



Universiteit
Leiden
The Netherlands

Science maps for information retrieval

Bascur Cifuentes, J.P.

Citation

Bascur Cifuentes, J. P. (2026, January 21). *Science maps for information retrieval*. Retrieved from <https://hdl.handle.net/1887/4287774>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4287774>

Note: To cite this publication please use the final published version (if applicable).

Bibliography

- [1] Muhammad Kamran Abbasi and Ingo Frommholz. Cluster-based polyrepresentation as science modelling approach for information retrieval. *Scientometrics*, 102(3):2301–2322, 2015. doi: 10.1007/s11192-014-1478-1.
- [2] Jimi Adams and Kate Vinita Fitch. Whose social capital?: Citation and co-citation patterns of a fragmented concept. *Socius*, 9:23780231231184766, 2023. doi: 10.1177/23780231231184766.
- [3] Tanay Aggarwal, Angelo Salatino, Francesco Osborne, and Enrico Motta. Large language models for scholarly ontology generation: An extensive analysis in the engineering field. *arXiv*, 2024. doi: 10.48550/arXiv.2412.08258.
- [4] Per Ahlgren, Yunwei Chen, Cristian Colliander, and Nees Jan van Eck. Enhancing direct citations: A comparison of relatedness measures for community detection in a large set of PubMed publications. *Quantitative Science Studies*, 1(2):714–729, 2020. doi: 10.1162/qss_a-00027.
- [5] Altmetric.com. About us. <https://www.altmetric.com/about-us/>, 2024. Accessed: 2024-11-01.
- [6] Miriam Arnold, Mascha Goldschmitt, and Thomas Rigotti. Dealing with information overload: a comprehensive review. *Frontiers in Psychology*, 14:1122200, 2023. doi: 10.3389/fpsyg.2023.1122200.
- [7] Jesús Arroyo, Carey E. Priebe, and Vince Lyzinski. Graph matching between bipartite and unipartite networks: To collapse, or not to collapse, that is the question. *IEEE Transactions on Network Science and Engineering*, 8(4):3019–3033, 2021. doi: 10.1109/tNSE.2021.3086508.
- [8] Akari Asai, Jacqueline He, Rulin Shao, Weijia Shi, Amanpreet Singh, Joseph Chee Chang, Kyle Lo, Luca Soldaini, Sergey Feldman, Mike D’arcy, David Wadden, Matt Latzke, Minyang Tian, Pan Ji, Shengyan Liu, Hao Tong, Bohao Wu, Yanyu Xiong, Luke Zettlemoyer, Graham Neubig, Dan Weld, Doug Downey, Wen tau Yih, Pang Wei Koh, and Hannaneh Hajishirzi. Openscholar: Synthesizing scientific literature with retrieval-augmented lms, 2024. URL <https://arxiv.org/abs/2411.14199>.
- [9] Ahmet Yasin Aytar, Kemal Kilic, and Kamer Kaya. A retrieval-augmented generation framework for academic literature navigation in data science, 2024. URL <https://arxiv.org/abs/2412.15404>.
- [10] Leif Azzopardi, Kalervo Järvelin, Jaap Kamps, and Mark D. Smucker. Report on the SIGIR 2010 workshop on the simulation of interaction. *ACM SIGIR Forum*, 44(2):35–47, 2011. doi: 10.1145/1924475.1924484.
- [11] Zhichao Ba and Zhentao Liang. A novel approach to measuring science-technology linkage: From the perspective of knowledge network coupling. *Journal of Informetrics*, 15(3):101167, 2021. doi: 10.1016/j.joi.2021.101167.

