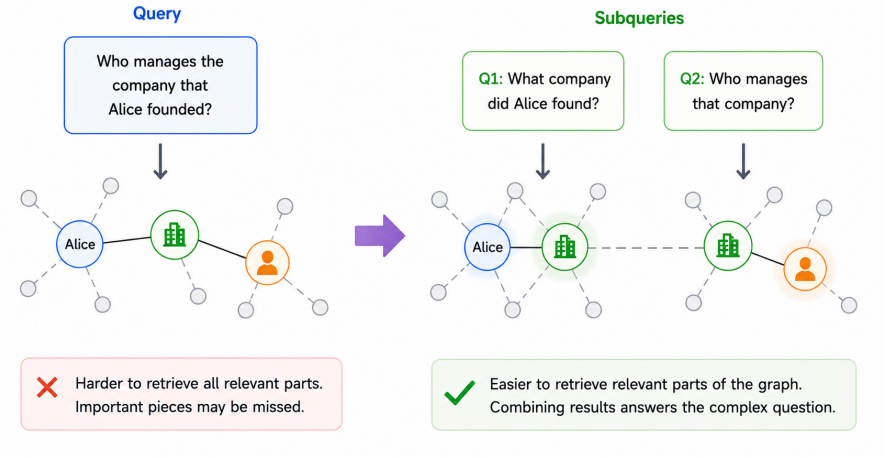
*"Which city is the birthplace of the author of the novel "1984"?"*

**Subquestions:**

1. Who is the author of the novel "1984"?
2. Where was this author born?

**Decomposition makes implicit reasoning explicit.**

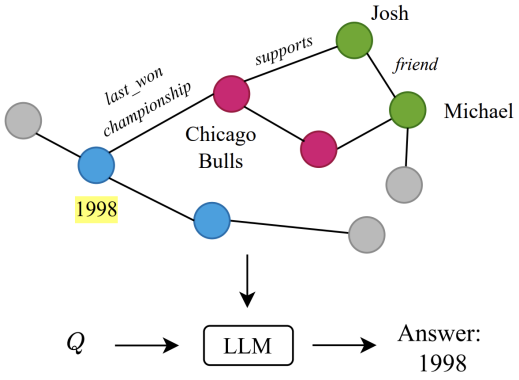
# Query decomposition helps local retrieval



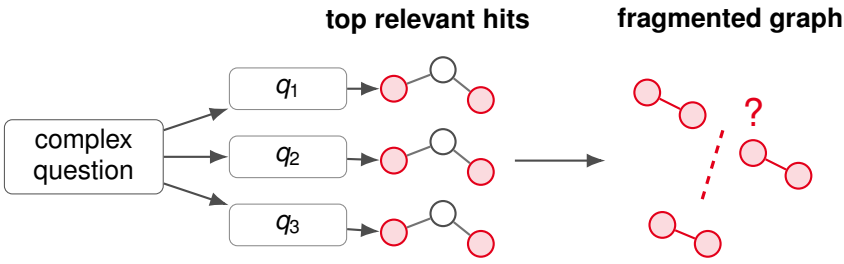
# End-to-end retrieval example

$Q$ : When did the team that Michael's best friend support last win the Championship ?

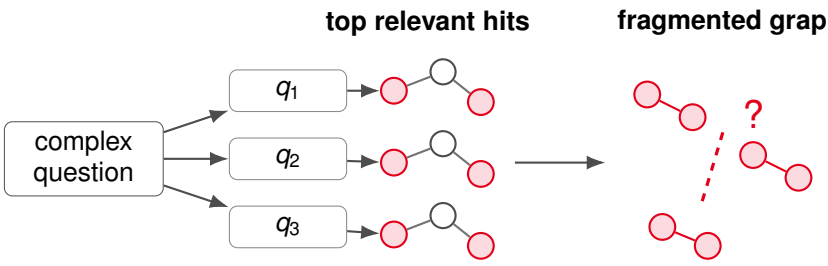
$q_1$ : Who is Michael's best friend ?
$q_2$ : What team does he support ?
$q_3$: When did that team last win the Championship ?



