J Data Sci. Info. Citation Studies., 2026; 5(3):237-253.
https://www.jcitation.org

Review Article

Mapping the Intellectual Landscape of Recommendation Systems Research for Academic and Research Libraries: A Scientometric Review

Srijan Dey[1,*], Parikshit Mondal[2]

[1]Central Library, National Institute of Technology Mizoram, Aizawl, Mizoram, INDIA.
[2]Department of Library and Information Science, University of Calcutta, Kolkata, West Bengal, INDIA.

ABSTRACT

Recommender Systems (RS) are central information-filtering technologies, yet their relevance to academic and research libraries remains weakly mapped. This study maps the growth, publication venues, collaboration patterns, thematic structure and intellectual base of RS research, and draws implications for library discovery, personalised reference and Selective Dissemination of Information (SDI) services. A Scopus cluster query for 2016-2025 retrieved 3,137 records; after removal of 146 duplicates, 2,991 documents were analysed in Biblioshiny (bibliometrix) using a workflow adapted from PRISMA 2020. Techniques comprised publication-growth analysis, Bradford's and Lotka's laws, country and keyword networks, thematic mapping and co-citation analysis. Annual output rose from 132 documents in 2016 to 457 in 2025. A Bradford core of 37 sources, led by IEEE Access and Expert Systems with Applications, yielded about one-third of the documents. China and India contributed the most documents, and China-United States was the strongest co-authorship tie (76 documents). Keyword analysis revealed two dominant clusters, centred on recommender systems and collaborative filtering and on deep learning and machine learning. Co-citation analysis anchored the intellectual base in the work of Koren, Sarwar, Adomavicius and Tuzhilin, and Ricci. RS research is large, fast-growing and concentrated in computing venues, with no library-science outlet among the most productive sources. Libraries can adopt content-based and hybrid recommendation first and NLP-based SDI profiling later, using the field's evaluation literature to benchmark systems. The map also guides researchers to productive authors, journals and research fronts.

Keywords: Academic libraries, Bibliometrics, Collaborative filtering, Personalized reference service, Recommender systems, Scientometrics.

Correspondence:
Mr. Srijan Dey
Library and Information Assistant, Central Library, National Institute of Technology Mizoram, Aizawl, Mizoram, INDIA.
Email: srijandeyphd2126@gmail.com

Received: 21-09-2026;
Revised: 13-10-2026;
Accepted: 03-11-2026.

INTRODUCTION

The exponential growth of scholarly and digital information has intensified the long-standing problem of information overload for researchers, students, and information professionals alike. Recommendation Systems (RS)-software agents that filter, rank, and surface content matched to a user's inferred interests-have become one of the principal computational responses to this problem, moving well beyond their origins in e-commerce and media streaming into education, health, tourism, and library science (Zhang *et al.*, 2019; Ko *et al.*, 2022). At their core, RS operationalize decades of expert-system reasoning, statistical inference, and, increasingly, Artificial Intelligence (AI) and deep learning, to predict which items-books, articles, courses, or services-a given user will find relevant (Aggarwal, 2016).

For academic and research libraries, this technological trajectory intersects with two much older professional practices: personalized reference service and Selective Dissemination of Information (SDI), the practice of proactively routing new information to users based on standing interest profiles (Luhn, 1961). Where SDI once relied on manually curated subject profiles, contemporary recommender architectures-powered by Natural Language Processing (NLP) for semantic understanding of queries and documents, and Artificial Neural Networks (ANN) for latent-preference modelling-offer libraries the possibility of automating and scaling personalized service delivery across discovery layers, institutional repositories, and digital reference desks. Research-paper recommender systems are the clearest precedent for this transfer (Beel *et al.*, 2016).

Despite the practical convergence of RS and library science, the two literatures have evolved largely in parallel: RS research is concentrated in computer science, artificial intelligence, and

OPEN ACCESS

Manuscript TECHNOMEDIA

DOI: 10.5530/jcitation.20260017

Copyright Information :
Copyright Author (s) 2026 Distributed under Creative Commons CC-BY 4.0

Publishing Partner : Manuscript Technomedia. [www.mstechnomedia.com]

information-systems venues, while Library and Information Science (LIS) scholarship on personalized service and SDI has developed with comparatively limited engagement with the technical RS literature. A systematic, quantitative account of how the RS field itself has grown, diversified, and organised intellectually is therefore a necessary first step toward more informed adoption by the LIS community.

This study addresses that gap through a bibliometric and scientometric review of the recommendation-systems literature indexed in Scopus between 2016 and 2025. Bibliometrics-the quantitative study of publications, citations, and authorship patterns (Pritchard, 1969)-and scientometrics, its closely related discipline concerned with mapping the structure and evolution of scientific fields and traced to the work of (Price, 1963), together provide the methodological toolkit used here. By combining publication-trend analysis, source (journal) productivity via Bradford's Law, author- and country-level analyses, thematic mapping, and co-citation analysis, the review offers both a descriptive map of a decade of RS scholarship and a practice-oriented discussion of its relevance to library and information practice.

Objectives and Scope of the Study

The study aims to map the intellectual landscape of recommendation-systems research and to translate that map into implications for personalized reference and SDI services in academic and research libraries. Four research questions guide the analysis: (RQ1) What are the publication and growth trends of RS research between 2016 and 2025? (RQ2) Which journals, authors, institutions, and countries dominate output and collaboration? (RQ3) What thematic and intellectual structures, revealed through keyword co-occurrence and co-citation analysis, characterise the field? (RQ4) What implications do these structures carry for the design of personalized reference and SDI services in academic and research libraries?

The scope is delimited as follows. The data source is Scopus alone; the publication window is 2016-2025; all document types are retained; and the search vocabulary combines core recommender-system terminology with library-service concepts (SDI service, academic library). The unit of analysis is the bibliographic record, and the review is descriptive and mapping-oriented: it characterises the literature rather than evaluating the effectiveness of individual recommendation algorithms.

LITERATURE REVIEW

A recommender system is broadly defined as an information-filtering technique that predicts the "rating" or preference a user would assign to an item, generating a ranked list of suggestions (Ricci *et al.*, 2011). Three canonical filtering paradigms dominate the literature. Content-Based Filtering (CBF) recommends items similar, in their descriptive attributes, to those a user has previously favoured, typically using term-weighting and vector-space representations of item content (Pazzani and Billsus, 2007). Collaborative Filtering (CF), by contrast, infers preferences from the behaviour of similar users (user-based CF) or the co-occurrence patterns among items (item-based CF), and has been substantially advanced by latent-factor and matrix-factorisation models (Koren *et al.*, 2009; Sarwar *et al.*, 2001). Hybrid recommendation systems combine CBF and CF, and increasingly knowledge-based and demographic signals-to mitigate the cold-start and sparsity problems that limit each approach (Burke, 2002).

Contemporary recommendation algorithms are built upon a broader artificial intelligence and Machine Learning (ML) substrate. Deep neural architectures-autoencoders, convolutional and recurrent networks, and, more recently, graph neural networks and transformer-based models-have increasingly supplemented or displaced purely memory-based CF, enabling the fusion of heterogeneous signals (text, images, sequential behaviour) into unified user and item embeddings (Zhang *et al.*, 2019). Natural language processing also supports content-aware and explainable recommendation, allowing systems to parse unstructured text such as abstracts, reviews, or reference queries into structured semantic profiles (Zhang and Chen, 2020).

Within library and information science, the conceptual antecedent of the modern recommender system is SDI, described by Luhn (1961) as a mechanism for automatically matching new documents to standing user-interest profiles. Personalized reference service extends this logic to interactive, one-to-one assistance, traditionally delivered by a human intermediary but increasingly supplemented by AI-driven discovery tools and recommendation-enabled catalogues. The research-paper recommender-systems literature surveyed by Beel *et al.* (2016) is the most direct bridge between the two traditions, because it addresses scholarly discovery rather than commerce or entertainment.

Conceptual Framework

The framework adopted here treats library recommendation as a layered technical stack serving three families of library service. At the base of the stack, machine learning supplies the general inferential machinery: supervised regression and classification for rating prediction, unsupervised clustering for user or item segmentation, and reinforcement learning for sequential, session-aware recommendation. Artificial neural networks, and specifically deep architectures such as autoencoders, convolutional, recurrent and long short-term memory networks, and graph neural networks, learn non-linear latent representations of users and items directly from raw interaction, text, or graph-structured data (Zhang *et al.*, 2019). Natural language processing intersects this stack wherever recommendation

must operate over unstructured text, for example in sentiment analysis of reviews and topic modelling of abstracts, and transformer-based large language models have begun to supply open-world knowledge and reasoning to recommenders (Xi *et al.*, 2024). Expert systems-rule-based reasoning engines that predate statistical machine learning-persist chiefly as knowledge-based components that encode explicit domain constraints, such as prerequisite relationships in course recommendation, which purely data-driven models struggle to capture reliably.

Academic and research libraries are natural adopters of this stack on at least three fronts: (a) discovery-layer recommendation ("readers who used this also used"); (b) institutional-repository and research-guidance recommendation, which echoes the research-paper recommender tradition documented by Beel *et al.* (2016); and (c) revived, automated SDI alerting built on NLP-based semantic profiling rather than manually maintained keyword lists. The present review treats the RS literature as the technical foundation on which these services can be built. Each bibliometric technique maps onto one question about that foundation: publication-growth analysis indicates how fast the foundation is expanding; Bradford's and Lotka's laws and the country and author analyses show who produces it and where; keyword co-occurrence and thematic mapping show which technical strands it contains; and co-citation analysis shows which works anchor it.

METHODOLOGY

This study followed a quantitative bibliometric and scientometric design. The Scopus database was selected as the sole data source for its broad multidisciplinary coverage of computer science, information systems, and library and information science venues, and its established use in large-scale bibliometric reviews (Aria and Cuccurullo, 2017).

