Open Research Online

The Open University's repository of research publications and other research outputs

Capturing and Exploiting Citation Knowledge for the Recommendation of Scientific Publications

Thesis

oro.open.ac.uk

CAPTURING AND EXPLOITING CITATION KNOWLEDGE FOR THE RECOMMENDATION OF SCIENTIFIC PUBLICATIONS

Anita Khadka

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Knowledge Media Institute

Faculty of Science, Technology, Engineering & Mathematics

The Open University

September 2020

ABSTRACT

With the continuous growth of scientific literature, it is becoming increasingly challenging to discover relevant scientific publications from the plethora of available academic digital libraries. Despite the current scale, important efforts have been achieved towards the research and development of academic search engines, reference management tools, review management platforms, scientometrics systems, and recommender systems that help finding a variety of relevant scientific items, such as publications, books, researchers, grants and events, among others.

This thesis focuses on **recommender systems for scientific publications**. Existing systems do not always provide the most relevant scientific publications to users, despite they are present in the recommendation space. A common limitation is the **lack of access to the full content of the publications** when designing the recommendation methods. Solutions are largely based on the exploitation of metadata (e.g., titles, abstracts, lists of references, etc.), but rarely with the text of the publications. Another important limitation is the lack of time awareness. Existing works have not addressed the important scenario of recommending the most recent publications to users, due to the challenge of recommending items for which no ratings (i.e., user preferences) have been yet provided. The **lack of evaluation benchmarks** also limits the evolution and progress of the field.

This thesis investigates the **use of fine-grained forms of citation knowledge**, extracted from the full textual content of scientific pub-

lications, to enhance recommendations: citation proximity, citation context, citation section, citation graph and citation intention. We **design and develop new recommendation methods** that incorporate such knowledge, individually and in combination.

By conducting offline evaluations, as well as user studies, we show how the use of citation knowledge does help enhancing the performance of existing recommendation methods when addressing two key tasks: (i) recommending scientific publications for a given work, and (ii) recommending recent scientific publications to a user. Two novel evaluation benchmarks have also been generated and made available for the scientific community.

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledge, the work presented is entirely my own.

_____

Anita Khadka

Somewhere, something incredible is waiting to be known.

— Carl Sagan

ACKNOWLEDGMENTS

I have received a great deal of support and assistance throughout my PhD journey. I would like to acknowledge those amazing people who have walked with me during this adventure. First of all, I would like to show my gratitude to my excellent supervisor Dr Miriam Fernandez for her immense patience towards me. Her continuous support and guidance in how to rigorously think and write in academia have reshaped my research thinking process. The meetings with her always had a calming effect during the years, when I am lost and in panic mode, she has always been there to help me, both technically and emotionally. I am very grateful to her believing in me and taking me under her wings and supporting me especially during the toughest time of my PhD study. I could not have completed this journey and this thesis without her exemplary supervision and support.

I would equally like to thank and show my gratitude to my external supervisor Dr Ivan Cantador, with whom I had the honour to work with. He is the most intelligent and diligent researcher I had a chance to work with and I am very grateful for such an opportunity. His expertise, guidance, feedback and invaluable discussion in the research topics have undoubtedly had a great impact on the quality of this work.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ACMDL Association for computing Machinery Digital Library

AP Average Precision

CBF Content Based Filtering

CF Collaborative Filtering

CP Citation Proximity

CPA Citation Proximity Analysis

CPC Citation Proximity-Context

CPI Citation Proximity Index

FM Factorisation Machine

GROBID GeneRation Of Bibliographic Data

IBCF Item-based Collaborative Filtering

LDA Latent Dirichlet Allocation

LSA Latent Semantic Analysis

MAE Mean Absolute Error

MAP Mean Average Precision

MF Matrix Factorisation

ML Machine Learning

MRR Mean Reciprocal Rank

nDCG Normalised Discounted Cumulative Graph

PDF Portable Document Format

PHOAKS People Helping One Another Know Stuff

RMSE Root Mean Square Error

RS Recommender Systems

SPRS Scientific Publication Recommender Systems

SVD Singular Value Decomposition

TEI Text Encoding Initiatives

TF-IDF Term Frequency Inverse Document Frequency

VSM Vector Space Model

Part I

INTRODUCTION AND LITERATURE REVIEW

INTRODUCTION

A general overview of the thesis is provided in this chapter. Section 1.1 motivates the research problem and summarises existing progress and limitations reported in the literature. Section 1.2 defines the scope of the study by stating the addressed research questions and key contributions of this research. Section 1.3 outlines the research methodology carried out in the thesis. Section 1.4 describes the structure of this document and Section 1.5 lists the terms and phrases used in the thesis interchangeably. Lastly, Section 1.6 lists the publications that resulted from the research undertaken in the thesis.

1.1 MOTIVATION

With the continuous growth of scientific literature, it is becoming more and more challenging to discover relevant scientific publications from the plethora of available academic digital libraries.

A recent report by the International Association of Scientific, Technical and Medical Publishers[1] claims the existence of 33,100 active scholarly peer-reviewed English-language journals in mid-2018 (plus a further 9400 non-English-language journals), collectively publishing over three million papers per year. The report also states that the production of scientific publications is steadily increasing at a 4% annual rate. March 2020 statistics from one of the well-known digital libraries

Information overload problem

---

[1] https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf

– the Association for computing Machinery Digital Library (ACMDL) [109]– show numbers in the order of 181,692 books, 74,545 theses and 26,251 proceedings. It is also relevant to highlight that more than 100 academic databases and search engines, including academic journals, institutional repositories, archives, or collections or scientific articles, are currently available.[2]

Existing solutions   Given the scale of digital scholarly information currently available, a wide range of efforts have been invested in the last decade to discover, analyse and exploit digital contents. Those efforts range from the development of academic search engines like Google Scholar[3], CiteSeer[4] and PubMed[5], reference management tools (e.g. JabRef[6], Zotero[7]), review management platforms [273], scientometrics systems -which analyse the impact of scientific literature [106]-, and Recommender Systems (RS) for scientific publications, books, authors, events and patents, among others [19, 84].

All these solutions are needed to facilitate the efficient discovery, management and analysis of relevant research [66, 183, 202, 254], helping researchers and practitioners to concentrate on existing approaches and resources, and avoid the unnecessary duplication of previous research efforts [34]. Search engines are designed to find relevant content for a given query [224]. Reference management tools help users to record bibliographic citations, generating databases of bibliographic references that can then be filtered, selected or exported to facilitate the creation of reference lists for scientific articles, reports, etc. Review management platforms help with the automatic assignment of sci-

---

2 https://en.wikipedia.org/wiki/List_of_academic_databases_and_search_engines
3 https://scholar.google.co.uk/
4 https://citeseerx.ist.psu.edu/index
5 https://www.ncbi.nlm.nih.gov/pubmed/
6 www.jabref.org
7 www.zotero.com

entific papers to reviewers [70]. Scientometrics systems measure and analyse scientific literature, including the impact of research papers and academic journals, or the impact of research institutions [106, 157]. Finally, RS help to provide personalised suggestions of relevant academic resources (publications, patents, events, etc.) to users [123].

In this thesis, our focus is on **the research and development of novel RS for scientific publications**. The recommendation of scientific publications is an important and timely problem. Personalised recommendations can help researchers to keep up-to-date with the latest scientific discoveries in their fields [19, 213], provide junior researchers with key papers to help them familiarise with the concepts of a particular research field [72, 126], or help researchers and practitioners to find relevant papers for their work, e.g. by finding relevant references for an ongoing manuscript.

Recommender Systems

Over the last decade, multiple works have emerged that address the recommendation of scientific publications. These works have focused on a variety of recommendation tasks including: (i) recommending relevant papers for a given user [189, 267], (ii) recommending relevant papers for a given paper [140, 160], (iii) recommending relevant papers for a particular snapshot of text (title, abstract, free research related text etc.) [30, 116], (iv) recommending relevant papers for a particular collection of papers [72, 236], and (v) recommending relevant papers for an undergoing manuscript (i.e., a paper yet to be published) [102, 248]. While the first task –recommending relevant papers for a given user– focuses on the traditional RS problem, where items (i.e., research publications) are recommended to users (researchers) based on their preferences (expressed as or extracted from e.g. previous publications, topics of interest, etc.), the rest of the tasks do not have users (but a paper, a set of papers, a snapshot of text, or an ongoing manuscript)

Various recommendation tasks

as **targets** of the recommendations, enabling the discovery of relevant scientific publications to help with a specific piece of work.

Addressed tasks　In this thesis, we aim to address both main scenarios: (i) the **recommendation of relevant scientific publications to a user** and (ii) the **recommendation of relevant scientific publications for a specific piece of work (defined in this thesis as a scientific publication)**.

Regarding the recommendation of relevant scientific publications to *The time aspect*　users, while most of the existing solutions recommend relevant papers to users independently of the time when such papers were published, **in this thesis we address the real-world problem of recommending recently published papers** [95]. This is a particularly challenging scenario where traditional recommendation methods such as collaborative filtering are not effective, since they are not able to recommend the latest, most recent papers which have not previously seen, rated or cited. In the recommender systems literature, this problem is known as "new item cold start." This problem has barely been addressed in the context of academic recommendations. The concept of **time** has been previously studied to better define and delimit long-term vs. short-term preferences when capturing the researchers' interests [252], and to suggest papers to researchers with no previous activity (i.e., new researcher) [112]. However, to the best of our knowledge, only a few works [95, 272] have addressed the problem of recommending the most recently published scientific work, i.e., new publications.

In addition, while existing recommendation methods for scientific *Citation knowledge*　publications explore a wide range of metadata to generate user and item profiles from which recommendations are generated (e.g., title, abstract, keywords, authors, publication venues, bibliographies, and citation graphs among authors and papers) [19, 26, 102], full-text has not been as widely used. It is important to highlight that full-text

of scientific articles is often stored behind paywalls, and hence, it is not possible to freely access it. However, with the emergence of the open access movement[8], many scientific publications have now become publicly available, providing an exciting opportunity for a fine-grained exploration and use of their content [71, 102].

Motivated by this situation, **the thesis explores the use of full-text of publications to capture and exploit the knowledge provided by citations**. Citations (or papers cited within a given paper) capture important and relevant work, as authors carefully choose their references [45, 254]. Therefore, citations can play an important role on comprehending the authors' preferences (research interests) [115]. In this context, our main hypothesis is that the exploitation of citation knowledge can help improving the recommendation of scientific publications.

We, therefore, investigate **how citation knowledge could be captured and exploited to support the discovery of relevant scientific publications**. Among the captured citation knowledge we include (i) the citation-graph (where nodes represent papers and edges represent relations between such papers based on their citations), (ii) citation section (section within the publication where the citation appears), (iii) citation proximity (distance between citations), (iv) citation context (text surrounding the citation), and (iv) citation intention (purpose with which the citation is made - compare, criticise, etc.). We explore such forms of citation knowledge, individually and in combination, providing an in-depth analysis of their strengths and limitations.

In the subsequent sections, we describe the main research questions and contributions behind this work in Section 1.2, the followed re-

---

8 https://open-access.net/en/information-on-open-access/
history-of-the-open-access-movement

search methodology in Section 1.3, Section 1.4 presents the outline of
this thesis, various terminologies interchangeably used this thesis are
stated in Section 1.5, and lastly the publications backing up this work
are presented in Section 1.6.

