



Recommendation Systems in Scholarly Publishing

Yannis Manolopoulos
Open University of Cyprus

15th IC3K Conference

November 2023, Rome

Presentation Menu

- **Recommendation Systems**
 - Information age and Social networks
 - Challenges and Data representation
- **Scientometric Entities**
 - Abundance of Scholarly Data
 - Taxonomy
 - Statistics
- **Approaches on the Intersection**
 - Citation
 - Supervisor
 - Journal
 - Venue
- **Bias in Recommenders**
- **Further Reading**

Information Age and Social Networks





- Plethora of available information online
- Increase in web traffic, new domains
- Emergence of social networks
- Emergence of recommender systems



How the Community Started







3rd RecSys 2009: New York, NY, USA

-     Lawrence D. Bergman, Alexander Tuzhilin, Robin D. Burke, Alexander Felfernig, Lars Schmidt-Thieme:
Proceedings of the 2009 ACM Conference on Recommender Systems, RecSys 2009, New York, NY, USA, October 23-25, 2009. ACM 2009, ISBN 978-1-60558-435-5 [contents]





82

2nd RecSys 2008: Lausanne, Switzerland

-     Pearl Pu, Derek G. Bridge, Bamshad Mobasher, Francesco Ricci:
Proceedings of the 2008 ACM Conference on Recommender Systems, RecSys 2008, Lausanne, Switzerland, October 23-25, 2008. ACM 2008, ISBN 978-1-60558-093-7 [contents]

47

1st RecSys 2007: Minneapolis, MN, USA

-     Joseph A. Konstan, John Riedl, Barry Smyth:
Proceedings of the 2007 ACM Conference on Recommender Systems, RecSys 2007, Minneapolis, MN, USA, October 19-20, 2007. ACM 2007, ISBN 978-1-59593-730-8 [contents]

38



How the Community Evolved

year	city	days	papers	volumes
2023	Singapore	5	197	2*
2022	Seattle	6	125	7
2021	Amsterdam	5	134	7
2020	Brazil	5	132	7
2019	Copenhagen	5	122	10
2018	Vancouver	6	102	7
2017	Como	5	96	10
2016	Boston	5	105	8
2015	Vienna	5	91	8
2014	Silicon Valley	5	86	4
2013	Hong-Kong	5	98	4
2012	Dublin	5	79	4
2011	Chicago	5	73	2
2010	Barcelona	5	78	3
2009	New York	3	82	1
2008	Lausanne	3	47	1
2007	Minneapolis	2	38	1

Data from DBLP.org

#papers in main volume of ACM proceedings

DIAMOND SUPPORTER



PLATINUM SUPPORTER



GOLD SUPPORTER



SILVER SUPPORTER



BRONZE SUPPORTER



CHALLENGE SPONSOR



SPECIAL SUPPORTERS



Why to Use Recommenders

Value for the customer

- Find products/items that are interesting
- Narrow down the set of choices
- Help customer explore the space of options

Value for the provider

- Personalized service for the customer
- Increase trust and customer loyalty
- Opportunities for promotion, persuasion

Industrial Implementations

Enterprise

- Google Search
- Netflix, YouTube
- Google News
- IMDb
- Last.fm
- Amazon.com
- Facebook
- Tripadvisor, Booking

Provision

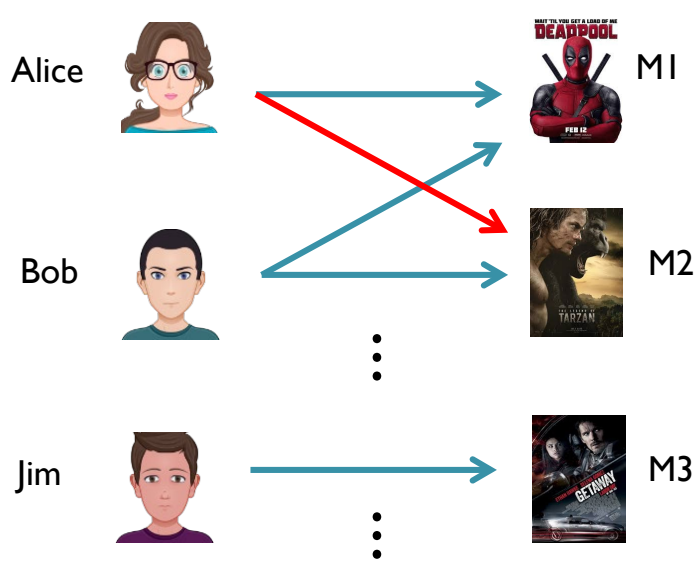
- Advertisements
- Video
- News
- Movies
- Music
- E-commerce
- Friends, Advertisements
- Travel products



Goal: *predict* how much a consumer will *like* a product/item

Challenges

- Recommenders rely on collaborative filtering
- Similar-minded users tend to get similar recommendations












Data sparsity:

users might not have common preferences with others

Cold start: users/items with no history record

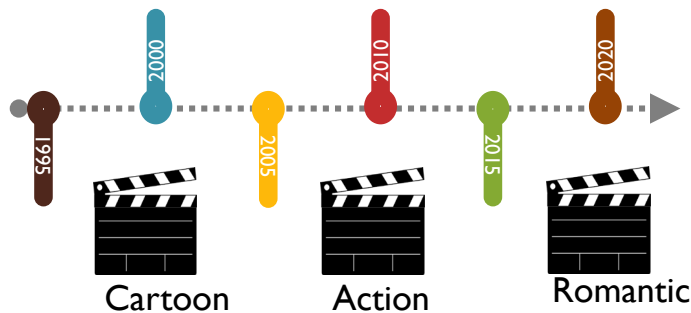
Challenges

- **User bias:**
 - a user can **criticize** items, assigning **low** ratings in general
 - a user can be **generous** and assign **high** ratings
- **Item bias:**
 - an item might get **low** or **high** ratings

						b_i
	4	4	?	3	5	4
	?	2	3	?	?	2.5
	5	3	?	?	?	4
	?	2	?	?	1	1.5
b_j	4.5	2.75	3	3	3	

Challenges

Long-term preferences



Short-term preferences

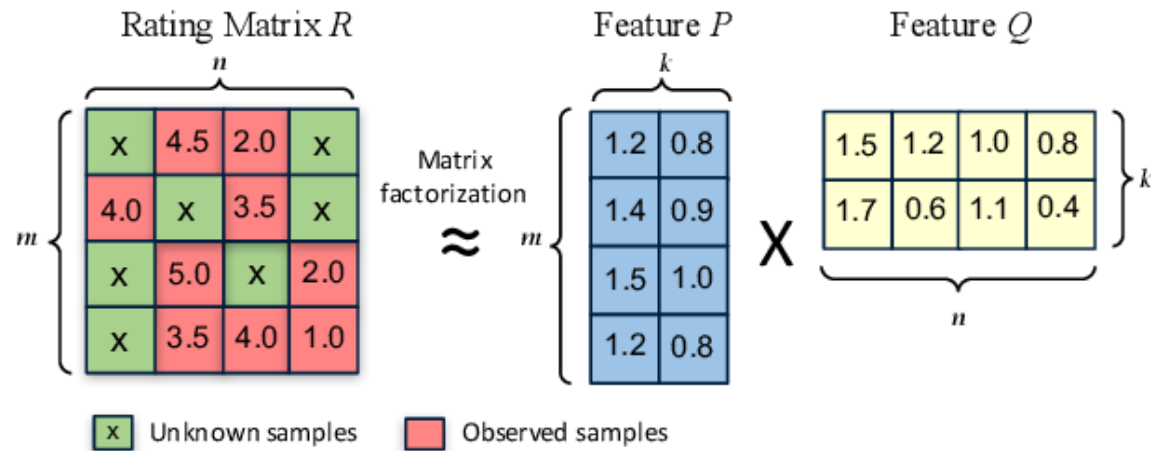


Capture users' time evolving preferences

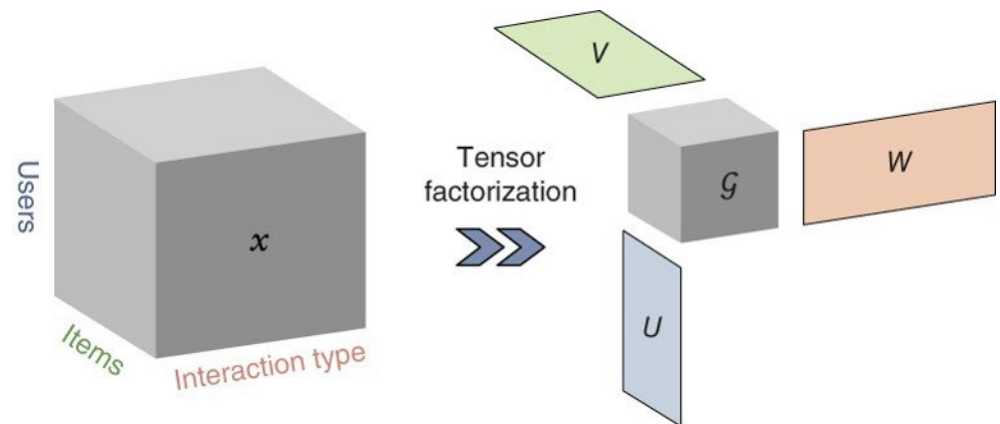


Data Representation

- Matrix-based
- Tensor-based
- Graph-based
- Hybrid



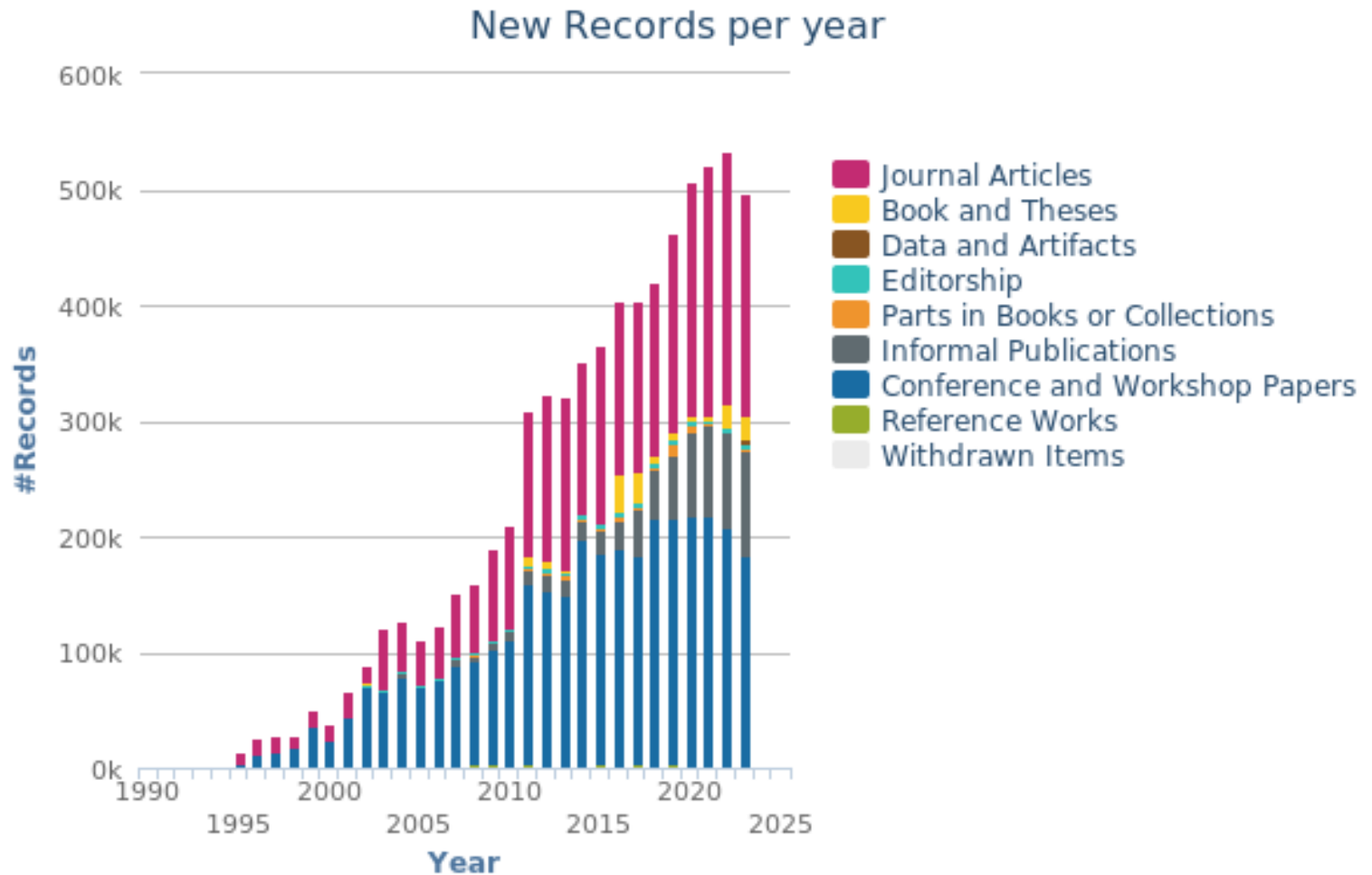
- **Matrix factorization**
- **Dimensionality reduction**