- [12] Juan Pablo Bascur. Which topics are best represented by science maps? An analysis of clustering effectiveness for citation and text similarity networks (data). *Zenodo*, 2024. doi: 10.5281/zenodo.11181030.
- [13] Juan Pablo Bascur. Use of diverse data sources to control which topics emerge in a science map. Supplementary material. *Zenodo*, 2024. doi: 10.5281/zenodo.14170721.
- [14] Juan Pablo Bascur. Academic information retrieval using citation clusters: In-depth evaluation based on systematic reviews (data). *Zenodo*, 2025. doi: 10.5281/zenodo.6702251.
- [15] Juan Pablo Bascur. Prototype of Scimacro. <https://github.com/jpbascur/Scimacro>, 2025. Accessed: 2025-04-11.
- [16] Juan Pablo Bascur, Nees Jan van Eck, and Ludo Waltman. An interactive visual tool for scientific literature search: Proposal and algorithmic specification. In *BIR@ ECIR*, pages 76–87, 2019. URL <https://ceur-ws.org/Vol-2345/paper7.pdf>.
- [17] Juan Pablo Bascur, Suzan Verberne, Nees Jan van Eck, and Ludo Waltman. Academic information retrieval using citation clusters: In-depth evaluation based on systematic reviews. *Scientometrics*, 128(5):2895–2921, 2023. doi: 10.1007/s11192-023-04681-x.
- [18] Juan Pablo Bascur, Rodrigo Costas, and Suzan Verberne. Use of diverse data sources to control which topics emerge in a science map. *arXiv*, 2024. doi: 10.48550/arXiv.2412.07550.
- [19] Juan Pablo Bascur, Suzan Verberne, Nees Jan van Eck, and Ludo Waltman. Which topics are best represented by science maps? An analysis of clustering effectiveness for citation and text similarity networks. *Scientometrics*, 130:1181–1199, 2025. doi: 10.1007/s11192-024-05218-6.
- [20] Christopher W. Belter. Citation analysis as a literature search method for systematic reviews. *Journal of the Association for Information Science and Technology*, 67(11):2766–2777, 2016. doi: 10.1002/asi.23605.
- [21] Christopher W. Belter. A relevance ranking method for citation-based search results. *Scientometrics*, 112(2):731–746, 2017. doi: 10.1007/s11192-017-2406-y.
- [22] Ravi Varma Kumar Bevara, Brady D Lund, Nishith Reddy Mannuru, Sai Pranathi Karedla, Yara Mohammed, Sai Tulasi Kolapudi, and Aashrith Mannuru. Prospects of retrieval augmented generation (rag) for academic library search and retrieval. *Information Technology and Libraries*, 44(2), 2025.
- [23] Nauman bin Ali and Binish Tanveer. A comparison of citation sources for reference and citation-based search in systematic literature reviews. *e-Informatica Software Engineering Journal*, 16(1):220106, 2022. doi: 10.37190/e-Inf220106.
- [24] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3:993–1022, 2003.
- [25] Katy Börner. Atlas of science: Visualizing what we know. *M Press*, 2010. doi: 10.5555/1995300.
- [26] Lutz Bornmann and Rüdiger Mutz. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11):2215–2222, 2015. doi: 10.1002/asi.23329.
- [27] Kevin W. Boyack and Richard Klavans. A comparison of large-scale science models based on textual, direct citation and hybrid relatedness. *Quantitative Science Studies*, 1(4):1570–1585, 2020. doi: 10.1162/qss.a.00085.

- [28] Kevin W. Boyack, Catherine Smith, and Richard Klavans. A detailed open access model of the PubMed literature. *Scientific Data*, 7(1):408, 2020. doi: 10.1038/s41597-020-00749-y.
- [29] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1):107–117, 1998. doi: 10.1016/S0169-7552(98)00110-X.
- [30] Guillaume Cabanac, Muthu Kumar Chandrasekaran, Ingo Frommholz, Kokil Jaidka, Min-Yen Kan, Philipp Mayr, and Dietmar Wolfram. Report on the joint workshop on bibliometric-enhanced information retrieval and natural language processing for digital libraries (BIRNDL 2016). *SIGIR Forum*, 50(2):36–43, 2017. doi: 10.1145/3053408.3053417.
- [31] Michel Callon, Jean-Pierre Courtial, William A. Turner, and Serge Bauin. From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22(2):191–235, 1983. doi: 10.1177/053901883022002003.
- [32] Zeljko Carevic, Maria Lusky, Wilko van Hoek, and Philipp Mayr. Investigating exploratory search activities based on the stratagem level in digital libraries. *International Journal on Digital Libraries*, 19:231–251, 2018. doi: 10.1007/s00799-017-0226-6.
- [33] David Carmel, Elad Yom-Tov, Adam Darlow, and Dan Pelleg. What makes a query difficult? In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06)*, page 390. ACM, 2006. doi: 10.1145/1148170.1148238.
- [34] Chiara Carusi and Giuseppe Bianchi. Scientific community detection via bipartite scholar/journal graph co-clustering. *Journal of Informetrics*, 13(1):354–386, 2019. doi: 10.1016/j.joi.2019.01.004.
- [35] Centre for Science and Technology Studies. Leiden ranking fields, 2023. URL <https://www.leidenranking.com/information/fields>. Accessed: 03-04-2023.
- [36] Chang Chang and Chao Tang. Community detection for networks with unipartite and bipartite structure. *New Journal of Physics*, 16(9):093001, 2014. doi: 10.1088/1367-2630/16/9/093001.
- [37] Chaomei Chen. 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):359–377, 2006. doi: 10.1002/asi.20317.
- [38] Chaomei Chen. Science mapping: A systematic review of the literature. *Journal of Data and Information Science*, 2(2):1–40, 2017. doi: 10.1515/jdis-2017-0006.
- [39] Chaomei Chen, Fidelia Ibekwe-SanJuan, and Jianhua Hou. The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for Information Science and Technology*, 61(7):1386–1409, 2010. doi: 10.1002/asi.21309.
- [40] Giulio Cimini, Alessandro Carra, Luca Didomenicantonio, and Andrea Zaccaria. Meta-validation of bipartite network projections. *Communications Physics*, 5(1):76, 2022. doi: 10.1038/s42005-022-00856-9.
- [41] Clarivate. Web of Science. <https://clarivate.com/products/web-of-science/>, 2018. Accessed: 2018-01-27.
- [42] Manuel J. Cobo, Antonio G. López-Herrera, Enrique Herrera-Viedma, and Francisco Herrera. Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7):1382–1402, 2011. doi: 10.1002/asi.21525.

