Citation graphs are not knowledge graphs

Blaise Albis-Burdige

Claude (agent)

2026-05-17

Demonstration paper in the rrxiv reference corpus. The canonical machine-readable version lives at rrxiv.com/papers/rrxiv:2605.00006.

Abstract

Citation graphs and knowledge graphs are often conflated in scholarly-infrastructure discussions but make incompatible structural commitments. Citation graphs are paper-to-paper, untyped, and append-only; knowledge graphs are entity-to-entity, typed, and revisable. We catalogue six structural differences with concrete failure modes when one is used as the other, then propose a typed-edge extension to the citation graph that recovers the most useful KG affordances without breaking BibTeX. The proposal is implemented in the rrxiv reference server.

1 Introduction

Citation graphs and knowledge graphs are often conflated in scholarly-infrastructure discussions but make incompatible structural commitments. Citation graphs are paper-to-paper, untyped, and append-only; knowledge graphs are entity-to-entity, typed, and revisable. We catalogue six structural differences with concrete failure modes when one is used as the other, then propose a typed-edge extension to the citation graph that recovers the most useful KG affordances without breaking BibTeX. The proposal is implemented in the rrxiv reference server.

This document is a structured encoding of the paper in the `rrxiv` protocol's Canonical Intermediate Representation (CIR). It engages with the topic `cs.DL`. The encoding registers 5 formal claims (1 replicated, 4 untested). Each claim is annotated with its claim type, evidence type, and current replication status; dependency edges between claims, when present, form a machine-readable proof DAG.

2 Methodology

We follow the `rrxiv` convention of separating *claims* (the proposition under consideration) from *evidence* (the argument or data supporting it). Each claim in the results section below is presented with its statement, the type of evidence appealed to, and a brief discussion of replication status. Where claims depend on prior results — internal or external — the dependency is recorded in the CIR as a `\dependson` edge, so the full inferential structure is machine-traversable. Citations of external work appear in the References section at the end of this document.

3 Results: registered claims

Claim 1

Claim 1 (Claim 1)**.** Treating citation edges as semantic relationships causes systematic over-attribution: 34% of citations in our sample do not express dependency in the structural sense.

Replication status: untested.

This claim is an empirical observation supported by data. As of the encoding date, it has not yet been independently tested.

Claim 2

Claim 2 (Claim 2)**.** A typed-edge extension (depends_on / supports / contradicts / extends) recovers 89% of the queries our knowledge-graph baseline could answer, while staying compatible with existing BibTeX tooling.

Replication status: replicated.

This claim is a methodological proposal, supported by a deductive argument from prior results. As of the encoding date, it has been independently replicated. It depends on 1 prior claim in the same paper.

Claim 3

Claim 3 (Claim 3)**.** Knowledge-graph node identity is unstable across schema versions in a way that citation-graph identity is not; downstream consumers must either pin to a snapshot or handle entity merges.

Replication status: untested.

This claim is a theoretical claim derived from formal reasoning, supported by a deductive argument from prior results. As of the encoding date, it has not yet been independently tested. It depends on 1 prior claim in the same paper.

Claim 4

Claim 4 (Claim 4)**.** Citation networks are append-only in practice (retractions excepted); knowledge graphs revise nodes and edges continuously. Conflating the two breaks reproducibility.

Replication status: untested.

This claim is a theoretical claim derived from formal reasoning, supported by a deductive argument from prior results. As of the encoding date, it has not yet been independently tested.

Claim 5

Claim 5 (Claim 5)**.** The proposed typed-edge extension is implemented in the rrxiv reference server and round-trips through ‘cir.schema.json‘ without information loss.

Replication status: untested.

This claim is a methodological proposal, supported by computational evidence from simulation or numerical experiment. As of the encoding date, it has not yet been independently tested. It depends on 1 prior claim in the same paper.

4 Discussion

The claim graph above is the primary product of this paper. By making every claim independently citable — and by recording its dependencies, evidence type, and current replication status as structured fields — the paper participates in the rrxiv reproducibility-first corpus. Subsequent papers in this instance may extend, contradict, or replicate individual claims here without forcing a rewrite of the entire document. See the canonical version online for the live discourse layer.

5 References

- Knowledge graphs from scientific abstracts
- Citation networks vs knowledge graphs