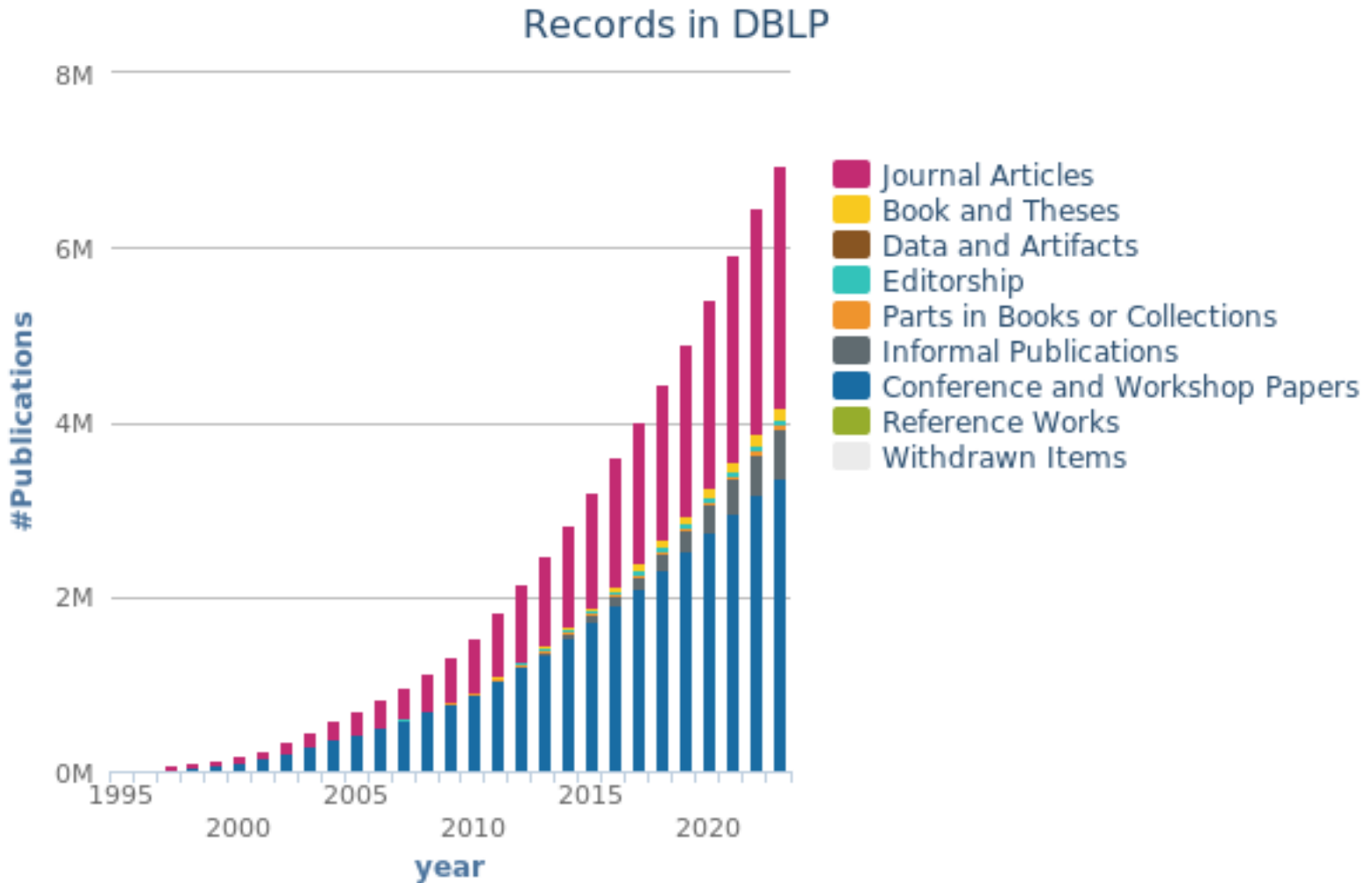
Presentation Menu

- Recommendation Systems
 - Information age and Social networks
 - Challenges and Data representation
- **Scientometric Entities**
 - Abundance of Scholarly Data
 - Taxonomy
 - Statistics
- Approaches on the Intersection
 - Citation
 - Supervisor
 - Journal
 - Venue
- Bias in Recommenders
- Further Reading

Abundance of Bibliographic Data



Abundance of Bibliographic Data



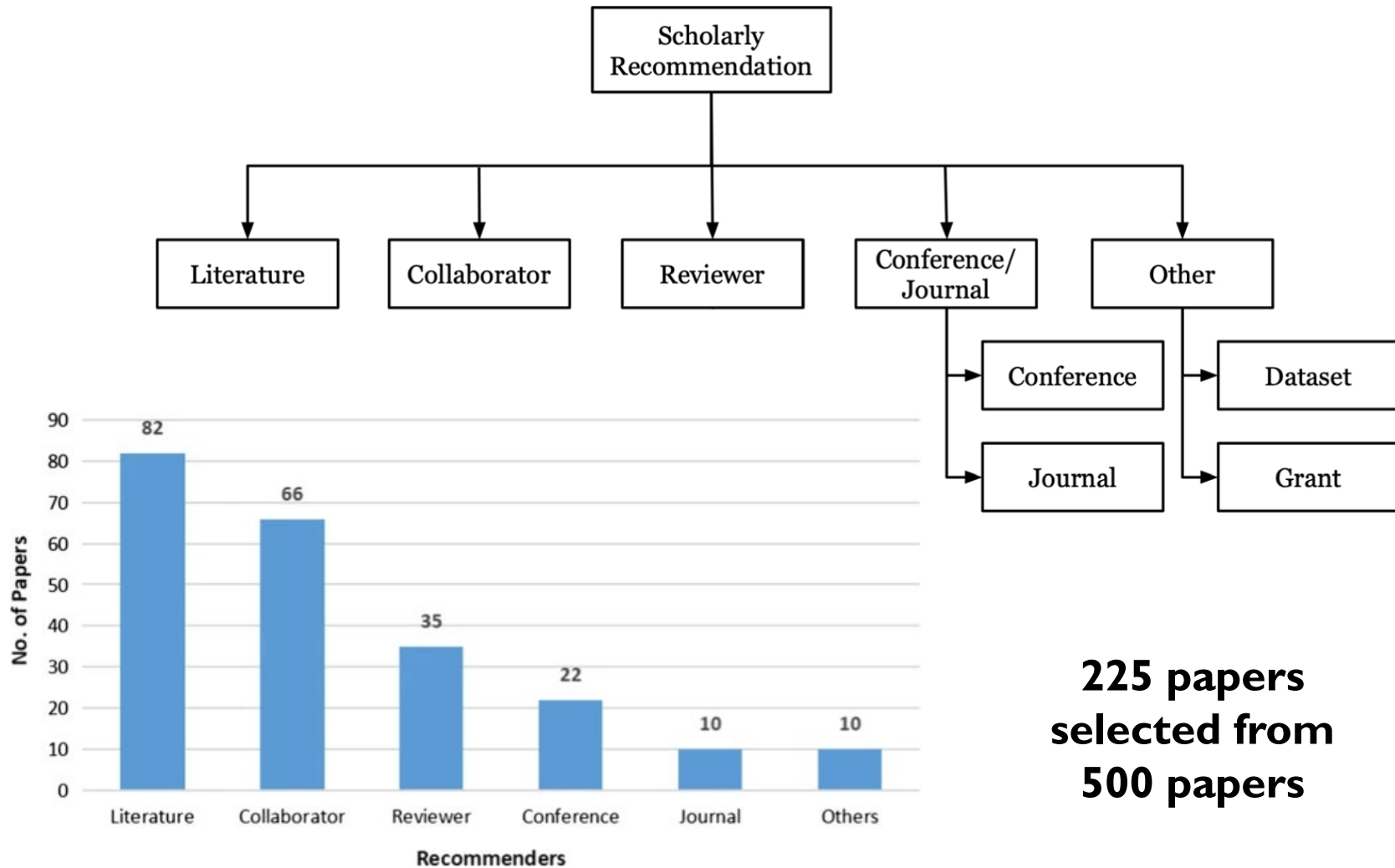
Abundance of Bibliographic Data

- “The rapid growth of scientific publications brings the problem of finding appropriate citations for authors”
- “The increase in the number of scholarly journals has made it difficult for researchers to choose the correct journal for publishing their articles”
- “Choosing a suitable academic venue for publishing one’s research can represent a challenging task considering the plethora of available conferences”
- “With the explosive growth of patent applications, how to recommend relevant patents from the massive number of patents has become an extremely challenging problem”

Abundance of Bibliographic Data

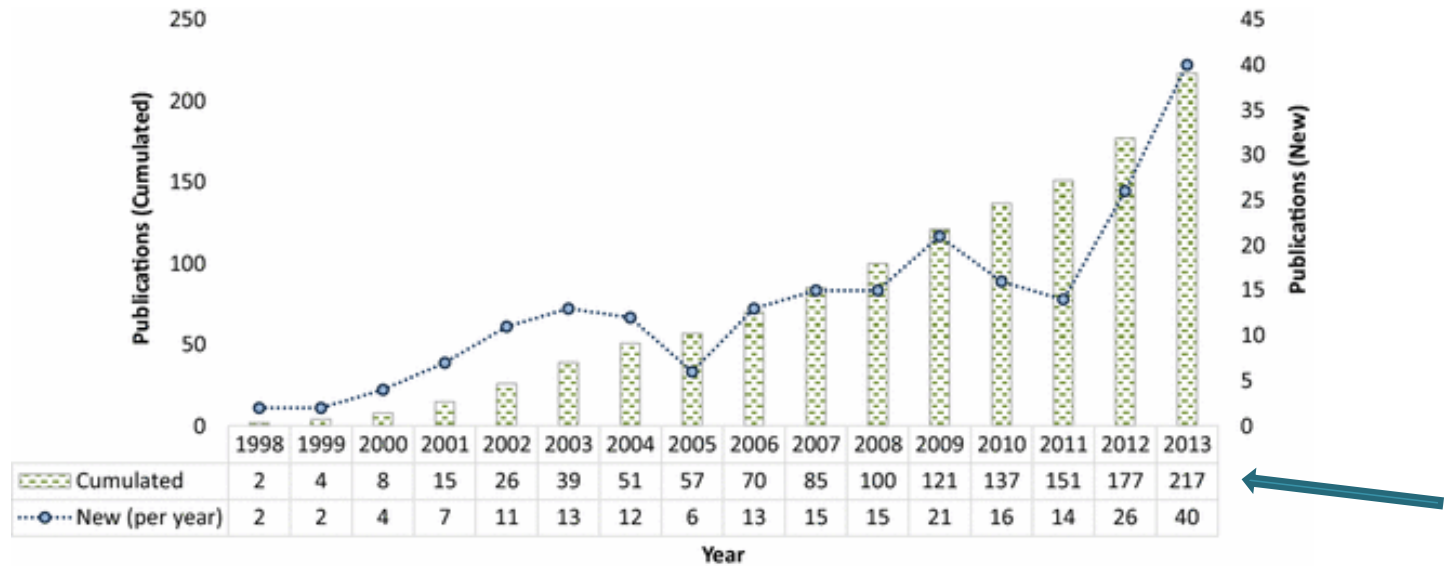
- “Finding the most suitable co-author is one of the most important ways to conduct effective research”
- “With the increasing number of scientific papers, it is difficult for researchers to locate the most relevant and important keywords from the vast majority of papers and establish the research focus and preliminaries”
- “With the world of academia growing at a tremendous rate, we have an enormous number of researchers on hosts of research topics”
- “The steadily rising number of datasets is making it increasingly difficult for researchers and practitioners to be aware of all datasets”

A Taxonomy of Scientometric Entities



Zhang Z., Patra B.G., Yaseen A., Zhu J., Sabharwal R., Roberts K., Cao T. and Wu H.: "Scholarly recommendation systems: A literature survey", *Knowledge and Information Systems*, 2023.