1.2   RESEARCH QUESTIONS, HYPOTHESES AND CONTRIBUTIONS

The main research goal investigated in this thesis is:

**Investigating whether and how citation knowledge could be
used to improve the performance of recommender systems for
scientific publications.**

The main focus of this thesis is enhancing the performance of RS
for scientific publications by investigating a series of recommenda-
tion methods that utilise a wider notion of citation knowledge. Our
research can be summarised in four key research questions. To invest-
igate the first two, we have conducted a systematic literature review
of existing works on RS for scientific publications (–see Chapter 3).
Research related with the last two questions is based on the empirical
investigations (–see Chapters 4 and 5).

**RQ1: Which types of citation knowledge have been used in RS
for scientific publications?**

This research question aims to investigate the different types of
citation knowledge that have been previously used in the literature.
Our hypothesis is that, due to the lack of access to the full content of
publications, previous works have explored only a limited notion of
citation knowledge and hence, more fine-grained characterisations of
this knowledge can be extracted, exploited and modelled to enhance
the performance of recommendation of scientific publications.

HP1: A limited notion of citation knowledge has been used so far in the research and development of RS for scientific publications.

To address this research question, we have conducted a systematic literature review of RS for scientific publications. We have defined and launched a number of formal queries to collect relevant literature from the Elsevier Scopus[9] and ISI Web of Knowledge[10] digital libraries. These queries are aimed to extract literature about *recommender systems* and *scientific publications*. Details about the queries are described in Chapter 3. In particular, from the queries we obtained 741 scientific publications. We conducted a manual filtering of such publications to ensure that they were related to our topic of research interest, reviewing a final number of 202 publications. Our contributions on this line of work are:

- Conducting a systematic literature review on RS for scientific publications.

- Identifying the different types of citation knowledge that have been used in the literature of RS for scientific publications.

- Proposing more fine-grained characterisations of citation knowledge by exploiting the full textual content of publications.

RQ2: Which are the different recommendation tasks that have been proposed in the literature of RS for scientific publications, and how citation knowledge has been applied for each of these tasks?

The purpose of RS is to suggest relevant items to a user (identified by a personal profile). In the context of RS for scientific publications,

---

a user profile sometimes captures the preferences of a particular individual, but it can also capture the characteristics of a piece of work for which recommendations are made. Based on the target of the recommendation, we can differentiate five different recommendation tasks in the literature: (i) recommending scientific publications for a specific user [47, 189, 252, 253, 267], (ii) recommending scientific publications for a specific paper [160, 183], (iii) recommending scientific publications for a particular snapshot of text (title, abstract, etc.) [30, 116], (iv) recommending scientific publications for a manuscript under development [102, 248] and, (v) recommending scientific publications for a set of specific papers [72, 150, 241]. While the first task focuses on the need of recommending relevant scientific publications to a user, the other four tasks focus on the need of recommending relevant scientific publications for a particular piece of work.

In this thesis, we investigate the use of citation knowledge to address both problems. Chapter 4 focuses on exploring the use of different types of citation knowledge to recommend scientific publications for a particular piece of work. In our case, this piece of work is identified by a specific paper. Chapter 5 addresses, on the other hand, the use of citation knowledge to recommend relevant scientific publications to a user.

Our hypothesis when investigating this research question is that different recommendation tasks exist in the literature of RS for scientific publications and that each of these tasks presents specific limitations, particularly related to the exploitation of citation knowledge.

HP2: Different recommendation tasks exist in the literature of RS for scientific publications, and particular limitations exist on how citation knowledge has been used for each of the such tasks.

To address the above research question and to test our hypothesis, we have conducted a systematic literature review following the same methodology specified in **RQ1**. Our contribution on this line of work are:

- Identifying the different recommendation tasks that exist in the literature of RS for scientific publications.

- Identifying key limitations on how citation knowledge has been used for each of the existing recommendation task.

RQ3: When addressing the task of recommending scientific publications for a particular piece of work, can citation knowledge help improving existing RS?

When investigating RQ1 and RQ2, we identified different recommendation tasks with two main objectives: (i) recommending scientific publications to a user, and (ii) recommending scientific publications for a specific piece of work, where this piece of work can be captured by a paper, a set of papers, a snapshot of text (title, abstract, free research related text, etc.), or an ongoing manuscript. In RQ3, we focus on the recommendation of scientific publications for a specific piece of work (defined in our case as a research paper).

For this particular use case, we propose two novel notions of citation knowledge, namely citation proximity and citation context, and use them to improve existing recommendation methods. We hypothesise that papers that are cited in close proximity within the content of a scientific publication are related. Relations captured via citation proximity can therefore be used to enhance the recommendation of scientific publications. Similarly, the text around a citation (i.e., citation context) can help to better capture the topic of a cited paper, and whether this topic is similar to the one of the target paper (or piece

of work for which recommendations are being made). Our hypothesis is that citation context can help establishing topic similarity, and hence also providing more relevant recommendations.

HP3: Citation proximity and citation context can be used to enhance existing recommendation methods when addressing the task of recommending scientific publications for a particular piece of work.

To address RQ3 and test HP3, we have designed two experiments where (i) we defined novel notions of citation knowledge (in this case, citation proximity and citation context), (ii) we proposed new recommendation methods that integrate such notions of citation knowledge, individually and in combination, and (iii) we tested whether those notions of citation knowledge do indeed help to enhance the performance of the existing recommendation methods. Evaluations have conducted by means of user studies. This research is extensively described in Chapter 4. Our key contributions can be summarised as:

1. Utilising two novel notions of citation knowledge (citation proximity and citation context).

2. Proposing novel recommendation approaches that incorporate citation proximity and citation context to address the task of recommending scientific publications for a particular piece of work (defined in our thesis as a scientific publication).

3. Assessing how our proposed methods compare against standard recommendation approaches by means of user studies.

RQ4: When addressing the task of recommending recent scientific publications to a user, can citation knowledge help improving existing RS?

This research question focuses on the objective of recommending scientific publications to a user. When investigating RQ1 and RQ2 we

observed that, while existing approaches had focus on recommending papers to a user independently on when those papers were published, very few approaches had focused on the problem of recommending recent scientific publications to a user. This is a particularly relevant problem since it is focused on keeping users up to date with the most recent literature. It is also a particularly difficult problem, since traditional recommendation methods, such as collaborative filtering, do not work in this particular case. These methods are not able to recommend the latest, most recent papers, since these papers have not been previously seen, rated or cited. In the RS literature, this problem is known as "new item cold start," and it has barely been addressed in the context of academic recommendations.

We propose to address this problem by means of defining and incorporating four types of citation knowledge to enhance existing recommendation methods. In particular, we propose the use of: (i) citation section, i.e. section of the paper where the citation is placed (introduction, related work, conclusions, etc.), (ii) citation context, i.e. text around the citation, (iii) citation graph, i.e. where nodes represent papers and edges represent relations between such papers based on citations, and (iv) citation intention, i.e. intention with which the citation is made –to provide background, to compare papers, etc. Each of these types of citation knowledge captures valuable information that may help improving the recommendation of scientific publications to users. User profiles are defined by considering the publication history of the users. Our key research hypothesis can be defined as:

HP4: Citation section, citation context, citation intention and citation graph are notions of citation knowledge that can help enhancing existing recommendation methods when addressing the task of providing personalised recommendations of recent scientific publications to a user.

To address RQ4 and test HP4, we have designed two experiments in which we (i) defined novel notions of citation knowledge (in this case, citation section, citation context, citation intention and citation graph), (ii) proposed new recommendation methods that integrate such notions of citation knowledge, individually and in combination, and (iii) tested whether those notions of citation knowledge do indeed help to enhance existing recommendation methods. Evaluations have conducted by creating a 'gold standard' dataset and testing the performance of our proposed recommendation methods against multiple baselines. This research is extensively described in Chapter 5. Our key contributions can be summarised as:

- Utilising novel notions of citation knowledge (citation section, citation context, citation graph and citation intention).

- Propose novel recommendation approaches that incorporate the above types of citation knowledge in isolation and in combination to address the task of recommending recent scientific publications to a user.

- Creating a new 'gold-standard' dataset to assess the performance of RS in the particular scenario of recommending recent and relevant scientific publications to a user.

- Conducting a rigorous evaluation of our proposed recommendation methods by comparing them against multiple baselines by using the previously generated gold-standard dataset.

1.3 RESEARCH METHODOLOGY

This section (see Figure 1.1) briefly summarises the research methodology followed to answer the research questions posed in this thesis. These research questions required both theoretical and empirical investigations.

For our theoretical investigations (RQ1 and RQ2), we conducted a literature review of RS (see chapter 2) and of the particularities of the area of RS for scientific publications (see chapter 3). We analysed the literature from different dimensions concentrating particularly on (i) the proposed recommendations tasks, (ii) how items and targets (e.g., users) profiles are built to target such recommendation tasks, (iii) the proposed recommendation methods, and (iv) their evaluation (evaluation methods, metrics, datasets, etc.). Within this analysis, we also concentrated our efforts on investigating the different notions of citation knowledge extracted and used for the recommendation of scientific publications. This literature review helped us to identify multiple gaps in the domain and hence, unexplored research directions.

For our empirical investigations, we concentrated on the two most popular recommendation tasks: (i) the recommendation of scientific publications for a given piece of work chapter 4, and (ii) the recommendation of scientific publications for a user chapter 5. To target these two tasks, we explored the use of different types of citation knowledge both, individually and in combination. To conduct this empirical research we considered the following steps:

1. Dataset Building: This is a first crucial step where data needs to be prepared before starting the empirical investigation. If

Chapter 3

Systematic Literature Review

Survey Scope and Methodology

Recommendation tasks

Citation Knowledge

Item and Target Modelling

Recommendation Approaches

Evaluation

Analysis of Recommender Systems for Scientific Publications

Chapter 4

Recommending to a piece of work

Chapter 5

Recommending to a user

Dataset Building

Proposed Approach

Evaluation

Collecting Data

Knowledge for Modelling

Items

Preferences for Modelling

Target

Citation Knowledge

Proposed Recommendation Methods

Baselines

Evaluation Method and Metric

Evaluation Results

Chapter 6

Conclusion

Discussion

Open Issues and Future Work

Figure 1.1: Pipeline of our work under each chapter. Arrows pointing to the same block mean two chapters are sharing same structure

existing datasets can not be reused, new datasets are created. This process contains three important steps, including:

a) Collecting and Parsing Data: In this step, digital copies of scientific publications are collected. In general, scientific publications are in Portable Document Format (PDF) format. Data extractors are created to parse such files and extract information, as well as citation knowledge from them.

b) Modelling Items: Features are extracted to model scientific publications and to present them in a machine readable format (vector, matrices, etc.).

c) Modelling Targets (e.g., user profiles): Targets' profiles are built by considering a variety of features and are also presented in a machine readable format.

2. Proposed Recommendation Approach: We proposed a variety of recommendation approaches that incorporate citation knowledge to address the above mentioned tasks. To develop such approaches we:

- Selected and extracted citation knowledge: Exploring and selecting various types of citation knowledge to address the specific recommendation task.

- Proposed Recommendation Method: Investigate and propose various recommendation methods that integrate the previously selected citation knowledge.

3. Evaluation: Measures the performance of the developed recommendation methods. When conducting evaluations we follow various steps:

- Baseline Selection: We selected the baselines to validate the performance of the proposed methods.

- Evaluation Method and Metric: We selected the most appropriate evaluation method and evaluation metrics.

- Evaluation Results: We analysed the obtained results after comparing our proposed recommendation methods against the selected baselines using the chosen evaluation method and metric. This analysis aims to answer our RQs.

We then reflect on the conducted theoretical and empirical research, discuss the strengths and limitations of the conducted work and propose new research directions.

1.4 THESIS OUTLINE

This section presents the structure of the thesis, which is divided into three major parts as follows.

