

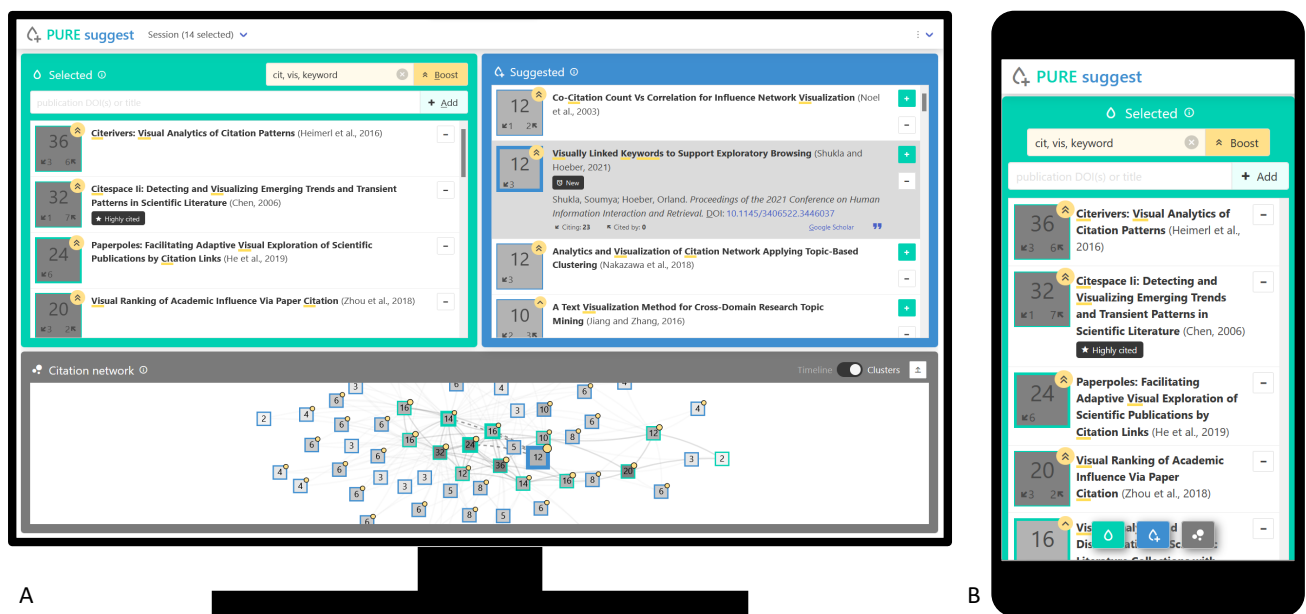


# Visually Explaining Publication Ranks in Citation-based Literature Search with PURE suggest

Fabian Beck  and Cedric Krause 

University of Bamberg, Germany



**Figure 1:** PURE suggest allows selecting publications (A, left) and then suggests related publications (A, right), while visualizing the citation network (A, bottom). A full-feature interface for small screens is also available (B).

## Abstract

Tracing citation links helps retrieve related publications. While most tools only allow the user to follow the citations of a single publication, some approaches support jointly analyzing the citations of a set of publications. Along similar lines, PURE suggest provides a detailed visual explanation of the ranking of suggested publications. The ranking is based on a score that combines citation numbers with keyword matching and is shown as a glyph for each publication. A citation network component references this glyph and visually embeds it into a timeline and cluster visualization.

## 1. Introduction

A citation represents a semantic link between two publications. When a set of publications already defines a research focus, and multiple citation links to or from these seed publications reference the same other paper, then, this work is likely also of high relevance for the given focus. Iteratively applying this idea now allows inter-actively building a coherent body of references through a small set

of seed publications, for instance, to assemble a section of related work for a publication. We refer to this concept as *citation-based literature search*.

Traditional search interfaces for academic literature, like *Google Scholar*, only partly support this approach—they only allow investigating the citations from and to a single selected paper. However, there exist specialized citation-based search tools, like *Cita-*

tion *Gecko* and *Research Rabbit*, that work with sets of selected seed papers to inform their suggestions as discussed. They also visualize the citations in the network and prioritize suggestions, but do not explain their prioritization or allow the user to steer it. Related techniques described in research literature may also visualize citation links, but have a different focus (cf. Section 3).