Search strategy

Records were retrieved using a Boolean cluster query (Cluster A) combining the domain's core technical and applied-service vocabulary:

Expert System AND (Recommendation System OR Recommender System) AND Collaborative filtering AND Content-based filtering AND Hybrid recommendation AND (Personalized recommendation OR Personalised recommendation) AND Algorithm AND SDI Service AND Academic Library

The search was restricted to the publication window 2016-2025 and imposed no restriction on document type. It identified 3,137 records.

Screening and eligibility

The records were processed following a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 flow diagram adapted for bibliometric studies (Page *et al.*, 2021), comprising identification (records retrieved from Scopus), screening (duplicate removal and metadata verification), eligibility (subject-area, document-type and publication-window confirmation), and inclusion. Screening removed 146 duplicate records, leaving 2,991 records. Because the subject-area, document-type and date criteria were embedded in the Scopus query and applied at database level, no further records were excluded at the eligibility stage, and the 2,991 records constituted the final analytic corpus (Figure 1). This is consistent with the study's aim of producing a comprehensive, full-spectrum map of the domain rather than a narrowly filtered systematic review.

Analytical tools and techniques

The corpus's bibliographic metadata (author, source, year, citations, keywords, affiliations, and references) were exported and analysed using Biblioshiny, the web interface for the bibliometrix R package (Aria and Cuccurullo, 2017). Five complementary techniques were applied: (a) annual publication-growth trend analysis; (b) source-productivity analysis via Bradford's Law of scattering (Bradford, 1934), which stratifies journals into zones of decreasing productivity; (c) author-productivity and collaboration analysis, informed by Lotka's Law of scientific productivity (Lotka, 1926); (d) keyword co-occurrence and strategic (thematic) mapping, following the density/centrality logic used in science-mapping software (Cobo *et al.*, 2011); and (e) co-citation analysis to reconstruct the field's intellectual structure (Small, 1973). Biblioshiny identifies authors by surname and initials, so homonymous authors may be merged in author-level analyses.

RESULTS AND ANALYSIS

Publication Growth and Descriptive Profile

Annual output rose from 132 documents in 2016 to 457 in 2025 (Figure 2). The largest year-on-year rises occurred between 2017 and 2018 (from 157 to 241 documents) and between 2021 and 2022 (from 279 to 361). Output dipped slightly between 2019 (284) and 2020 (271) and was nearly flat in 2021 (279) before growth resumed. The 2018 rise is consistent in timing with the diffusion

Table 1: Documents by type.

Document type	Number of documents
Article	1829
Conference Paper	802
Review	165
Book Chapter	120
Book	42
Retracted	21
Short Survey	9
Editorial	3

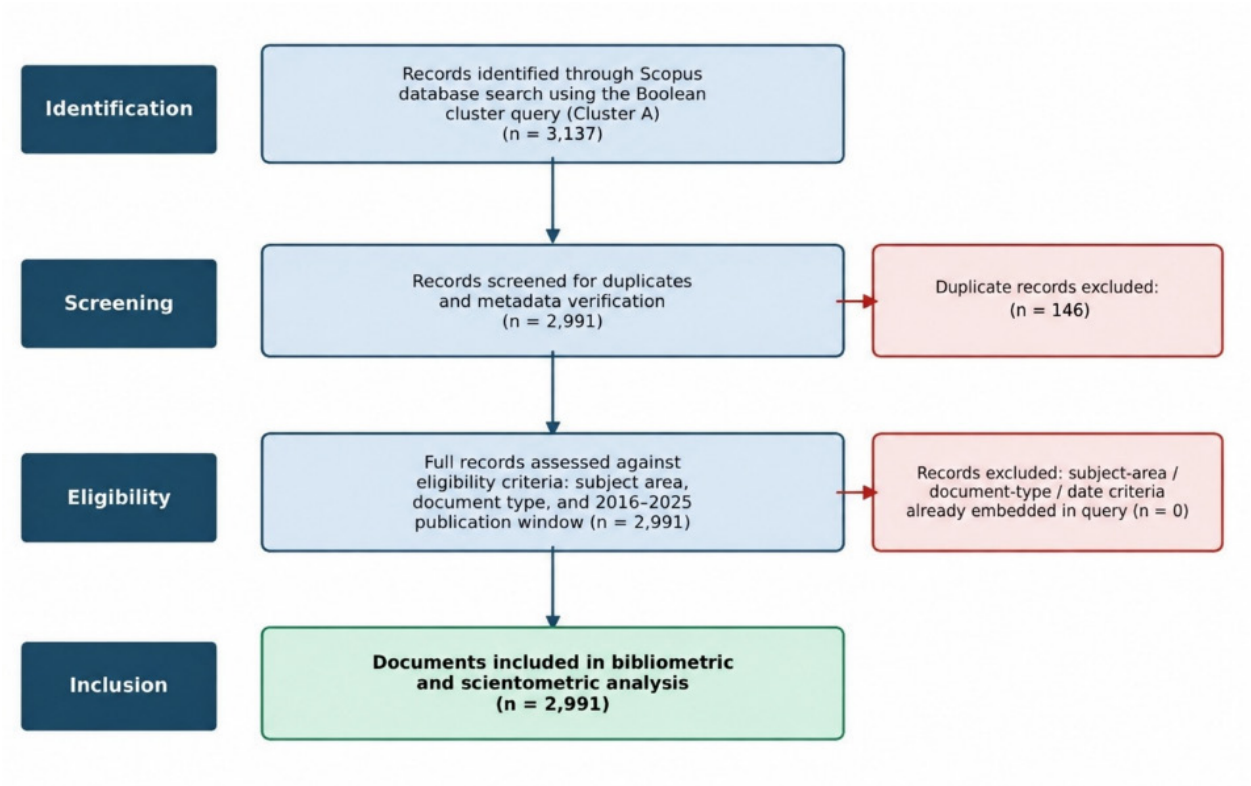

Figure 1: PRISMA flow diagram of the study identification, screening, and inclusion process for the bibliometric analysis.

of deep-learning methods into recommender-system design (Zhang *et al.*, 2019), and the growth after 2021 with the spread of hybrid, explainable, and large-language-model-augmented approaches (Xi *et al.*, 2024), although the present analysis does not test these links. The trajectory indicates that recommendation systems remain an expanding, rather than plateauing, research front. Several of the corpus's most-cited items are survey and review articles (Section 6.7), a pattern typical of fields that consolidate a growing primary literature into synthesis works.

Figure 3 shows the ten most productive authors. Margaris, D. has the highest document count (17), followed by Vassilakis, C. (16). Three further authors (Katarya, R.; Niu, Z.; and Yera, R.) have contributed 12 documents each, whereas the remaining five authors have contributed 10 or 11.

Figure 4 presents the ten most productive institutional affiliations. Vellore Institute of Technology leads, followed by the Ministry of Education of the People's Republic of China and Delhi Technological University. Other prominent affiliations include Mohammed V University Rabat, University of Technology Sydney, Beijing Institute of Technology, Chinese Academy of Sciences, SASTRA Deemed University, Beihang University and Universiti Sidi Mohamed Ben Abdellah.

Figure 5 shows the distribution of documents by country or territory. China contributed the most documents (about 815), followed by India (about 665) and the United States (about 240). Iran contributed about 125 documents; Australia, Malaysia, Spain and South Korea each produced approximately 105-115; and the United Kingdom and Italy recorded about 98 and 75, respectively. Output is therefore concentrated in China and India.

Table 1 shows that articles are the dominant document category (61.2% of output), followed by conference papers (26.8%). Reviews constitute 5.5% and book chapters 4.0%. Books represent 1.4% of the documents, whereas retracted documents, short surveys, and editorials account for 0.7%, 0.3%, and 0.1%, respectively. Articles and conference papers together constitute 88.0% of the corpus, indicating that RS research is disseminated predominantly through journal articles and conference papers.

As shown in Figure 6, the most frequent subject area is Computer Science (40.8%), followed by Engineering (18.4%) and Mathematics (9.7%). Other subject areas account for 6.7%, while Decision Sciences and Social Sciences contribute 6.2% and 5.4%, respectively. Business, Management and Accounting account for 4.2%, followed by Materials Science (3.7%), Physics and Astronomy (2.6%), Energy (1.3%), and Medicine (1.1%). Computer science, engineering, and mathematics together form 68.9% of the documents.

As Figure 7 shows, only a few funding bodies are recorded. The National Natural Science Foundation of China is the most frequent sponsor (14 documents), and every other sponsor appears in seven or fewer documents; the remaining sponsors in the top ten come from Europe, the United States, China, Taiwan, Portugal and Brazil. Sponsor counts are small relative to the size of the corpus, so Figure 7 should be read as indicative only.

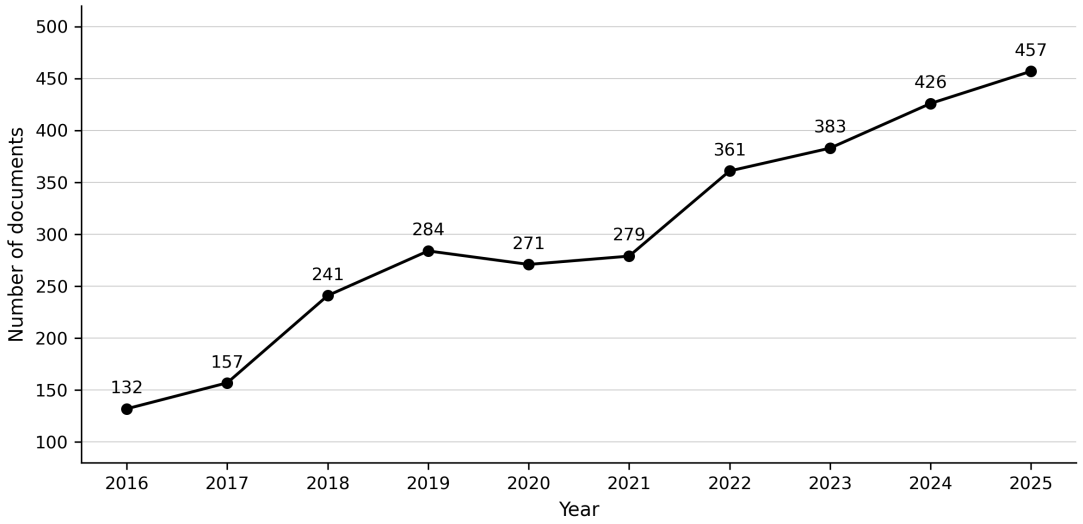

Figure 2: Annual number of Scopus-indexed documents, 2016-2025 (*n*=2,991).