Part I: Background and Literature Review In this first part of the thesis, we introduce the addressed problem, as well as our research questions and hypothesis. We also summarise existing works in the literature and present (i) an overview of the area of RS, and (ii) a systematic review of the specific field of RS for scientific publications. **Chapter 2** provides fundamental background knowledge for understanding RS in general. We start with a brief history of RS, followed by the essential components of RS, existing popular recommendation approaches and methods, and traditional evaluation methods and metrics.

Chapter 3 presents a systematic literature review of the field of RS for scientific publications. A comprehensive overview of existing works,

as well as an in-depth analysis of their strengths and limitations is provided in this chapter, addressing the first two research questions of this thesis (RQ1 and RQ2).

Part II: Proposed Empirical Solutions

This part of the thesis focuses on presenting our proposed approaches for the two key recommendation scenarios identified in the literature of RS for scientific publications: (i) recommending scientific publications for a given piece of work, and (ii) recommending scientific publications for users.

Chapter 4 addresses RQ3, the use of citation knowledge for the recommendation of scientific publications for a particular piece of work (characterised as a paper). Two notions of citation knowledge are proposed to address this task (citation proximity and citation context), capturing relations among publications based on citation distance and topical similarity. Novel recommendation methods that incorporate this knowledge are proposed and compared against existing methods by means of user studies.

Chapter 5 addresses RQ4, the use of citation knowledge for the recommendation of recent scientific publications to a user. Four notions of citation knowledge are proposed to address this task (citation section, citation graph, citation context and citation intention). These types of citation knowledge capture a variety of information about users' preferences and the relations among publications. Novel recommendation methods that incorporate this knowledge are proposed and compared against baseline methods by means of a previously generated evaluation benchmark. This benchmark, along with the proposed notions of citation knowledge, and the novel recommendation methods that incorporate this knowledge, are the key contributions of our work.

Part III: Discussion, Conclusion and Future Work In this last part of the thesis, we reflect on our proposed notions of citation knowledge, and present our proposed recommendation methods and conducted experiments. We discuss how our experiments answer our research questions and support or reject our hypotheses. Existing challenges and open lines of work are also discussed, paving the way for new research in the field.

Chapter 6 presents an in-depth discussion of our work, its strengths and limitations. We describe how the conducted work has helped us to answer the research questions, and the key challenges we met along the way. We also discuss the main outcomes of this thesis as well as the key future research directions that could be further explored.

1.5   TERMINOLOGY

In terms of nomenclature, please note that the following terms and expressions are used interchangeably along the thesis (see table 1.1). We particularly acknowledge that academic recommender systems refer to a wider variety of systems focused not only on scientific articles, but on the recommendation of reviewers, scientific venues, etc. However, for simplicity, when we mention academic recommender systems in this manuscript explicitly refer to those systems that recommend scientific publications.

Table 1.1: Terms and phrases used in the thesis interchangeably

Terms/Phrase	Interchangeable used terms/phrases
Scientific publication	Research paper, scientific paper, scholarly paper, scholarly publication, and research publication
Item	Scientific publication
User	Author, researcher, practitioner
Scientific publication recommender system	Academic recommender system

1.6 PUBLICATIONS

The work in this thesis have been published in the following peer-reviewed venues. These publications are linked to the two main technical chapters presented in this thesis.

Publication related to Chapter 4:

- Petr Knoth and Anita Khadka. 2017. **Can we do better than co-citations? - Bringing citation proximity analysis from idea to practice in research article recommendation**. In Proceedings of the 2_{nd} Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries, (BIRNDL '17), at the 40th ACM SIGIR conference.

- Anita Khadka and Petr Knoth. 2018. **Using citation-context to reduce topic drifting on pure citation-based recommendation**. In Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18), 362--366.
  DOI: https://doi.org/10.1145/3240323.3240379

Publications related to Chapter 5:

- Anita Khadka, Iván Cantador and Miriam Fernández. 2020. **Exploiting citation knowledge in personalised recommendation**

of recent scientific publications. In Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC '20).

- Anita Khadka, Iván Cantador and Miriam Fernández. 2020. **Capturing and Exploiting Citation Knowledge for Recommending Recently Published Papers**. In Proceedings of the Semantic technologies for smart information sharing and web collaboration Conference Track at 29th IEEE WETICE Conference.

RECOMMENDER SYSTEMS

Recommender Systems (RS) are proactive software tools that intend to help users discover relevant items, by narrowing down a large collection to the selective relevant item(s) [38, 223]. As mentioned in Chapter 1, information overload is a challenging problem for different domains. As a result, the use of RS is growing in a diverse and wider set of communities, including e-commerce (e.g., Amazon.com[1], eBay[2], Netflix[3], Spotify[4], YouTube[5]), academia (e.g. researchers, publishers), education (e.g., students, teachers, schools, university), broadcasting (e.g. e-newspapers – BBC[6], Google news[7], Reuters news[8]) and others.

The use of recommender systems benefits to both service providers and end users [123]. With regards to service providers, RS can help them by recommending products to potential customers, improving user satisfaction [73, 82, 125, 223] and user fidelity [223] (e.g. providing extra care for returning customers by rewarding with loyalty vouchers, for example). It can also help diversifying the products offered to the customers and increasing their sales [223]. As for the users, RS can help them to discover relevant items (e.g., cheaper travel packages) [222], or obtain personalised recommendations [223].

Purposes of RS

---

1 www.amazon.com
2 www.ebay.com
3 www.netflix.com
4 www.spotify.com
5 www.youtube.com
6 https://www.bbc.co.uk/news
7 https://news.google.com/?hl=en-GB&gl=GB&ceid=GB%3Aen
8 https://uk.reuters.com/

This chapter provides an overview of RS. The structure and content of this chapter is inspired by the following references [41, 266].

We start with the formalisation of the recommendation problem in Section 2.1, followed by a brief history of RS in Section 2.2 where we describe the key milestones in the development of RS. In Section 2.3, we present the different components of RS including: modelling of users and items, recommendation tasks, etc. Sections 2.4 and 2.5 present a variety of existing recommendation techniques and various evaluation settings for RS respectively. We list multiple open challenges yet to be resolved in the domain in Section 2.6, followed by Section 2.7 which concludes this chapter.

2.1 OVERVIEW OF RECOMMENDER SYSTEMS

We formulate the recommendation problem based on [3] as follows:

The recommendation problem and its formulation

Let $u, u_1, u_2, ..., u_m \in U$ be the set of all the users of a recommender system, where m is the total number of users and $i, i_1, i_2, ..., i_n \in I$ be the set of all items in the system's database containing n the number of items, which are accessible to users in U. Let $f : U \times I \to R$ be a utility function for measuring usefulness of an item from the set I to a user who belongs to U and R is an ordered set of real numbers . Then, for each user $u \in U$, the objective of the recommender system is to find an item $i' \in I$ which is yet to be known by the user u and that maximises the user's utility function. More formally,

$$\forall u \in U, \quad i'_u = \arg\max_{i \in I} f(u, i) \tag{2.1}$$

where i'_u is an item which is not known by the user u yet.

The utility of an item is represented as an interaction which indicates the usefulness of an item to a user. It can vary based on the application however, such interaction is usually a *rating* either given by the user or assigned by the system. It is important to highlight that the utility function f is not defined on the entire user-item space ($U \times I$) but on some subset of the data. This is due to the fact that not all users assign ratings to all the items that they have interacted previously. Thus, f needs to be extrapolated to the whole user-item space ($U \times I$). For this, the recommendation system needs to estimate the ratings for the items which have not been rated yet. Once the unknown ratings are known, recommendations to a user can be made in two ways 1) by choosing an item with highest rating score or 2) by providing an ordered list of items based on the rating score [3]. Section 2.4 discusses various techniques used to define such utility functions and to assess the unknown ratings and Section 2.5 shows how to validate the performance of such techniques.

2.2 BRIEF HISTORY OF THE RECOMMENDER SYSTEMS

In the early 90s, Goldberg et al. [87] introduced the idea of a system to act as a mediator where users do not need to know each other to receive recommendations. They developed a system named *Tapestry* which filters through annotated items (e.g. news articles, emails) and recommends related items to the matching mailing list of like-mined users. This system disseminates users' interests and computes similarities among users by following the heuristic of *like-minded people share similar interests*. Following this idea, several recommendation based projects have been developed. The *GroupLens system* [220] suggested netnews (i.e. news on the internet) based on the intuition that every

Early works on recommender systems

time a user reads a Usenet News article, this shows the user has a preference towards that article. By capturing those preferences as ratings and using the ratings of similar readers, personal preferences were generated as a part of the article header. The *Ringo system* [239] provided recommendations for music artists (or musicians) using a similar technique of [220]. The Video recommender [105] employed a similar approach of like-minded users as [220] to support recommendations through e-mails and the web among a virtual community of movie lovers. The People Helping One Another Know Stuff (PHOAKS) system recommended news articles from the internet [262].

By the late 90s, RS had gained popularity in both academia and industry. In academia, several research workshops and conferences were established. Examples of these conferences include Recommender Systems (RecSys)[9], which quickly became the premium conference on recommender systems, and other mainstream highly ranked conferences, such as Special Interest Group on Information Retrieval (SIGIR) or The Web conference[10], which started to include tracks and sessions on recommender systems.

Progress in the RS research community

On the Industry side, companies started to deploy RS to attract customers and increase their revenues [205]. In addition, several competitions became prominent. A key competition for the development of RS has been the Netflix challenge.[11] A million dollar prize incentive was offered as a part of this challenge where participants were asked to improve the performance of the Netflix's in-house software application *CineMatch* system by 10%. This competition boosted the design and development of recommendation methods which are now embedded in a wide range of commerce and content applications

Netflix Challenge

---

9 https://www.recsys.acm.org/
10 https://www.thewebconf.org/
11 https://www.netflixprize.com/

Table 2.1: Examples of diverse recommender systems and their corresponding domains

Domain	Some examples
Movie	MovieLens[12], NetFlix[3], Amazon.com[1],
Music	Ringo[239], Spotify[4], YouTube[5]
News	Reuters[13] , Google News[7]
Scholarly	Suggest[14], Google Scholar[15]
Travel	Expedia[16], booking.com[17]
Fashion	Online fashion retailers (e.g. ASOS.com[18])
Jobs	LinkedIn Jobs[19], Glassdoor Jobs[20]
MatchMaking	Tinder[21]
e-Learning courses	Coursera[22], Udemy[23]
Games	Google play[24], Apple store[25]

(both online and offline). In Table 2.1, we list some of the existing RS and the domains on which they have been applied.

Even though these systems initially started based on the idea of *like-mined users* (i.e. user similarity), approaches have now expanded to include a broader range of methods including content similarity, and hybrid methods combining content and user similarities. The research field keeps evolving not only with the development of new recommendation methods, but also with the refinement of evaluation

---

12 https://movielens.org/
13 https://uk.reuters.com/
14 https://www.mendeley.com/
15 https://scholar.google.com/
16 https://www.expedia.com/
17 https://www.booking.com
18 https://www.asos.com/
19 https://www.linkedin.com/jobs/
20 https://www.glassdoor.co.uk/Job/
21 https://tinder.com/
22 https://www.coursera.org/
23 https://www.udemy.com/
24 https://play.google.com/
25 https://apps.apple.com/

methodologies and metrics and the application of RS to new domains
and diverse forms of data. We discuss the emergence of recommend-
ation problems, the evaluation of approaches and the progress on
evaluation methods in the following sections.

2.3 COMPONENTS OF RECOMMENDER SYSTEMS

A recommender system helps its users by discovering potential relev-
ant items for them. In the process, several components play a vital part
in understanding users' requirements for delivering potential desir-
able items to the users. Hence, users, items and their interactions are
an important part of any RS. In the following sections, we introduce
the various components involved in the recommendation process.

