Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mapping the Convergence of Information Retrieval and Large Language Models

Prince Opoku Saaluon

kay.saaluon@gmail.com

Ghana Cocoa Board https://orcid.org/0009-0007-8921-0918

Research Article

Keywords: information retrieval, large language models, bibliographic coupling, bibliometrics, science mapping, network analysis

Posted Date: August 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10552766/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Large language models have changed how information retrieval research is done, but the bibliometric shape of that change has not been described in detail. This paper asks three things: how the information retrieval (IR) and large language model (LLM) literatures have grown into each other, which publications hold the two together, and how the structure of the field has shifted year on year. We assembled a corpus of retrieval-related work published between 2015 and 2025 from OpenAlex, built a bibliographic coupling network with bibliometrix, and analysed it in Gephi. After thresholding by coupling strength and removing several off-topic clusters that entered through the ambiguity of the word “retrieval,” we were left with 4,064 documents and 71,534 coupling links. These fall into six main communities, spanning multimodal and vision–language retrieval, neural and conversational document retrieval, transformers and LLMs for IR, neural ranking, graph and code retrieval, and hashing-based retrieval. Betweenness centrality points to the neural-IR survey literature as the work that does most of the bridging between these groups. Splitting the network into four time periods shows a pattern easy to miss in a single static map: the field first diversifies, with modularity rising through 2019 to 2021 as transformer methods spawn new sub-areas, then converges, with modularity dropping sharply in 2022 to 2023 ($Q = 0.342$) as LLM and retrieval-augmented generation work pulls the separate communities together. The results suggest that IR has not simply borrowed LLM techniques but reorganised around them, with neural ranking research at the centre as the connective tissue.

1. Introduction

Fields converge. Methods that grew up apart start citing the same foundational papers, vocabularies blend, and after a few years what looked like two literatures reads like one. We tend to notice this only in hindsight, once the merger is complete and a review article announces it. Catching convergence while it happens, and saying when it happened, is harder, because the thing that changes is the structure of a literature rather than any single paper in it.

Information retrieval over the last decade is a good place to look for this. The arrival of pretrained transformers, dense vector representations, and more recently large language models has changed both the methods IR researchers use and the words they use to describe them. Dense passage retrieval, neural ranking, and retrieval-augmented generation now sit somewhere between classical IR and the wider machine-learning and natural-language-processing communities. Anyone who has followed the field can feel that something has shifted. What has been missing is a way to show it, to point at the moment the separate strands of retrieval research began drawing on a common base rather than simply asserting that they did.

Bibliographic coupling is well suited to that task. Two papers are coupled when they cite the same earlier work, so a coupling network groups documents by their shared intellectual foundations. If previously distinct literatures start to converge, that shows up as a change in the network's community structure, and because coupling can be measured as soon as a paper appears, the change can be tracked year by

year. Science-mapping studies in information science have long used coupling, co-citation, and co-word analysis to chart fields and their shifts over time (Chang et al., 2015), but the IR and LLM intersection has not been mapped this way, and we found nothing that follows how its structure has moved as the technology advanced.

This paper does that, and reports a convergence that can be dated. Working from a coupling network of retrieval-related work published between 2015 and 2025, we split the literature into four periods and track its community structure across them. The headline result is a sharp drop in modularity between 2019–2021 and 2022–2023, the signature of separate communities folding into one another, which coincides with the rise of large language models and retrieval-augmented generation. Along the way the analysis shows which sub-areas of retrieval merged and which stayed distinct, and which works did most of the connecting. The contribution is therefore twofold: a substantive account of how the IR–LLM literature came together, and a demonstration that temporal coupling analysis can register convergence as a measurable, dated event rather than an impression after the fact.

We organise this around three questions:

RQ1. What does the intellectual structure of the IR–LLM literature look like when viewed through bibliographic coupling?

RQ2. Which publications act as bridges between the communities that make up this literature?

RQ3. How has the community structure changed over time, and does it show convergence?

2 Background and related work

2.1 Science mapping and bibliographic coupling

Science mapping is the use of bibliometric data to reveal the structure of a research field, and it has a long history in information science. Maps are typically built from one of a few relational signals: co-citation, where documents are linked by being cited together; co-word or keyword co-occurrence, where they are linked by shared terms; and bibliographic coupling, where they are linked by shared references (Kessler, 1963). Each captures a slightly different notion of relatedness, and the choice among them shapes what a map can show. Boyack and Klavans (2010) compared co-citation, bibliographic coupling, and direct citation for their ability to represent the research front, and found coupling to perform at least as well as the alternatives, which matters here because the IR–LLM intersection is recent and still forming. Coupling has a further practical advantage: unlike co-citation, which only takes shape once later authors cite a pair of papers together, it can be measured the moment a paper is published, making it well suited to a fast-moving literature.

