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:  

## 2nd RecSys 2008: Lausanne, Switzerland

Pearl Pu, Derek G. Bridge, Bamshad Mobasher, Francesco Ricci:  
ISBN 978-1-60558-093-7 [contents]

## 1st RecSys 2007: Minneapolis, MN, USA

Joseph A. Konstan, John Riedl, Barry Smyth:  
## How the Community Evolved

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

Data from DBLP.org

#papers in main volume of ACM proceedings
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.
Challenges

**Long-term preferences**

- Cartoon
- Action
- Romantic

**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

https://dblp.org/statistics/newrecordsperyear.html
Abundance of Bibliographic Data

Records in DBLP

- Journal Articles
- Book and Theses
- Data and Artifacts
- Editorship
- Parts in Books or Collections
- Informal Publications
- Conference and Workshop Papers
- Reference Works
- Reference Works
- Withdrawn Items

http://dblp.org/statistics/recordsindblp
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”
“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

Scholarly Recommendation

- Literature
- Collaborator
- Reviewer
- Conference/Journal
- Other
  - Conference
  - Journal
  - Dataset
  - Grant

225 papers selected from 500 papers

Production over the Years

<table>
<thead>
<tr>
<th>Year</th>
<th>Publications (Cumul.)</th>
<th>New (per year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2001</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2002</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>2003</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>2004</td>
<td>26</td>
<td>17</td>
</tr>
<tr>
<td>2005</td>
<td>39</td>
<td>22</td>
</tr>
<tr>
<td>2006</td>
<td>51</td>
<td>25</td>
</tr>
<tr>
<td>2007</td>
<td>70</td>
<td>36</td>
</tr>
<tr>
<td>2008</td>
<td>85</td>
<td>64</td>
</tr>
<tr>
<td>2009</td>
<td>100</td>
<td>79</td>
</tr>
<tr>
<td>2010</td>
<td>121</td>
<td>87</td>
</tr>
<tr>
<td>2011</td>
<td>137</td>
<td>97</td>
</tr>
<tr>
<td>2012</td>
<td>151</td>
<td>110</td>
</tr>
<tr>
<td>2013</td>
<td>177</td>
<td>135</td>
</tr>
<tr>
<td>2014</td>
<td>217</td>
<td>156</td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td>191</td>
</tr>
<tr>
<td>2016</td>
<td></td>
<td>227</td>
</tr>
<tr>
<td>2017</td>
<td></td>
<td>257</td>
</tr>
<tr>
<td>2018</td>
<td></td>
<td>270</td>
</tr>
</tbody>
</table>

My collection: 270 papers

Production per Country

Production/country during 2010-2023 (270 papers)

Production/country during 2018-2023 (160 papers)
Production per Country

Comparison of China-USA-Germany-India

Year

Percentage/year

Series 1
Series 2
Series 3
Series 4
Key-word Cloud

After processing; 75 out of 465 words
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

Recommending Citations

Scientometrics

Recommending Citations

Recommending Citations

- Key methodologies: deep learning based approaches.

![Graph showing the cumulative number of total publications and new publications over the years 2015 to 2020. The graph indicates a steady increase in total publications and a consistent increase in new publications.]

Recommending Citations

- Key methodologies: Collaborative filtering vs Content-based filtering

Recommending Citations

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

https://iq.opengenus.org/introduction-to-recommendation-system/
Recommending Citations

- Key methodologies: Similarity calculation

Recommending Citations

- Key methodologies: graph based approaches

Recommendating Citations

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

Recommending Citations

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

---

Recommending Citations

- A graph-based approach

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://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

<table>
<thead>
<tr>
<th>Manuscript Matchers</th>
<th>Title and Abstract</th>
<th>Keyword</th>
<th>Field of research</th>
<th>Open Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elsevier Journal Finder</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Springer Journal Suggester</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>IEEE Journal Recommender</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Edanz Journal Selector</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Enago Open Access Journal Finder</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Journal/Author Name Estimator</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Journal Guide</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

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

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.

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

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

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
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 viewpoints.
- 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

## Gender-level comparison of single-author publications

<table>
<thead>
<tr>
<th></th>
<th>NAS subgroup</th>
<th>AS subgroup</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Women</td>
<td>Men</td>
<td>Women</td>
</tr>
<tr>
<td>BHS</td>
<td>-4,7</td>
<td>1,6</td>
<td>-4,9</td>
</tr>
<tr>
<td>LES</td>
<td>-8,2</td>
<td>2,6</td>
<td>-4,6</td>
</tr>
<tr>
<td>PSE</td>
<td>-45,5</td>
<td>9,4</td>
<td>-13,5</td>
</tr>
<tr>
<td>MCS</td>
<td>-23,6</td>
<td>5,1</td>
<td>-9,2</td>
</tr>
<tr>
<td>SSH</td>
<td>-12,5</td>
<td>6,6</td>
<td>-2,8</td>
</tr>
</tbody>
</table>

## Gender-level comparison of multi-author publications

<table>
<thead>
<tr>
<th></th>
<th>NAS subgroup</th>
<th>AS subgroup</th>
<th>Mix-sphere group</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Women</td>
<td>Men</td>
<td>Mix-gender</td>
<td>Women</td>
</tr>
<tr>
<td>BHS</td>
<td>-9,7</td>
<td>6,8</td>
<td>-3,1</td>
<td>-6,3</td>
</tr>
<tr>
<td>LES</td>
<td>-8,1</td>
<td>5,3</td>
<td>-3</td>
<td>-3,5</td>
</tr>
<tr>
<td>PSE</td>
<td>-17,3</td>
<td>8,2</td>
<td>-8,7</td>
<td>4,6</td>
</tr>
<tr>
<td>MCS</td>
<td>-36,6</td>
<td>1,2</td>
<td>-15,1</td>
<td>7,4</td>
</tr>
<tr>
<td>SSH</td>
<td>-10,7</td>
<td>5,8</td>
<td>-1,4</td>
<td>-4,2</td>
</tr>
</tbody>
</table>

- 200K Wiki citations
- 1.9M papers

Bias in Wikipedia citations

<table>
<thead>
<tr>
<th>Country-level comparison of single-author publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women subgroup</td>
</tr>
<tr>
<td>NAS</td>
</tr>
<tr>
<td>BHS</td>
</tr>
<tr>
<td>LES</td>
</tr>
<tr>
<td>PSE</td>
</tr>
<tr>
<td>MCS</td>
</tr>
<tr>
<td>SSH</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country-level comparison of multi-author publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women subgroup</td>
</tr>
<tr>
<td>NAS</td>
</tr>
<tr>
<td>BHS</td>
</tr>
<tr>
<td>LES</td>
</tr>
<tr>
<td>PSE</td>
</tr>
<tr>
<td>MCS</td>
</tr>
<tr>
<td>SSH</td>
</tr>
</tbody>
</table>

Redundancy vs. Variety

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

# Redundancy vs. Variety

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

- **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


Further Reading

Personal Involvement
Thank You