2.3.1 *Modelling Users and Items*

Every recommender system needs to comprehend its users and items
for the provision of personalised recommendations. RS model users
by creating user profiles based on their preferences and model items
by creating item profiles based on their features. Below we introduce
both items and users along with the essential information required to
create their profiles.

2.3.1.1 *Item and Item representation*

'Item' is the general term that is used to denote what the system
recommends to users [171]. Items vary based on the objectives of
RS. For example, movies in movie recommendations, books in book
recommendations etc. RS capture features of items and based on

these features, items' profiles are created. For instance, for a movie, movie features such as actors, genre, title, synopsis of the movie can be considered. To build items' profiles, features are represented in variety of formats including vectors [134], matrices [49, 97, 183] etc.

2.3.1.2 *User and User profiles*

'User' is the general term that is used to denote the target of the recommendation. RS create users' profiles by inferring users' preferences (interests). One common approach of gathering users' interests is to collect interactions between users and items in the system such as purchasing, downloading, commenting, and rating. Users' profiles are created using such users' interactions [223].

2.3.2 *The Recommendation Task*

This section defines different ways of recommendation delivering mechanisms to users, usually known as the recommendation task. In the early days of RS development, RS were viewed as tools to predict ratings for items provided by users. With the continuous development in the domain, the focus has shifted towards delivering a list of ranked items that are relevant to users, so that users can pick the most relevant items to them. We introduce both types of recommendation tasks below:

2.3.2.1 *Rating Prediction*

As already mentioned, early works concentrated towards estimating ratings for items which are yet to have rated by the users and recommend the item with the highest rating score. One of the notable works that focus on the accurate rating prediction is the Netflix challenge.

The competition was sponsored by the media streaming company Netflix, where they offered a million dollar incentive to those (researchers and industry practitioners) who can develop a recommendation algorithm that is able to increase the performance of the Netflix system *CineMatch* by 10% by predicting the ratings of users to items.

2.3.2.2 *Item Ranking*

Analogous to the rating prediction task, the item ranking task aims to generate an ordered list of relevant items that are useful to the users. When a user wants to discover something then one expects to get a list of relevant items rather that one item which has a highest match [56]. For instance, if the predicted item is already known by the user then the recommendation has clearly failed to help the user. However, if the user is given a list (– Top N ranked list where N is the number of matched (high score) items) then the user has a choice to discover relevant item(s) from many (i.e. a recommendation list of items).

2.3.3 *Sources of User Preferences*

Data source for user modelling

To generate personalised recommendations of items that are tailored to an individual's preferences, RS require information about users, specifically their past interactions with the system. Such interactions are considered as evidence of the users' choices or users' feedback. To accurately extrapolate the utility function in Equation (2.1) , RS need enough user-item interaction data [125, 223]. These interactions can be categorised into two types, namely explicit feedback and implicit feedback.

2.3.3.1   *Explicit Feedback*

Preferences towards items explicitly expressed by the users are termed as explicit feedback [124]. Such feedback is usually provided in different forms [223, 229], such as:

- *scalar* ratings (– e.g. 1-5 stars with 1 being negative feedback and 5 being the highest positive feedback),

- *ordinal* ratings (– e.g. Likert scale selections : strongly disagree, disagree, neutral, agree and strongly agree)

- *Binary* ratings (– e.g. Good or bad, positive or negative, like or dislike)

Explicit ratings are considered to be difficult to collect as systems have to rely on their users to provide feedback. It is likely that without rewards users may not be motivated to provide feedback explicitly. Even when users explicitly provide feedback they tend to provide it for those items that they like (i.e., positive but not negative feedback is generally provided) [179]. In addition, providing explicit feedback requires a considerable amount of cognitive effort by the users and, summing their experiences in a single rating, or even in a ordinal rating, may be challenging for some users [124, 280].

2.3.3.2   *Implicit Feedback*

Feedback inferred from users' actions is referred as implicit feedback. Some examples of implicit feedback include browsing sessions [118, 188, 189], purchase history, reading lists [210, 214] etc. Unary rating (– e.g. if a user buys an item then 1 otherwise 0, which means no information or preference is provided) is one of the widely applied implicit feedback methods in the domain [223].

Even though this type of feedback is comparatively easier to collect (– generally if the recommender systems are already in the production stage), collecting implicit feedback also faces substantial challenges. For example, *no negative preferences can be gathered from implicit feedback*: Since this feedback is collected based on users' interactions and these interactions are generally considered as an indication of the users' positive preferences, no feedback is gathered about the users' negative preferences. If a user does not interact with a particular item that does not necessarily mean the user dislikes the item. Some applications, however, allow blocking or banning certain content, which may be considered as negative feedback. Nonetheless, such feedback is not available on all recommendation scenarios and/or applications. In addition, the fact that a user interacts with a particular item does not necessarily mean that the user has a positive preference towards such item, such interactions may be the result of inadvertent actions (e.g. browse, download, click-bait etc.) [237, 252]. It may also be the case that the user dislikes the item after having interacted with it (i.e after purchasing, downloading etc.). Such interaction can inject noise into the data.

2.4   RECOMMENDATION TECHNIQUES

Since the first proposition of automated RS in the early 90s, the domain is progressing continuously. As a result, multiple recommendation techniques have been proposed and developed. This section presents various recommendation techniques categorised based on two perspectives. They are an *algorithmic perspective* and a *data exploitation perspective*. While the algorithmic perspective depends on the underlying type of algorithm to estimate the relevance of the items to the target

user, the data exploitation perspective relies on the type of information that has been exploited concerning the user-item interactions to compute item relevance.

2.4.1   *Recommendation Methods: From the Algorithmic Perspective*

From the algorithmic perspective, recommendation methods are classified into two types, namely *memory-based* and *model-based*.

2.4.1.1   *Memory-based Recommendation Methods*

Memory-based recommendation methods are heuristic-based methods where a recommendation task (e.g item relevance computation) is performed based on the entire collection of previously rated items by the users [3, 103, 228]. They rely on heuristics to directly estimate item relevance. This heuristic nature makes these methods easier to implement and maintain [103, 220]. Due to the use of entire collection while generating recommendations, the results are assumed to be more accurate. However, the use of these methods may not always be feasible, particularly when it comes to large amounts of data. Examples of memory-based techniques include: cosine similarity [35, 228] and Pearson correlation [220, 239].

Memory-based recommendation methods

2.4.1.2   *Model-based Recommendation Methods*

Model-based recommendation methods can work with a subset of the data. They use Machine Learning techniques to learn a model which is then applied to the recommendation tasks (e.g. item relevance computation) [125].

Model-based recommendation methods

Some of the prominent model-based methods include: Bayesian networks [35], Singular Value Decomposition (SVD) [31], Latent factor

models [148], and Latent Dirichlet Allocation (LDA) [32] and others. Most model-based techniques learn latent factors by encoding a set of assumptions about the generative process of the observed feedback. This is done by introducing parameters aimed to interpret the user-item interactions. The optimal values for such parameters are learned in the training phase, by applying a suitable loss function to minimise errors between the predicted values and the actual values (preferred by the users). The fine-tuning of different parameters is needed to make the model learn latent preferences from the data. With better tuning and minimum error, model-based methods have the flexibility to achieve efficient results. However, they may suffer from various problems: (i) overfitting, which occurs when the training data fits too well while learning, (ii) under-fitting, which occurs when the model is unable to capture the underlying features of the data and, (iii) lack of interpretability, as most Machine Learning models are a black box, making it difficult to understand the decisions made and the outputs provided by the model.

2.4.2   *Recommendation Methods: From the Data Exploitation Perspective*

From the perspective of data exploitation (e.g. user-item interactions, user and item features), recommendation methods are classified into three main types, namely, Collaborative Filtering (CF), Content Based Filtering (CBF), and Hybrid. We explain and review each method in detail in the following sub-sections:

2.4.2.1   *Content Based Filtering (CBF)*

CBF approach analyses characteristics (features) of items that were previously rated by a user and builds a profile of the user based on

those rated items' descriptions. Likewise, items profiles are created based on their own characteristics. Then the recommendation process match up the features of the user profile and the features of an item profile. A higher similarity between the user and the item profiles shows a stronger level of the user's interest towards the item [171, 206, 207]. In the case of Scientific Publication Recommender Systems (SPRS) domain, items are scientific publications and their features are commonly represented by means of the entities such as title, abstract, keywords etc.

More formally, an item i_n, i.e., a scientific publication, is represented in a vector format $i_n = w_{n,1}, w_{n,2}, ..., w_{n,l} \in \mathbb{R}^L$ where $w_{n,l}$ denotes the relative relevance (weight) of feature f_l for i_n, and L is the number of existing textual features. Similarly, a user u_m consists of a set of texts from items that they have previously interacted and u_m is also represented as a vector $u_m = w_{m,1}, w_{m,2}, ..., w_{m,l} \in \mathbb{R}^L$, where $w_{m,l}$ denotes the relative relevance (weight) of feature f_l for u_m. The recommendation score of an item i for a target user u is then estimated by means of similarity computing approaches. As mentioned in Section 2.4.1, these approaches can be either memory based or model based. For the memory-based, heuristic based learning such as the cosine similarity $score(u,i) = cos(u,i)$ is widely preferred. As for the model-based, learning using statistical and machine learning techniques such as Bayesian classifiers [208], clustering algorithms [4, 155], neural-network [21] based methods are more commonly considered.

There are multiple ways to represent and weight features. One of the most common ways to encapsulate features is by means of a vector representation, also called Vector Space Model (VSM) [178]. VSM assigns weights to features by using some weighing scheme. This weighing scheme determines the relevancy of a particular feature

Feature representation and weighing scheme

over the whole collection. A widely used weighing scheme is Term Frequency Inverse Document Frequency (TF-IDF), proposed by [135]. The underlying idea of TF-IDF is that the terms with the highest weight occur more often in a particular document than in the other documents, but more infrequently in the full collection. They are therefore more central and discriminate to the topic of the document. More formally,

$$w(t,d) = \frac{tf_{t,d} \, log(\frac{N}{df_t})}{\sqrt{\sum_i (tf_{t_i,d})^2 log(\frac{N}{df_{t_i}})^2}} \tag{2.2}$$

where $w_{(t,d)}$ is a weighing function for a term t in relevance to a document d, $tf_{t,d}$ is the number of times t appear in the document d, N is the number of documents in the collection.

After items are represented and weighted based on a suitable weighing scheme, user profiles $\bar{u}_m \in \mathbb{R}^N$ are built by aggregating features from all the items that the user interacted with. A utility function of item i for user u is heuristically computed by using a similarity metrics such as cosine similarity, formulated in Equation (2.3).

Memory based techniques for CBF

$$cos(\vec{u}, \vec{i}) = \frac{\vec{u}.\vec{i}}{\|\vec{u}\|\|\vec{i}\|} \tag{2.3}$$

where $cos(\vec{u}, \vec{i})$ is the utility function for computing the similarity between user u and item i, $\vec{u}$ is the vector representations of features for the user u and $\vec{i}$ is the vector representations of features for the item i. We discuss this metric further in this chapter, see Section 2.4.2.2 (–Equations 2.5 and 2.7).

Even though CBF is widely popular within heuristic based recommendation, it is also used within model-based. In this case, to identify the relevancy of an item to users, the recommendation task is generally tackled as a classification problem where the possible classes are binary, such as *relevant* or *irrelevant*, *like* or *dislike* or even 1 or 0 [125, 208]. Once the task is categorised as classification task, various machine learning models can be applied. One of the most prominent ones is the probabilistic model. We refer the reader to [125, 171, 223] for more details.