The tooling has matured alongside the methods. VOSviewer (van Eck & Waltman, 2010) is widely used for distance-based maps in which related items are placed close together, and it is a natural reference

point for the present study because it represents the dominant style of bibliometric visualisation in information science. Bibliometrix (Aria & Cuccurullo, 2017) provides a scriptable R environment for building the underlying networks, while Gephi (Bastian et al., 2009), with the ForceAtlas2 layout (Jacomy et al., 2014) and Louvain community detection (Blondel et al., 2008), supports the node-level structural analysis, betweenness and modularity in particular, that distance-based tools capture less directly. Applied work has used these methods to track how research subjects evolve; Chang et al. (2015), for example, combined keyword, coupling, and co-citation analysis to follow shifts in library and information science across several periods, an approach close in spirit to the temporal slicing used here.

2.2 Neural and LLM-based information retrieval

The technical literature mapped in this study has moved quickly. The transformer architecture (Vaswani et al., 2017) and pretrained language models such as BERT (Devlin et al., 2019) reshaped natural language processing and, soon after, information retrieval. A central development for retrieval was the shift from sparse, term-based matching to dense vector representations: Karpukhin et al. (2020) showed that a simple dual-encoder trained on relatively few examples could outperform a strong BM25 baseline for passage retrieval, and Khattab and Zaharia (2020) introduced a late-interaction scheme that improved the accuracy–efficiency trade-off. These approaches are documented in a series of surveys that the present analysis recovers as central nodes (Trabelsi et al., 2021; Mitra & Craswell, 2018; Onal et al., 2018), and in more recent reviews of pretrained transformers and semantic models for first-stage retrieval (Lin et al., 2021; Guo et al., 2022; Fan et al., 2022).

The most recent turn is the integration of large language models with retrieval. Retrieval-augmented generation (Lewis et al., 2020) couples a retriever with a generative model so that outputs can be grounded in retrieved evidence, and this pattern has become a major locus of activity at the boundary between IR and the wider machine-learning community. It is precisely this period that the temporal analysis below identifies as the point of structural convergence.

2.3 Position of the present study

Two gaps motivate this paper. First, while science mapping is well established in information science, the IR–LLM intersection has not been mapped in this way, and its evolution over time has not been traced. Second, much existing bibliometric work in the field relies on distance-based mapping, which is well suited to showing clusters but less suited to identifying the individual works that bridge them or to quantifying how community structure changes year on year. By combining bibliographic coupling with node-level centrality and temporal slicing, this study addresses both gaps for a fast-moving and timely case. To find the bridging papers we use betweenness centrality (Freeman, 1977); the data come from OpenAlex (Priem et al., 2022).

3 Data and methods

3.1 Data source and search strategy

We retrieved the records from OpenAlex through the openalexR package on 26 June 2026. OpenAlex was the natural choice here. The IR–LLM intersection lives largely in computer-science conference proceedings and preprints, which OpenAlex covers more fully than the subscription databases, and its open API makes the whole pull reproducible. We defined the corpus using eight bridging terms, each of which already sits across the IR and neural/LLM boundary: dense retrieval, dense passage retrieval, neural information retrieval, neural ranking, retrieval-augmented generation, semantic search, bi-encoder, and cross-encoder. Each term was matched against document abstracts over the window 2015 to 2025. In practice this meant eight queries against the OpenAlex works endpoint, each of the form:

https://api.openalex.org/works?filter=abstract.search:<phrase>,from_publication_date:2015-01-01,to_publication_date:2025-12-31

with <phrase> running over the eight terms above. A second query combined IR-side title terms (information retrieval, document ranking, learning to rank, passage retrieval) with LLM-side abstract terms (large language model, transformer, BERT, GPT), in all sixteen combinations, over the same window. The first query, after de-duplication and the inclusion filter described below, yielded a corpus of 14,492 records. The second query returned 1,012 records (812 unique), of which 645 were new to the corpus and 532 survived the inclusion filter. Combining the two gives an analytic corpus of 15,024 records (14,492 + 532). The inclusion filter required both an IR and an LLM signal in the abstract and the presence of at least one reference. The full collection and filtering chain is given in Table 1, and Fig. 1 shows it as a flow diagram.