# Relevant evidence is not enough

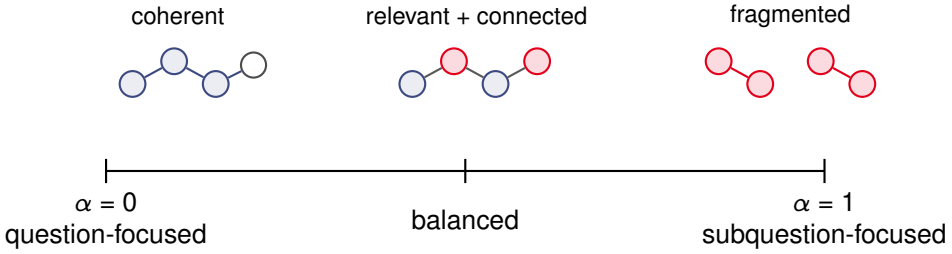


# Relevant evidence is not enough



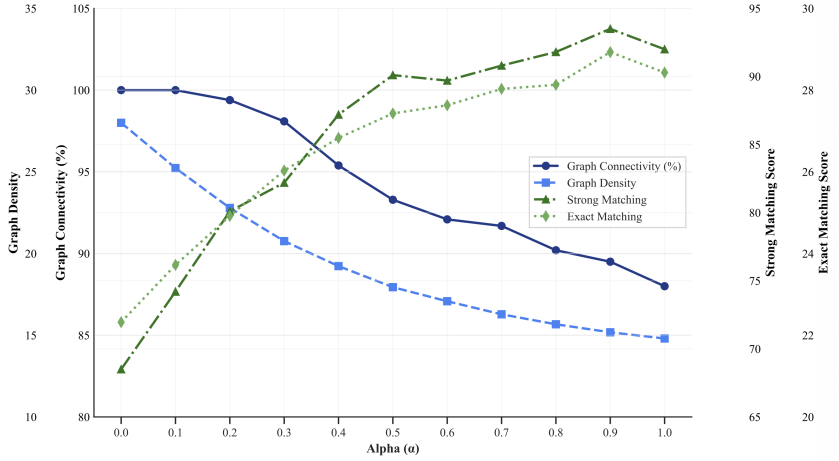

**The LLM may get the facts, but miss the chain.**

# A simple trick: balance query and subqueries

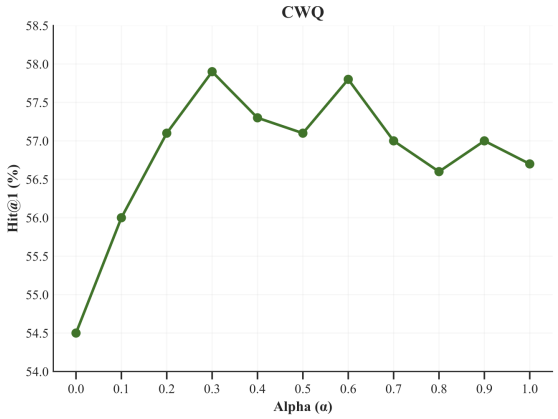


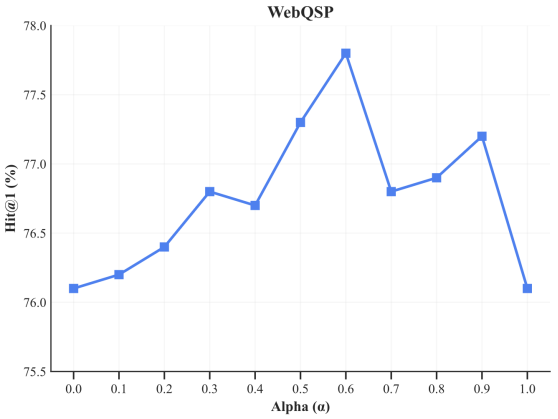

$$\text{score}_\alpha = (1 - \alpha)\,\text{original question} + \alpha\,\text{subquestion}$$

# Trade-off: relevance up, structure down



# LLM accuracy peaks when both are balanced



# What have we learned?

**1**

**Relevance is not enough**

The retrieved graph must also be usable.

**2**

**Fragmentation is the risk**

Subqueries can retrieve good pieces in isolation.

**3**

**Balance works best**

LLM accuracy peaks between the two extremes.

# What have we learned?

**1**

**Relevance is not enough**

The retrieved graph must also be usable.

**2**

**Fragmentation is the risk**

Subqueries can retrieve good pieces in isolation.

**3**

**Balance works best**

LLM accuracy peaks between the two extremes.

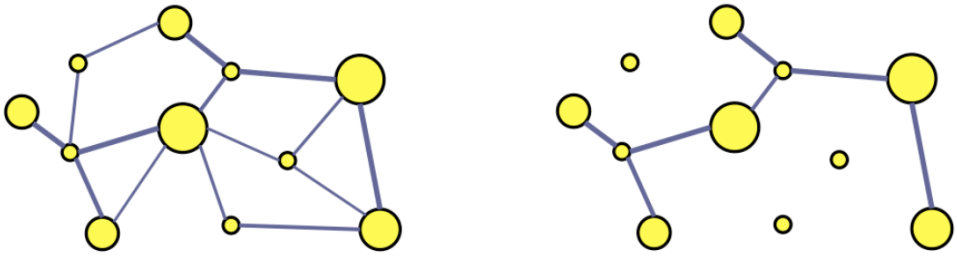
**For KG retrieval, topology is part of the context.**

# Shameless promotion :)


Interested in NLP or GraphRAG?

Feel free to connect and reach out to the CEA-LIST / LASTI lab!

# Backup: Prize-Collecting Steiner Tree



Akhmedov et al. (2018)

**PCST keeps useful evidence connected and compact.**

High-prize nodes are kept when their value outweighs edge costs. PCST keeps the links needed for a usable subgraph and removes low-value branches.