One of the advantages of CBF is that it can recommend items by matching item contents and therefore it does not suffer from the item cold start problem, which refers to the situation where items recently added to a system that have not yet been rated, hence cannot not be recommended. In addition, recommendations based on CBF have a high degree of explainability based on their features [171].

Strengths of CBF techniques

Despite their strengths, CBF systems suffer from several limitations, first issue is overspecialisation, which refers to recommending items that are too similar to the ones that are previously rated by the users (or items in the users' profiles) [171, 223]. For example, in a news recommender system, an overspecialised system may recommend the same news across different sites that have already been read by the user. This may affect users to lose interest in the recommendations and eventually on the system. Second, expert domain knowledge may be needed for content analysis specifically for extracting meaningful content features from items. Lastly, user cold start problem, CBF are unable to recommend items for those users that are new and have not yet interacted with or rated any items.

Limitations of CBF techniques

2.4.2.2 *Collaborative Filtering (CF)*

Collaborative filtering exploits similarity among users (user-based) and items (item-based) to make recommendations. It is based on the assumptions that users with similar taste have similar interaction patterns with items in the system (user-based), and that users will enjoy items that are similar to the ones they have already rated positively (item-based) [219, 223, 229]. Similarities are computed based on the user-item rating history, not on content, as opposed to CBF. Collaborative Filtering (CF) relies on ratings provided by users to items. It is, therefore, a content-independent recommendation technique [229]. CF has been extensively studied in the literature, where we can find multiple CF techniques, including memory-based and model-based methods. For memory-based the entire user-item matrix needs to be loaded into memory to produce recommendations.

2.4.2.2.1 *Memory based Collaborative Filtering*   Two types of memory based CF methods are widely adopted in the literature: (i) user-based collaborative filtering (UBCF) and (ii) item-based collaborative filtering (IBCF) [103, 220].

User Based Collaborative Filtering (UBCF) recommends items to users based on like-minded users [220]. The method finds the neighbourhood of like-minded users for the target user u. Ratings are then analysed for each item i rated by the neighbourhood $N(u)$ of the target user. User-based CF make use of the entire set of user-item rating matrix as input and typically produce outputs in two ways: (i) a numerical score indicating to what degree the target user will like or dislike an item and, (ii) a list of n recommendations where the list contains relevant items to the target user [125]. The neighbourhood

User-based Collaborative Filtering

size may vary (5, 10, 50, 100 etc.). Some of the most popular metrics to compute similarities among users include: (i) Pearson correlation [220, 239], depicted in Equation (2.4), and cosine similarity [35, 228], depicted in Equation (2.5).

$$sim(u,v) = \frac{\sum_{i \in I_{(u,v)}} (r_{(u,i)} - \bar{r}_u)(r_{(v,i)} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{(u,v)}} (r_{(u,i)} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{(u,v)}} (r_{(v,i)} - \bar{r}_v)^2}} \qquad (2.4)$$

$$sim(u,v) = \frac{\sum_{i \in I_{(u,v)}} r_{(u,i)} r_{(v,i)}}{\sqrt{\sum_{i \in I_{(u,v)}} r_{(u,i)}^2} \sqrt{\sum_{i \in I_{(u,v)}} r_{(v,i)}^2}} \qquad (2.5)$$

where $sim(u,v)$ is the similarity score between users u and v, $I_{(u,v)}$ denotes the items that both u and v have rated and $\bar{r}_u$ is the mean ratings of the items interacted by user u. The predicted rating for an unknown item i for the target user u is then computed as shown in Equation (2.6):

$$r_{(\hat{u,i})} = \frac{\bar{r}_u + \sum_{v \in N(u)} (r_{(v,i)} - \bar{r}_v) sim(u,v)}{\sum_{v \in N(u)} |sim(u,v)|} \qquad (2.6)$$

where $N(u)$ denotes a neighbourhood of the user u which contains a set of chosen number of neighbours as most similar users to the user u. The items with the highest predicted ratings are then recommended to the target user or a list of top n (– where n is the number of items on the recommendation list).

Item-based Collaborative Filtering (IBCF) recommends items to users based on item-similarity patterns. Lets say a user u_1 likes an item i_a which is similar to i_b based on the other users' opinions who liked both items i_a and i_b. Also the user u_1 has not rated the item

Item-based Collaborative Filtering

i_b therefore, based on IBCF, the item i_b should be recommended to the user u_1. IBCF examines each item on the target user's list of interacted items and discovers other items that are similar to these items based on the user-item rating matrix. It is important to highlight that items that the user already interacted with are discarded from the recommendation list.

IBCF is proposed to reduce the computation of scanning a vast number of potential neighbours which makes the computation impossible to calculate predictions in real time. This is a usual case in e-commerce industry where the number of users is likely to be higher than the number of items and their tastes can change whereas items' features remain same [228]. This allows systems to pre-compute the item based score (i.e. based on their ratings) and thus allows for the computation of recommendations in real-time without any high computation load [125, 228].

To find similar items, *cosine similarity* is established as the widely utilised metric in the domain. It measures the similarity between items represented as n-dimensional vectors (usually composed by the ratings given to items) by computing the angle between them. Let us examine this mechanism with the continuation of the example mentioned earlier in this section, let U be a set of users that rated both items i_a and i_b and the item-based cosine similarity looks like in Equation (2.7).

Popular similarity measures

$$sim(i_a, i_b) = \frac{\sum_{u \in U} r_{(u,i_a)} \cdot r_{(u,i_b)}}{\sqrt{\sum_{u \in U} r_{(u,i_a)}^2} \sqrt{\sum_{u \in U} r_{(u,i_b)}^2}} \qquad (2.7)$$

where $r_{(u,i_a)}$ is the rating given to the item i_a by the user u and $r_{(u,i_b)}$ is the rating given to the item i_b by the user u. However, the standard

cosine similarity does not consider differences in the average rating behaviours of the users (e.g. the differences in rating scale between users, as different users can have different perceptions over items). This can be solved by using the adjusted cosine measure which deducts the average ratings of the user [3, 125], see Equation (2.8).

$$sim(i_a, i_b) = \frac{\sum_{u \in U}(r_{(u,i_a)} - \bar{r}_u)(r_{(u,i_b)} - \bar{r}_u)}{\sqrt{\sum_{u \in U}(r_{(u,i_a)} - \bar{r}_u^2)}\sqrt{\sum_{(u,i_b)} - \bar{r}_u^2}} \qquad (2.8)$$

where $\bar{r}_a$, $\bar{r}_b$ are the average ratings given to the item i_a and i_b respectively. Likewise, *Pearson correlation coefficient* is also commonly applied in the domain which is formulated in Equation (2.9).

$$sim(i_a, i_b) = \frac{\sum_{u \in U}(r_{(u,i_a)} - \bar{r}_a)(r_{(u,i_b)} - \bar{r}_b)}{\sqrt{\sum_{u \in U}(r_{(u,i_a)} - \bar{r}_a^2)}\sqrt{\sum_{(u,i_b)} - \bar{r}_b^2}} \qquad (2.9)$$

2.4.2.2.2 *Model based CF methods*   Model-based CF methods use Machine Learning (ML) to predict user ratings for unrated items. The model-based methods built on a subset of the rating matrix. It offers the benefits of both efficiency and scalability if large amounts of data need to be processed to generate recommendations. Some examples of model-based CF are Latent factor models (e.g. Matrix Factorisation (MF)[148], LDA[32], latent model[108]), Probabilistic models (e.g. Probabilistic latent semantic model [107]) among others. We discuss some of the most popular model-based techniques below.

Model-based Collaborative Filtering (CF)

Matrix Factorisation: Recommender systems based on Matrix Factorisation (MF) identify latent features from the data (– user-item rating matrix) by performing dimensionality reduction over a highly sparse rating matrix. MF decomposes user and item rating matrix into a subspace of latent factors, aiming to capture implicit preferences

Matrix Factorisation (MF) model

of users and items. This is possible by characterising both users and items by vectors of hidden factors [125].

The idea of exploiting latent factors is not a recent observation, Deerwester et al. [62] discover the latent factors in documents in the information retrieval domain applying Latent Semantic Analysis (LSA) method. LSA attempts to infer concepts from documents implicitly and makes it possible to retrieve relevant documents even if it does not contain many matching words of the user's query [125]. A large matrix of document vectors can be collapsed into a smaller-rank estimation in which highly correlated and co-occurring terms are captured in a single factor. [227] applied such concept for rating prediction task. In contrast to LSA, the SVD method cannot perform well with a large sparse matrices which are common in RS. One of the notable works in the early MF based recommender system is by Funk et al. [78] where they factorise only observed ratings instead of the whole matrix, in which each user u is assigned a vector of latent features automatically inferred from the data and each item i is assigned a vector in the same feature subspace. Intuitively, latent features aim to capture latent relationships in the data. Then, by applying a dot product of latent feature vectors, ratings for the item given by the user are estimated.

More formally, in a user-item rating matrix $U \times I$, each user u is assigned a vector $p_u \in \mathbb{R}^f$ of latent features automatically inferred from the data and each item i is assigned a vector $q_i \in \mathbb{R}^f$ in the same subspace. By conducting a dot product between latent feature vectors q_i and p_u, the overall interests of the user u on the item i can be calculated in the form of rating as shown in Equation (2.10).

$$\hat{r}_{ui} = q_i^T p_u \tag{2.10}$$

where $\hat{r}_{ui}$ is the predicted rating for the item i given by the user u. Even though computing the mapping of each item and user to factor vectors $q_i, p_u \in \mathbb{R}^f$ is a computationally expensive task, once it is completed, it can be used to predict the ratings that users will give to items (see Equation (2.10)). Also MF focuses on a subset data to learn latent factors (p_u and q_i), minimising the squared errors on the set of known ratings. This process is depicted in Equation (2.11) [148].

$$\min_{q*,p*} = \sum_{(u,i)\in K} (r_{ui} - q_i^T p_u)^2 + \lambda(\|q_i\|^2 + \|p_u\|^2) \qquad (2.11)$$

where k is the set of the user-item (u, i) pair for which ratings r_{ui} are available. λ is a constant which prevents the model from overfitting. Multiple approaches are available to minimise overfitting through a regularisation such as Stochastic gradient descent [78] and alternating least squares (ALS) [27] among others [148].

As we observed by our previous descriptions, CF methods display a series of **advantages**. These methods are content independent, since they are based on the user-item rating matrix. As they are content independent, they can recommend any items even the ones that are dissimilar to those items that are previously interacted. Despite the above listed strengths CF techniques also suffer from multiple **limitations**. Referring to [3, 38], we list some of the widely recognised CF weaknesses below:

Strengths and Weaknesses of CF techniques

- **The cold start problem** affects both, new users and new items. Recommendations can not be provided for users and items for which no ratings are associated.

- Not all users provide ratings to items that they have interacted with, which may derive on **data sparsity**. CF rely on the overlap

in ratings across users or items, they can suffer when the rating matrix is sparse. There may be many items that are rated by only few users in such case the sparsity of rating can be high and finding overlap between users can be challenging in such scenario [4].

- **Gray sheep problem** implies users with unusual preferences, and who do not fit with any groups, may not obtain benefit from CF based recommendations as CF methods will struggle to find similar users with unique preferences [3].

- Extra precautions need to be made to validate that the ratings obtained from users are genuine. Some users may deliberately provide positive or negative ratings to obtain a benefit or to discredit other users/items [223].

2.4.2.3   *Hybrid Approaches*

Hybrid methods jointly exploit the strengths of various approaches. They aim to alleviate the shortcomings of one individual method by combining the strengths of other methods [38, 39]. For instance, the cold start problem suffered by CF approaches can be mitigated by using the contents of items, following the CBF approach. Following this premise, [38] proposed a series of hybrid methods combining different approaches and categorised hybrid methods into seven types:

- *Weighted:* The scores of different recommendation components are combined numerically. Recommendations are generated based on the highest combined scores.