Table 1. Data collection and filtering (reproducibility summary)

Step	Detail / count
Data source	OpenAlex, via openalexR (R)
Retrieval date	26–28 June 2026
Publication window	2015-01-01 to 2025-12-31
Query 1 (bridge terms)	8 abstract.search phrases (see text); 14,492 records after de-duplication and filter
Query 2 (intersection)	16 IR-title × LLM-abstract combinations; 1,012 raw, 812 unique, 645 new, 532 retained after filter
Inclusion filter	Abstract present, ≥1 referenced work, IR + LLM signal in abstract
Analytic corpus	15,024 records
Working coupling network (≥5 threshold + giant component)	5,312 nodes; 82,063 edges
Off-topic clusters removed	neuroscience, remote sensing, re-identification (term polysemy); see 3.2
Final network	4,064 nodes; 71,534 edges
Community detection	Louvain, resolution 1.0, weighted
Centrality	betweenness (bridging), eigenvector (within-community)
Software	R (openalexR, bibliometrix); Gephi 0.11.2 (ForceAtlas2)

3.2 Building and cleaning the network

From the 15,024 records we built a bibliographic coupling network with the biblioNetwork() function in bibliometrix, taking documents as nodes and weighting edges by the number of shared references. Coupling networks are dense by nature, since any two papers that share even a single reference are joined, so we kept only edges representing at least five shared references and restricted the network to its giant connected component. This produced a working network of 5,312 documents and 82,063 links.

Inspecting the resulting communities, several had nothing to do with information retrieval. They had entered the corpus on the back of ambiguous terms. In cognitive neuroscience “retrieval” means memory retrieval; in remote sensing it means recovering a physical quantity from sensor data; and in computer vision “re-identification” and architecture “search” form their own literatures. Community detection placed each of these in its own cluster, which made them straightforward to identify by reading the most central papers in each. We removed them, together with a small number of scattered same-topic stragglers, and recomputed the network on what remained. The final coupling network has 4,064 documents and 71,534 links. We report this cleaning openly because the polysemy of broad retrieval

vocabulary is itself a methodological finding of general relevance: each removed cluster was a coherent, nameable literature from another field, not a set of individually chosen documents.

It is worth noting that this corpus is substantially larger than an earlier version of the analysis built from the bridge terms alone (about 3,300 documents), yet the community structure and the temporal pattern reported below are essentially unchanged. That the findings survive a roughly one-quarter expansion of the corpus is reassuring evidence that they reflect the structure of the field rather than the particularities of one search.

3.3 Analysis and community labelling

All the metrics and figures were produced in Gephi 0.11.2. We detected communities with the Louvain algorithm at resolution 1.0, using edge weights. To find the bridging papers we used betweenness centrality, which counts how often a node sits on the shortest path between other nodes, and we added eigenvector centrality as a second measure that picks out influence within a community rather than between communities. Networks were laid out with ForceAtlas2.

The author designed the study, specified the search and analysis strategy, and verified all outputs. The data-collection and network-analysis scripts (OpenAlex retrieval via openalexR and network construction in bibliometrix) were executed with the assistance of an AI coding tool (Claude Code), under the author’s direction.

Communities were labelled through a content reading of their most central documents. For each community returned by the Louvain algorithm, member documents were ranked by eigenvector centrality, which surfaces the works most representative of a community's internal citation structure, and the titles of the highest-ranked documents were read (the top five in each dominant community, and a larger sample where a community's theme was not immediately clear). A theme label was then assigned on the basis of the recurring topical content of these representative works, for example “neural ranking / neural IR” for the community anchored by Trabelsi et al. (2021) and Mitra and Craswell (2018), or “pretrained transformers and LLMs for IR” for the community anchored by Guo et al. (2022) and Lin et al. (2021). Only the dominant communities, those accounting for more than three per cent of nodes, were labelled and carried into the analysis; smaller communities were treated as peripheral and left unlabelled. Where a community's representative documents pointed clearly to a literature outside information retrieval, it was flagged as off-topic and handled through the corpus-refinement step in Section 3.2 rather than given a substantive label.

3.4 Temporal analysis

To follow how the structure changed, we split the corpus into four periods chosen to line up with real shifts in the field: before 2019, before transformers were widely used in IR; 2019 to 2021, when dense

retrieval and transformer-based ranking took hold; 2022 to 2023, when LLMs and retrieval-augmented generation arrived; and 2024 to 2025. For each period we built a separate coupling network with the same threshold and giant-component rule, and computed its modularity and community count. One caveat is worth stating up front. Because coupling depends on references accumulating over time, recent papers have had less chance to pick up couplings, so the later slices are smaller and the periods are not strictly comparable. For that reason we read the trend in modularity rather than the individual numbers, and we treat the most recent slice with extra caution.

4 Results

4.1 The overall structure (RQ1)

The final network has 4,064 nodes and 71,534 edges and is undirected. Its community structure is clear, with a modularity of 0.627 across 33 communities, though six communities dominate and between them hold close to nine in ten of all the documents (Table 2). The remaining communities are small and peripheral. The structure is notably stable: an earlier, smaller version of the corpus produced almost the same modularity (0.631) and the same six dominant themes, so the picture below is not an artefact of how the corpus was drawn.

