Crystal: Characterizing Relative Impact of Scholarly Publications

Hannah Collison, Benjamin Van Durme, Daniel Khashabi

hgonza13, vandurme, danielk @ jhu.edu

Paper

Introduction

- Funding agencies, hiring committees, and researchers routinely rely on **measures of scientific impact** to make consequential decisions.
- **Citation count** alone provides a **poor measure of impact** because not all citations are equally important.
- Citation content analysis captures **how** and **why** a paper is cited.
- Focusing on each citation context in **isolation** is a natural choice: it concentrates on the most relevant text.
- However, the **relative importance** of a reference compared to the other works cited alongside it is disregarded.
- We propose CRISP, a method that **jointly ranks** all cited papers within a citing paper using LLMs to characterize their relative impact.

Results

- Our method outperforms prior state-of-the-art impact classifier by **+9.5%** in accuracy and **+8.3%** in F1 score on average across all three models.
  - **Qwen3-30B** is a strong alternative to GPT-5.1, offering competitive performance at **lower cost**.

Model	Method	Acc.	P	R	F1
Baseline	Random	49.9±1.2	33.1	50.9	40.1
GPT-5.1	UKP	66.7±1.1	49.6	63.5	55.7
	CRISP	78.6±1	72.2	63.7	**67.7**
	Δ	+11.9	+22.6	+0.2	+12.0
o4-mini	UKP	63.0±1.1	46.1	73.6	56.7
	CRISP	65.4±1.1	76.3	57.1	65.3
	Δ	+2.4	+30.2	−16.5	+8.6
Qwen3	UKP	60.8±1.2	44.5	76.7	56.3
	CRISP	75.1±1	70.0	53.3	60.5
	Δ	+14.3	+25.5	−23.4	+4.2

Table 1: Comparison of model performance between CRISP and prior state-of-the-art method (Arnaout et al., 2025) for identifying impactful citations.

Methods

- **Data:** 442 citing papers and 1,338 cited papers. Human-labeled dataset of citation contexts, in which each instance is annotated if the citation is impact-revealing or other.

CRISP:

- **(1) Downstream corpus retrieval:** Given a target paper p*, we use the Semantic Scholar API to retrieve all papers that cite it.
- **(2) Citation context extraction:** For each downstream paper q, we extract its full reference list and its associated citation contexts via the same API.

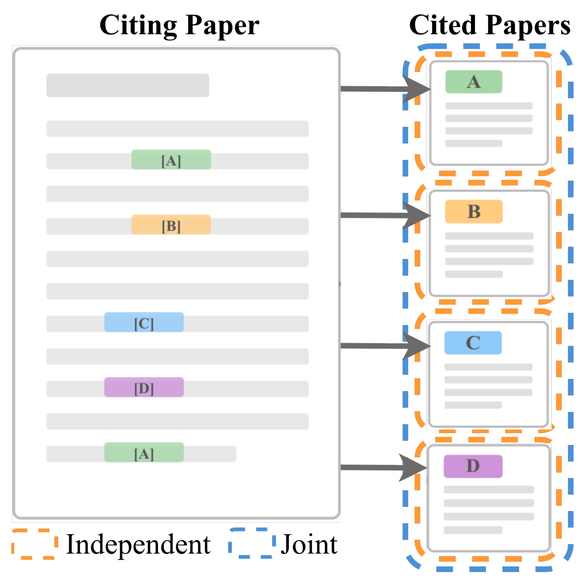

Figure 1: CRISP considers all cited papers *jointly* to assess the relative impact that some paper q had on p, unlike prior work that considers only p and q in isolation, evaluating impact independently.

- **(3) Citation-calibrated impact labeling:** We assign an impact label to each citation edge. The key modeling choice in CRISP is that the impact label for a citation (q → p) is predicted relative to the **full citation environment** of the citing paper q.

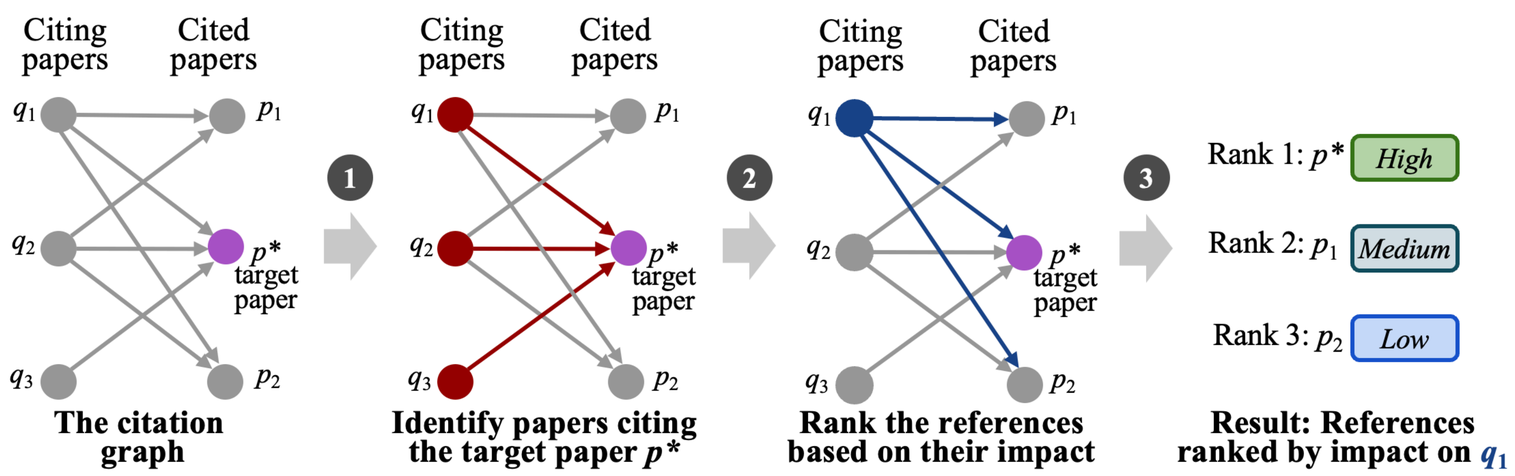

Figure 2: Overview of CRISP, which characterizes the relative impact of a target paper p* on each of its citing papers, e.g. q1. **1.** Identify all citing papers of p* (q1, q2, q3). **2.** For each citing paper, rank its references by impact (e.g., p1, p2, p* for q1). **3.** The result is an impact-based ranking of q1's references.

Future Work

- Build a **weighted citation graph** capturing how authors and papers are connected (ACL publications).
- Create more nuanced metrics.
- **Generate impact reports** for grant evaluation or author's research statements.

Acknowledgements

This research is based upon work supported in part by Office of Naval Research (N00014-24-1-2089). Moreover, we thank the Johns Hopkins University community for their helpful feedback.