- *Switching:* The system chooses among various recommendation components and applies the selected one to provide recommendations.

- *Mixed:* Recommendations from different RS are presented together.

- *Feature Combination:* Features derived from different knowledge sources are combined together and given to a single recommendation algorithm. This thesis follows this type of hybrid method where we capture and explore different features (citation knowledge) for generating recommendations.

- *Feature Augmentation:* One recommendation method is used to compute a set of features. These features are used by a different recommendation method to provide the final recommendation.

- *Cascade: RS* are given strict priority, with the lower priority ones breaking ties in the scoring of the higher ones.

- *Meta-level:* One recommendation method is applied to produce a recommendation model, which is then input to another method.

Some of the early hybrid methods are focused on the combination of CBF and CF methods. For example, P-Tango recommends news by combining CBF and CF using a weighted average function [52]. In the domain of scientific publications, hybrid approaches have also emerged [72, 267], focused on recommending scientific publications to users by combining CBF and CF methods.

Within the hybrid approaches, the last few years have seen the emergence of **Factorisation Machine (FM)** to provide recommendations[218]. Factorisation Machine is a supervised ML technique [218] famous for its ability to reduce the dimensionality problem. In addition to help reducing the user-item rating matrix sparsity problem. More importantly, FM can help to incorporate other features in the original user-item rating matrix by extending the matrix [218]. In case

of movies, these features can include genres, actors, categories, etc. In case of research paper recommendations features such as title, abstract, citations, etc. could be incorporated as features in the original matrix. Since it combines other features with rating matrix, it can be classified as feature combination type of hybrid methods from [38]'s hybrid categorisations.

As mentioned earlier, FM model allows to include extra features, from which higher order latent interactions between users and items can be modelled. According to [218], a second order FM model suffices for sparse matrix and is formulated as below:

FM for recommendation

$$\hat{y}(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle v_i, v_j \rangle x_i x_j \tag{2.12}$$

where the model parameters that have to be estimated as: $w_0 \in \mathbb{R}$, $w \in \mathbb{R}^n$, $\mathbf{V} \in \mathbb{R}^{n \times k}$ the bracket operator represents the inner product of two vectors of size k as shown below:

$$\langle v_i, v_j \rangle = \sum_{f=1}^{k} v_{i,f} . v_{j,f} \tag{2.13}$$

where v represents k-dimensional latent vectors associated with each variable (i.e., users and items) and $k \in \mathbb{N}_0^+$ is a hyperparameter that defines the dimensionality of the factorisation. w_0 is a global bias, w_i models the strength of the ith variable, $\hat{w}_{i,j} = \langle v_i, v_j \rangle$ models the interaction between the ith and j^{th} variable.

Note that, Equation (2.12) is similar to MF model, it contains a global bias as well as user/item specific biases and includes user-item interactions. According to [218], if we assume that each $x(j)$ vector is

only non-zero at positions u and i, we can get traditional MF model
as shown below:

$$\hat{y}(x) = w_0 + w_i + w_u + \langle v_i, v_u \rangle \tag{2.14}$$

The main difference between the previous two 2.12 and 2.14 is that
FM introduces higher order interactions in terms of latent vectors that
are also affected by categorical data. This means that the models go
beyond co-occurrences in order to find stronger relationships between
the latent representations of each feature. For more detail, how FM is
applied in SPRS where rating matrix is combined with other features
is explained in Section 5.4.1.

2.5 EVALUATION

Evaluation is a key part of designing and validating recommendation
systems. Since the early study of recommender systems developed in
the 90s, the research on the evaluation of RS has been ongoing and
is still an active research topic in the domain [238]. Some validation
approaches quantify how close the predicted ratings match the ac-
tual ratings given by the user while others consider the frequency
with which a recommender system makes the correct classification of
whether an item is relevant or not for the user [41]. Despite different
evaluation strategies, one noticeable focus in the community is a shift
from rating prediction to item ranking and other related prominent
tasks are to benchmark evaluation methods and metric and compare
different recommendation algorithms in the domain.

In this section, we review three specific elements of the evaluation of RS namely, evaluation methods, evaluation metrics and user-centric evaluation aspects. Below we discuss each of these elements in detail.

2.5.1   *Evaluation Methods*

Two types of evaluation methods are widely applied in RS: offline and online.

2.5.1.1   *Offline Evaluation*

Offline evaluation is usually performed without the active participation of users. It is based on the development of evaluation benchmarks that are used to automatically compare recommendation approaches.

These benchmarks usually contain: (i) a set of users, (ii) a set of items and, (iii) a user-item rating matrix encapsulating users' preferences. To develop such benchmarks approaches typically rely on real-world RS (i.e., systems that are already in production stage), from where users, items and preferences can be collected. One example of such benchmarks is MovieLens[26], which contains ratings given by actual users to movies. The collected data for the benchmark is split into a training set and a test set. The training set is used to train the recommendations algorithms, as well as to tune parameters. The test set contains the ground truth (i.e., user preferences) that are used to assess the results provided by the trained recommendation algorithm.

One important aspect of offline evaluations is how the data is split for evaluation. Data splits should emulate as much as possible the reality of the recommendation system. For this purpose different

---

26 https://grouplens.org/datasets/movielens/latest/

methods have been proposed to generate training and test sets. We list here some of the most popular methods.

- Random split: The majority of methodologies split the data at random, e.g., selecting 80% of the user-item interactions for training, and the remaining 20% for test. Usually this procedure is repeated several times, averaging performance values in order to get a more robust evaluation. For which, k-fold cross-validation approach is widely used where the data is split into k roughly equal sets (usually consider as folds). Then, k-1 folds are used for training and the remaining one is held out as ground truth. The process is repeated k times, selecting each time a different fold as test set [228].

- Time split: When time is relevant for the recommendation, and the timestamp for each interaction is available, the dataset is split based on time. In general, older interactions are used for training while the newest ones are used for test. [40]. This approach will be applied for the evaluation present in Chapter 5.

A key advantage of using evaluation benchmarks is that evaluation can be conducted in an automatic manner and it is easy to reproduce. Hence, it is a prominent approach used in research, since it helps to compare approaches and selecting the best performing algorithms. We have indeed selected this approach to evaluate the work conducted in Chapter 5. It is however not advisable to completely rely on offline testing and deploy recommendation methods in production without online testing [24]. Evaluation benchmarks may not contain up-to-date information (e.g., users' interests may have changed since the creation of the benchmark) and recommendation algorithms may perform differently on the live system.

2.5.1.2   *Online Evaluation*

Online evaluation requires active participation of users to assess the RS. There are two main types of online evaluation methods: user-studies, and real-world systems.

Evaluation of real-world systems is used for already deployed systems with a sufficient large base of users. These systems are generally found in industry. To assess these systems, businesses analyse the behaviour of real customers using techniques such as A/B testing [28, 146]. In A/B testing, a subset of users interacts with the system to be evaluated, while the rest of the users interact with the normal system, which acts as baseline. The performance of the system to be evaluated is measured relative to the baseline. Metrics such as number of visited links, click-through rate, downloaded items, product revenue etc. are applied in these type of assessments. This type of evaluation allows businesses to collect information directly from real users, and test their recommendation approaches.

User studies, on the other hand, evaluate RS by monitoring the interactions of a set of test users with the system in a control environment. The users' actions are recorded and used to assess the quality of the recommendations [142]. Quality is assessed both, quantitatively (based on metrics) and qualitatively (asking users about their experience). When conducting a quantitative assessment one can measure, for example, the time to complete a certain task with the help of the proposed RS. Qualitative assessments, on the other hand, gather information directly from the users. This includes judgements over the recommended items (e.g. relevant or irrelevant, good or bad, Likert scale -1 to 5, strongly agree to strongly disagree-) based on the users' perception of the different aspects (accuracy, novelty, diversity, etc.)

of such items. User studies allow collecting fine-grained information from users but are expensive to conduct. This evaluation approach has been selected for the work conducted in Chapter 4.

2.5.2 *Evaluation Metrics*

Several evaluation metrics have been proposed over the years to assess the performance of RS. In this section, we present some of the most popular metrics used for evaluation. These metrics can be divided into two main types: (i) Error-based metrics and (ii) Ranking-based metrics. In the early days of RS, the key task was to predict ratings. Error-based metrics were, therefore, used for evaluation. Once the focus shifted towards producing item rankings as the results of the recommendation, ranking-based metrics have been commonly applied evaluation. In the following subsections, we describe some of the most popular error-based and ranking-based evaluation metrics.

2.5.2.1 *Error-based Evaluation Metrics*

To evaluate the ability of a system to correctly predict a user's preference for a specific item, error-based metrics are applied. Let's imagine we have a RS that generates predicted ratings $\hat{r}_{ui}$ for a test set T of user-item pairs (u, i) for which the 'true' ratings r_{ui} are known. r_{ui} are the ground truth ratings obtained through an online or an offline evaluation method. Error-based metrics measure the error between the predicted and the 'true' ratings.

Two of the most popular error-based metrics include: Root Mean Square Error (RMSE), presented in Equation (2.15) and Mean Absolute Error (MAE), presented in Equation (2.16).

Root Mean Square Error (RMSE) measures the error between the predicted and the actual ratings. This metric penalises large errors.

$$RMSE = \sqrt{\frac{1}{|T|} \sum_{(u,i) \in T} (\hat{r}_{ui} - r_{ui})^2} \tag{2.15}$$

MAE is an alternative to RMSE. It measures the average deviation between the predicted ratings $\hat{r}_{ui}$ and the actual rating values r_{ui} for all evaluated users $u \in U$ and all items in the test set (T). The smaller the MAE value, the more accurate the predicted ratings for items are. It is formulated in Equation (2.16).

$$MAE = \frac{1}{|T|} \sum_{(u,i) \in T} (|\hat{r}_{ui} - r_{ui}|) \tag{2.16}$$

2.5.2.2 *Ranking-based Metric*

Ranking-based metrics assess the performance of recommendation methods that output a ranked list of items as recommendation. Some of the most popular ranking-based evaluation metrics include: Precision, Recall, F1 measure, Mean Average Precision (MAP) and Normalised Discounted Cumulative Graph (nDCG).

Precision (P) is the fraction of retrieved documents that are relevant [178] (see Equation (2.17)).

$$P = \frac{TR}{TR + FR} \tag{2.17}$$

where TR is the number of true relevant (– recommended items that are actually relevant), FR is the number of false relevant, that is, representing wrong items as relevant. Since users may loose interest after the first recommended items, different cutoffs are also used

for evaluation, e.g., P@5, P@10, P@20 etc., which evaluates precision for the first five, ten, and twenty items respectively. This is called Precision@N (see Equation (2.18)) where N is the number of highly ranked recommendations for the target user.

$$P@N = \frac{1}{|U|} \sum_{u \in U} \frac{|Rel_u@N|}{N} \tag{2.18}$$

Recall (R) is the fraction of relevant items that are retrieved [178] (see Equation (2.19))

$$R = \frac{TR}{TR + FN} \tag{2.19}$$

where TR is the number of true relevant (recommended items that are actually relevant), FN is the number of false irrelevant. As for precision, recall at different levels, Recall@N is also considered.

$$R@N = \frac{1}{|U|} \sum_{u \in U} \frac{|Rel_u@N|}{Rel_u} \tag{2.20}$$

where Rel_u is the set of relevant items for the user u and $Rel_u@N$ is the set of relevant items for the user u available at the top N rank of the recommendation list.