Table 2. The six dominant communities, with representative documents

Theme	Share	Representative documents (highest eigenvector centrality)
Multimodal / vision–language retrieval	25.0%	Lei et al. (2021, CLIPBERT); Johnson et al. (2016, DenseCap); Reed et al. (2016)
Neural & conversational document retrieval	19.2%	Sevgili et al. (2022, neural entity linking); Gao et al. (2019, neural approaches to conversational IR)
Pretrained transformers & LLMs for IR	16.0%	Guo et al. (2022); Fan et al. (2022); Lin et al. (2021)
Neural ranking / neural IR	14.9%	Trabelsi et al. (2021); Mitra & Craswell (2018)
Graph / code / item-ranking retrieval	9.0%	Zhang et al. (2021, hierarchical multi-view graph); Zamani et al. (2020)
Hashing-based retrieval	6.0%	Zhai et al. (2020, deep transfer hashing); Zhe et al. (2019, deep class-wise hashing)

Shares are percentages of the 4,064 documents. Themes were assigned by reading the highest eigenvector-centrality papers in each community (Section 3.3).

The two largest groups are multimodal and vision–language retrieval and a broad neural and conversational document-retrieval community; transformer- and LLM-based IR, neural ranking, graph and

code retrieval, and hashing-based retrieval make up the rest. These are recognisably the main strands of modern retrieval research.

4.2 The bridges (RQ2)

Betweenness centrality picks out the papers that most often lie between communities. Nearly all of the top ones are neural-IR surveys (Table 3). The strongest bridge is Trabelsi et al. (2021), with Mitra and Craswell (2018) close behind. Both are reviews, which fits: surveys cite widely across sub-areas, so they end up sitting on a lot of shortest paths. One detail is worth drawing out. The papers with the highest betweenness are not the ones with the highest eigenvector centrality. In other words, the works that bridge the field are not the same as the works that dominate any single part of it. Bridging and prestige come apart here, and that is itself a small finding.

Table 3. The leading bridges by betweenness centrality

Document (first author, year, short title)	Community / theme
Trabelsi et al., 2021, Neural ranking models for document retrieval	Neural ranking / neural IR
Mitra & Craswell, 2018, An introduction to neural IR	Neural ranking / neural IR
Guo et al., 2022, Semantic models for first-stage retrieval	Transformers & LLMs for IR
Onal et al., 2018, Neural information retrieval	Neural ranking / neural IR
Lei et al., 2021, CLIPBERT	Multimodal / vision–language
Fan et al., 2022, Pre-training methods in IR	Transformers & LLMs for IR

Titles are shortened; full details are in the reference list. Ranked by betweenness centrality in the final network.

Put simply, neural ranking research, and the survey literature in particular, is what holds the modern retrieval field together. It links classical text IR, transformer- and LLM-based methods, image retrieval, and multimodal retrieval.

4.3 How the field changed over time (RQ3)

The four period networks show a pattern that a single static map would hide (Table 4, Figs. 2 and 3). Before 2019 the field is moderately clustered, with a modularity of 0.585. From 2019 to 2021 modularity rises to 0.628 and the number of communities grows, as transformers and dense retrieval open up new and fairly distinct sub-areas. Then in 2022 to 2023 modularity drops sharply to 0.342. The previously separate communities become heavily interlinked as LLM and retrieval-augmented-generation work starts drawing on a shared base of references. The 2024 to 2025 slice shows modularity back up at

0.635, but this is much the smallest slice and the one most affected by the recency problem with coupling, so we would not read too much into the apparent rebound.

This pattern held when the analysis was repeated on the larger rebuilt corpus: the sharp 2022 to 2023 dip is, if anything, slightly deeper than in the earlier version, and its position is unchanged. A diversify-then-converge trajectory, with convergence landing in 2022 to 2023, is the clearest single result of the study.

Table 4. Structure of the period networks

Period	Nodes	Edges	Modularity (Q)	Communities
Before 2019	817	7,517	0.585	19
2019–2021	1,667	19,277	0.628	29
2022–2023	822	9,800	0.342	12
2024–2025	358	1,114	0.635	17

All four networks use the same coupling threshold (at least five shared references) and are restricted to the giant component. Slice sizes differ; see the limitations.

5 Discussion

The clearest result is the one that can be dated. Across the four periods, modularity rises through 2019 to 2021 and then falls steeply in 2022 to 2023, from 0.628 to 0.342. A drop of that size means the boundaries between communities have weakened: papers that would once have clustered apart are now coupled through a shared body of references. This is what convergence looks like in a coupling network, and here it has a timestamp. The window in which it happens is the same window in which large language models and retrieval-augmented generation moved to the centre of retrieval research. We stop short of calling that a cause, because coupling structure cannot establish one, but the coincidence is hard to ignore and it is exactly where an explanation would have to start.