For efficiently deciding whether a publication is relevant, it is key to understand why it is suggested, specifically:

- Q1:** *Is the publication linked by incoming or outgoing citations?*
- Q2:** *Is the publication covering a certain topic of interest?*
- Q3:** *Does the publication have special (citation) characteristics?*
- Q4:** *How does the publication relate to the selected publications?*

Focusing on these questions, we have developed *PURE suggest* as shown in Figure 1. It computes and visually explains a score to rank the suggested publications based on incoming and outgoing citations links (Q1), as well as on a boost degree to consider the topics of interest (Q2). Using citation statistics and other indicators, certain characteristics (*highly cited*, *literature survey*, and *new*) of a paper can be inferred, which are then used as tags for the respective publications (Q3). A citation network visualization and interactive highlighting support exploring the relations between suggested and selected seed publications (Q4).

*PURE suggest* is implemented as a web application using *Vue.js* and *D3.js*. Citation and publication meta-data is retrieved through public APIs of *OpenCitations* and *Crossref*. The interface is split into three sections: the selected seed publications on the left, the suggested publications on the right, and the citation network at the bottom (cf. Figure 1A). On mobile screens or when resizing the window to smaller width, the layout switches to a vertical scrolling layout (cf. Figure 1B), with quick access buttons at the bottom to jump between the three sections; this supports using the tool also without window switching aside an enlarged document. The implemented tool is publicly available at <https://fabian-beck.github.io/pure-suggest/> and published as open source on GitHub; this paper describes the tool in version v0.2.1.<sup>†</sup>

## 2. Visually Explaining Publication Ranks

The publications are ranked by a score  $s$ , both to order the selected publications and to prioritize suggestions. The score is based on the number of outgoing (*citing*,  $o$ ) and incoming (*cited by*,  $i$ ) citation links referencing the set of currently selected publications. While the sum  $(o + i)$  is a good basis already, a search can even better be steered when having the option to boost certain keywords. Given a user-specified set of keywords, we compute boost degree  $b$  for every publication as the number of distinct matched keywords in the title. To make a significant difference in the score, we raise 2 to the power of  $b$  as a factor for the final score  $s = (o + i) \cdot 2^b$ .

Throughout the interface, we consistently use cyan-green ■ for discerning selected publications from suggested ones colored in blue ■, while information related to the boost keywords is shown in yellow ■. In both the selected and suggested publications, each publication carries a rectangular glyph providing the ranking score

as a number (large number at the top), encoded also as background color from white to gray. The glyph further visually explains the score, showing the number of outgoing citation links  $o$  on the lower left (e.g., 1) and the number of incoming ones  $i$  on the lower right (e.g., 2); the boost degree is visually indicated by a chevron icon and a yellow circle, growing from one chevron and a small circle for  $b = 1$  to three chevrons and a larger circle for  $b \geq 3$  (e.g., 3 for  $b = 2$ ). Hovering (desktop version) or clicking (mobile version) the glyph also displays a tooltip with a textual explanation. On click, a publication gets activated, and further details are shown, such as the full reference with general citation counts and external links. Boost keywords are marked with a thick yellow underline if matched in the title. Tags (e.g., ★ Highly cited) appear if the publication matches the respective heuristic; they are explained in tooltips on demand.

The citation network visualization at the bottom visually references the publication glyphs in a simplified version, using the same colors and shapes, as well as displaying the overall score as a label. The visualization can be switched between a *timeline* and *cluster* mode, both implemented based on the force-directed layout of *D3.js* with adapted forces (e.g., to place publications at the correct x-position on the timeline). Publications can also be activated via clicking in the visualization and are then highlighted in the respective list of publications as well. Citation links of the currently active publication are shown as dashed lines in the visualization, and linked publications are drawn with a thicker border in all views.

## 3. Application Example

As a demonstration, we used *PURE suggest* to collect related approaches for this work. We started our literature search from three relevant seed papers known to the authors [BKW16, HHKE16, HPLC19]. We set the keywords *cit* and *vis* to focus the search on visualizations of citation data. We summarize the application example and detected relevant publications in the following; the complete process is described in our supplemental material.