To get the global precision and recall score at the top N, the individual scores are averaged across all the test users as shown in Equations 2.18 and 2.20 respectively.

F1 measure (F1) is defined as a harmonic mean between precision (P) and recall (R) metrics [178]. It can be computed using the formula exhibited in Equation (2.21).

$$F1@N = 2 * \frac{P@N * R@N}{P@N + R@N} \tag{2.21}$$

Average Precision (AP) measures how good the rank is by running the precision from 1 to N where N is the number of items that are recommended (see Equation (2.22)).

$$AP@N = \frac{1}{m} \sum_{k=1}^{N} P(k) \cdot rel(k) \tag{2.22}$$

where N is the cutoff point which determines the number of items to be recommended, and m is the number of relevant items in the full space of items $rel(k)$ is an indicator for noting whether k^{th} item was relevant ($rel(k)$=1) or not ($rel(k)$=0). AP is applied to a single user.

Mean Average Precision (MAP) derives from applying AP to all the users in the test set, as shown in Equation (2.23).

$$MAP@N = \frac{1}{|U|} \sum_{u \in U} AP(u) \tag{2.23}$$

where U is a set of all users in the test set.

Normalised Discounted Cumulative Graph (nDCG) considers multiple levels of relevance in the ground truth. The more relevant and item, the more it contributes to the quality if it is recommended, but

adjusted to its relative position in the ranking. This metric is defined in Equation (2.24).

$$nDCG@N = \frac{1}{|U|} \sum_{u \in U} \frac{DCG_u@N}{IDCG_u@N} \qquad (2.24)$$

where, $DCG_u@N$ is defined as:

$$DCG_u@N = \sum_{k=1}^{N} \frac{2^{rel_u[i]-1}}{log(k+1)} \qquad (2.25)$$

Other relevant evaluation metrics not listed here include: Hit Rank (HR), Average Reciprocal Hit Rank (ARHR), Receiver Operating Characteristics (ROC), Area Under the ROC curve (AUC), and Mean Reciprocal Rank (MRR), among others. More about these metrics, and the evaluation of RS can be found in [104, 238].

2.5.3 *User-centric Evaluation Aspects*

Evaluating relevancy is sometimes not sufficient, since users may have different goals and objectives when interacting with a RS. They may want to discover the most novel items, or the most diverse ones. To address this, a variety of user-centric aspects of evaluations have been considered. We will briefly introduce here some of these aspects, but for more information about user-centric evaluation metrics the reader is referred to the following works [28, 150, 270, 299].

- **Serendipity** measures how unexpected and useful the recommended items are to the users [54]. Serendipitous items are both *unexpected* and *useful* to the user [54, 82]. To increase serendipity in the recommendations, Ge et al. [296] focused on the incor-

poration of time rareness, which is achieved by considering less popular items (e.g. with less ratings) and dissimilarity (items that match least with the target user). Lu et al. [172] explored the long-tail approach, recommending items from the bottom of the recommendation list (i.e., choosing items with lowest scores). [252] claimed dissimilar users can provide serendipitous recommendations.

- **Novelty** This aspect encapsulates how new and relevant an item is for a user [43, 299]. Analogous to serendipity, novelty may not cater for unexpectedness, for instance, say a recommender system suggests movies that were written by a user's favourite director. If the system recommends a movie (directed by the user's favourite director) which the user was unaware of, then the system will be considered as a novel but maybe not serendipitous, since the user is likely to discover the movie on their own.

- **Coverage** quantifies the percentage of items for which a recommender system is able to make predictions [103, 226]. Coverage can be further categorised into two types, namely, item coverage and user coverage.

  - Item coverage is the percentage of items included in the recommendation list over the number of potential items.

$$Coverage_{item} = \frac{n}{N} * 100 \tag{2.26}$$

– User coverage is the percentage of users for whom the recommender system was able to generate a recommendation list over the number of potential users.

$$Coverage_{user} = \frac{u}{U} * 100 \qquad (2.27)$$

- **Diversity** measures how diverse or different, items are in the recommended list. There is no clear understanding of how much diversity is actually desirable for any domain or how large of a compromise on accuracy should be tolerable. As a result, this calls for more standardised multi-metric optimisation and evaluation schemes in the context of a specific recommendation task [123].

2.6 OPEN CHALLENGES

Multiple challenges are present in the area of RS, and although some solutions to address these challenges have been proposed in the last two decades [3, 216, 221, 238], many of them still remain open. In this section, we cite some of the most common challenges present in the field of RS.

2.6.1 *Understanding and Gathering User Preferences*

Understanding users' preferences is a very complex task. Since the early development of RS, ratings are used to capture users' preferences. These ratings are captured via explicit or implicit feedback. However, as mentioned earlier, preferences gathered via implicit feedback tend to be noisy, since not always users click or download items they

actually like them. To better comprehend users' preferences multiple works have added to the use of ratings and the consideration of other aspects, including: demographic features [25, 197], social information [8, 162, 211], context (e.g. time [150], location [66]), etc. Users may have different preferences at different times or in different contexts (e.g., office environment vs. house environment). In our thesis, we explore the use of citation knowledge, as an additional source of information that can help us to better understand and capture user preferences.

2.6.2 *Data Sparsity*

Data sparsity refers to the amount of available interactions between users and items. It happens when the number of user-item pairs (i.e., interactions between a user and an item) is very small compared to the number of possible user-item pairs. First of all, not all users interact with all items in a collection (e.g., not all users have seen all Netflix movies), and users not always provide preferences for items they have interacted with. As shown in Table 2.2, the number of empty cells is substantially higher than the filled cells with rating entries. This phenomenon can be measured as Equation (2.28) as mentioned in [125].

$$sparsity = 1 - \frac{|R|}{|U|.|I|} \tag{2.28}$$

where $\{u_1, u_2, ..., u_m\} \in U$ be all the users and $\{i_1, i_2, ..., i_n\} \in I$ be all the items in the matrix shown in Table 2.2, and R be all the ratings provided by users to items in the matrix of size $m \times n$.

Table 2.2: user-item rating matrix

item / user	i_1	i_2	i_3	i_4	...	i_n
u_1	x					x
u_2						
u_3		x				
u_4						
$\vdots$						
u_m						

2.6.3 *Cold Start*

When a new user starts interacting with a recommender system, the user will not have any rating associated with them. Any recommender system relying on users' preferences to generate recommendations may therefore not be able to generate any recommendations for the user. This problem is known as the **user cold-start problem**. Similarly, if a new item is introduced into the system (a new product, a new movie, etc.) and it hasn't been rated by any user yet, it is likely that the item does not appear in any recommendation list. This problem is known as the **item cold-start problem** [3].

Several approaches have been proposed in the literature to mitigate these problems. These includes: (i) recommending popular items to new users, (ii) exploiting additional information from users (such as demographic data, etc.) [112, 206], (iii) recommending unrated items to users by considering their content and matching it with the users' profiles, etc. We address the item cold start problem in this thesis (see Chapter 5) when addressing the problem of recommending the latest scientific publications to a user. Note that, in this real use-case scenario (where users need to be up to date with the latest scientific

publications), new publications may not have received any rating yet. We propose the use of contextual knowledge to target this problem.

2.6.4 *Trust*

If users do not trust the RS then there is a high chance they might not return [238]. In order to improve the trust, the RS needs to be consistently useful to the users and should not intrude the users' information without informing them. However, both building and measuring trust in RS is still an open problem. While we think is important to highlight this issue, we don't address the challenge of trust in this work.

2.6.5 *Explainability*

Explaining why the system is providing certain recommendations, and no others is needed for multiple reasons including: building trust into the RS, ensuring that biases are not captured and perpetuated by the underlying algorithms, etc. We hypothesise that the use of citation knowledge could help enhancing the explainability of existing RS for scientific publications. While we do not assess this hypothesis in our work, we consider it an important element of our future line of work.

2.6.6 *User-centric evaluation*

The fact that the recommended items are relevant may not always be the most important aspect for a user. Users may seek other aspects such as usefulness [104, 182], novelty [270], etc. Conducting user-

centric evaluations is key to increase the overall user satisfaction with RS. Although we do not target this particular challenge in this thesis we also consider it to be a relevant aspect of our future work.

2.7 CONCLUSION

In this chapter, we have provided an overview of the area of RS. As we have seen, these systems are essential tools in modern days, since they help us addressing the information overload problem in many different domains.

We have presented in this chapter: (i) a brief overview of the history of RS, (ii) a description of their key components, including users, items and ratings, as well as the different techniques used to gather such ratings, (iii) a summary of the most popular recommendation techniques, both from the algorithmic perspective as well as from the data exploitation perspective, (iv) a description of some of the most popular evaluation methodologies and metrics for RS, and (v) a summary of some of the challenges that we aim to address in this thesis, as well as some that we consider as part of our future work.

3

SCIENTIFIC PUBLICATION RECOMMENDER SYSTEMS

As pointed out in the introductory chapter, a key part of this thesis has been to conduct a systematic literature review of the landscape of academic Recommender Systems (RS). Academic RS involve the recommendation of a wide variety of items including research papers, books, scientific events, and patents among others [26, 152]. In this chapter, we specifically review those works that focused on the recommendation of scientific publications, which is our primary goal. On reviewing the literature, we aim to answer two key research questions of this thesis:

RQ1 : Which types of citation knowledge have been used in RS for scientific publications?

This first question aims to explore the types of citation knowledge that have been used so far for the recommendation of scientific publications. As mentioned in Chapter 1, until very recently, the full textual content of scientific papers have been hidden behind paywalls, hence the use of fine-grained notions of citation knowledge for the recommendation of scientific papers is still under explored.

RQ2 : Which are the different recommendation tasks that have been proposed in the literature of RS for scientific publications, and how citation knowledge has been applied for each of these tasks?

This second research question aims to investigate the variety of recommendation tasks that have been addressed in the literature (e.g., recommending papers for an ongoing piece of work, recommending papers to a user, etc.) and how, the above identified citation knowledge has been used for each of these tasks.

The rest of the chapter is structured as follows. The survey scope and the methodology followed to conduct this literature review is presented in Section 3.1. The different recommendation tasks that have been tackled in this domain are presented in Section 3.2, followed by the different notions of citation knowledge that have been considered in the literature - see Section 3.3. How recommendation targets and items are modelled is presented in Section 3.4. Note that, while the target of a recommendation is generally a user in this domain, various works have also considered a given piece of work as the target of the recommendation. We follow by describing the different recommendation approaches that have been used for this domain (see Section 3.5) and how they have been evaluated (see Section 3.6). We conclude this chapter by reflecting on the open challenges of the domain in Section 3.7.

3.1   SURVEY SCOPE AND METHODOLOGY

To understand and analyse the landscape of recommender systems for scientific publications, we conducted a systematic literature review. In this section, we present a methodology followed for conducting this review.

To gather papers about the domain, we selected two well-known academic databases, namely Elsevier Scopus[1] and Web of Science[2].

Queries to search relevant publications

To retrieve relevant scientific publications from these databases, we compose a set of queries including the following concepts, namely, recommender systems and scientific publications, and their linguistic variances.

For **recommender systems**, we included variances such as: recommend*, recommendation systems*, recommender system*, recommendation service*, recommender service*, recommendation approach*, recommender approach*, recommendation model*, recommender model*, recommendation method*, recommender method*, recommendation algorithm*, recommender algorithm*, recommendation application*, recommender application*, recommendation engine*, recommender engine*, recommendation framework*, recommender framework*.