- [43] Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel S. Weld. SPECTER: Document-level representation learning using citation-informed transformers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2270–2282. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.acl-main.207.
- [44] Connected Papers. Connected papers: Visual tool for literature discovery. <https://www.connectedpapers.com/>, 2025. Accessed: 2025-04-11.
- [45] Rodrigo Costas, Sarah de Rijcke, and Noortje Marres. “Heterogeneous couplings”: Operationalizing network perspectives to study science-society interactions through social media metrics. *Journal of the Association for Information Science and Technology*, 72(5):595–610, 2021. doi: 10.1002/asi.24427.
- [46] Sarah E. Cousins, Elizabeth Tempest, and David J. Feuer. Surgery for the resolution of symptoms in malignant bowel obstruction in advanced gynaecological and gastrointestinal cancer. *Cochrane Database of Systematic Reviews*, 2016. doi: 10.1002/14651858.CD002764.pub2.
- [47] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal, Complex Systems*, 2006. URL <https://igraph.org>.
- [48] Douglass R. Cutting, David R. Karger, Jan O. Pedersen, and John W. Tukey. Scatter/Gather: A cluster-based approach to browsing large document collections. *SIGIR '92: Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 318 – 329, 1992. doi: 10.1145/133160.133214.
- [49] Christiaan M. De Vries, Shlomo Geva, and Andrew Trotman. Document clustering evaluation: Divergence from a random baseline. *arXiv*, 2012. doi: 10.48550/arXiv.1208.5654.
- [50] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics, 2019. doi: 10.18653/v1/N19-1423.
- [51] Digital Science. Dimensions. <https://www.dimensions.ai/>, 2018. Accessed: 2018-01-27.
- [52] Ying Ding. Community detection: Topological vs. topical. *Journal of Informetrics*, 5(4): 498–514, 2011. doi: 10.1016/j.joi.2011.02.006.
- [53] Pablo Dorta-González, Alejandro Rodríguez-Caro, and María Isabel Dorta-González. Societal and scientific impact of policy research: A large-scale empirical study of some explanatory factors using Altmetric and Overton. *Journal of Informetrics*, 18(3):101530, 2024. doi: 10.1016/j.joi.2024.101530.
- [54] Ciriaco Andrea D’Angelo and Nees Jan Van Eck. Collecting large-scale publication data at the level of individual researchers: A practical proposal for author name disambiguation. *Scientometrics*, 123:883–907, 2020. doi: 10.1007/s11192-020-03410-y.
- [55] David Ellis. Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *The Library Quarterly*, 63(4):469–486, 1993. doi: 10.1086/602622.
- [56] Elsevier. Scopus. <https://www.scopus.com/>, 2018. Accessed: 2018-01-27.
- [57] Elsevier. Topic prominence in science—Scival. <https://www.elsevier.com/solutions/scival/features/topic-prominence-in-science>, 2023. Accessed: 2023-01-25.

- [58] Martin J. Eppler. A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Information Visualization*, 5(3):202–210, 2006. doi: 10.1057/palgrave.ivs.9500131.
- [59] Zhichao Fang, Rodrigo Costas, Wencan Tian, Xianwen Wang, and Paul Wouters. An extensive analysis of the presence of altmetric data for Web of Science publications across subject fields and research topics. *Scientometrics*, 124(3):2519–2549, 2020. doi: 10.1007/s11192-020-03564-9.
- [60] Zhichao Fang, Jonathan Dudek, Ed Noyons, and Rodrigo Costas. Science cited in policy documents: Evidence from the Overton database. *arXiv*, 2024. doi: 10.48550/arXiv.2407.09854.
- [61] Matilde Ficozzi, Mathieu Jacomy, Dario Rodighiero, Anne Beaulieu, and Anders Kristian Munk. Grounding ai map: The consequences of living with the trouble of an irreductionist map. *Design et abstractions, Revue Design Arts Medias*, 2025(6), 2025.
- [62] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010. doi: 10.1016/j.physrep.2009.11.002.
- [63] Ingo Frommholz, Philipp Mayr, Guillaume Cabanac, and Suzan Verberne. Bibliometric-enhanced information retrieval: 11th international BIR workshop. In *Advances in Information Retrieval*, pages 705–709. Springer International Publishing, 2021. doi: 10.1007/978-3-030-72240-1_85.
- [64] Jochen Gläser. Opening the black box of expert validation of bibliometric maps. In *Lockdown Bibliometrics: Papers not submitted to the STI Conference 2020 in Aarhus*, pages 27–36, 2020.
- [65] Audilio Gonzales-Aguilar and María Ramírez-Posada. Carot2: Búsqueda y visualización de la información. *Profesional de la información*, 21(1):105–112, 2012. doi: 10.3145/epi.2012.ene.14.
- [66] Google. Google scholar. <https://scholar.google.com/>, 2018. Accessed: 2018-01-27.
- [67] Christian Sengstock Hamed Abdelhaq and Michael Gertz. EvenTweet: Online localized event detection from Twitter. *Proceedings of the VLDB Endowment*, 6(12):1326–1329, 2013. doi: 10.14778/2536274.2536307.
- [68] Robin Haunschild and Werner Marx. Discovering seminal works with marker papers. *Scientometrics*, 125(3):2955–2969, 2020. doi: 10.1007/s11192-020-03358-z.
- [69] Robin Haunschild, Hermann Schier, Werner Marx, and Lutz Bornmann. Algorithmically generated subject categories based on citation relations: An empirical micro study using papers on overall water splitting. *Journal of Informetrics*, 12(2):436–447, 2018. doi: 10.1016/j.joi.2018.03.004.
- [70] Frank Havemann, Jochen Gläser, and Michael Heinz. Memetic search for overlapping topics based on a local evaluation of link communities. *Scientometrics*, 111:1089–1118, 2017. doi: 10.1007/s11192-017-2302-5.
- [71] Jiangen He, Qing Ping, Wen Lou, and Chaomei Chen. PaperPoles: Facilitating adaptive visual exploration of scientific publications by citation links. *Journal of the Association for Information Science and Technology*, 70(8):843–857, 2019. doi: 10.1002/asi.24171.
- [72] Jiyin He, Edgar Meij, and Maarten de Rijke. Result diversification based on query-specific cluster ranking. *Journal of the American Society for Information Science and Technology*, 62(3):550–571, 2011. doi: 10.1002/asi.21468.

