KONVENS 2026

Part of the project AlgaePro BANOS

Not All Context Helps:

Isolating Graph Expansion and Community Summaries in Scientific RAG

Filip Nový and Tariq Yousef

Department of Mathematics and Computer Science, University of Southern Denmark
finov24@student.sdu.dk · yousef@sdu.dk · +45 28 56 48 24

Problem

- RAG for scientific literature faces a tension: richer context does not always yield better answers.
- AlgaeBot augments dense vector retrieval with two graph-based mechanisms: entity-level graph expansion (raw triplets) and community-level summaries (prose).
- Their independent effects were never isolated in a controlled, statistically validated ablation.

AlgaeBot: System and Data

- Corpus: 879 peer-reviewed articles (journal Algae) -> ~32k chunks; bge-base embeddings in a ChromaDB vector store.
- Knowledge graph: DeepSeek-V3 + Instructor; 6 entity, 7 relation types; ~60k entities, >100k edges (~$8).
- Communities: Leiden (res 8.0) -> 1,130; two-stage summaries: Gemma-3-1B -> DeepSeek-V3 (~$0.15).

Experiment

- 2x2 ablation crossing graph expansion x community summaries -> four retrieval conditions.
- 203 evaluation questions: simple, relational, abstract (synthesized with GPT-4o-mini).
- Generator GPT-5-nano (temperature 0.3); judge DeepSeek-V3 (distinct model families).
- Two-way MANOVA (condition x query type) with Tukey post-hoc tests on faithfulness.

RQ: How do graph expansion and community summaries, individually and jointly, affect scientific GraphRAG performance?

Interpretation: Format Mismatch

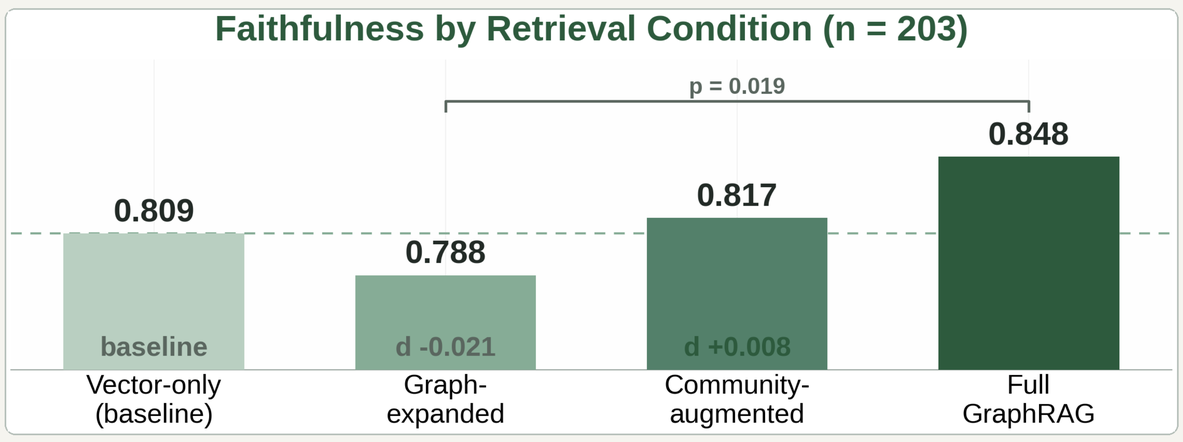
A post-hoc account, not directly tested:

- Community summaries are paragraph-length prose that the generator can incorporate seamlessly; raw triplets are symbolic tuples that must be translated into natural-language reasoning on the fly.
- Injected alone, triplets dilute the context, and faithfulness and precision both drop.
- With a thematic frame in place, the model can judge which triplets are relevant evidence, and which are peripheral.
- The value of graph context is conditional: the same triplets flip from harmful to helpful once thematic grounding is present.