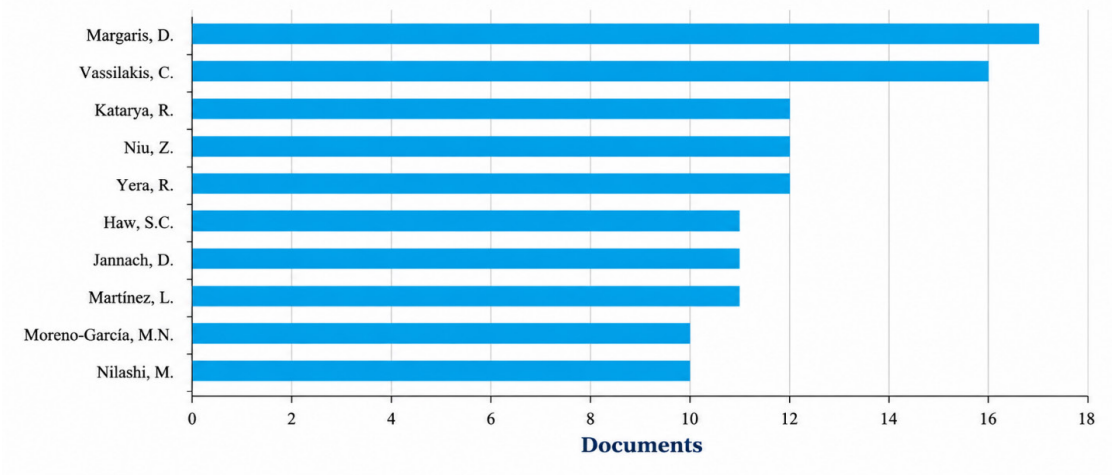

Figure 3: Documents by author (ten most productive authors).

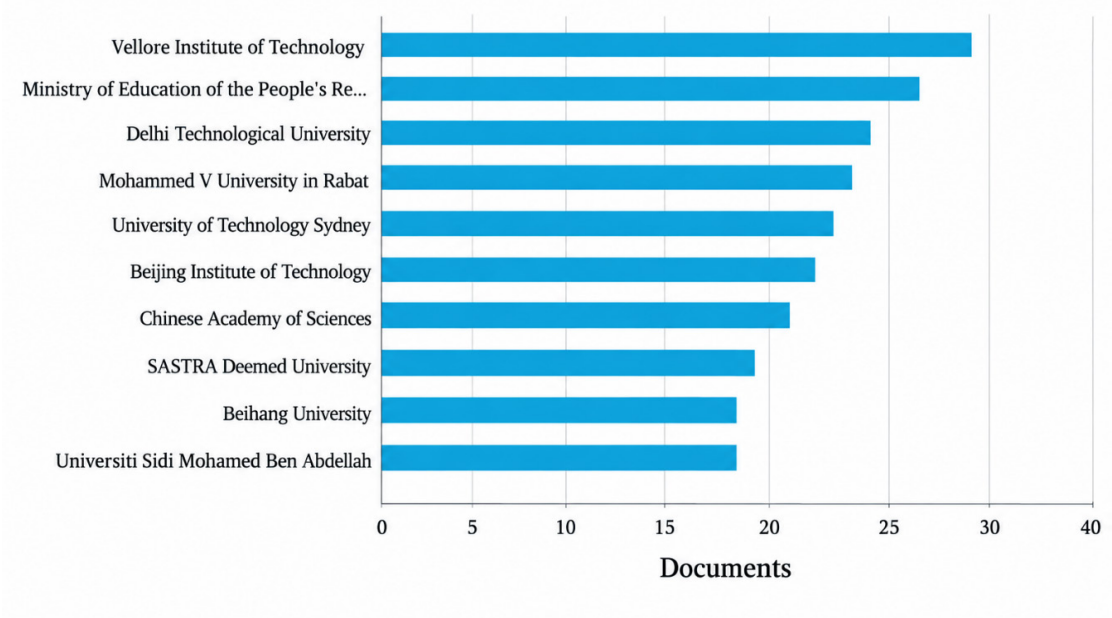

Figure 4: Documents by affiliation (ten most productive institutions).

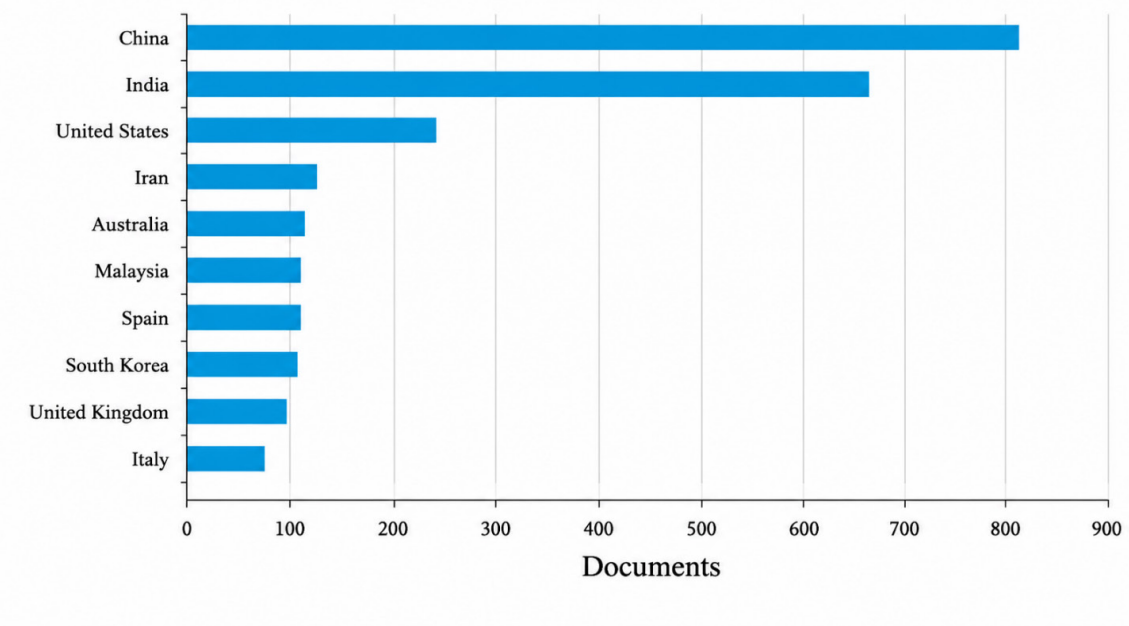

Figure 5: Documents by country or territory (ten most productive countries).

Table 2: Top ten most productive sources (Bradford Zone 1).

Rank	Source	Number of documents
1	IEEE Access	153
2	Expert Systems with Applications	105
3	Lecture Notes in Networks and Systems	59
4	Multimedia Tools and Applications	49
5	Applied Sciences (Switzerland)	45
6	Lecture Notes in Computer Science	43
7	Knowledge-Based Systems	40
8	ACM International Conference Proceeding Series	34
9	Electronics (Switzerland)	30
10	Information Sciences	28

Journal Productivity and Bradford's Law

Applying Bradford's Law of scattering to the 1,215 source titles that carried the 2,991 documents identifies a compact nucleus of highly productive journals. Zone 1, the core zone yielding roughly one-third of total output (33.7%), comprises 37 sources, headed by *IEEE Access* (153 documents), *Expert Systems with Applications* (105), *Lecture Notes in Networks and Systems* (59), *Multimedia Tools and Applications* (49), and *Applied Sciences* (45). Zone 2 comprises 239 sources and Zone 3 a further 939 (Figure 8 and Table 2 list the ten most productive sources, which together account for 586 documents (19.6% of the corpus).

The prominence of *IEEE Access, Expert Systems with Applications*, and *Knowledge-Based Systems* indicates that the field privileges applied, algorithm-oriented contributions over purely theoretical treatments, and the presence of open-access outlets (*IEEE Access, Applied Sciences, Electronics*) suggests that rapid open dissemination matters to this community. No library or information-science journal appears among the ten most productive sources. Beyond the Zone 1 nucleus, the long Bradford tail-extending across hundreds of specialised and conference-proceeding sources-confirms the classic scattering pattern typical of a technically diversifying, interdisciplinary field (Figure 8) (Bradford, 1934).

Author Productivity

The author-level results (Figures 3 and 9) indicate a field organised around many moderately sized author groups rather than a small number of globally dominant individuals: the most productive author contributed 17 documents (Figure 3), against a corpus of 2,991. This pattern is consistent with Lotka's Law, which predicts a very large population of low-productivity contributors and a much smaller elite (Lotka, 1926). The ten authors with the highest output over time in Figure 9 are all recorded under common Chinese surnames, which is consistent with the country-level productivity patterns reported above; because Biblioshiny keys authors by surname and initials, however, these labels may merge several individuals and should not be read as the output of single persons.

Country Collaboration Network

International co-authorship mapping (Figure 10) shows China as the field's most connected node, with an aggregate collaboration frequency of 352 across its partnerships, followed by the United States (222), India (166), Australia (137), and the United Kingdom (127). The single strongest bilateral tie is the China-United States dyad (co-authorship frequency =

76), followed by China-Australia (44), China-United Kingdom (19), and China-Malaysia and China-Pakistan (18 each). This structure suggests a collaboration regime anchored by China and the United States, with China additionally linking several South and Southeast Asian and Middle Eastern research communities (for example, Pakistan-Saudi Arabia, frequency = 14) into the global RS network.