- [73] Marti A. Hearst and Jan O. Pedersen. Reexamining the cluster hypothesis: Scatter/Gather on retrieval results. In *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 1996. doi: 10.1145/243199.243216.
- [74] Matthias Held. Know thy tools! Limits of popular algorithms used for topic reconstruction. *Quantitative Science Studies*, 3(4):1054–1078, 2022. doi: 10.1162/qss.a_00217.
- [75] Matthias Held and Jochen Gläser. Exploring publication networks with a local cohesion-maximizing algorithm. *Quantitative Science Studies*, 5(3):681–703, 2024. doi: 10.1162/qss.a_00314.
- [76] Matthias Held and Theresa Velden. How to interpret algorithmically constructed topical structures of scientific fields? A case study of citation-based mappings of the research specialty of invasion biology. *Quantitative Science Studies*, 3(3):651–671, 2022. doi: 10.1162/qss.a_00194.
- [77] Matthias Held, Grit Laudel, and Jochen Gläser. Topic reconstruction from networks of papers may not be possible if only one algorithm is applied to only one data model. In *Lockdown Bibliometrics: Papers not submitted to the STI Conference 2020 in Aarhus*, pages 18–26, 2020.
- [78] Matthias Held, Grit Laudel, and Jochen Gläser. Challenges to the validity of topic reconstruction. *Scientometrics*, 126:4511–4536, 2021. doi: 10.1007/s11192-021-03920-3.
- [79] Bradley M. Hemminger, Dongrong Lu, K. T. L. Vaughan, and Susan J. Adams. Information seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology*, 58(14):2205–2225, 2007. doi: 10.1002/asi.20686.
- [80] César A Hidalgo. Disconnected, fragmented, or united? A trans-disciplinary review of network science. *Applied Network Science*, 1:1–19, 2016. doi: 10.1007/s41109-016-0010-3.
- [81] Katja Hofmann, Shimon Whiteson, and Maarten de Rijke. Balancing exploration and exploitation in listwise and pairwise online learning to rank for information retrieval. *Information Retrieval*, 16(1):63–90, 2013. doi: 10.1007/s10791-012-9197-9.
- [82] Thomas E. Ferrin Holly J. Atkinson, John H. Morris and Patricia C. Babbitt. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS ONE*, 4(2):e4345, 2009. doi: 10.1371/journal.pone.0004345.
- [83] Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. spaCy: Industrial-strength natural language processing in Python. *Zenodo*, 2020. doi: 10.5281/zenodo.1212303.
- [84] Daniel W. Hook, Simon J. Porter, and Christian Herzog. Dimensions: Building context for search and evaluation. *Frontiers in Research Metrics and Analytics*, 3:23, 2018. doi: 10.3389/frma.2018.00023.
- [85] Tanya Horsley, Owen Dingwall, and Margaret Sampson. Checking reference lists to find additional studies for systematic reviews. *Cochrane Database of Systematic Reviews*, 2011. doi: 10.1002/14651858.MR000026.pub2.
- [86] Darko Hric, Richard K. Darst, and Santo Fortunato. Community detection in networks: Structural communities versus ground truth. *Physical Review E*, 90(6):062805, 2014. doi: 10.1103/PhysRevE.90.062805.
- [87] Inciteful. Inciteful: Interactive academic literature discovery. <https://inciteful.xyz/>, 2025. Accessed: 2025-04-11.