The static structure makes the same point from a different angle. The literature is not one undifferentiated mass; it falls into several clear communities spanning text, image, multimodal, and structured-data retrieval (RQ1). What holds them together is mainly the neural-IR survey literature, which sits in the most central bridging positions (RQ2). Read alongside the temporal result, this suggests the field has not simply absorbed LLM techniques and carried on as before. It has reorganised around neural and LLM-based methods, with neural ranking research acting as the hinge that connects retrieval traditions, including visual and multimodal ones, that used to sit apart.

One detail is worth drawing out for its own sake. The papers with the highest betweenness are not the ones with the highest eigenvector centrality. The works that bridge the field are not the works that

dominate any single part of it. Integration and prestige are different roles here, and the distinction is easy to miss in a method that reports only one kind of centrality.

For a science-mapping reader, the wider point is methodological. Convergence is usually claimed in retrospect and in prose. Splitting a coupling network into periods turns it into something narrower and more testable: a trajectory of modularity values, with a date attached to the shift. The IR–LLM case is a clean example because the convergence is recent and sharp, but the approach is not specific to it. Any field where two literatures may be merging can be examined the same way, and the result is a claim other researchers can check against updated data rather than take on trust.

6 Limitations

A few things temper the results. First, the corpus is built from a defined set of search terms, and although both the bridge-term and the IR-by-LLM queries were run, the strategy still leans toward precision over recall. It reliably captures work that uses the convergence vocabulary explicitly, but it will miss work that belongs to the intersection without naming it. Second, broad retrieval terms are subject to ambiguity. “Retrieval,” “dense,” “re-identification,” and “search” pulled in neighbouring literatures from neuroscience, remote sensing, and computer vision, which we identified as coherent off-topic clusters and removed; their appearance is a reminder of how easily this happens and is itself a methodological point worth recording. Third, coupling favours recent work only in the narrow sense that newer papers accumulate couplings over time, which is why the time slices differ in size and why we lean on the trend in modularity rather than any single value, and treat the smallest, most recent slice with particular caution. Fourth, force-directed layouts depend on network size and on where the algorithm starts, so we treat the period figures as illustrations of the quantitative trend rather than as independent evidence. Finally, the node labels come from first-author and year metadata, and while OpenAlex does a good job with author disambiguation and references, it is not perfect.

7 Conclusion

This study mapped the convergence of information retrieval and large language models through a temporal bibliographic coupling analysis of work published between 2015 and 2025. The field falls into six main communities, is held together mostly by neural-IR survey literature, and follows a diversify-then-converge path whose convergence lands in 2022 to 2023. On the methods side, the study shows what node-level centrality and time-slicing can add, since neither comes through as clearly in distance-based mapping, and it shows how much care the ambiguity of broad retrieval terms demands when a corpus is being defined. A natural follow-up would pair this analysis of intellectual structure with a co-authorship analysis of social structure, to see whether the people are collaborating across the same lines along which the ideas have converged.

Declarations

Competing interests The author has no competing interests to declare that are relevant to the content of this article.

Funding No funding was received for conducting this study.

Data availability The bibliographic records analysed in this study were retrieved from OpenAlex (https://openalex.org), an openly available index, using the queries and date window reported in Section 3.1 and Table 1. The processed datasets generated and analysed during this study, including the final and period coupling networks (GraphML), the node-level community and centrality table, and the data-collection provenance files, are openly available in the Zenodo repository at https://doi.org/10.5281/zenodo.21006279

Code availability The R scripts used for data collection (OpenAlex retrieval via openalexR, both query arms), network construction and thresholding (bibliometrix), and the temporal slicing, together with the query logs, are openly available in the same Zenodo repository at https://doi.org/10.5281/zenodo.21006279 The Gephi workflow (Louvain community detection, ForceAtlas2 layout, and centrality computation) is described in Section 3.3.

Author contributions The sole author conceived the study, collected and analysed the data, and wrote the manuscript.

References

1. Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007
2. Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media*, 3(1), 361–362. https://doi.org/10.1609/icwsm.v3i1.13937
3. Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008
4. Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389–2404. https://doi.org/10.1002/asi.21419
5. Chang, Y.-W., Huang, M.-H., & Lin, C.-W. (2015). Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics*, 105(3), 2071–2087. https://doi.org/10.1007/s11192-015-1762-8

6. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019*, pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423
7. Fan, Y., Xie, X., Cai, Y., Chen, J., Ma, X., Li, X., Zhang, R., & Guo, J. (2022). Pre-training methods in information retrieval. *Foundations and Trends in Information Retrieval*, 16(3), 178–317. https://doi.org/10.1561/1500000100
8. Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 35–41. https://doi.org/10.2307/3033543
9. Guo, J., Cai, Y., Fan, Y., Sun, F., Zhang, R., & Cheng, X. (2022). Semantic models for the first-stage retrieval: A comprehensive review. *ACM Transactions on Information Systems*, 40(4), Article 80, 1–42. https://doi.org/10.1145/3486250
10. Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLOS ONE*, 9(6), e98679. https://doi.org/10.1371/journal.pone.0098679
11. Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6769–6781. https://doi.org/10.18653/v1/2020.emnlp-main.550
12. Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25. https://doi.org/10.1002/asi.5090140103
13. Khattab, O., & Zaharia, M. (2020). ColBERT: Efficient and effective passage search via contextualized late interaction over BERT. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 39–48. https://doi.org/10.1145/3397271.3401075
14. Lei, J., Li, L., Zhou, L., Gan, Z., Berg, T. L., Bansal, M., & Liu, J. (2021). Less is more: CLIPBERT for video-and-language learning via sparse sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7331–7341. https://doi.org/10.1109/CVPR46437.2021.00725
15. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems (NeurIPS) 33*, pp. 9459–9474.
16. Lin, J., Nogueira, R., & Yates, A. (2021). Pretrained transformers for text ranking: BERT and beyond. *Synthesis Lectures on Human Language Technologies*, 14(4), 1–325. Morgan & Claypool. https://doi.org/10.2200/S01123ED1V01Y202108HLT053
17. Mitra, B., & Craswell, N. (2018). An introduction to neural information retrieval. *Foundations and Trends in Information Retrieval*, 13(1), 1–126. https://doi.org/10.1561/1500000061

18. Onal, K. D., Zhang, Y., Altingovde, I. S., Rahman, M. M., Karagoz, P., Braylan, A., … Lease, M. (2018). Neural information retrieval: At the end of the early years. *Information Retrieval Journal*, 21(2–3), 111–182. https://doi.org/10.1007/s10791-017-9321-y
19. Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint* arXiv:2205.01833. https://doi.org/10.48550/arXiv.2205.01833
20. Trabelsi, M., Chen, Z., Davison, B. D., & Heflin, J. (2021). Neural ranking models for document retrieval: A survey. *Information Retrieval Journal (Discover Computing)*, 24, 400–444. https://doi.org/10.1007/s10791-021-09398-0
21. van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3
22. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS) 30*, pp. 5998–6008.