Keyword Co-occurrence and Thematic Structure

Keyword co-occurrence analysis of the merged keyword field (250 terms, minimum frequency 14) yields four clusters (Table 3 and Figure 11), of which two are dominant. The larger cluster, labelled recommender systems (8,471 total occurrences), is anchored by recommender systems (1,663 occurrences), collaborative filtering (1,093), recommendation system (330), personalized recommendation (218), recommendation algorithms (199), and content-based filtering (151 and 198 occurrences under its two spelling variants), together with problem-oriented terms such as cold start problems (126) and information overloads (143). The second cluster, labelled deep learning (4,231 total occurrences), is led by deep learning (360), learning systems (313), machine learning (263), data mining (204), artificial intelligence (180), and big data (109), with deep neural networks (89) marking the cluster's link to neural recommendation approaches. Two small

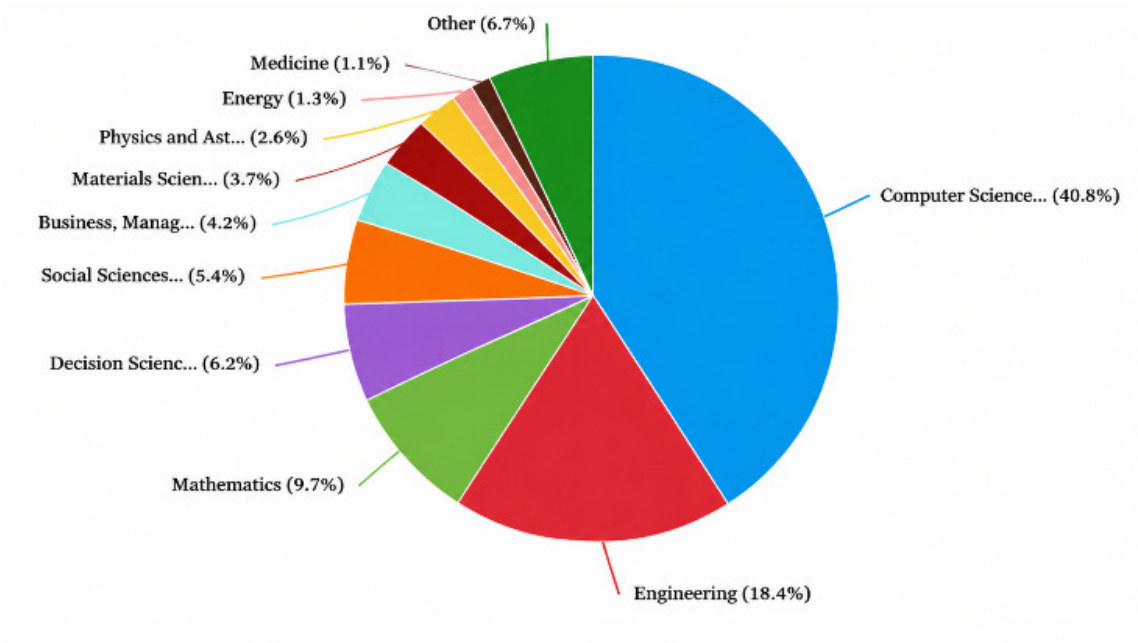

Figure 6: Documents by subject area (percentage share).

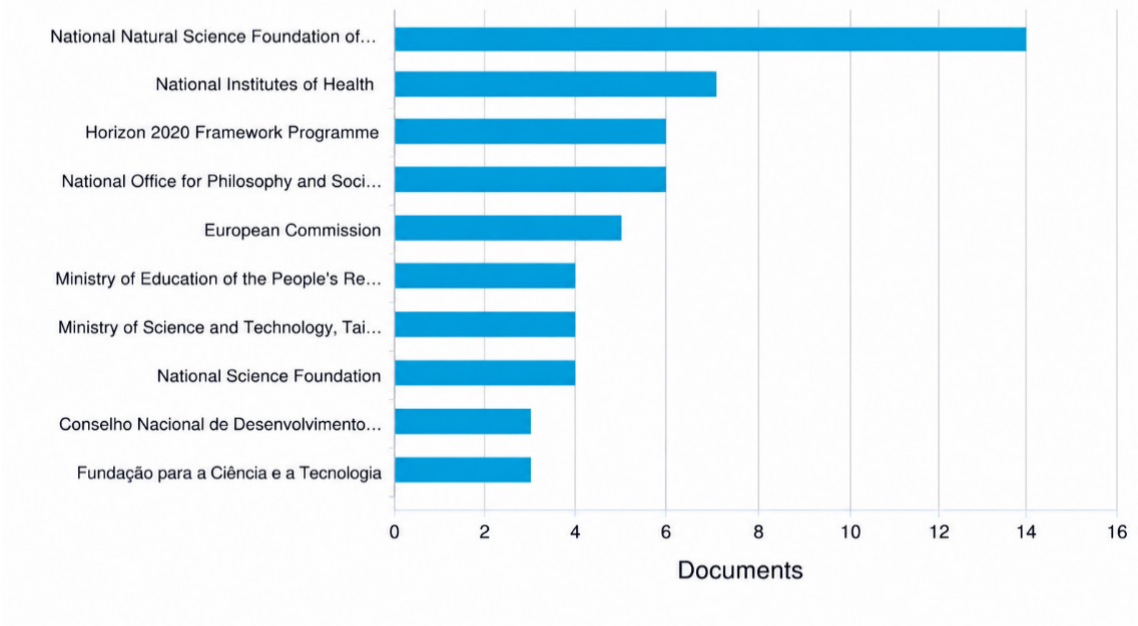

Figure 7: Documents by funding sponsor (ten most frequent sponsors).

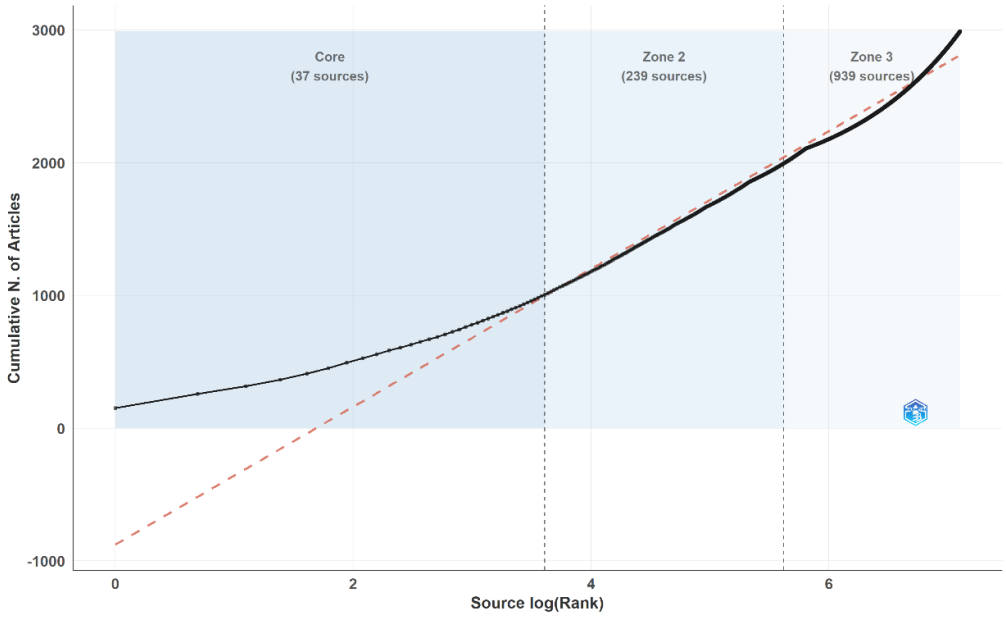

Figure 8: Bradford scattering curve showing the Zone 1-3 boundaries across the corpus's source records. Fitted model: $C(r) = -875.6 + 519.1 \log(r)$, $R^2 = 0.9724$, Bradford multiplier $k = 5.19$.

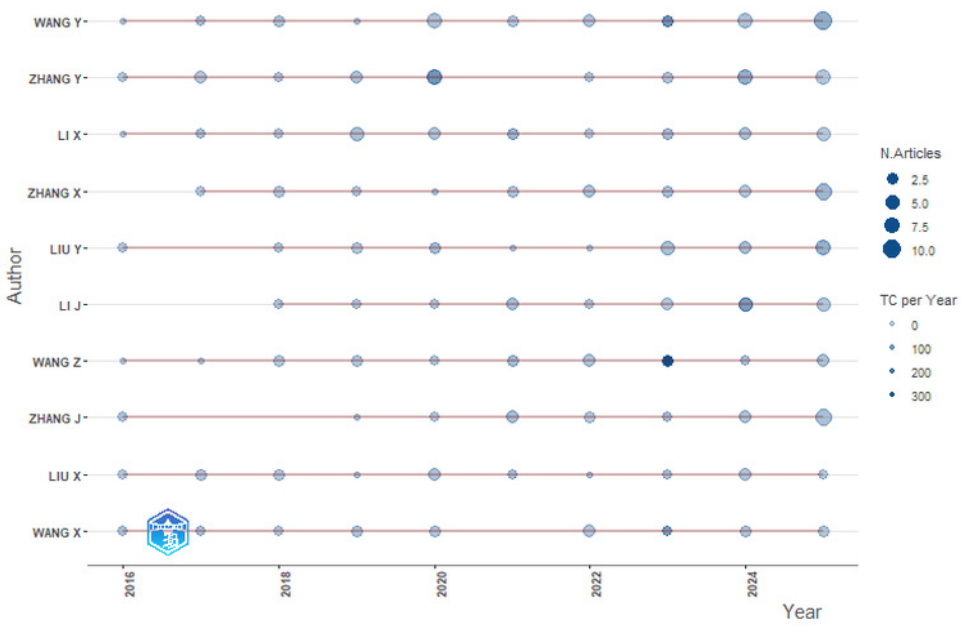

Figure 9: Authors' production over time, showing the number of articles and total citations per year.

clusters complete the structure: one around matrix factorization and matrix algebra, and one around movie recommendations, motion pictures, and sentiment analysis (87 occurrences), the latter linking the network to NLP-driven approaches.