- [88] Iris AI AS. Iris.ai: Ai-powered research assistant. <https://iris.ai/>, 2025. Accessed: 2025-04-11.
- [89] A. Cecile J. W. Janssens and Marta Gwinn. Novel citation-based search method for scientific literature: Application to meta-analyses. *BMC medical research methodology*, 15:1–11, 2015. doi: 10.1186/s12874-015-0077-z.
- [90] A. Cecile J. W. Janssens, Marta Gwinn, John E. Brockman, Kenneth Powell, and Melody Goodman. Novel citation-based search method for scientific literature: A validation study. *BMC Medical Research Methodology*, 20(1):1–11, 2020. doi: 10.1186/s12874-020-0907-5.
- [91] Frizo Janssens, Wolfgang Glänzel, and Bart De Moor. A hybrid mapping of information science. *Scientometrics*, 75(3):607–631, 2008. doi: 10.1007/s11192-007-2002-7.
- [92] Nick Jardine and Cornelis Joost van Rijsbergen. The use of hierarchic clustering in information retrieval. *Information Storage and Retrieval*, 7(5):217–240, 1971. doi: 10.1016/0020-0271(71)90051-9.
- [93] Byeongwoo Kang and Gianluca Tarasconi. PATSTAT revisited: Suggestions for better usage. *World patent information*, 46:56–63, 2016. doi: 10.1016/j.wpi.2016.06.001.
- [94] Diane Kelly. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, 3(1–2):1–224, 2009. doi: 10.1561/15000000012.
- [95] Max Kemman, Martijn Kleppe, and Stef Scagliola. Just Google it. Digital research practices of humanities scholars. *Studies in the Digital Humanities*, 2014. URL <http://hdl.handle.net/1765/50779>.
- [96] Maxwell Mirton Kessler. Bibliographic coupling between scientific papers. *American Documentation*, 14(1):10–25, 1963. doi: 10.1002/asi.5090140103.
- [97] Richard Klavans and Kevin W. Boyack. Quantitative evaluation of large maps of science. *Scientometrics*, 68(3):475–499, 2006. doi: 10.1007/s11192-006-0125-x.
- [98] Richard Klavans and Kevin W. Boyack. Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology*, 68(4):984–998, 2017. doi: 10.1002/asi.23734.
- [99] Ben Slater Krishnan Chandra and Mike Ma. Research rabbit. <https://www.researchrabbit.ai/>, 2025. Accessed: 2025-04-11.
- [100] Carol C. Kuhlthau. Inside the search process: Information seeking from the user’s perspective. *Journal of the American Society for Information Science*, 42(5):361–371, 1991.
- [101] Sameer Kumar. Co-authorship networks: A review of the literature. *Aslib Journal of Information Management*, 67(1):55–73, 2015. doi: 10.1108/ajim-09-2014-0116.
- [102] Kuei-Kuei Lai and Shiao-Jun Wu. Using the patent co-citation approach to establish a new patent classification system. *Information processing & management*, 41(2):313–330, 2005. doi: 10.1016/j.ipm.2003.11.004.
- [103] Kiwon Lee and Suchul Lee. Knowledge structure of the application of high-performance computing: A co-word analysis. *Sustainability*, 13(20):11249, 2021. doi: 10.3390/su132011249.
- [104] Loet Leydesdorff. Theories of citation? *Scientometrics*, 43:5–25, 1998. doi: 10.1007/BF02458391.

- [105] Yujie Liang, Qian Li, and Tianrui Qian. Finding relevant papers based on citation relations. In *Web-Age Information Management*, volume 6897 of *Lecture Notes in Computer Science*, pages 403–414. Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-23535-1_35.
- [106] Suzanne K. Linder, Geetanjali R. Kamath, Gregory F. Pratt, Smita S. Saraykar, and Robert J. Volk. Citation searches are more sensitive than keyword searches to identify studies using specific measurement instruments. *Journal of clinical epidemiology*, 68(4):412–417, 2015. doi: 10.1016/j.jclinepi.2014.10.008.
- [107] Litmaps. Litmaps. <https://www.litmaps.com/>, 2025. Accessed: 2025-04-11.
- [108] Ziyang Liu, Sivaramakrishnan Natarajan, and Yi Chen. Query expansion based on clustered results. *Proceedings of the VLDB Endowment*, 4(6):350–361, 2011. doi: 10.14778/1978665.1978667.
- [109] Maria Angeles Lopez-Olivo, Matxalen Amezaga Urruela, Lynda McGahan, Eduardo N. Pol-lono, and Maria E. Suarez-Almazor. Rituximab for rheumatoid arthritis. *Cochrane Database of Systematic Reviews*, 2015. doi: 10.1002/14651858.CD007356.pub2.
- [110] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [111] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. Scoring, term weighting and the vector space model. In *Introduction to Information Retrieval*, chapter 6. Cambridge University Press, 2008. URL <https://nlp.stanford.edu/IR-book/html/htmledition/scoring-term-weighting-and-the-vector-space-model-1.html>.
- [112] Philipp Mayr and Andrea Scharnhorst. Scientometrics and information retrieval: Weak-links revitalized. *Scientometrics*, 102(3):2193–2199, 2015. doi: 10.1007/s11192-014-1484-3.
- [113] Martin Meyer. Does science push technology? Patents citing scientific literature. *Research policy*, 29(3):409–434, 2000. doi: 10.1016/S0048-7333(99)00040-2.
- [114] Jose A. Moral-Munoz, Antonio G. López-Herrera, Enrique Herrera-Viedma, and Manuel J. Cobo. Science mapping analysis software tools: A review. *Springer Handbook of Science and Technology Indicators*, pages 159–185, 2019. doi: 10.1007/978-3-030-02511-3_7.
- [115] Erwan Moreau. Literature-based discovery: Addressing the issue of the subpar evaluation methodology. *Bioinformatics*, 39(2):btad090, 2023. doi: 10.1093/bioinformatics/btad090.
- [116] Peter Mutschke and Philipp Mayr. Science models for search: A study on combining scholarly information retrieval and scientometrics. *Scientometrics*, 102:2323–2345, 2015. doi: 10.1007/s11192-014-1485-2.
- [117] National Institutes of Health. Medical subject headings. Available at <https://www.nlm.nih.gov/mesh/meshhome.html>, 2024. Accessed: 2024-11-01.
- [118] Zachary Neal. The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. *Social Networks*, 39:84–97, 2014. doi: 10.1016/j.socnet.2014.06.001.
- [119] Rommert Dekker Nees Jan van Eck, Ludo Waltman and Jan van den Berg. A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology*, 61(12):2405–2416, 2010. doi: 10.1002/asi.21421.
- [120] Mark E.J. Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the national academy of sciences*, 101(suppl.1):5200–5205, 2004. doi: 10.1073/pnas.0307545100.