Production over the Years



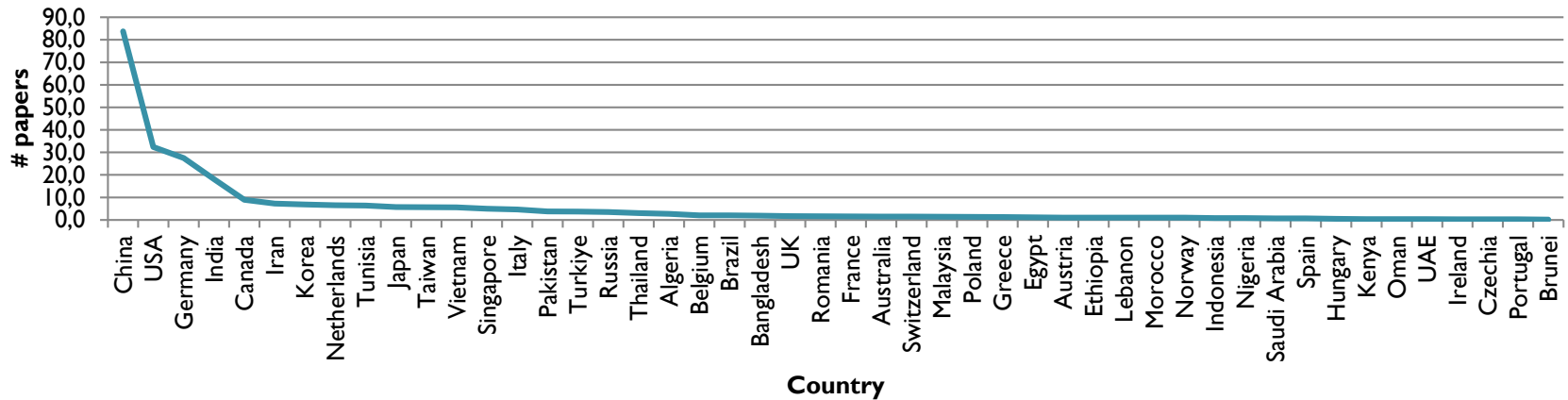
2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
1		2		5			7	11	17	22	25	36	64	79	87	97	110	135	156	191	227	257	270
1		1		3			2	4	6	5	3	11	28	15	8	10	13	25	21	35	36	30	13

My collection : 270 papers

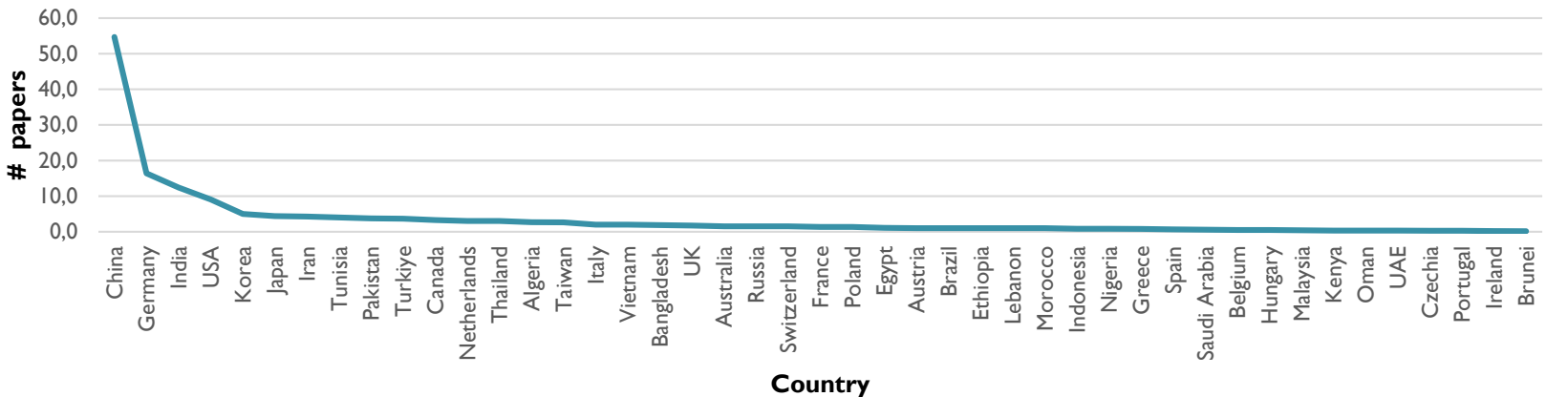
Beel J., Gipp B., Langer S. and Breitinger C.: "Research-paper recommender systems: A literature survey", *International Journal on Digital Libraries*, 2016.

Production per Country

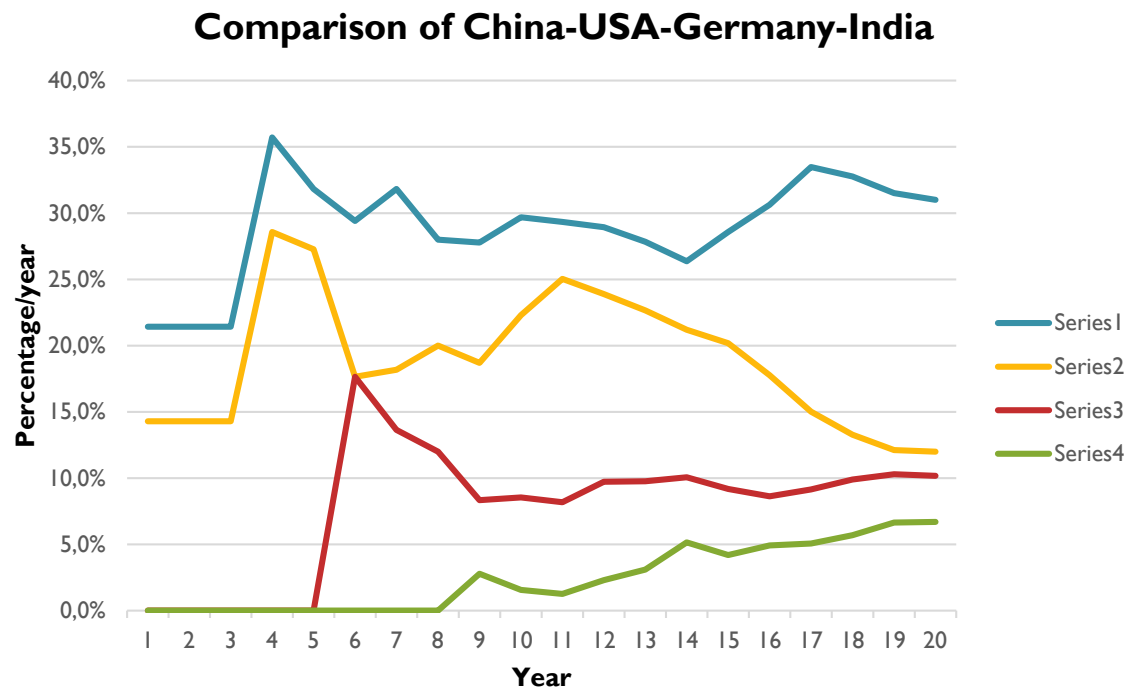
Production/country during 2010-2023 (270 papers)



Production/country during 2018-2023 (160 papers)