The coexistence of a classical filtering cluster and a deep-learning cluster is consistent with the narrative, visible in the field's most-cited works, of a growing role for deep, hybrid, and explainable AI-driven recommendation alongside classical memory-based collaborative and content-based filtering (Zhang *et al.*, 2019; Zhang and Chen, 2020). The strategic (thematic) diagram in Figure 12 locates each research theme by its centrality and density. *Recommender systems* and *deep learning* are the motor themes, with high centrality and density; *semantics* occupies the niche quadrant; and *matrix algebra* and *electronic commerce* lie in the lower-density half of the map. The hybrid terms (*hybrid recommendation, hybrid recommender systems*) are assigned to the recommender-systems cluster in Table 3 rather than to a separate theme, so the data show hybridisation as a feature of the core cluster; they do not by themselves demonstrate a bridging role between clusters (cf. Burke, 2002; Portugal *et al.*, 2018).

Table 3: Keyword co-occurrence network: cluster membership and node-level centrality measures (betweenness, closeness, and PageRank).

Node	Cluster	Betweenness	Closeness	PageRank
Recommender systems	1	25.984643	0.020833	0.127617
Collaborative filtering	1	12.216563	0.020833	0.09011
Recommender system	1	0.8964386	0.020408	0.031962
Recommendation system	1	0.7127855	0.019608	0.027216
Personalized recommendation	1	0.3573797	0.020833	0.020778
Recommendation algorithms	1	0.3382349	0.020833	0.019317
Content based filtering	1	0.496383	0.020833	0.02494
Electronic commerce	1	0.5167728	0.020833	0.021789
Semantics	1	0.1987555	0.020833	0.0161
Recommendation systems	1	0.1320892	0.018519	0.014222
Social networking (online)	1	0.1762343	0.020833	0.015378
Content-based filtering	1	0.1826108	0.02	0.017068
Information overloads	1	0.123143	0.02	0.015619
User's preferences	1	0.1916763	0.020408	0.014943
Decision making	1	0.1282897	0.019608	0.013184
Behavioural research	1	0.1766228	0.020833	0.015847
Online systems	1	0.2249499	0.020408	0.016334
Recommendation methods	1	0.1356765	0.020408	0.013121
Cold start problems	1	0.1306537	0.02	0.014736
Hybrid recommendation	1	0.2390142	0.02	0.015745
Cold-start	1	0.1492287	0.02	0.012992
User profile	1	0.1159326	0.020408	0.012958
Content-based	1	0.1212699	0.018868	0.012711
Search engines	1	0.0861971	0.02	0.011714
Sales	1	0.0775548	0.018868	0.012139
Clustering algorithms	1	0.0554818	0.019608	0.010879
Data sparsity	1	0.0688105	0.018868	0.011212
Performance	1	0.0614924	0.018868	0.009688
Hybrid recommender systems	1	0.0592761	0.018868	0.011129
Recommendation	1	0.0286598	0.018182	0.007932
Deep learning	2	1.1494313	0.020833	0.033254
Learning systems	2	1.3801639	0.020833	0.034606
Machine learning	2	0.6442622	0.020833	0.026234
Data mining	2	0.364037	0.020408	0.021054
Artificial intelligence	2	0.1634621	0.020408	0.01549
Learning algorithms	2	0.2910925	0.020833	0.019054
Machine-learning	2	0.1793637	0.02	0.017042
E-learning	2	0.058423	0.018519	0.012421
Personalizations	2	0.1380292	0.02	0.013106
Big data	2	0.0839143	0.02	0.011989
Deep neural networks	2	0.0656769	0.02	0.011474
Matrix algebra	3	0.262784	0.02	0.019217

Node	Cluster	Betweenness	Closeness	PageRank
Matrix factorizations	3	0.1603888	0.02	0.019209
Matrix factorization	3	0.1675917	0.020408	0.017626
Factorization	3	0.0946415	0.019231	0.017084
Forecasting	3	0.1056769	0.019608	0.012992
Movie recommendations	4	0.0948568	0.019231	0.013776
Motion pictures	4	0.1332204	0.020408	0.013915
Sentiment analysis	4	0.0801646	0.019608	0.01108

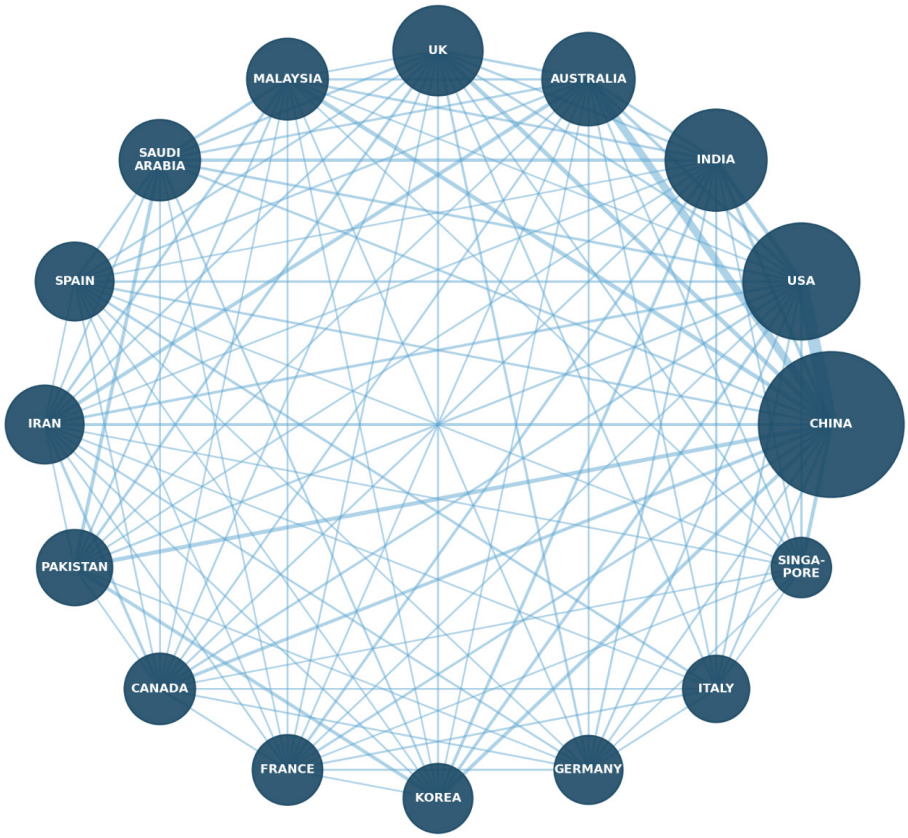

Figure 10: Country collaboration network for the top 16 most active countries. Edge width shows co-authorship frequency; node size shows total ties.

Co-citation Network and Intellectual Structure

Co-citation analysis of the corpus's cited references reconstructs the field's intellectual base as three interlocking clusters (Table 4 and Figure 13). The first, anchored by the matrix-factorization work of Koren *et al.* (2009) (the network's highest-betweenness node), together with (Sarwar *et al.*, 2001; Resnick *et al.*, 1994; Breese *et al.*, 1998), represents the classical collaborative-filtering and latent-factor lineage. The second cluster, centred on (Adomavicius and Tuzhilin, 2005; Burke, 2002; Goldberg *et al.*, 1992), captures the hybrid- and knowledge-based filtering tradition together with early information-filtering prototypes. The third cluster, organised around (Ricci *et al.*, 2011; Jannach *et al.*, 2010; Shani and Gunawardana, 2011; Adomavicius *et al.*, 2011), corresponds to the field's methodological and evaluation literature, principally the *Recommender Systems Handbook* tradition. This tripartite structure offers a citation roadmap for the field. It suggests, although the present analysis does not test it, that newer deep-learning-oriented publications draw on all three clusters: the classical CF lineage for baseline comparisons, the hybrid lineage for architectural choices, and the handbook lineage for evaluation protocols.

Highly Cited Documents

Table 5 lists the corpus's ten most-cited recommender-systems-focused documents (Zhang *et al.*, 2019; Aggarwal, 2016; Zhang and Chen, 2020; Beel *et al.*, 2016; Ko *et al.*, 2022; Portugal *et al.*,

2018; Roy and Dutta, 2022; Cui *et al.*, 2020; Wang *et al.*, 2022; Zhang *et al.*, 2021), which collectively define the field's most visible reference points for both computer-science and, increasingly, library-oriented audiences.

The prominence of Beel *et al.* (2016), a survey specifically of research-paper recommender systems, is particularly relevant to the present review's library focus, as it shows that the RS community has already produced substantial, highly cited work directly applicable to scholarly discovery and, by extension, to academic-library reference contexts.