- [121] Mark E.J. Newman and Aaron Clauset. Structure and inference in annotated networks. *Nature Communications*, 7(1):11863, 2016. doi: 10.1038/ncomms11863.
- [122] Open Knowledge Maps. Open knowledge maps: A visual interface to the world’s scientific knowledge. <https://openknowledgemaps.org>, 2019. Accessed: 2025-04-16.
- [123] OpenAlex. OpenAlex topic classification. <https://github.com/ourresearch/openalex-topic-classification>, 2024. Accessed: 20-04-2025.
- [124] Francisco M. Ortuño, Ignacio Rojas, Miguel A. Andrade-Navarro, and Jean-François Fontaine. Using cited references to improve the retrieval of related biomedical documents. *BMC Bioinformatics*, 14(1):113, 2013. doi: 10.1186/1471-2105-14-113.
- [125] Leto Peel, Daniel B Larremore, and Aaron Clauset. The ground truth about metadata and community detection in networks. *Science Advances*, 3(5):e1602548, 2017. doi: 10.1126/sciadv.1602548.
- [126] Frank Peinemann, Carmen Bartel, Ulrich Grouven, and Frank Berthold. Retinoic acid post consolidation therapy for high-risk neuroblastoma. *Cochrane Database of Systematic Reviews*, 2013. doi: 10.1002/14651858.CD010685.
- [127] Eugenio Petrovich. Science mapping and science maps. *Knowledge Organization*, 48(7-8): 535–562, 2022. doi: 10.5771/0943-7444-2021-7-8-535.
- [128] Henrique Pinheiro, Etienne Vignola-Gagné, and David Campbell. A large-scale validation of the relationship between cross-disciplinary research and its uptake in policy-related documents, using the novel Overton altmetrics database. *Quantitative Science Studies*, 2(2):616–642, 2021. doi: 10.1162/qss.a_00137.
- [129] Peter Pirolli, Patricia Schank, Marti Hearst, and Christopher Diehl. Scatter/Gather browsing communicates the topic structure of a very large text collection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '96): Common Ground*, pages 213–220. ACM, 1996. doi: 10.1145/238386.238489.
- [130] Ian Potter. Introducing citation topics in incites. <https://clarivate.com/blog/introducing-citation-topics/>, 2020. Accessed: 2025-04-16.
- [131] Ismael Rafols. Towards multiple ontologies in science mapping. A tribute to Loet Leydesdorff. *SocArXiv*, 2025. doi: 10.31235/osf.io/47umh_v2.
- [132] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992. Association for Computational Linguistics, 2019. doi: 10.18653/v1/D19-1410.
- [133] Stephen Robertson and Hugo Zaragoza. The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends in Information Retrieval*, 3(4):333–389, 2009. doi: 10.1561/15000000019.
- [134] Karen A. Robinson, Adam G. Dunn, Guy Tsafnat, and Paul Glasziou. Citation networks of related trials are often disconnected: Implications for bidirectional citation searches. *Journal of Clinical Epidemiology*, 67(7):793–799, 2014. doi: 10.1016/j.jclinepi.2013.11.015.
- [135] Giulio Rossetti, Luca Pappalardo, and Salvatore Rinzivillo. A novel approach to evaluate community detection algorithms on ground truth. In *Complex Networks VII: Proceedings of the 7th Workshop on Complex Networks CompleNet 2016*, pages 133–144. Springer, 2016. doi: 10.1007/978-3-319-30569-1_10.