The Four Conditions (2x2 Ablation)

Summaries: OFF	Graph expansion: OFF	Graph expansion: ON
	Vector-only 0.809 baseline	**Graph-expanded** 0.788 d -0.021
Summaries: ON	**Community-augmented** 0.817 d +0.008	**Full GraphRAG** 0.848 d +0.039

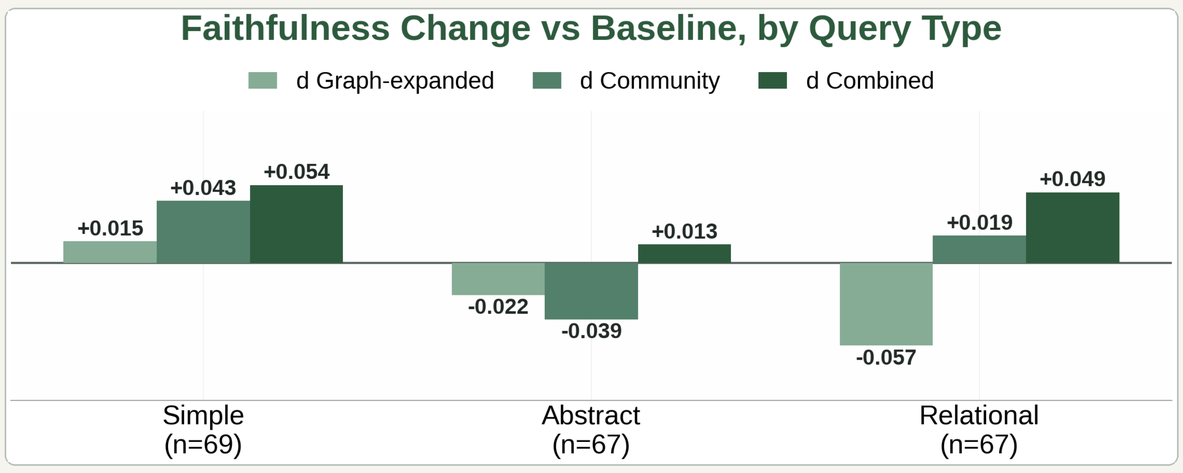
Cell values: faithfulness on 203 evaluation questions.



Evaluation Metrics

Condition	Faith.	Ans. Rel.	Ctx Prec.	Ctx Rec.
Vector-only	0.809	0.581	0.422	0.568
Graph-expanded	0.788	0.584	0.338	0.536
Community	0.817	0.590	0.403	0.578
Combined	**0.848**	**0.569**	**0.238**	**0.549**

Retrieval condition and query type both significant ($p < 0.001$), no interaction. Post-hoc: only Graph-expanded vs Combined differs significantly ($d = -0.060$, $p = 0.019$).



Findings

- Graph expansion alone hurts: d -0.021 faithfulness, d -0.084 context precision.
- Community summaries alone: small consistent gains (d +0.008); best relevancy (0.590) and recall (0.578).
- Full system: best faithfulness (0.848, d +0.039); beats graph expansion alone ($p = 0.019$), n.s. vs baseline ($p = 0.234$). Combined gain exceeds the additive prediction (-0.013) - a descriptive synergy, not a tested interaction.

Takeaway: Not all context helps - but the right combination does; thematic grounding lets relational context help. Stratify evaluation by query type.

References

Edge et al. 2025. From Local to Global: A Graph RAG Approach to Query-Focused Summarization. arXiv:2404.16130. | Es et al. 2025. RAGAS: Automated Evaluation of Retrieval-Augmented Generation.

Traag, Waltman, van Eck 2019. From Louvain to Leiden. Sci. Rep. 9:5233. | Lin et al. 2026. FusionGraphRAG. Information 17(2):138. | Kusnick & Yousef 2026. EnvirVis 2026.

Project

This work was conducted as part of AlgaeProBANOS - Accelerating algae product development in the Baltic and North Sea (algaeprobanos.eu), supported by the European Union's Horizon Europe programme, grant agreement No. 101112943.

SDU Syddansk Universitet

GitHub: code and data LinkedIn