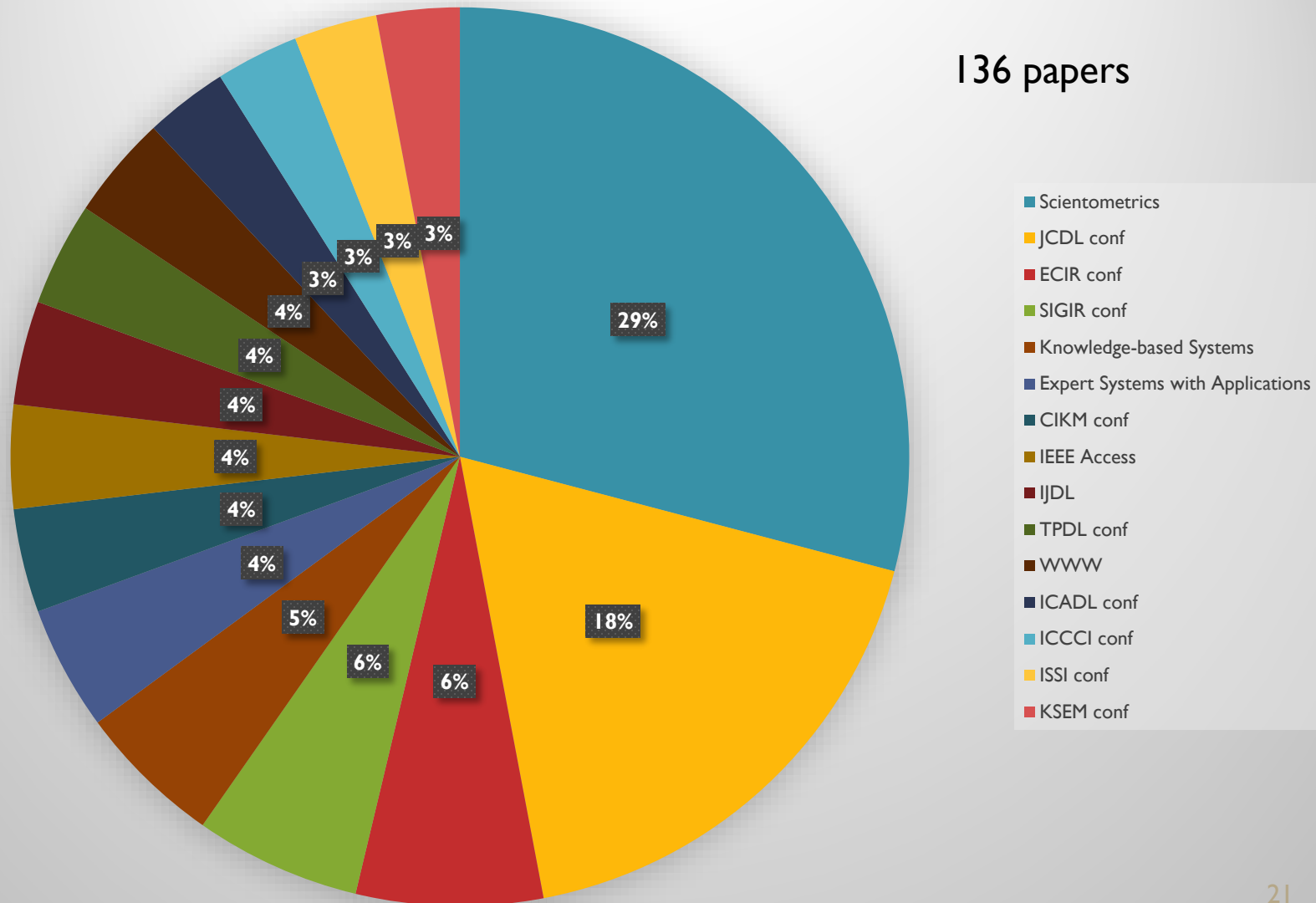
Production per Country



Main Outlets

Chart Title

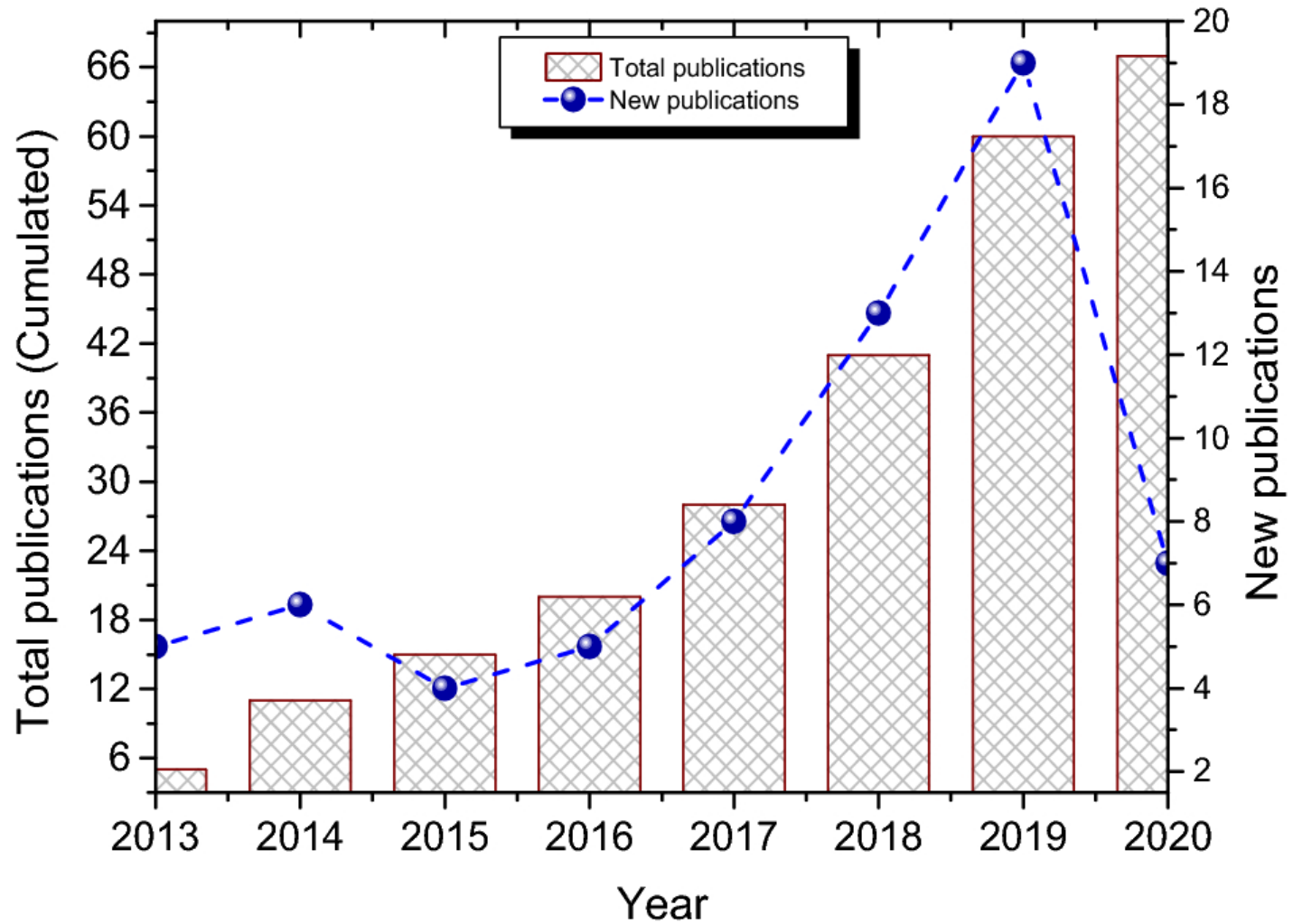
136 papers



Presentation Menu

- Recommendation Systems
 - Information age and Social networks
 - Challenges and Data representation
- Scientometric Entities
 - Abundance of Scholarly Data
 - Taxonomy
 - Statistics
- **Approaches on the Intersection**
 - Citation
 - Supervisor
 - Journal
 - Venue
- Bias in Recommenders
- Further Reading

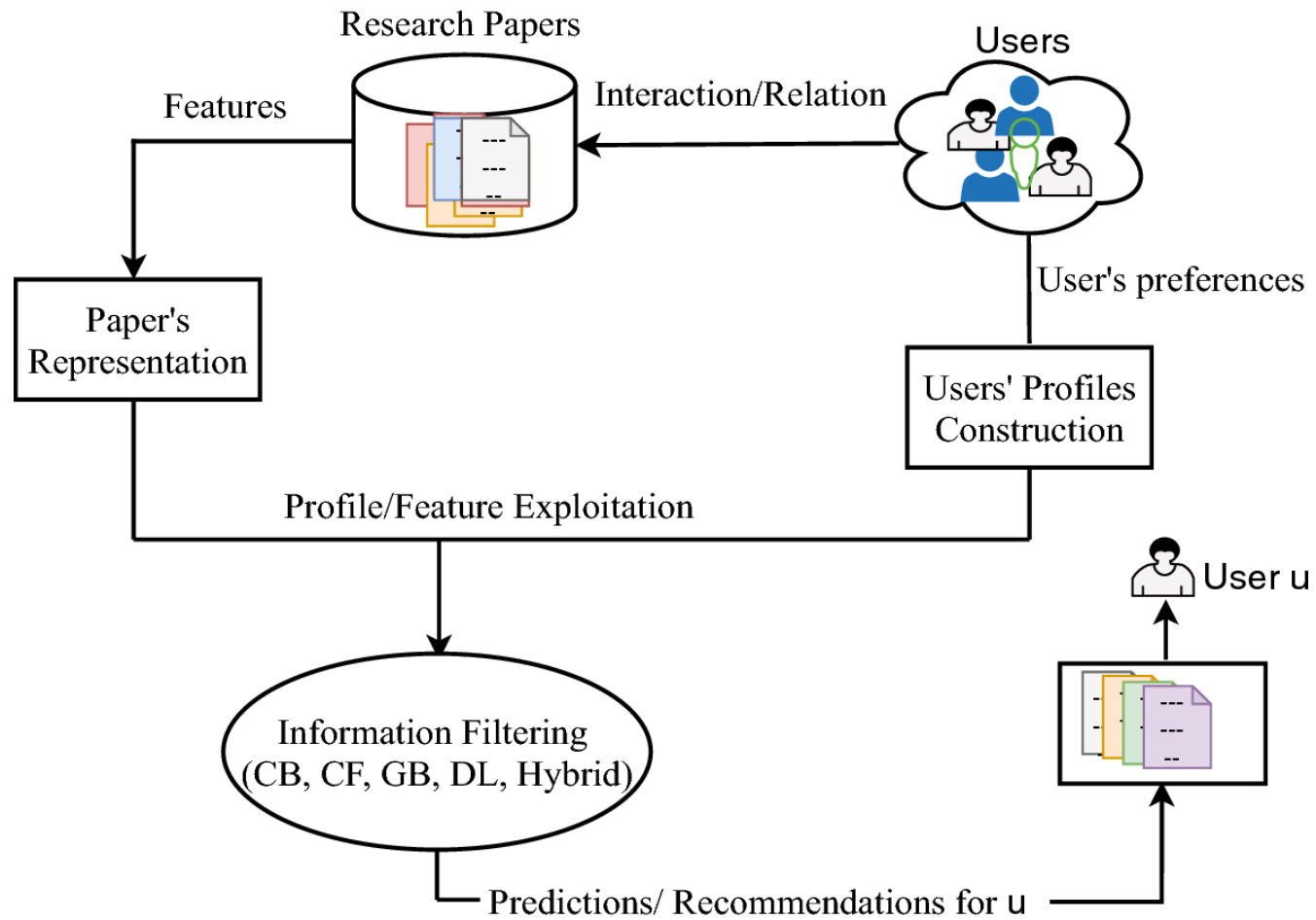
Recommending Citations



Ali Z., Ullah I., Khan A., Jan A.U. and Muhammad K.:“An overview and evaluation of citation recommendation models”, *Scientometrics*, 2021.

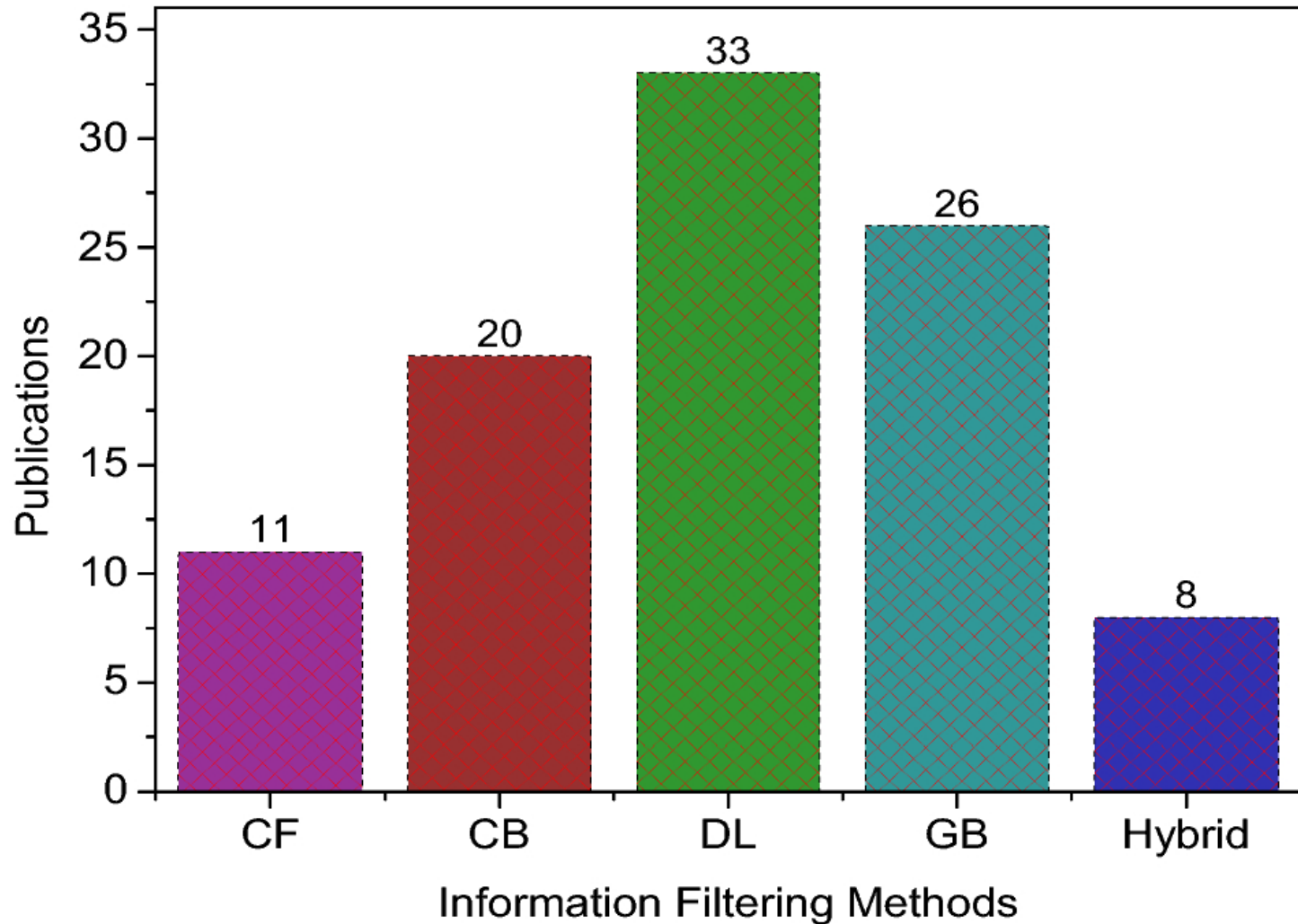
Recommending Citations

Scientometrics



Ali Z., Ullah I., Khan A., Jan A.U. and Muhammad K.: "An overview and evaluation of citation recommendation models", *Scientometrics*, 2021.

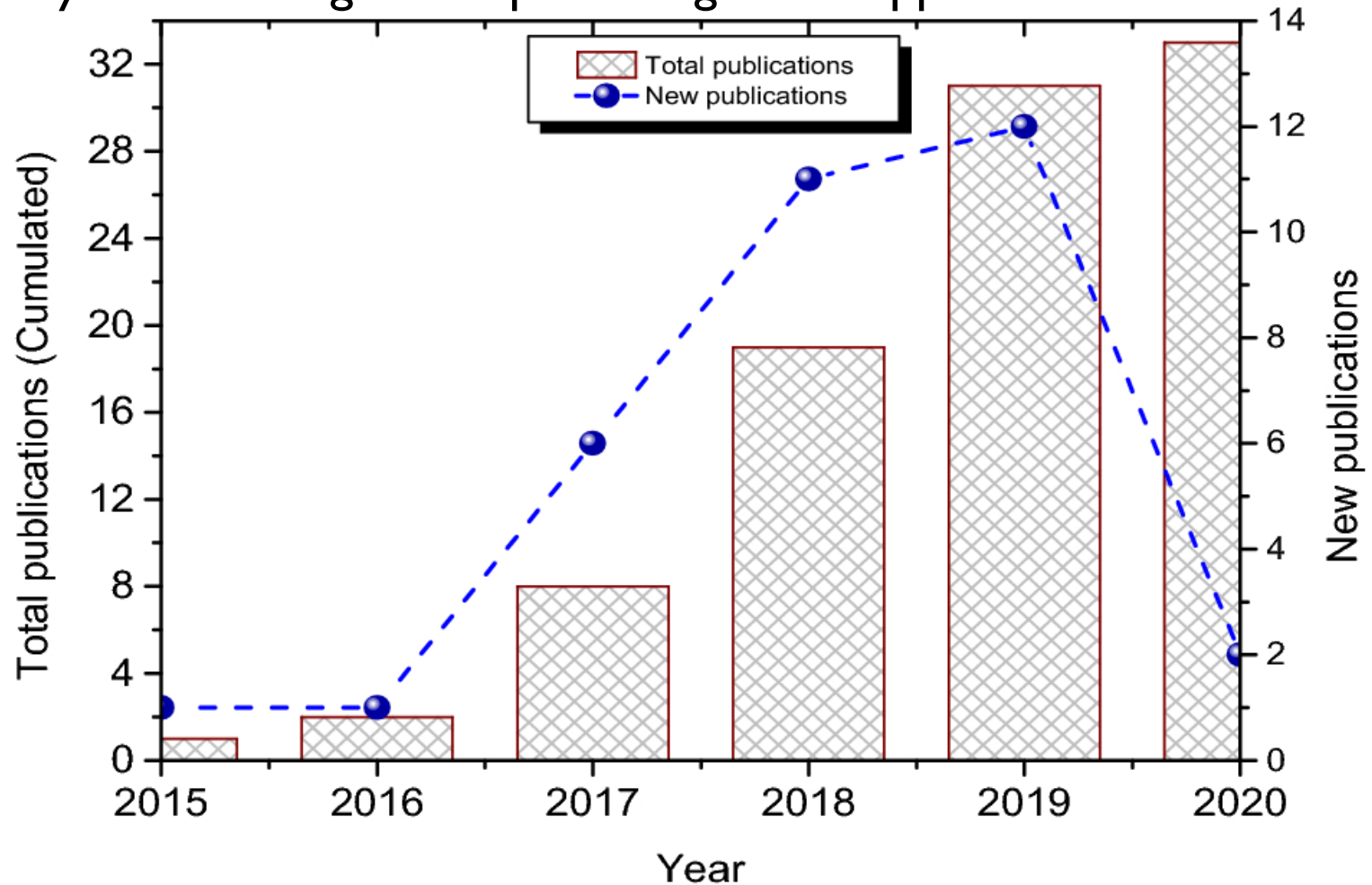
Recommending Citations



Ali Z., Ullah I., Khan A., Jan A.U. and Muhammad K.:“An overview and evaluation of citation recommendation models”, *Scientometrics*, 2021.