- [136] Tony Russell-Rose, Jon Chamberlain, and Leif Azzopardi. Information retrieval in the workplace: A comparison of professional search practices. *Information Processing & Management*, 54(6):1042–1057, 2018. doi: 10.1016/j.ipm.2018.07.003.
- [137] SAS Institute. Clustering methods. Available at https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_cluster_sect012.htm, 2009. Accessed: 2024-11-01.
- [138] Eric Sayers. E-utilities quick start. <https://www.ncbi.nlm.nih.gov/books/NBK25500/>, 2008. Accessed: 2025-04-16.
- [139] Harrisen Scells, Guido Zuccon, Bevan Koopman, Anthony Deacon, Leif Azzopardi, and Shlomo Geva. A test collection for evaluating retrieval of studies for inclusion in systematic reviews. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, page 1237–1240. Association for Computing Machinery, 2017. doi: 10.1145/3077136.3080707.
- [140] Harrisen Scells, Daniel Locke, and Guido Zuccon. An information retrieval experiment framework for domain specific applications. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1281–1284. ACM, 2018. doi: 10.1145/3209978.3210167.
- [141] Cara Seitz, Marion Schmidt, Nathalie Schwichtenberg, and Theresa Velden. A case study of the epistemic function of citations—implications for citation-based science mapping. In *Proceedings of the 18th International Conference of the International Society for Scientometrics and Informetrics (ISSI)*, pages 1027–1032, 2021.
- [142] Peter Sjögarde and Per Ahlgren. Granularity of algorithmically constructed publication-level classifications of research publications: Identification of topics. *Journal of Informetrics*, 12(1): 133–152, 2018. doi: 10.1016/j.joi.2017.12.006.
- [143] Peter Sjögarde and Per Ahlgren. Granularity of algorithmically constructed publication-level classifications of research publications: Identification of specialties. *Quantitative Science Studies*, 1(1):207–238, 2020. doi: 10.1162/qss_a.00004.
- [144] Peter Sjögarde, Per Ahlgren, and Ludo Waltman. Algorithmic labeling in hierarchical classifications of publications: Evaluation of bibliographic fields and term weighting approaches. *Journal of the Association for Information Science and Technology*, 72(7):853–869, 2021. doi: 10.1002/asi.24452.
- [145] Henry Small. Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the Association for Information Science and Technology*, 24(4): 265–269, 1973. doi: 10.1002/asi.4630240406.
- [146] Henry Small and Eugene Garfield. The geography of science: Disciplinary and national mappings. *Journal of Information Science*, 11(4):147–159, 1985. doi: 10.1177/016555158501100402.
- [147] Jerzy Stefanowski and Dawid Weiss. Carrot2 and language properties in web search results clustering. In *Advances in Web Intelligence*, volume 2663 of *Lecture Notes in Computer Science*, pages 240–249. Springer Berlin Heidelberg, 2003. doi: 10.1007/3-540-44831-4.25.
- [148] Lovro Šubelj, Nees Jan Van Eck, and Ludo Waltman. Clustering scientific publications based on citation relations: A systematic comparison of different methods. *PloS one*, 11(4):e0154404, 2016. doi: 10.1371/journal.pone.0154404.

- [149] Richard Sullivan, Seth Eckhouse, and Grant Lewison. Using bibliometrics to inform cancer research policy and spending. *Monitoring financial flows for health research*, pages 67–78, 2007.
- [150] Martin Szomszor and Euan Adie. Overton: A bibliometric database of policy document citations. *Quantitative Science Studies*, 3(3):624–650, 2022. doi: 10.1162/qss.a_00204.
- [151] Bart Thijs. Science mapping and the identification of topics: Theoretical and methodological considerations. *Springer Handbook of Science and Technology Indicators*, pages 213–233, 2019. doi: 10.1007/978-3-030-02511-3_9.
- [152] Anastasios Tombros, Rita Villa, and Cornelis Joost van Rijsbergen. The effectiveness of query-specific hierarchic clustering in information retrieval. *Information Processing & Management*, 38(4):559–582, 2002. doi: 10.1016/S0306-4573(01)00048-6.
- [153] Vincent A. Traag, Ludo Waltman, and Nees Jan Van Eck. From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, 2019. doi: 10.1038/s41598-019-41695-z.
- [154] Nees Jan van Eck. Methodological advances in bibliometric mapping of science. *Erasmus Research Institute of Management*, 2011. URL hdl.handle.net/1765/26509.
- [155] Nees Jan van Eck. Examples – VOSviewer Online Docs, 2022. URL <https://app.vosviewer.com/docs/examples/>. Accessed: 2025-04-11.
- [156] Nees Jan van Eck and Ludo Waltman. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2):523–538, 2010. doi: 10.1007/s11192-009-0146-3.
- [157] Nees Jan van Eck and Ludo Waltman. CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of Informetrics*, 8(4):802–823, 2014. doi: 10.1016/j.joi.2014.07.006.
- [158] Nees Jan van Eck and Ludo Waltman. Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, 111(2):1053–1070, 2017. doi: 10.1007/s11192-017-2300-7.
- [159] Nees Jan van Eck and Ludo Waltman. An open approach for classifying research publications. *Leiden Madtrics*, 2024. doi: 10.59350/qc0px-76778.
- [160] Cornelis Joost van Rijsbergen. Information retrieval: Theory and practice. In *Proceedings of the joint IBM/University of Newcastle Upon Tyne seminar on Data Base Systems*, pages 1–14, 1979.
- [161] Cornelis Joost van Rijsbergen and W. Bruce Croft. Document clustering: An evaluation of some experiments with the cranfield 1400 collection. *Information Processing & Management*, 11(5–7):171–182, 1975. doi: 10.1016/0306-4573(75)90006-0.
- [162] Theresa Velden, Kevin W. Boyack, Jochen Gläser, Rob Koopman, Andrea Scharnhorst, and Shenghui Wang. Comparison of topic extraction approaches and their results. *Scientometrics*, 111(2):1169–1221, 2017. doi: 10.1007/s11192-017-2306-1.
- [163] Barnabas James Walker. Citationgecko. *Zenodo*, 2022. doi: 10.5281/zenodo.7068284. Accessed: 2025-04-16.
- [164] Ludo Waltman and Nees Jan van Eck. A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, 63(12):2378–2392, 2012. doi: 10.1002/asi.22748.