Application of Lotka's, Bradford's and Zipf's Laws

The author- and source-productivity patterns documented above conform closely to the classical scientometric laws underpinning this review's design. Lotka's Law (Lotka, 1926) predicts that the number of authors publishing n papers is inversely proportional to n^2, producing a small elite of prolific contributors atop

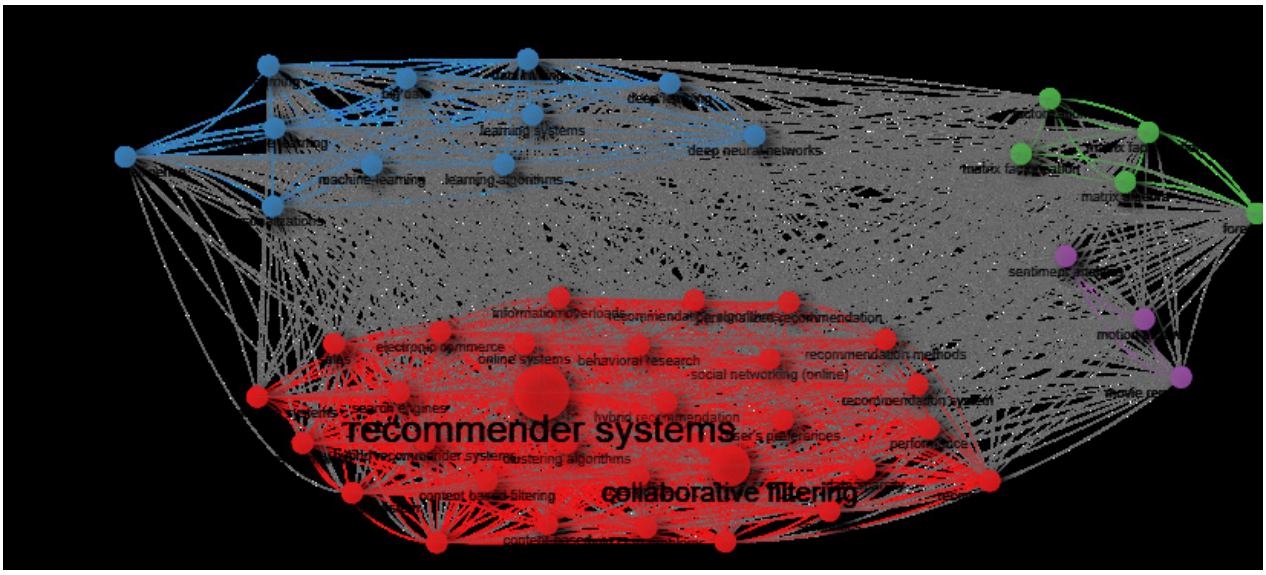

Figure 11: Keyword co-occurrence network of the merged keyword field (250 terms, minimum frequency 14).

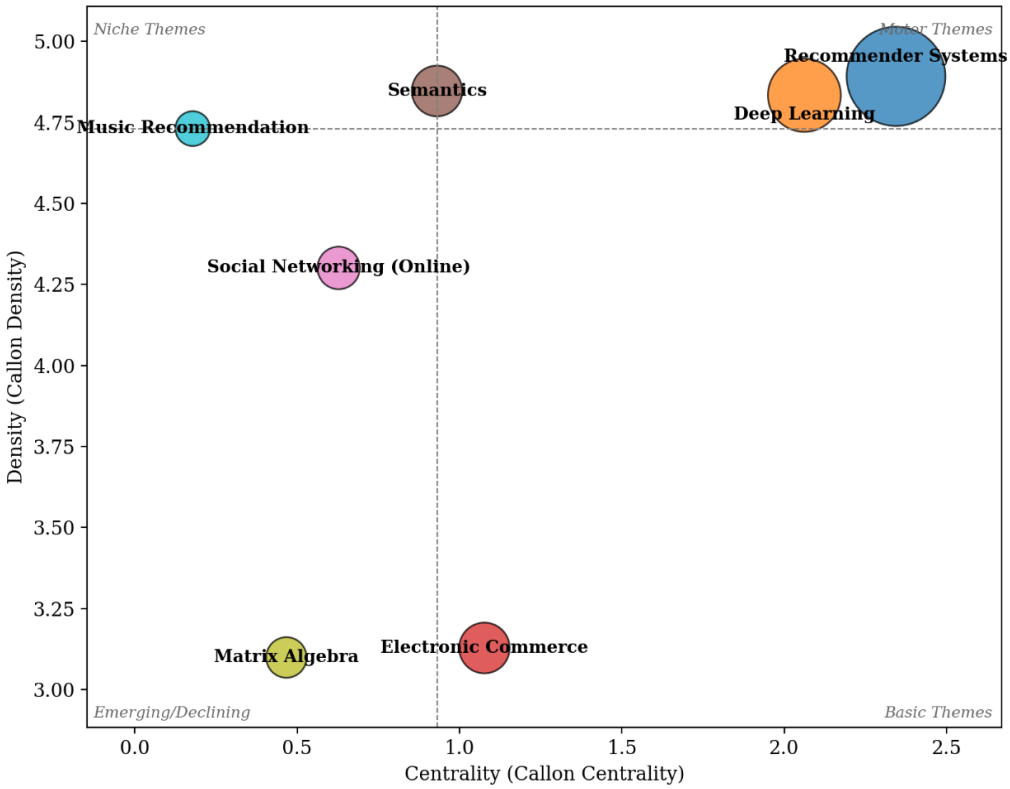

Figure 12: Strategic (thematic) diagram plotting each research theme by centrality and density; bubble size shows cluster keyword frequency.

Table 4: Co-citation network: cluster membership and node-level centrality measures (betweenness, closeness, and PageRank).

Node	Cluster	Betweenness	Closeness	PageRank
Koren Y. 2009-1	1	381.9118	0.038539	0.036488
Sarwar B. 2001	1	67.0734	0.035036	0.026419
Koren Y. 2008	1	41.50958	0.028905	0.016745
Resnick P. 1994	1	5.701507	0.028548	0.012283
Breese J.S. 1998	1	9.920182	0.029646	0.012491
Koren Y. 2009-2	1	17.34362	0.024864	0.008855
Adomavicius G. 2005-3	1	0.300883	0.021611	0.009467
Pazzani M.J. 2007-2	1	1.498706	0.0246	0.006981
Rendle S. 2012	1	1.038153	0.024864	0.011071
He X. 2017	1	2.558137	0.023839	0.010989
Adomavicius G. 2005-1	2	50.25847	0.030031	0.062711
Bobadilla J. 2013-1	2	11.62097	0.0282	0.052669
Burke R. 2002-1	2	31.10391	0.029646	0.046655
Lu J. 2015-1	2	13.84543	0.028548	0.0317
Goldberg D. 1992	2	1.278526	0.026277	0.023092
Linden G. 2003	2	2.78812	0.027528	0.022811
Isinkaye F.O. 2015	2	0.25093	0.024864	0.017558
Ko H. 2022	2	5.311891	0.025411	0.008305
Lika B. 2014	2	0.956087	0.026579	0.016003
Balabanovic M. 1997-1	2	5.037103	0.027528	0.020875
Nilashi M. 2018	2	0.071225	0.0246	0.011412
Wei J. 2017	2	0.694859	0.025134	0.012905
Park D.H. 2012	2	0.93096	0.027204	0.014847
Harper F.M.	2	1.751557	0.025411	0.009001
Herlocker J.L.	2	1.809879	0.026888	0.015741
Herlocker J.L. 2004-2	2	0.452699	0.025982	0.016789
Ricci F. 2011-2	2	0.309661	0.024864	0.008654
Burke R. 2002-2	2	1.977515	0.025693	0.008679
Su X. 2009	2	1.465359	0.026277	0.016513
Zhang S.	2	0.660287	0.025693	0.009179
Resnick P. 1997	2	0.328587	0.025411	0.017665
Lops P. 2011	3	57.61159	0.031248	0.042811
Pazzani M.J. 2007-1	3	72.1257	0.032569	0.033007
Adomavicius G. 2005-2	3	1.152861	0.024087	0.03401
Ricci F. 2011-1	3	22.98563	0.030031	0.03192
Shani G. 2011	3	29.41736	0.030426	0.030478
Jannach D. 2010	3	21.25696	0.029646	0.023915
Adomavicius G. 2011	3	21.17367	0.030031	0.024745
Bobadilla J. 2013-2	3	1.391823	0.023839	0.023858
Burke R. 2007	3	15.54621	0.028905	0.020327
Schafer J.B. 2007	3	11.95153	0.028548	0.022983
Ricci F. 2015	3	18.85143	0.029646	0.019178

Node	Cluster	Betweenness	Closeness	PageRank
Lu J. 2015-2	3	0.576515	0.023357	0.014884
Desrosiers C. 2011	3	8.202788	0.027204	0.014743
Herlocker J.L. 2004-1	3	0.14463	0.023124	0.016943
Koren Y. 2015	3	5.506043	0.025982	0.014572
Aggarwal C.C. 2016	3	2.134371	0.025134	0.009142
Balabanovic M. 1997-2	3	0.092084	0.024341	0.01308
Blei D.M. 2003	3	0.058984	0.022234	0.007498
Adomavicius G. 2005-4	3	0.059861	0.018954	0.006349

Note. Numeric suffixes (for example, -1, -2) distinguish separate reference records that share the same first author and year in the exported Scopus reference strings.

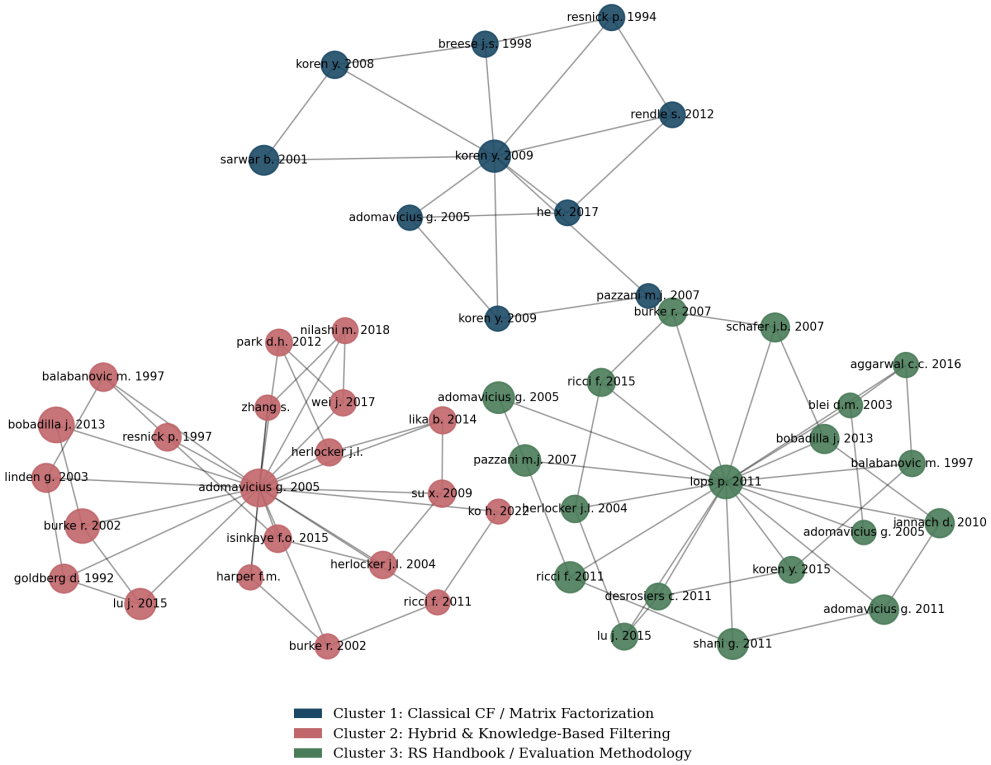

Figure 13: Co-citation network showing the field's three intellectual clusters; node size shows PageRank centrality.

a much larger population of one- or two-paper authors. In the corpus, most authors published one or just a few papers (roughly three-quarters published a single document; Figure 15) while a small fraction generated many. The estimated Lotka parameters are $\beta = 2.46$ and $C = 0.5278$, with $R^2 = 0.9516$; the Kolmogorov-Smirnov test returned a p-value of 0.22, so the fitted distribution is not rejected at the 5% significance level. As noted in Section 5, authors are identified by surname and initials, so the longest tail of the distribution may include merged homonyms.