Recommending Citations

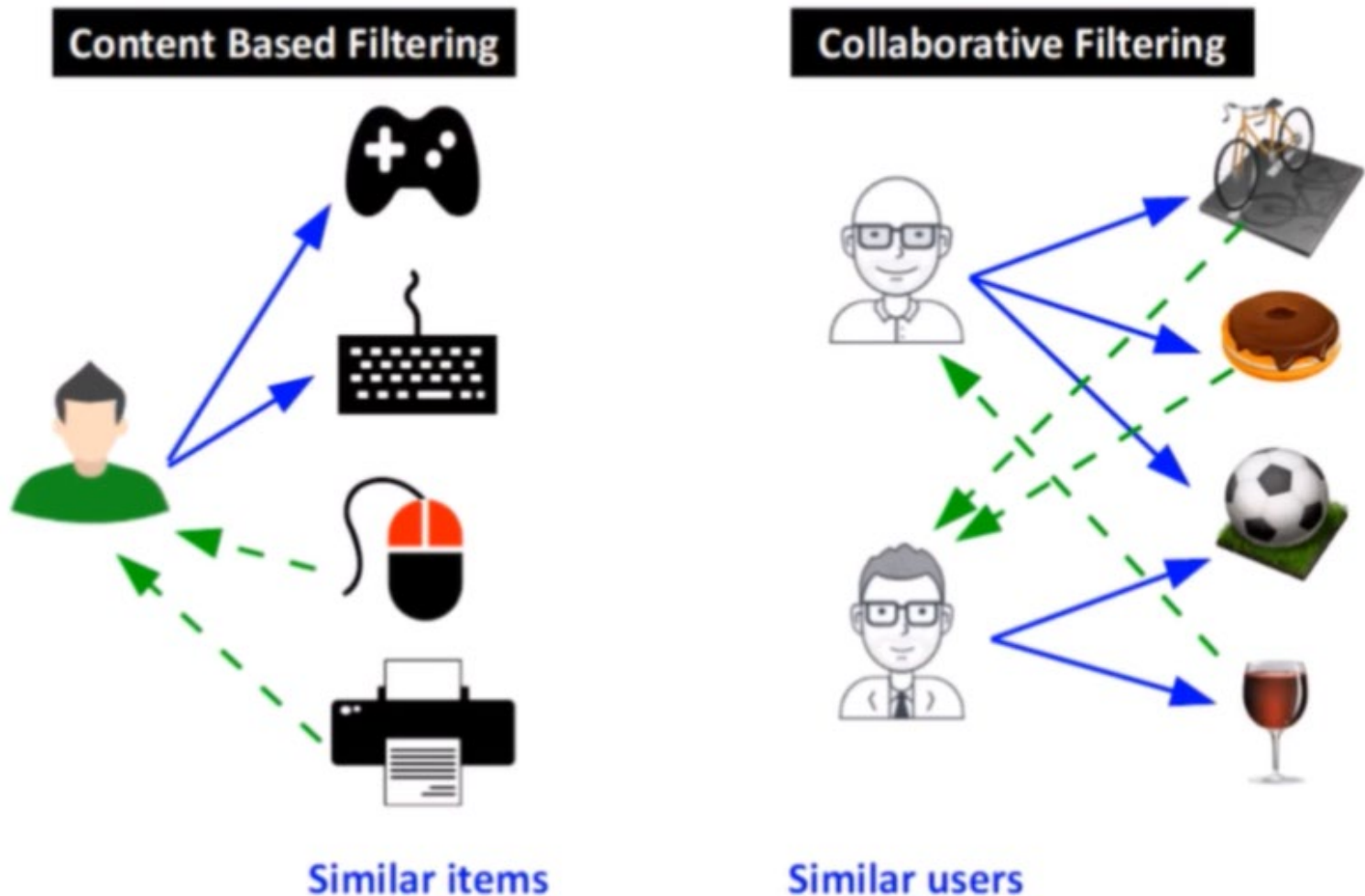
- Key methodologies: deep learning based approaches.



Ali Z., Ullah I., Khan A., Jan A.U. and Muhammad K.:“An overview and evaluation of citation recommendation models”, *Scientometrics*, 2021.