- [165] Ludo Waltman, Kevin W. Boyack, Giovanni Colavizza, and Nees Jan van Eck. A principled methodology for comparing relatedness measures for clustering publications. *Quantitative Science Studies*, 1(2):691–713, 2020. doi: 10.1162/qss.a.00035.
- [166] Feifei Wang, Chenran Jia, Xiaohan Wang, Junwan Liu, Shuo Xu, Yang Liu, and Chenyuyan Yang. Exploring all-author tripartite citation networks: A case study of gene editing. *Journal of Informetrics*, 13(3):856–873, 2019. doi: 10.1016/j.joi.2019.08.002.
- [167] Qi Wang, Bentao Zou, Jialin Jin, and Yuefen Wang. Studying the linkage patterns and incremental evolution of domain knowledge structure: A perspective of structure deconstruction. *Scientometrics*, 129:4249–4274, 2024. doi: 10.1007/s11192-024-05059-3.
- [168] Yanshan Wang, Liwei Wang, Majid Rastegar-Mojarad, Sungrim Moon, Feichen Shen, Naveed Afzal, Sijia Liu, Yuqun Zeng, Saeed Mehrabi, Sunghwan Sohn, and Hongfang Liu. Clinical information extraction applications: A literature review. *Journal of Biomedical Informatics*, 77:34–49, 2018. doi: 10.1016/j.jbi.2017.11.011.
- [169] Florian Wätzold, Bartosz Popiela, and Jonas Mayer. Methodology for AI-based search strategy of scientific papers: Exemplary search for hybrid and battery electric vehicles in the Semantic Scholar database. *Publications*, 12(4):49, 2024. doi: 10.3390/publications12040049.
- [170] Michael E. Weinblatt, Roy Fleischmann, Tom W. J. Huizinga, Paul Emery, Janet Pope, Elisa M. Massarotti, Ronald F. van Vollenhoven, Joerg Wollenhaupt, Clifton O. Bingham, Brian Duncan, Namit Goel, Owen R. Davies, and Maxime Dougados. Efficacy and safety of certolizumab pegol in a broad population of patients with active rheumatoid arthritis: Results from the REALISTIC phase IIIb study. *Rheumatology*, 51(12):2204–2214, 2012. doi: 10.1093/rheumatology/kes150.
- [171] Peter Willett. Recent trends in hierarchic document clustering: A critical review. *Information Processing & Management*, 24(5):577–597, 1988. doi: 10.1016/0306-4573(88)90027-1.
- [172] Kate Williams. What counts: Making sense of metrics of research value. *Science and Public Policy*, 49(3):518–531, 2022. doi: 10.1093/scipol/scac004.
- [173] Dietmar Wolfram. The symbiotic relationship between information retrieval and informetrics. *Scientometrics*, 102(3):2201–2214, 2015. doi: 10.1007/s11192-014-1479-0.
- [174] Seokkyun Woo and John P. Walsh. On the shoulders of fallen giants: What do references to retracted research tell us about citation behaviors? *Quantitative Science Studies*, 5(1):1–30, 2024. doi: 10.1162/qss.a.00303.
- [175] Katie Wright, Su Golder, and Rocio Rodriguez-Lopez. Citation searching: A systematic review case study of multiple risk behaviour interventions. *BMC Medical Research Methodology*, 14(1):73, 2014. doi: 10.1186/1471-2288-14-73.
- [176] Weili Wu, Hui Xiong, and Shashi Shekhar. *Clustering and information retrieval*. Springer Science & Business Media, 2003. doi: 10.1007/978-1-4613-0227-8.
- [177] Qianqian Xie and Ludo Waltman. A comparison of citation-based clustering and topic modeling for science mapping. *Scientometrics*, 2025. doi: 10.1007/s11192-025-05324-z.
- [178] Shuo Xu, Junwan Liu, Dongsheng Zhai, Xin An, Zheng Wang, and Hongshen Pang. Overlapping thematic structures extraction with mixed-membership stochastic blockmodel. *Scientometrics*, 117:61–84, 2018. doi: 10.1007/s11192-018-2841-4.
- [179] Puyu Yang and Giovanni Colavizza. A map of science in Wikipedia. In *Companion Proceedings of the Web Conference 2022*, pages 1289–1300, 2022. doi: 10.1145/3487553.3524925.

- [180] Zhiguo Yu. Understanding PubMed search results using topic models and interactive information visualization. *The University of Texas School of Biomedical Informatics at Houston*, 2017. URL https://digitalcommons.library.tmc.edu/uthshis_dissertations/42/.
- [181] Ming Yuan, Justin Zobel, and Jimmy Lin. Measurement of clustering effectiveness for document collections. *Information Retrieval Journal*, 25:239–268, 2022. doi: 10.1007/s10791-021-09401-8.
- [182] Jinhyuk Yun, Sejung Ahn, and June Young Lee. Return to basics: Clustering of scientific literature using structural information. *Journal of Informetrics*, 14(4):101099, 2020. doi: 10.1016/j.joi.2020.101099.
- [183] Zeta Alpha. Zeta alpha. <https://www.zeta-alpha.com>, 2025. Accessed: 2025-09-29.
- [184] Michel Zitt. Meso-level retrieval: IR-bibliometrics interplay and hybrid citation-words methods in scientific fields delineation. *Scientometrics*, 102(3):2223–2245, 2015. doi: 10.1007/s11192-014-1482-5.
- [185] Michel Zitt, Alain Lelu, Martine Cadot, and Guillaume Cabanac. Science mapping and the identification of topics: Theoretical and methodological considerations. *Springer Handbook of Science and Technology Indicators*, pages 25–68, 2019. doi: 10.1007/978-3-030-02511-3_2.