We employed different strategies. Oftentimes, we considered the suggested publications with the highest score. They mostly concerned the long-term evolution of research areas [Che06, Che99, ET07] and approaches that visualize author and citation networks for a single publication [GBP21, DSG\*12]. Finding publications with balanced high values for incoming and outgoing citations in the timeline view (Q1) yielded a paper that addresses the previously missing aspect of citation context [BMS17]. By inspecting further titles of high-ranked publications, we identified the field of publication ranking [ZSHL18]. This topic is relevant, but we did not identify other papers in that direction when using the keywords *scor* (stem of *score/scoring*) and *rank* (Q2). Analyzing connections of single publications in the cluster view (Q4) led to the relevant integration of keyword-based literature search, which was extended through the usage of the keyword *keyword* [IIS\*17, AYA22] (Q2). Lastly, adding related publications tagged as literature surveys [LTW\*18, FHKM17] (Q3) resulted in many new suggestions. This led to the addition of the only other publication we found that concerned the iterative buildup of a literature selection, but focuses on augmenting search engine results [SH21] (highlighted in Figure 1). We conclude that our combination of the above aspects addresses a gap in research literature.

<sup>†</sup> <https://github.com/fabian-beck/pure-suggest/releases/tag/v0.2.1>

## References

- [AYA22] ALMUHANNA A. A., YAFOOZ W. M. S., ALSAEEDI A.: An interactive scholarly collaborative network based on academic relationships and research collaborations. *Applied Sciences* 12, 2 (2022), 915. doi:10.3390/app12020915. 2
- [BKW16] BECK F., KOCH S., WEISKOPF D.: Visual analysis and dissemination of scientific literature collections with SurVis. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 180–189. doi:10.1109/tvcg.2015.2467757. 2
- [BMS17] BERGER M., MCDONOUGH K., SEVERSKY L. M.: cite2vec: Citation-driven document exploration via word embeddings. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 691–700. doi:10.1109/tvcg.2016.2598667. 2
- [Che99] CHEN C.: Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing & Management* 35, 3 (1999), 401–420. doi:10.1016/s0306-4573(98)00068-5. 2
- [Che06] CHEN C.: CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology* 57, 3 (2006), 359–377. doi:10.1002/asi.20317. 2
- [DSG\*12] DUNNE C., SHNEIDERMAN B., GOVE R., KLAVANS J., DORR B.: Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization. *Journal of the American Society for Information Science and Technology* 63, 12 (2012), 2351–2369. doi:10.1002/asi.22652. 2
- [ET07] ELMQVIST N., TSIGAS P.: CiteWiz: A tool for the visualization of scientific citation networks. *Information Visualization* 6, 3 (2007), 215–232. doi:10.1057/palgrave.ivs.9500156. 2
- [FHKM17] FEDERICO P., HEIMERL F., KOCH S., MIKSCH S.: A survey on visual approaches for analyzing scientific literature and patents. *IEEE Transactions on Visualization and Computer Graphics* 23, 9 (2017), 2179–2198. doi:10.1109/tvcg.2016.2610422. 2
- [GBP21] GUILARTE O. F., BARBOSA S. D. J., PESCO S.: RelPath: An interactive tool to visualize branches of studies and quantify the expertise of authors by citation paths. *Scientometrics* 126, 6 (2021), 4871–4897. doi:10.1007/s11192-021-03959-2. 2
- [HHKE16] HEIMERL F., HAN Q., KOCH S., ERTL T.: CiteRivers: Visual analytics of citation patterns. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 190–199. doi:10.1109/tvcg.2015.2467621. 2
- [HPLC19] HE J., PING Q., LOU W., CHEN C.: PaperPoles: Facilitating adaptive visual exploration of scientific publications by citation links. *Journal of the Association for Information Science and Technology* 70, 8 (2019), 843–857. doi:10.1002/asi.24171. 2
- [IIS\*17] ISENBERG P., ISENBERG T., SEDLMIR M., CHEN J., MÖLLER T.: Visualization as seen through its research paper keywords. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 771–780. doi:10.1109/tvcg.2016.2598827. 2
- [LTW\*18] LIU J., TANG T., WANG W., XU B., KONG X., XIA F.: A survey of scholarly data visualization. *IEEE Access* 6 (2018), 19205–19221. doi:10.1109/access.2018.2815030. 2
- [SH21] SHUKLA S., HOEBER O.: Visually linked keywords to support exploratory browsing. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (2021), ACM, p. 273–277. doi:10.1145/3406522.3446037. 2
- [ZSHL18] ZHOU Z., SHI C., HU M., LIU Y.: Visual ranking of academic influence via paper citation. *Journal of Visual Languages & Computing* 48 (2018), 134–143. doi:10.1016/j.jvlc.2018.08.007. 2