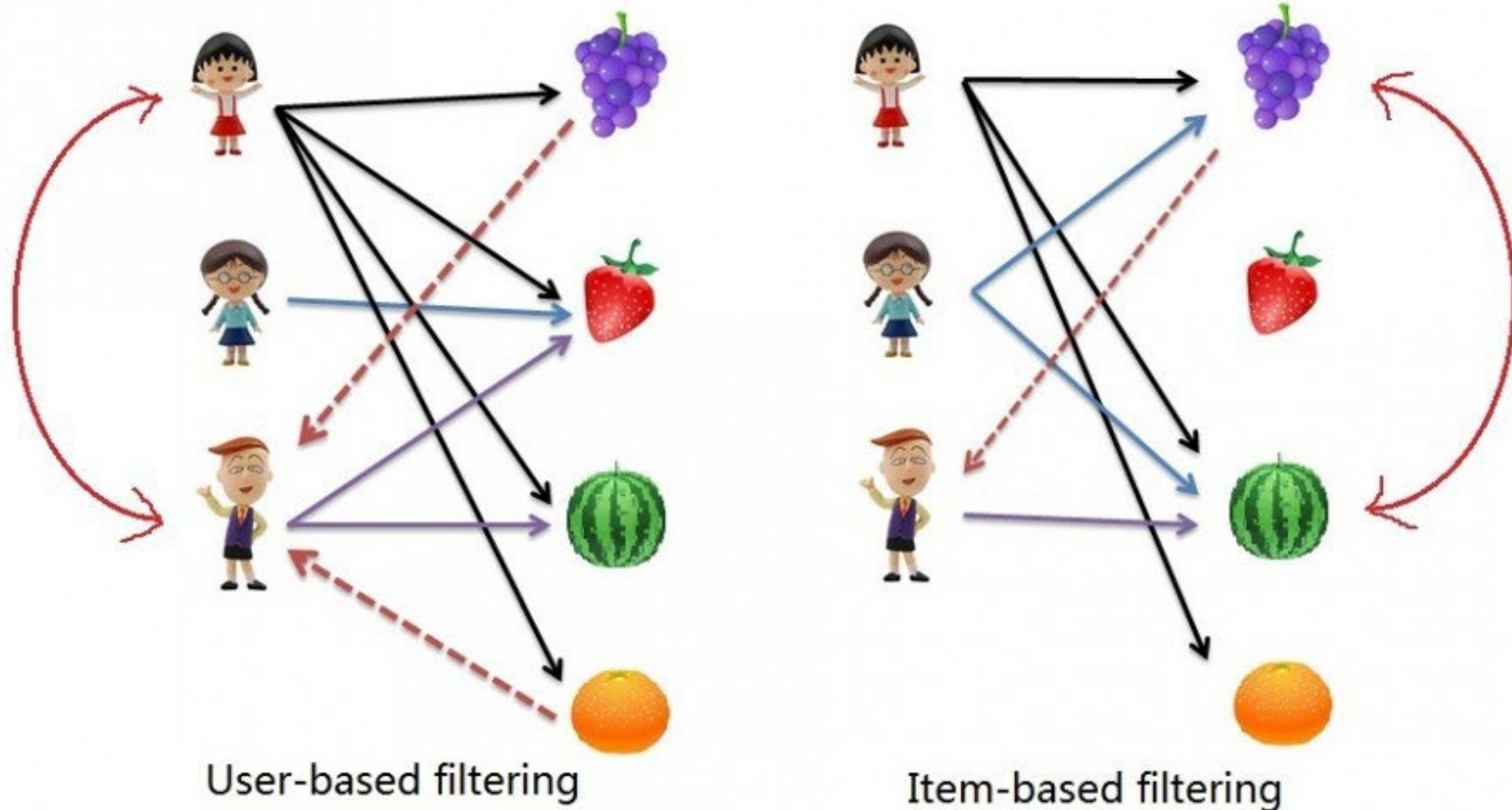
Recommending Citations

- Key methodologies: Collaborative filtering vs Content-based filtering



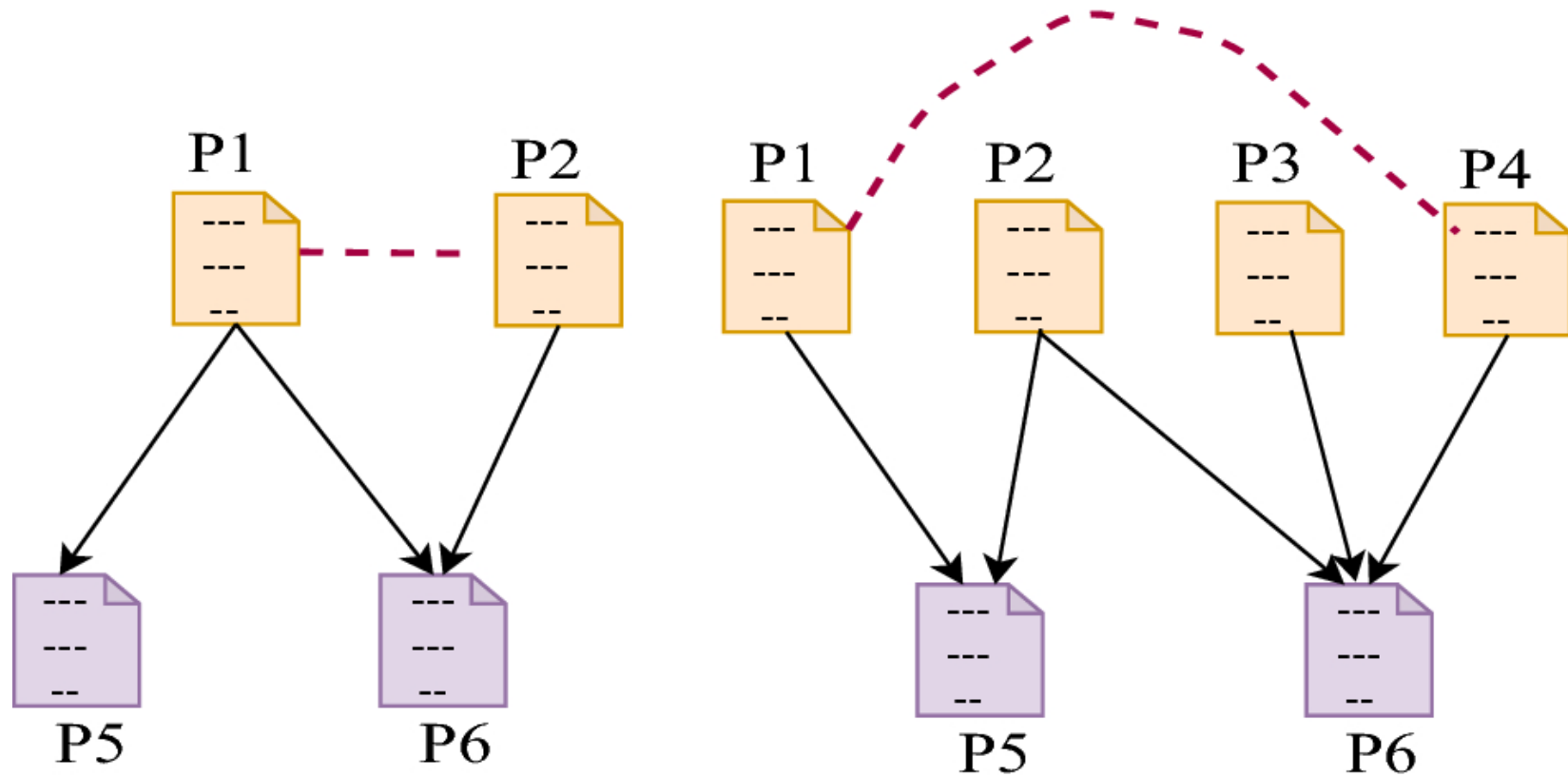
Recommending Citations

- Key methodologies: user-item filtering vs item-item filtering



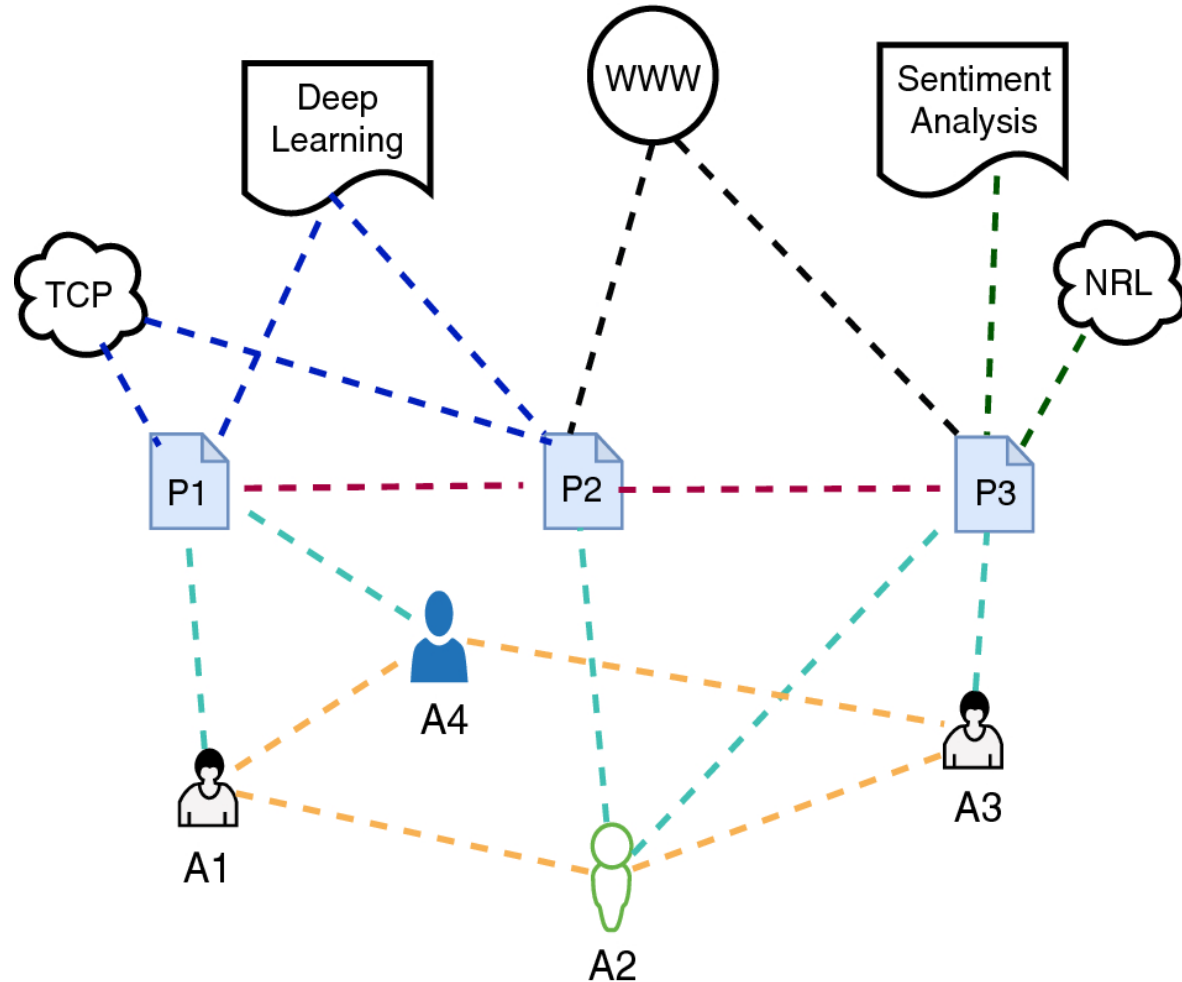
Recommending Citations

- Key methodologies: Similarity calculation



Recommending Citations

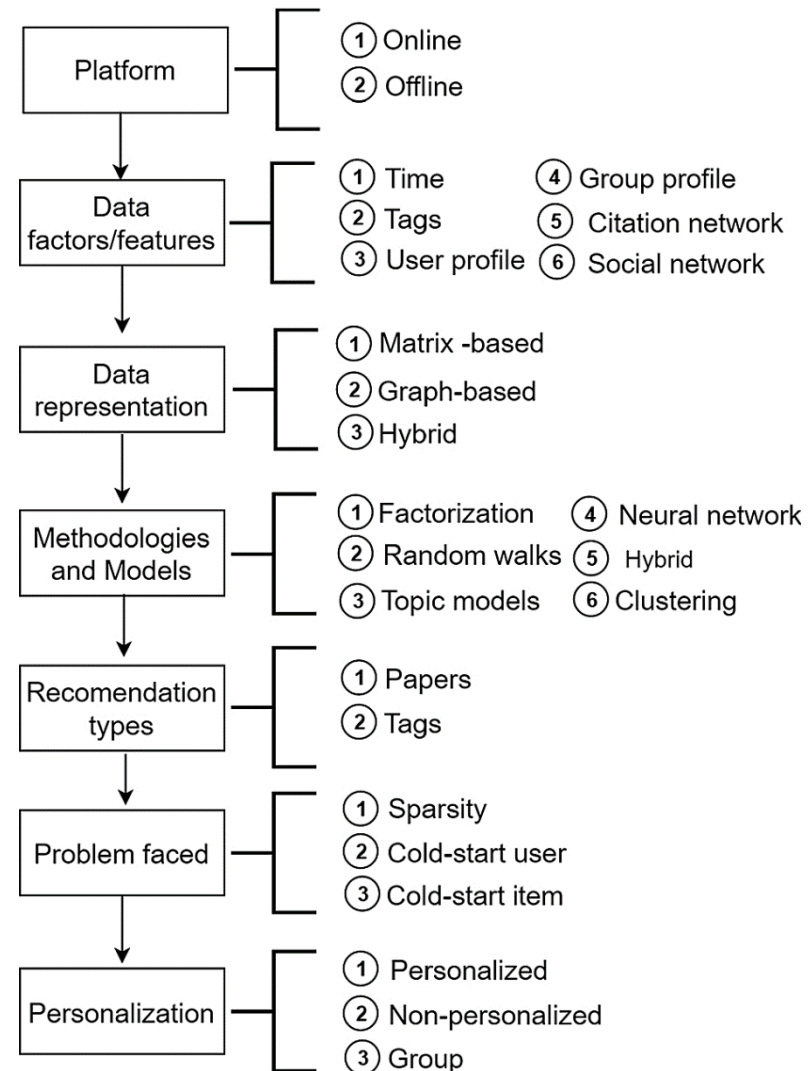
- Key methodologies: graph based approaches



Ali Z., Ullah I., Khan A., Jan A.U. and Muhammad K.: "An overview and evaluation of citation recommendation models", *Scientometrics*, 2021.

Recommending Citations

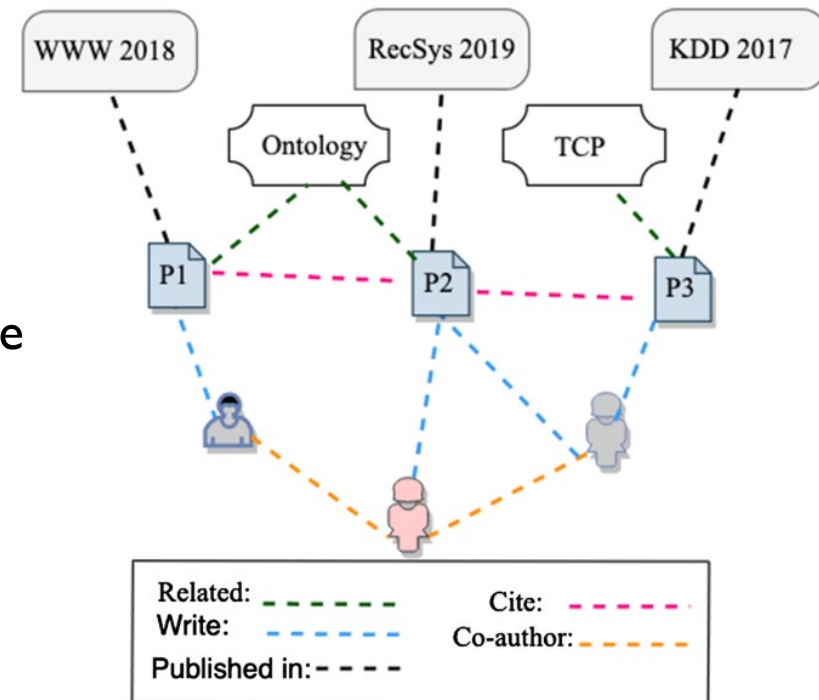
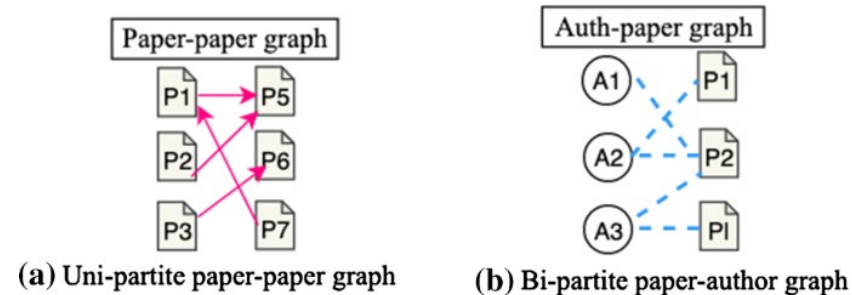
- Algorithms taxonomy
 - Off-line majority of platforms
 - “Time flies”
 - Tags \approx keywords
 - User vs group profiles
 - Citation vs social networks
 - Matrix vs graph representation
 - Random walk $P(i,j)=I/OutDegree$
 - With restart $r = cWr+ (1-c)e$
 - NN, Clustering, Classification
 - Recommendation types
 - Problems faced
 - Personalization



Ali Z., Qi G., Kefalas P., Abro W.A. and Ali B.: “A graph-based taxonomy of citation recommendation models”, *Artificial Intelligence Review*, 2020.

Recommending Citations

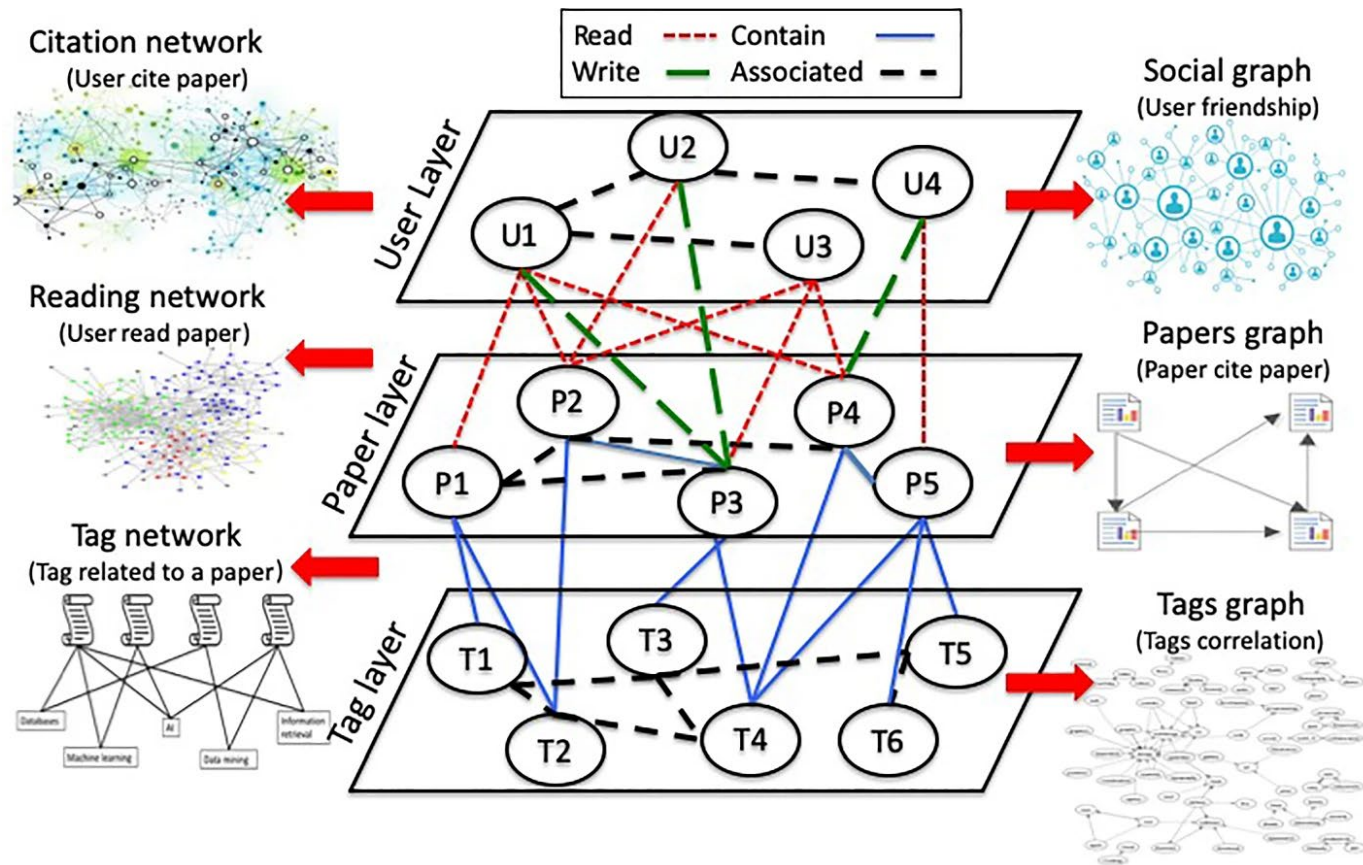
- Algorithms taxonomy
 - Off-line majority of platforms
 - “Time flies”
 - Tags \approx keywords
 - User vs group profiles
 - Citation vs social networks
 - Matrix vs graph representation
 - Random walk $P(i,j) = I / \text{OutDegree}$
 - With restart $r = cW_r + (1-c)e$
 - NN, Clustering, Classification
 - Recommendation types
 - Problems faced
 - Personalization