Figures

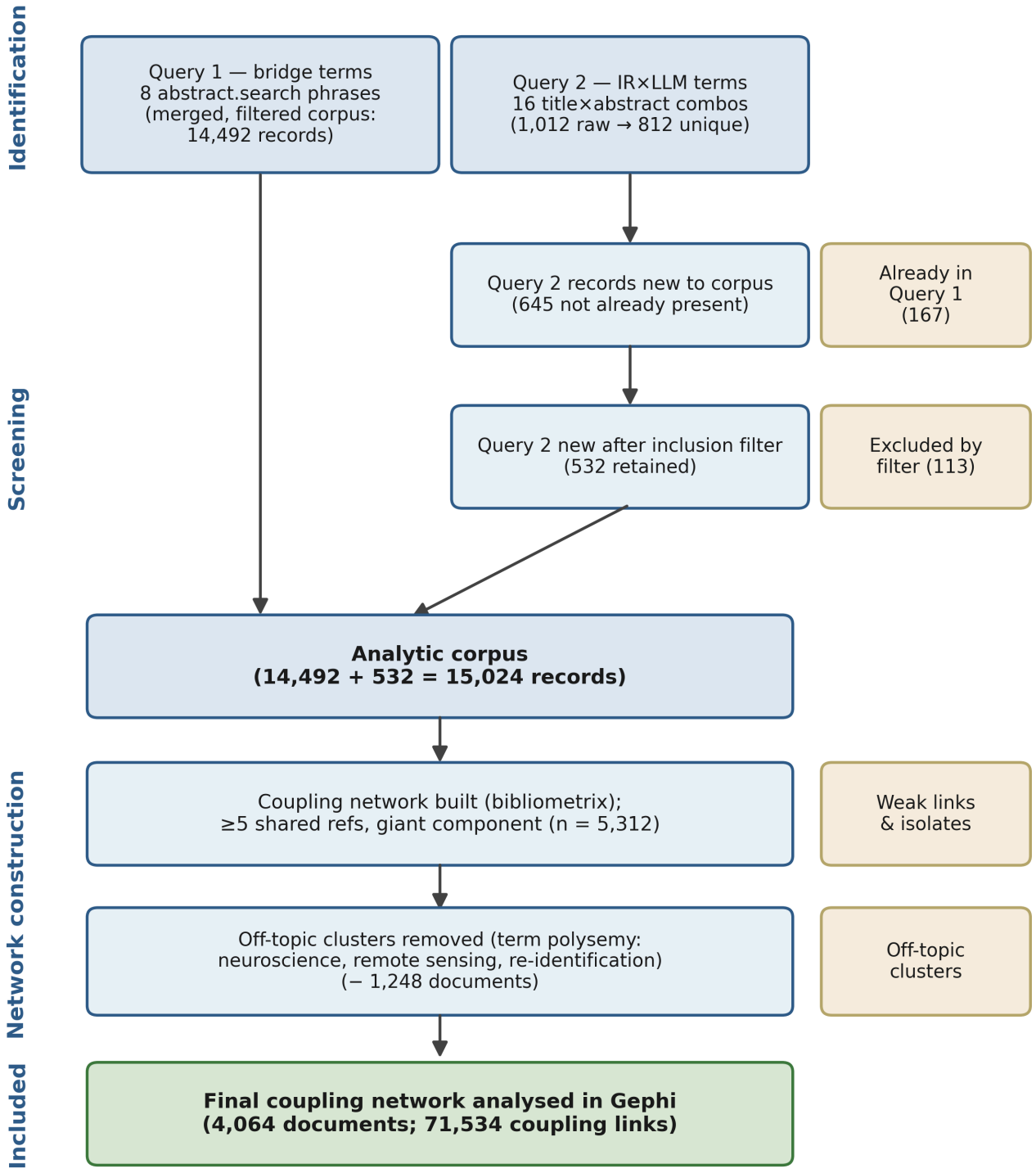


Figure 1

Flow of records from retrieval to the final coupling network

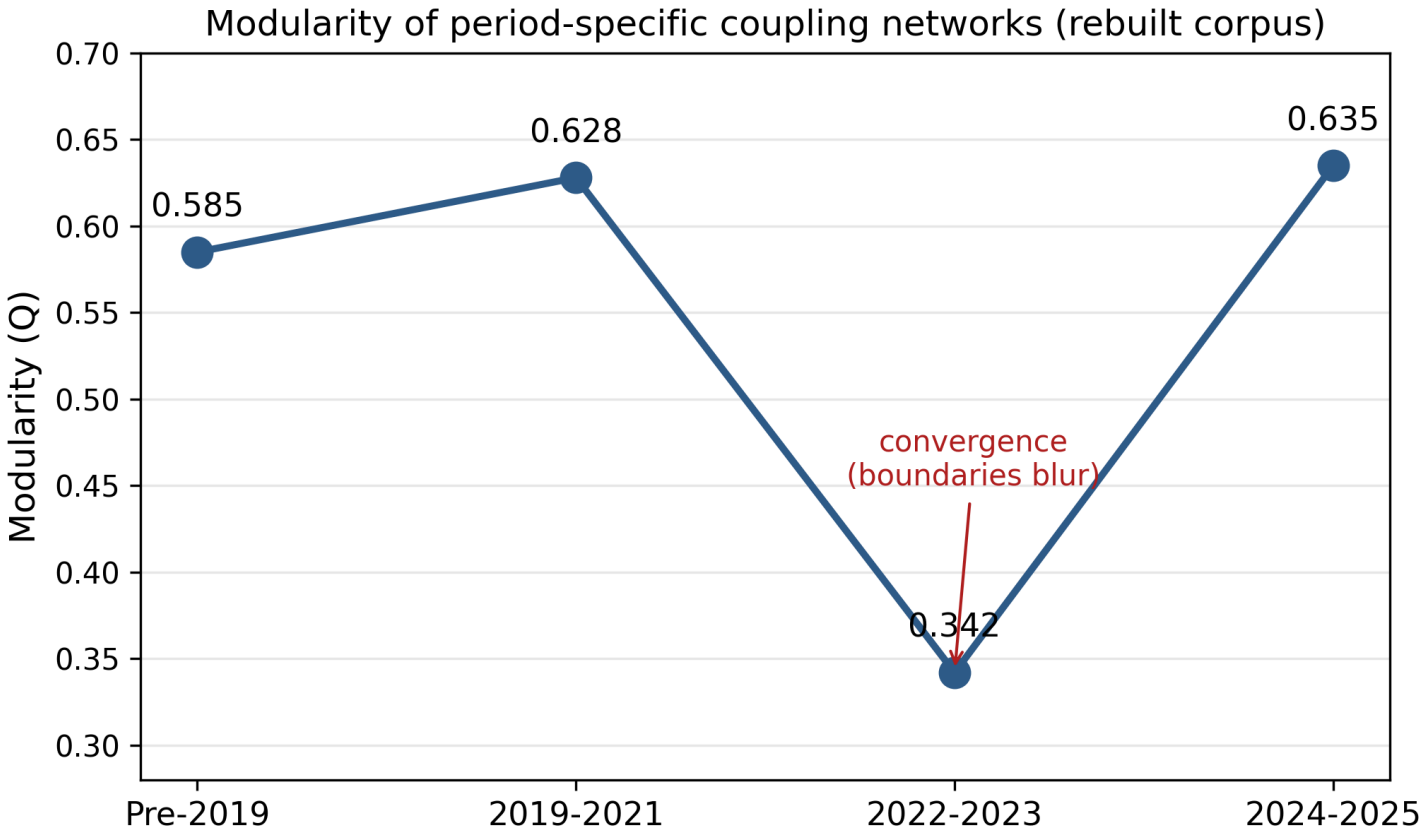


Figure 2

Modularity of the four period networks (rebuilt corpus). The sharp drop in 2022 to 2023 is the point at which the previously separate communities interlink