Bradford's Law of scattering (Bradford, 1934) is visible in the Zone 1-2-3 structure of Table 2; Figures 8 and 14: a compact nucleus of 37 Zone 1 journals yields about a third of total output (33.7%), a substantially larger Zone 2 population of 239 moderately productive sources yields a further third (33.1%), and a very long Zone 3 tail of 939 low-frequency, often highly specialised or conference-proceeding sources yields the remainder (33.3%), the textbook scattering signature of an applied, rapidly diversifying technical field.

Table 6 maps each field of the three-field plot to the bibliometric law it illustrates, and Figure 16 shows the resulting relationships between cited references, authors, and merged keywords.

DISCUSSION

Implications for Academic and Research Libraries

The bibliometric structure mapped above carries several implications for library and information practice. First, the prominence of deep-learning and hybrid terms alongside classical filtering (Section 6.5) suggests that library discovery layers and institutional repositories seeking personalized recommendation

should look beyond simple "users who borrowed this also borrowed" widgets toward hybrid, content-aware models able to exploit the rich semantic metadata (abstracts, subject headings, full text) that libraries already curate. Second, the visibility of NLP-related keywords (sentiment analysis, semantics) alongside core RS terms indicates growing attention to semantic and text-based techniques, which underpin any revived SDI service; whether they are ready for production use in library systems should be established through local pilots.

Third, the co-citation roadmap's emphasis on evaluation methodology (the *Recommender Systems Handbook* cluster) offers librarians and LIS researchers an established framework for benchmarking any locally developed or vendor-supplied recommendation feature, rather than relying on vendor-reported

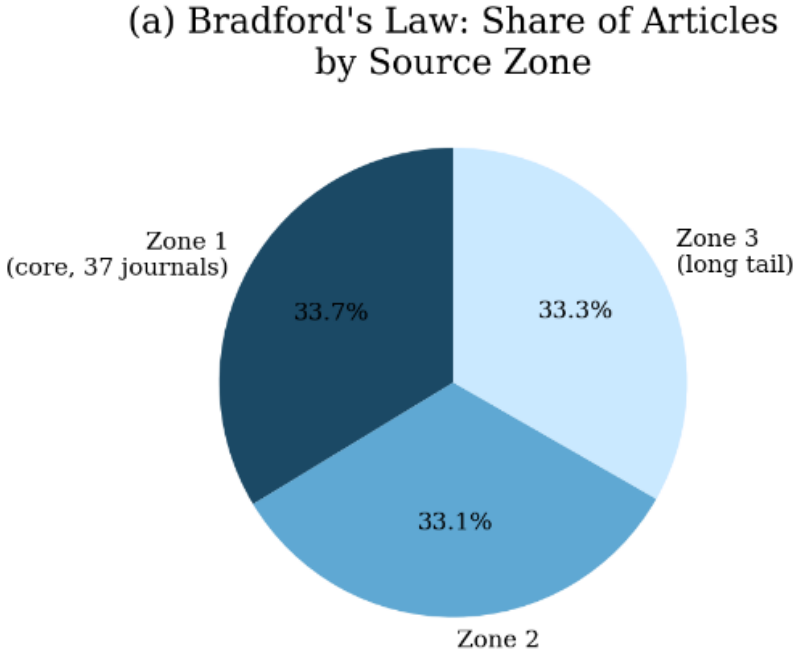

Figure 14: Bradford's Law: share of articles by source zone (Zone 1=core, 37 journals).

Table 5: Ten most-cited recommender-systems documents in the corpus.

Authors	Year	Title (abridged)	Source	Citations
Zhang, Yao, Sun, and Tay	2019	Deep learning-based recommender system: A survey and new perspectives	ACM Computing Surveys	2,563
Aggarwal	2016	Recommender Systems: The Textbook	Springer	1,413
Zhang and Chen	2020	Explainable recommendation: A survey and new perspectives	Foundations and Trends in Information Retrieval	866
Beel, Gipp, Langer, and Breitinger	2016	Research-paper recommender systems: A literature survey	International Journal on Digital Libraries	694
Ko, Lee, Park, and Choi	2022	A survey of recommendation systems	Electronics	668
and Cowan	2018	The use of machine learning algorithms in recommender systems	Expert Systems with Applications	654
Roy and Dutta	2022	A systematic review and research perspective on recommender systems	Journal of Big Data	600
Cui *et al.*	2020	Personalized recommendation system based on collaborative filtering for IoT	IEEE Transactions on Services Computing	544
Wang *et al.*	2022	A survey on session-based recommender systems	ACM Computing Surveys	535
Zhang, Lu, and Jin	2021	Artificial intelligence in recommender systems	Complex & Intelligent Systems	459

Note. Citation counts are those recorded in the Scopus export retrieved on 9 September 2026.

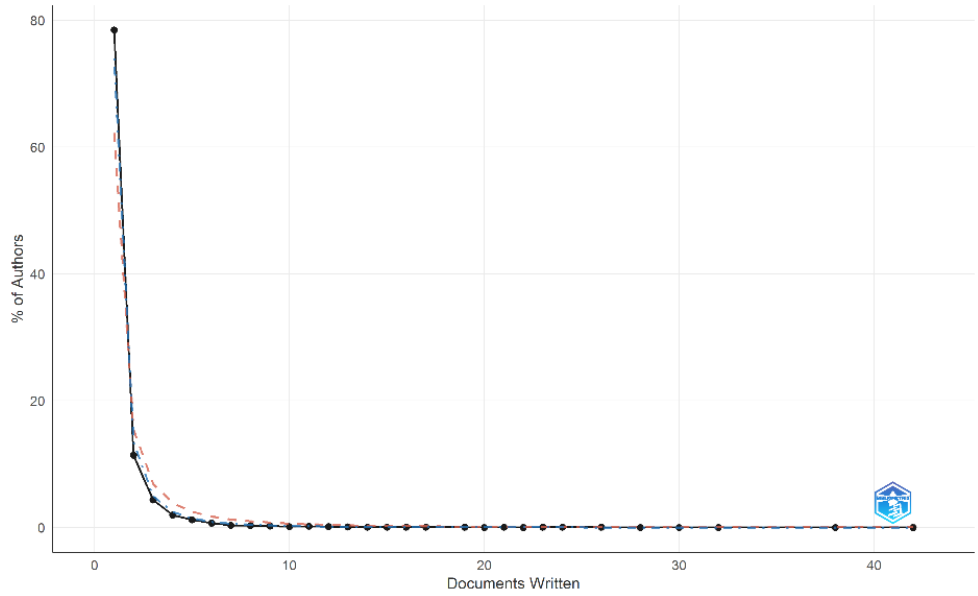

Figure 15: Lotka's Law fit to the author productivity distribution (estimated β = 2.46, C = 0.5278, R^2 = 0.9516).

accuracy claims alone (Shani and Gunawardana, 2011). Fourth, the concentration of RS research in computer-science and engineering venues (Section 6.2), with no LIS-specific outlet among the ten most productive sources, signals an opportunity and an obligation for LIS researchers to engage more with the technical RS literature, translating algorithmic advances into service models grounded in professional values such as intellectual freedom, privacy, and equitable access. This is especially pressing because personalization technologies raise concerns about algorithmic bias (Chen *et al.*, 2023), filter bubbles, and patron-data privacy, concerns that LIS ethics frameworks are well positioned to inform. Finally, the country-collaboration findings (Section 6.4) suggest that academic libraries in the United States, United Kingdom, and Australia, countries well represented in the global RS collaboration network, may be well positioned to form applied research partnerships with computer-science departments to pilot library-specific recommendation and SDI systems.

Operationally, these findings can be translated into a staged adoption pathway. In the near term, discovery-layer and repository platforms can adopt content-based and lightweight hybrid filtering approaches of the kind reported in the Zone 1 outlets (Section 6.2), using existing catalogue metadata as the content signal with minimal infrastructure investment. In the medium term, libraries with sufficient patron-interaction data (circulation, downloads, database sessions) can layer collaborative-filtering signals onto this content-based baseline, consistent with the layered architecture surveyed in the most-cited works (Table 5). In the longer term, NLP-based semantic profiling of documents and patron queries can support a revived, automated SDI service that alerts researchers to newly indexed material matching dynamically updated interest profiles rather than static keyword lists, operationalizing the SDI vision of Luhn (1961) with contemporary neural and transformer-based tools. Throughout this pathway, the co-citation-derived evaluation literature (Section 6.6) offers libraries an established set of metrics (precision, recall, novelty, diversity, and user satisfaction) with which to assess any locally built or vendor-procured system (Shani and Gunawardana, 2011), guarding against adoption decisions driven by marketing claims rather than evidence.