Ali Z., Qi G., Kefalas P., Abro W.A. and Ali B.: “A graph-based taxonomy of citation recommendation models”, *Artificial Intelligence Review*, 2020.

Recommending Citations

- A graph-based approach



Ali Z., Qi G., Kefalas P., Abro W.A. and Ali B.: "A graph-based taxonomy of citation recommendation models", *Artificial Intelligence Review*, 2020.

Recommending Supervisor

- Previous approaches were based on Multiple Criteria Decision Making
- Subjective criteria as they are reflecting judgements and opinions of both students and professors
- RecAdvisor: a tool based on objective criteria. Data selected from MAS, CORE, CVs (e.g. info about previous students, grants etc), doctoral dissertation information etc.
- First, users build their own profile into the system
- Then, the system provides a list of professors with info about publications, grant record, etc

Recommending Journals

- Rejection is the norm
- From Scopus we know that
 - 12 millions of peer reviewed papers between 1992-2002
 - this number doubled during 2003-2012
- The Elsevier recommender algorithm is divided in two parts
 - First, match the user query to the existing papers in the database, based on Okapi BM25 algorithm for IR
 - Then, rank the journals according to the average BM25 score per journal for the top X papers

Recommending Journals

- All major publishers have a tool
 - <https://journalfinder.elsevier.com/>
 - <https://journalsuggester.springer.com/>
 - <https://publication-recommender.ieee.org/home>
 - <https://www.edanzediting.com/journal-selector>
 - <https://www.enago.com/researcher-hub/journal-finder.htm>
 - <https://jane.biosemantics.org/>
 - <https://www.journalguide.com/>
 - <https://www.mdpi.com/about/journalselector>
 - <https://journalfinder.wiley.com/>
 - <https://journal-recommender.sagepub.com/>
 - <https://www.cambridge.org/universitypress/author-services/services/journal-recommendation/>

Recommending Journals

- All major publishers have a tool

Manuscript Matchers	Title and Abstract	Keyword	Field of research	Open Access
Elsevier Journal Finder	Yes	Yes	Yes	Yes
Springer Journal Suggester	Yes	Yes	Yes	Yes
IEEE Journal Recommender	Yes	Yes	No	No
Edanz Journal Selector	Yes	Yes	Yes	Yes
Enago Open Access Journal Finder	Yes	No	Yes	Yes
Journal/Author Name Estimator	Yes	Yes	No	Yes
Journal Guide	Yes	Yes	No	Yes

Manuscript Matchers	Acceptance Rate	CiteScore	Impact Factor	Publication Time	Database
Elsevier Journal Finder	Yes	Yes	Yes	Yes	Elsevier (2900)
Springer Journal Suggester	Yes	No	Yes	Yes	Springer, Biomed (2600)
IEEE Journal Recommender	No	No	No	Yes	IEEE (1670)
Edanz Journal Selector	Yes	No	Yes	No	(28651)
Enago Open Access Journal Finder	Yes	No	No	Yes	Web of Science, Scopus (10700)
Journal/Author Name Estimator	Yes	Yes	Yes	No	Medline, DOAJ
Journal Guide	Yes	Yes	Yes	Yes	Research Square

Nam N.D., Trung T., Trung N.T. and Thao T.P.T.: "Manuscript matcher: A tool for finding the best journal", *Proceedings 13th IMCIC Conference, 2022.*

Recommending Journals

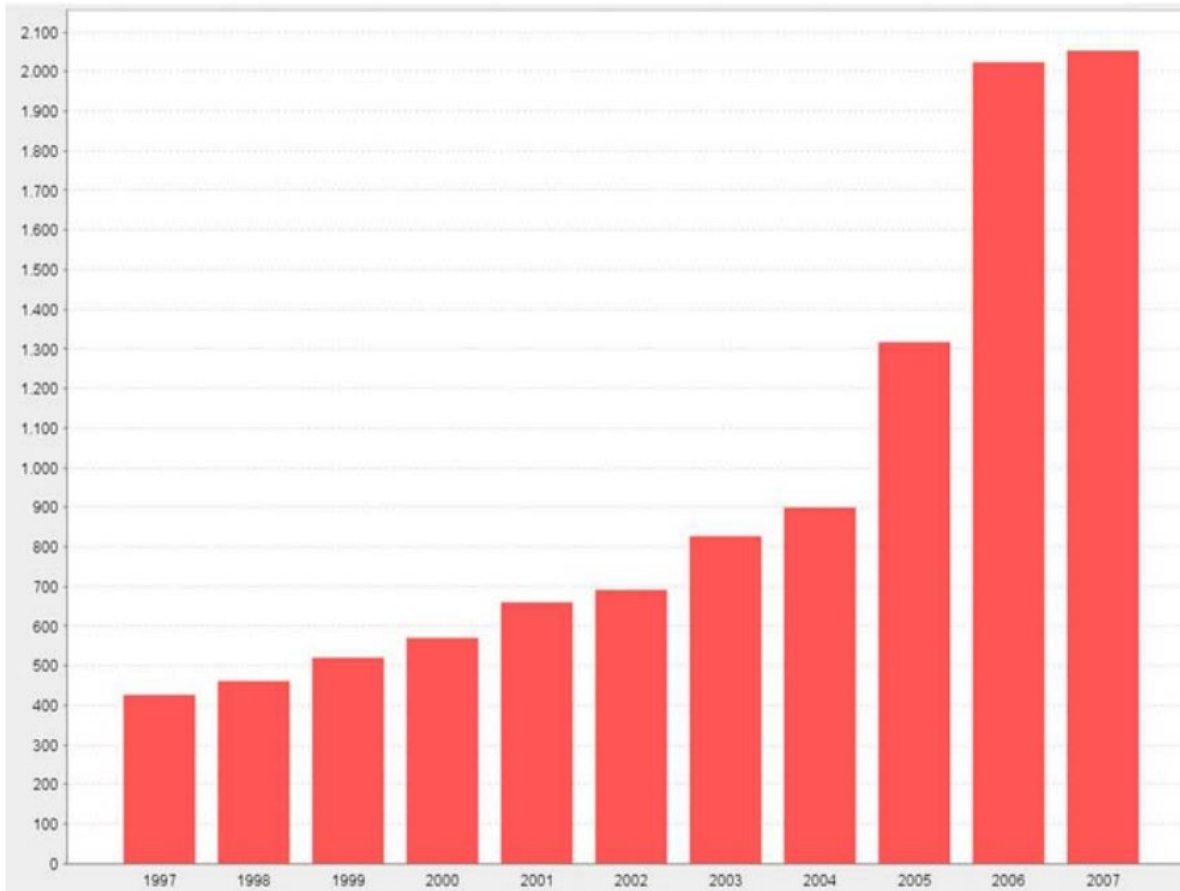
- Although the manuscript matchers are connected to particular publishers/databases, they have similar vital functions, a matching score based on impact factor, acceptance rate, production times, etc

Priority Criteria	Percentage
Acceptance rate	30
Publication speed	25
Publication fees	15
Accessibility (OA), number of readers	10
ISI/Scopus list of journals	7
Impact Factor, Cite Score	5
Field of research	3
Others	3

Recommending OA Journals

- The Open Access landscape keeps growing, making harder to choose a suitable journal to publish research findings
- Bibliometric & Semantic Open Access Recommender Network
- B!SON is build on top of
 - DOAJ: 17,669 journals, 7,489,975 articles
 - Open Citations: 72,268,850 articles, 1,294,283,603 citations
 - Journal Checker Tool: is a journal compliant with Plan-S?
 - Other data sources: Crossref, OpenAlex
- Technologies used
 - Data integration: PostgreSQL, Elasticsearch, Django backend, Vue.js frontend
 - Recommendation: Text similarity (OKAPI BM25), bibliographic coupling, combination of text and bibliographic similarity

Recommending Conferences



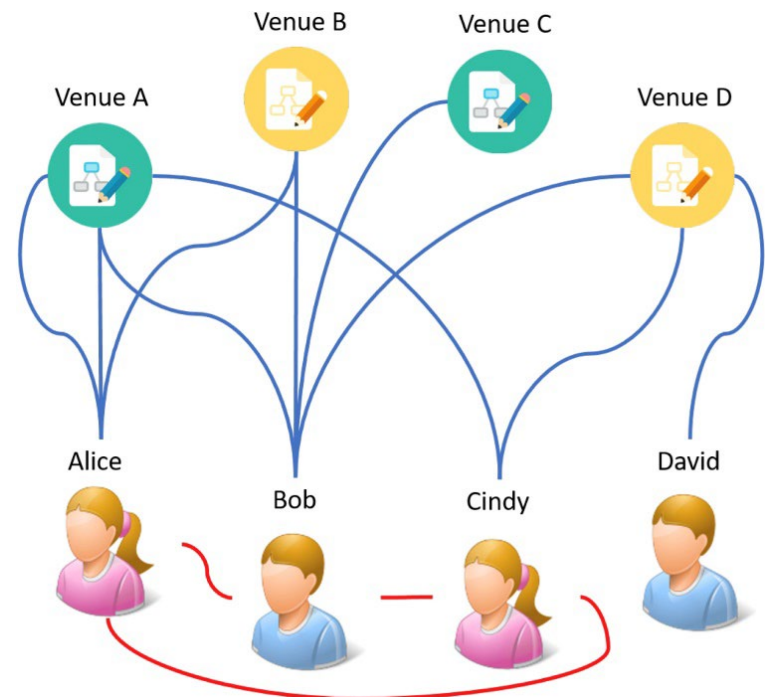
Number of events in DBLP (by distinct proceedings)

- 2000+ conferences in 2006
- 3711 conferences in 2015
- 6400 conferences as of today, 15/11/2023