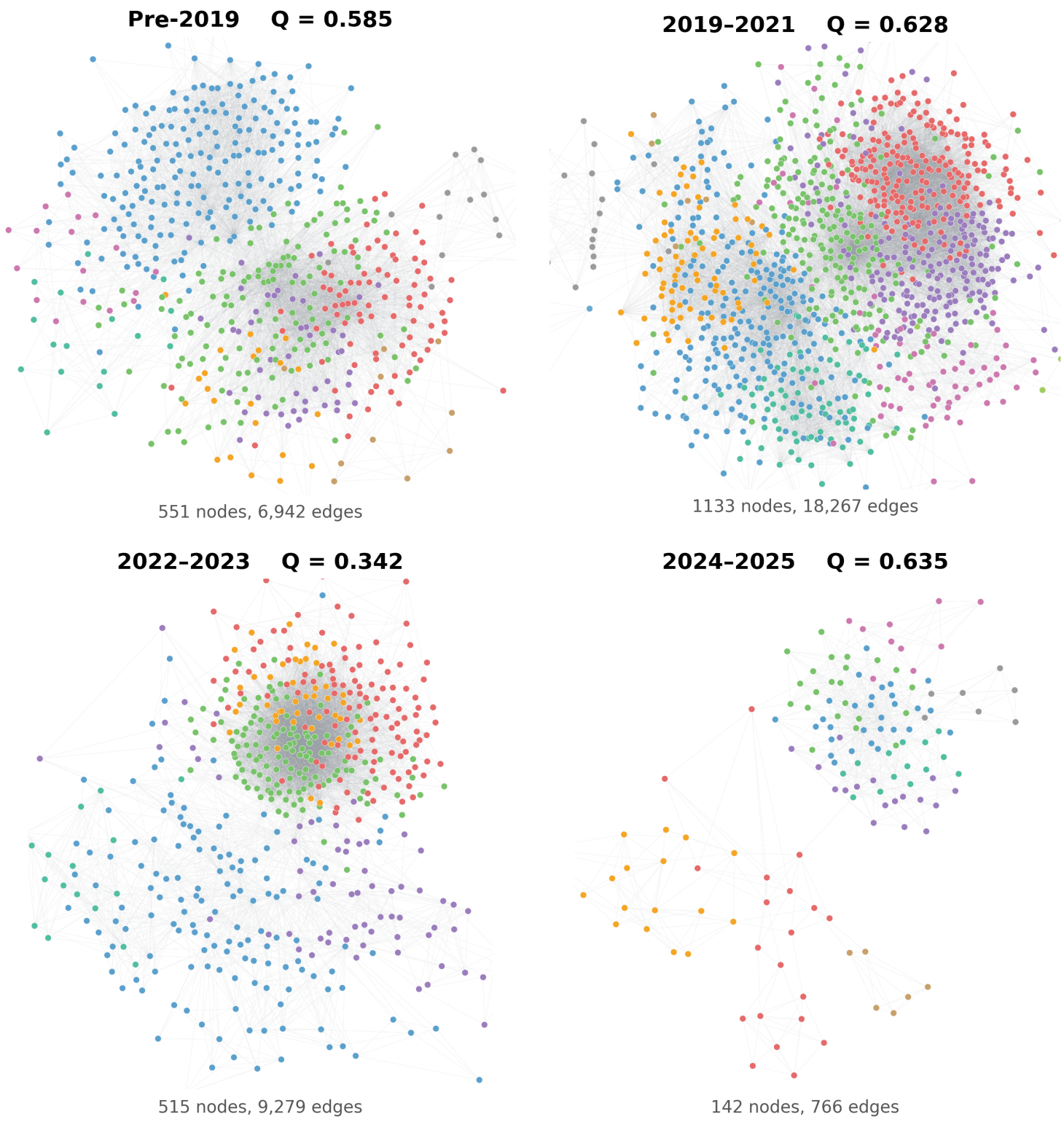


Figure 3

The four period networks, with nodes coloured by Louvain community. Communities are well separated in 2019 to 2021 and collapse into a single interlinked core in 2022 to 2023. Colours distinguish communities within each panel and are not comparable across panels; node positions are layout-dependent and not comparable in absolute terms. Panel size reflects the number of coupled documents

in each period; later periods are smaller because coupling accumulates over time. Weakly connected nodes (degree ≤ 3) were trimmed for legibility