Originality and Value

This review's principal contribution is methodological and integrative rather than purely descriptive. To the authors' knowledge, it is one of few studies to apply a full bibliometric and scientometric toolkit-publication trends, Bradford's Law, Lotka's Law, country collaboration analysis, thematic mapping, and co-citation analysis-to a Scopus cluster search explicitly designed to bridge recommender-systems engineering literature with LIS service constructs (SDI, personalized reference, academic libraries). Where prior RS literature reviews (for example, Zhang *et al.*, 2019; Roy and Dutta, 2022) map the field from a computer-science standpoint, this review's search strategy and discussion explicitly connect that literature to library practice, offering a template that future LIS-oriented bibliometric studies of AI-driven services can replicate.

Limitations

This review is subject to several limitations. First, reliance on a single database (Scopus), while standard for large-scale bibliometric work, excludes relevant records indexed only in Web of Science, Google Scholar, or specialised LIS databases such as LISTA. Second, the compound Boolean cluster query, while designed to capture the intersection of technical and library-service vocabulary, may under-represent purely LIS-authored work that does not use computer-science terminology. Third, co-citation and keyword co-occurrence analyses are sensitive to keyword

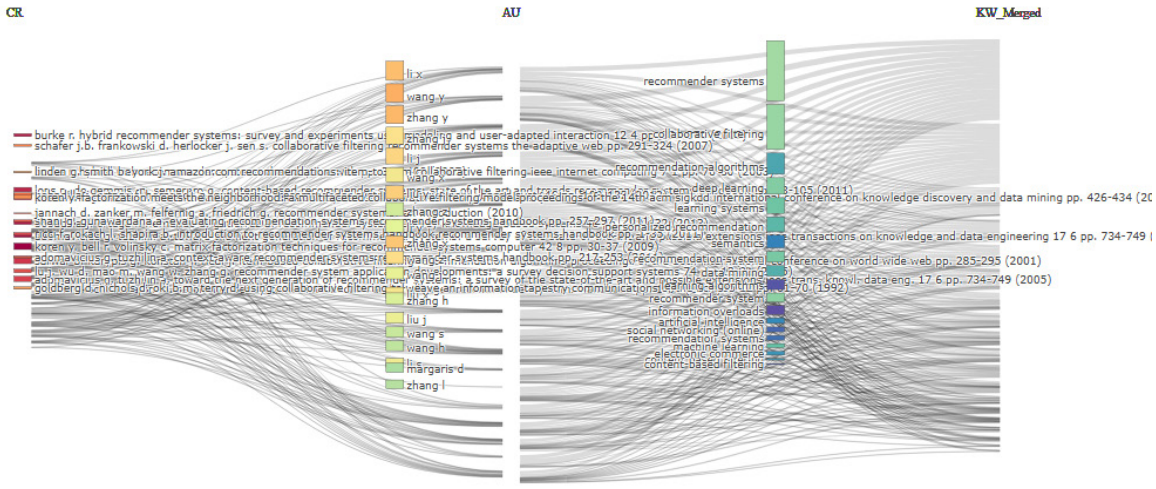

Figure 16: Three-field plot illustrating the relationships between cited references (left), authors (middle), and merged keywords (right) in the analysed literature.

Table 6: Mapping of the three-field plot to the bibliometric laws applied.

Field position	Map variable	Applicable bibliometric law	Core concept
Left field	Cited References (CR)	Bradford-type concentration (Bradford's Law of Scattering itself concerns journals)	A small core of references or journals accounts for a disproportionate share of citations or articles; in Bradford's formulation, a core of journals yields about as many articles as each of the successive, larger zones
Middle Field	Principal Authors (AU)	Lotka's Law of Scientific Productivity	A small nucleus of highly prolific authors produces a disproportionately large share of the literature.
Right Field	Merged Keywords (KW_Merged)	Zipf's Law of Word Occurrence	The frequency of a word (or keyword) is inversely proportional to its rank in the literature

normalisation choices (for example, merging "recommender system" and "recommender systems"), and while default Biblioshiny settings were used for consistency, alternative thresholds could yield different cluster boundaries. Fourth, author-level analyses rely on surname-and-initial identification and may merge homonymous authors. Finally, citation counts are time-sensitive and favour earlier-published works, so the most recent (2023-2025) contributions are under-represented in the highly-cited-document analysis (Section 6.7) despite their thematic importance.

CONCLUSION

This bibliometric and scientometric review of 2,991 Scopus-indexed documents (2016-2025) shows that recommendation-systems research is a large, rapidly growing field with prominent deep-learning vocabulary, concentrated in a compact nucleus of applied computer-science journals and anchored intellectually in a well-established co-citation lineage running from classical collaborative filtering through hybrid architectures to evaluation methodology. For academic and research libraries, the findings chart an evidence-based path: the breadth of hybrid, NLP-based, and neural methods documented here offers a sound technical basis for a new generation of personalized reference and SDI services, provided that library and information science researchers and practitioners engage more directly with this technical literature and evaluate systems locally. Future bibliometric work should extend this analysis beyond 2025 and complement it with qualitative content analysis of library-specific recommender-system implementations, to track how, and how well, these technical advances are translated into practice.

ABBREVIATIONS

AI: Artificial Intelligence; **ANN:** Artificial Neural Networks; **CBF:** Content-Based Filtering; **CF:** Collaborative Filtering; **LIS:** Library and Information Science; **ML:** Machine Learning; **NLP:** Natural Language Processing; **PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses; **RQ:** Research Question;

RS: Recommender Systems; **SDI:** Selective Dissemination of Information.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering, 17*(6), 734-749. https://doi.org/10.1109/TKDE.2005.99

Adomavicius, G., Manouselis, N., & Kwon, Y. (2011). Multi-criteria recommender systems. In *Recommender Systems Handbook* (pp. 769-803). Springer US. https://doi.org/10.1007/978-0-387-85820-3_24

Aggarwal, C. C. (2016). *Recommender systems*. Springer International Publishing. https://doi.org/10.1007/978-3-319-29659-3

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

Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2016). Research-paper recommender systems: A literature survey. *International Journal on Digital Libraries, 17*(4), 305-338. https://doi.org/10.1007/s00799-015-0156-0

Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering, 137*(3550), 85-86.

Breese, J. S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence* (pp. 43-52). Morgan Kaufmann.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction, 12*(4), 331-370. https://doi.org/10.1023/A:1021240730564

Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. (2023). Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*, *41*(3), 1-39. https://doi.org/10.1145/3564284

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). 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. https://doi.org/10.1002/asi.21525

Cui, Z., Xu, X., Xue, F., Cai, X., Cao, Y., Zhang, W., & Chen, J. (2020). Personalized recommendation system based on collaborative filtering for IoT scenarios. *IEEE Transactions on Services Computing, 13*(4), 685-695. https://doi.org/10.1109/TSC.2020.2964552

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM, 35*(12), 61-70. https://doi.org/10.1145/138859.138867

Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). *Recommender systems*. Cambridge University Press. https://doi.org/10.1017/cbo9780511763113

Ko, H., Lee, S., Park, Y., & Choi, A. (2022). A survey of recommendation systems: Recommendation models, techniques, and application fields. *Electronics*, *11*(1), 141. https://doi.org/10.3390/electronics11010141

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer, 42*(8), 30-37. https://doi.org/10.1109/MC.2009.263

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences, 16*(12), 317-323.

Luhn, H. P. (1961). Selective dissemination of new scientific information with the aid of electronic processing equipment. *American Documentation, 12*(2), 131-138. https://doi.org/10.1002/asi.5090120209

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71 PMID: 33782057

Pazzani, M. J., & Billsus, D. (n.d.). Content-based recommendation systems. In *Lecture Notes in Computer Science* (pp. 325-341). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72079-9_10

Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications, 97*, 205-227. https://doi.org/10.1016/j.eswa.2017.12.020

Price, D. J. (1963). *Little science, big science*. Columbia University Press. https://doi.org/10.7312/pric91844

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation, 25*(4), 348-349.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens. *Proceedings of the 1994 ACM conference on Computer supported cooperative work- CSCW '94*, 175-186. https://doi.org/10.1145/192844.192905

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender Systems Handbook* (pp. 1-35). Springer US. https://doi.org/10.1007/978-0-387-85820-3_1

Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, *9*(1), 59. https://doi.org/10.1186/s40537-022-00592-5

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th international conference on World Wide Web*, 285-295. https://doi.org/10.1145/371920.372071

Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems. In *Recommender Systems Handbook* (pp. 257-297). Springer US. https://doi.org/10.1007/978-0-387-85820-3_8

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265-269. https://doi.org/10.1002/asi.4630240406

Wang, S., Cao, L., Wang, Y., Sheng, Q. Z., Orgun, M. A., & Lian, D. (2022). A survey on session-based recommender systems. *ACM Computing Surveys*, *54*(7), 1-38. https://doi.org/10.1145/3465401

Xi, Y., Liu, W., Lin, J., Cai, X., Zhu, H., Zhu, J., Chen, B., Tang, R., Zhang, W., & Yu, Y. (2024). Towards open-world recommendation with knowledge augmentation from large language models. *18th ACM Conference on Recommender Systems*, 12-22. https://doi.org/10.1145/3640457.3688104

Zhang, Q., Lu, J., & Jin, Y. (2021). Artificial intelligence in recommender systems. *Complex & Intelligent Systems*, *7*(1), 439-457. https://doi.org/10.1007/s40747-020-00212-w

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system. *ACM Computing Surveys*, *52*(1), 1-38. https://doi.org/10.1145/3285029

Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. *Foundations and Trends in Information Retrieval, 14*(1), 1-101. https://doi.org/10.1561/1500000066

Cite this article: Dey S, Mondal P. Mapping the Intellectual Landscape of Recommendation Systems Research for Academic and Research Libraries: A Scientometric Review. Journal of Data Science, Informetrics, and Citation Studies. 2026;5(3):237-53.