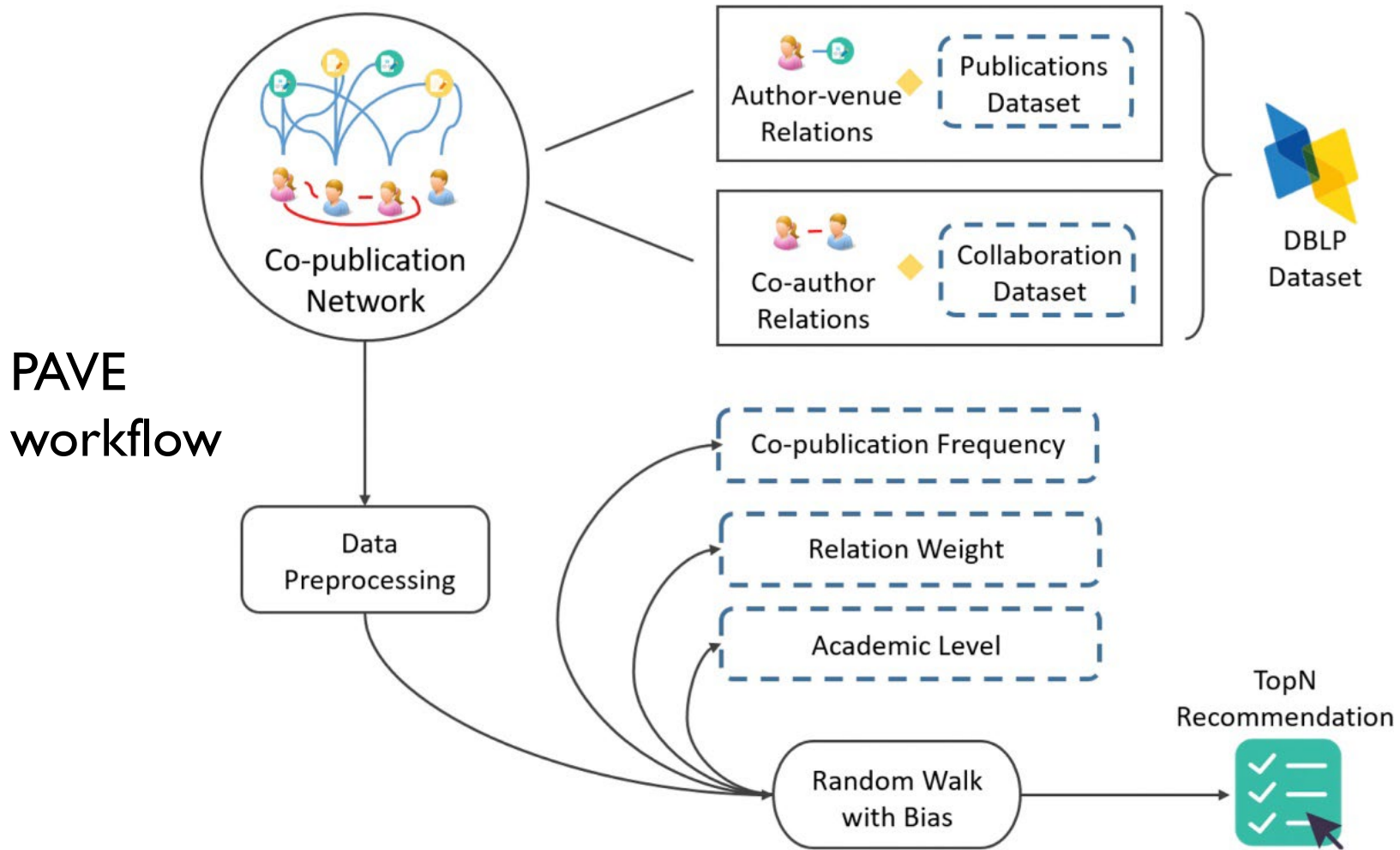
Klamma R., Cuong P.M. and Cao Y.: "You never walk alone: Recommending academic events based on social network analysis", *Proceedings 1st COMPLEX Conference*, 2009.

Recommending Conferences

- How researchers choose conferences?
 - Where can I obtain high quality venues?
 - Which venues are most suitable for me to contribute papers?
 - What are the most relevant conferences I should participate?
- Integrate authors, venues and publications into a co-publication network
- Introduce metrics
 - Co-publication frequency
 - Relation weight
 - Academic level
- Random Walk with Restart



Recommending Conferences



Presentation Menu

- Recommendation Systems
 - Information age and Social networks
 - Challenges and Data representation
- Scientometric Entities
 - Abundance of Scholarly Data
 - Taxonomy
 - Statistics
- Approaches on the Intersection
 - Citation
 - Supervisor
 - Journal
 - Venue
- **Bias in Recommenders**
- Further Reading

Bias of Scientific Recommenders

- Scientific recommender systems risk isolating scholars in *information bubbles* and insulating them from exposure to different view points.
- They also risk suffering from popularity biases, which lead to a *winner-takes all* scenario and reinforce discrepancies in recognition received by eminent scientists.
- Key-words
 - Filter bubbles
 - Echo chambers
 - Exposure problem
 - Preferential attachment
 - Cumulative advantage
 - Mathew effect

“For to all those that who have,
more will be given”

Bias in Scientific Recommenders

- Human-caused bias
 - Conformation bias
 - Position bias
 - Selection bias
- System-caused bias
 - Inductive bias
 - Exposure bias
 - Popularity bias
 - Unfairness
- All biases concerning recommendation systems in general also are relevant to scholarly recommender systems

Bias in Wikipedia citations

	NAS subgroup		AS subgroup		All	
	Women	Men	Women	Men	Women	Men
BHS	-4,7	1,6	-4,9	2,3	-4,9	2,1
LES	-8,2	2,6	-4,6	1,4	-6,1	1,9
PSE	-45,5	9,4	-13,5	1,8	-31,7	5,3
MCS	-23,6	5,1	-9,2	2,3	-15,1	3,5
SSH	-12,5	6,6	-2,8	1,7	-4,8	2,8

- 200K Wiki citations
- 1.9M papers

- Gender-level comparison of single-author publications

	NAS subgroup			AS subgroup			Mix-sphere group			All		
	Women	Men	Mix-gender	Women	Men	Mix-gender	Women	Men	Mix-gender	Women	Men	Mix-gender
BHS	-9,7	6,8	-3,1	-6,3	5,1	-3,1	-11,5	11,7	-12,5	-7,8	5,9	-3,3
LES	-8,1	5,3	-3	-3,5	3,6	-3,5	-1,5	6,1	-8,8	-6,2	4,6	-3,5
PSE	-17,3	8,2	-8,7	4,6	0,8	-2,4	-7,2	-2,2	6,3	-9,5	4,1	-5,5
MCS	-36,6	1,2	-15,1	7,4	5,6	-1,3	-1,6	0,3	1,1	-15,8	8,2	-13,3
SSH	-10,7	5,8	-1,4	-4,2	5,1	-2,1	-6,2	7,2	-6,3	-5,7	5,4	-2,1

- Gender-level comparison of multi-author publications

Bias in Wikipedia citations

	Women subgroup		Men subgroup		All	
	NAS	AS	NAS	AS	NAS	AS
BHS	-1,3	3,9	-14,8	6,7	-14,5	6,0
LES	-9,3	8,0	-9,2	8,0	-9,3	8,0
PSE	-16,2	19,9	-15,7	19,0	-15,7	19,1
MCS	-9,2	7,1	-17,2	19,7	-16,4	18,0
SSH	-12,8	4,1	-10,0	4,1	-10,9	4,1

- Country-level comparison of single-author publications

	Women subgroup			Men subgroup			Mix-gender group			All		
	NAS	AS	Mix-gender	NAS	AS	Mix-gender	NAS	AS	Mix-gender	NAS	AS	Mix-gender
BHS	-9,5	7,5	-2,9	-11,2	7,9	14,8	-10,0	8,3	13,4	-10,5	8	13
LES	-8,5	14,9	15,4	-9,5	9,7	13,5	-8,5	10,9	13,4	-9	10,5	13,5
PSE	-18,4	45,2	-7,7	-15,8	21,5	4,7	-17,5	27,8	20,7	-16,5	24	8,5
MCS	-24,8	30,1	-34,8	-14,7	15,8	18,1	-13,2	14,5	22	-14,6	16	17,9
SSH	-22,8	9,0	-3,8	-15,3	9,2	0	-17,4	9,3	2,4	-17,3	9,2	0.5

- Country-level comparison of multi-author publications

Zheng X., Chen J., Yan E. and Ni C.: "Gender and country biases in Wikipedia citations to scholarly publications", *Journal of the Association for Information Science & Technology*, 2023.

Redundancy vs. Variety

1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1

Variety

0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	100	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0

Redundancy

4	0	0	1	2	0	1	0	0	1	
2	4	1	1	0	0	0	0	0	0	
0	1	0	0	0	3	0	1	0	0	
1	0	0	2	0	1	1	0	3	1	
0	4	0	0	1	0	0	0	0	6	
0	0	1	7	1	2	0	1	0	0	
1	0	1	0	5	0	0	0	1	1	
0	8	1	0	0	1	1	0	0	1	
0	0	1	0	6	0	0	0	0	0	
2	0	0	0	1	0	2	1	1	1	

Balance

- Variety: each paper is cited once
- Redundancy: all papers cite one paper
- In practice: Pareto distributions

Andreas Pacher: “A typology of research discovery tools: Bibliometric redundancy vs. bibliometric variety”, *Proceedings 18th ISSI Conference, 2021*.

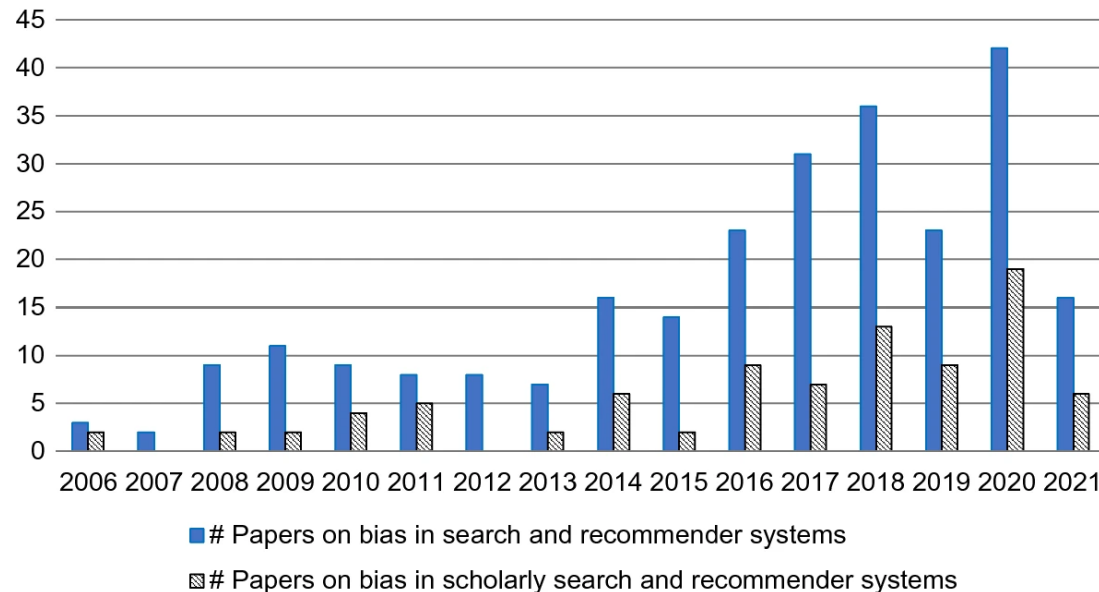
Redundancy vs. Variety

Cited Paper	Property before it is cited	When a new paper on topic X cites it, this leads to
Paper I	Already well-cited in the field X	Redundancy (enhances the citation pattern)
Paper II	A little-cited paper	Variety (it open up a novel reference)
Paper III	Well-cited paper on the topic Y, but never been cited on topic X	Variety (it bridges two topics in an original way)

- Redundancy: Scholar, MAS, WoS, Semantic Scholar etc.
- Variety: CVs, institutional repositories, journal ToCs, etc.

Recommendations for Scientific Recommenders

- Bias affects which research ideas get promoted and are finally used in the industry, impacting the welfare of industrial countries in the long run.
- Further research
 - Explainability
 - Diversifiability
 - Findability
- Awards
 - SIGIR 2020
 - KDD 2021



Färber M., Coutinho M. and Yuan S.: “Biases in scholarly recommender systems: Impact, prevalence, and mitigation”, *Scientometrics*, 2023.

Andrea Polonioli: “The ethics of scientific recommender systems”, *Scientometrics*, 2021.

Further Reading

1. Zhang A. et al.:“Scholarly recommendation systems:A literature survey”, *Knowledge and Information Systems*, 2023.
2. Jebari C. et al.:“Context-aware citation recommendation of scientific papers: Comparative study, gaps and trends”, *Scientometrics*, 2023.
3. Aymen A.T.M. et al.:“Scientific paper recommender systems:A review”, *Proceedings IC-AIRES Conference*, 2022.
4. El Alaoui et al.:“Overview of the main recommendation approaches for the scientific articles”, *Proceedings CBI Conference*, 2021.
5. Ali Z. et al.:“An overview and evaluation of citation recommendation models”, *Scientometrics*, 2021.
6. Ali Z. et al.:“A graph-based taxonomy of citation recommendation models”, *Artificial Intelligence Review*, 2020.
7. Ma S.: et al.:“A review of citation recommendation: From textual content to enriched content”, *Scientometrics*, 2020.
8. Farber M. and Jatowt A.:“Citation recommendation:Approaches and datasets”, *International Journal on Digital Libraries*, 2020.
9. Krutz C.K. and Schenkel R.:“Scientific paper recommendation systems:A literature review of recent publications”, *International Journal on Digital Libraries*, 2020.
10. Bai X. et al.:“Scientif paper recommendation:A survey”, *IEEE Access*, 2019.
11. Beel J. et al.:“Research-paper recommender systems:A literature survey”, *International Journal on Digital Libraries*, 2016.

Personal Involvement