For **scientific publication**, we included variances such as: "research paper*", research publication*", "research article*", "research document*", "research literature*", "scientific paper*" , "scientific publication*", "scientific document*", "scientific article*", "scientific literature*", "scholarly publication*", "scholarly paper*", "scholarly document*" , "scholarly literature*", "scholarly article*" , "academic publication*", "academic paper*", "academic document*", "academic article*", "academic literature*", "related publication*", "related paper*", "related document*", "related literature*" , "related article*", "digital librar*" , "citation recommend*" , "citation-based*" are included for the search of publications.

After running these queries, we obtained a total of 741 papers. These papers were manually filtered by reading through title and abstract. This helped us to discard those papers that, although mentioning the

---

above concepts, were not related with the domain of interest. This list was updated progressively during the course of the PhD, reviewing a final list of 202 papers for this survey. Note that, some of these papers are survey papers, or papers that refer to the creation of datasets and other resources. Hence, although the full list of 202 papers have been considered for our analysis, the statistics presented in the following subsections are extracted from 176 papers.

We review these papers considering various dimensions including: (i) the recommendation task they target, (ii) how items and recommendation targets (e.g., user) are modelled, (iii) which recommendation approaches are used, and (iii) which citation knowledge is used and how is it incorporated.[3]

3.2   RECOMMENDATION TASKS

This section presents the main recommendation tasks identified in the literature. Tasks are defined based on the target of the recommendation. Broadly we have identified two main tasks: (i) **recommending scientific publications for a given user**, (ii) **recommending scientific publications for a given piece of work**. The second task can be further categorised in several sub-tasks, where the piece of work can be represented: (i) as a paper, (ii) as a set of papers, (iii) as a snapshot of text (titles, abstracts, etc.), or (iv) as an ongoing (-yet to be published-) manuscript.

The categorisation of the analysed works in these tasks is presented in Table 3.1 and briefly explained in the sections below.

---

3 It is important to highlight that, some papers were vague in providing information, but, to the best of our knowledge, we extract knowledge appropriately.

Table 3.1: List of reviewed papers that are categorised based on different recommendation tasks

Recommendation task		References
A piece of work	A paper	[2, 5, 15, 16, 29, 42, 45, 50, 64, 72, 80, 83–85, 94, 97, 98, 113, 117, 120, 128–131, 138, 140, 143, 144, 151, 152, 159, 160, 164, 177, 182, 183, 195, 196, 199, 202, 217, 232, 234, 236, 240, 244, 247, 259, 279, 282, 288, 297]
	A set of papers	[50, 64, 72, 120, 128–130, 138, 151, 152, 232, 234, 236, 240]
	A manuscript	[88, 92, 101, 102, 127, 168, 248, 284]
	A snapshot of text	[18, 30, 34, 37, 44, 69, 71, 74, 79, 96, 114, 116, 132, 154, 165–167, 180, 181, 184, 192, 203, 215, 233, 235, 274, 289, 291, 298]
A user		[6–14, 17, 21–23, 33, 46–49, 51, 58–61, 63, 66, 68, 75, 90, 91, 93, 95, 110, 112, 118, 119, 121, 134, 145, 150, 155, 163, 169, 173, 175, 176, 188, 189, 191, 197, 198, 201, 204, 210, 211, 214, 231, 242, 245, 250–258, 260, 261, 265, 267, 268, 271, 272, 275–278, 281, 283, 285–287, 290, 292, 294, 295]

We can also observe, how from the analysed publications, 51% works focus on recommendations of scientific publications for a given piece of work, while 49% are focused on the recommendation of scientific publications for a user (see Figure 3.1).
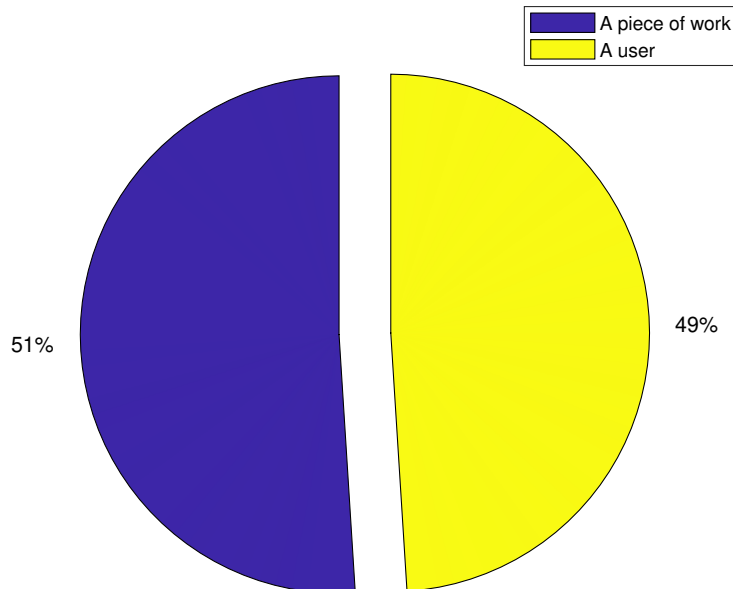
Figure 3.1: Recommendation tasks distribution

3.2.1   *Recommending Scientific Publications For a Given Piece of Work*

As mentioned before, recommending scientific publications for a given piece of work covers various targets. Each of these sub-tasks is driven by a different purpose. For example, users may be trying to find appropriate references for an ongoing paper (hence the target of the recommendation is an on-going manuscript), or they may be interested in finding scientific publications that are similar to a paper of their interest (in this case, the target of the recommendation is a paper).

Moreover, recommending scientific publications for a given piece of work can also be understood as a 'retrieval problem' since scientific publications are recommended/retrieved for a given work - or query - (represented as a paper, a set of papers or a snapshot of text). In this thesis, we are treating this task as a recommendation problem, where the final goal is to recommend relevant items to a user, whose preferences are represented from a piece of work provided by the user to the system. It is relevant to notice that, as opposed to user preferences, preferences captured from a piece of work are static in nature, i.e., they do not change over time [174, 251].

We can also observe that, among the 51% of the studied works that focus on recommending scientific publications for a given piece of work, 43% focus on recommending scientific publications for a given paper, 16% for a set of papers, 9% for an ongoing manuscript and 32% for snapshot of text or textual content. The distributions of the various recommendation sub-tasks can be seen in Figure 3.2.

It is important to highlight that, while some of these works have explored the use of citation knowledge to provide recommendations, due to the lack of access to the full-content of scientific publications,

Recommendation for a paper
Recommendation for a set of papers
Recommendation for a manuscript
Recommendation for a snapshot of text

32%

43%

9%

16%
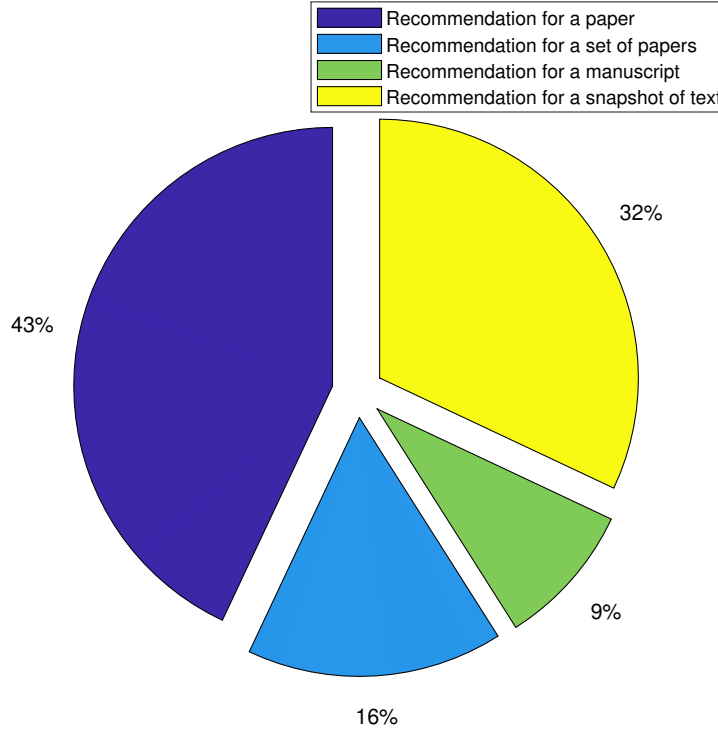
Figure 3.2: Different recommendation tasks distribution within a piece of work type

many of these works focus on shallow notions of citation knowledge extracted out of meta data [98, 167, 244, 279, 291].

Moreover, in the studied works, we observe how different notions of citation knowledge have been studied in isolation, but rarely in combination with one another. More information about the types of citation knowledge that have been so far explored to target this task, and how they have been used, is provided in Section 3.3.

In Chapter 4, we address this task focusing on the use of fine-grained notions of citation knowledge, extracted after parsing the full content of a large collection of scientific publications. We also focus on experimenting with different types of citation knowledge both, in isolation, and in combination.

3.2.2   *Recommending Scientific Publications For a User*

In this type of recommendation task, the target of the recommendation is a user. User preferences are generally captured based on their interests, which may change over time [174, 253]. For example, researchers may change the focus of their research over the years, or they may be working on various topics simultaneously.

Some of the works that have attempted to target this problem, we can highlight the following ones for their use of citation knowledge [49, 95, 163, 251–254, 267] (see Table 3.2 for more reviewed works).

However, there are two issues that can be pointed out about these works. First, as with works targeting the above task (i.e., the recommendation of scientific publications for a given piece of work), these works either explore shallow notions of citation knowledge where this knowledge has been extracted out of the metadata of the scientific publications and not out of their content, or they explore more fine-grained notions of citation knowledge but explored mainly individually, but rarely in combination. For example, [69, 160] exploit citation intention but not citation section, [253] explored citation section but not citation intention.

The other important issue that we observe when analysing these works is that, while they focus on the recommendation of scientific publications to a user, they do so independently on when those papers were published. However, to the best of our knowledge, only the works of [95, 272] have focused on the recommendation of the newly published scientific publications to a user. Note that, this reflects an important real-life problem where researchers and practitioners need to be up to date with the latest developments in their scientific fields. However, this is also a difficult problem to target since, as previously

mentioned in Chapter 1, traditional recommendation methods, such as Collaborative Filtering (CF), do not work in this setting. These methods are not able to recommend items for which ratings (preferences) have not yet been provided (also known as the item cold-start problem, as explained in Chapter 2).

In Chapter 5, we address the task of recommending newly published papers focusing on: (i) the recommendation of recent scientific publications to a user and, (ii) the use of deeper notions of citation knowledge, extracted after parsing the full content of a large collection of scientific publications. We also focus on experimenting with different types of citation knowledge both, in isolation, and in combination.

3.3 CITATION KNOWLEDGE

With the aim to answer RQ1, this section provides a summary list of the different types of citation knowledge that have so far been used in the literature of RS for scientific publications. Existing works capture citation knowledge in the following forms:

- **Citation graph:** A citation graph captures citation relations between papers. Nodes represent citing papers and edges represent relations between such papers based on their citations. Relations in the citation graph can be either directed –i.e., they capture the explicit source and target papers of the citations [183, 259] –, or undirected –i.e., they do not consider which paper is the one citing and which paper is the one being cited [95]. The hypothesis behind the use of the citation graph for recommendation is that if a user u authors a paper p_i that cites a paper p_j, then u is expressing a preference for p_j and p_i is

also expressing a preference for p_j. As we will see later on in Chapter 5 this is not always correct, since papers can be cited for different purposes, including criticisms. However, because of the availability of metadata, the citation graph is one of the notions of citation knowledge most frequently used in the literature.

- **Citation proximity:** Citation proximity is the distance between co-cited papers in a scientific publication [85]. The idea of citation proximity is that the shorter the distance between two papers that are co-cited, the strongly relevant they are to one another. This notion of citation knowledge was first conceptualised in 2009 by Gipp et al. [85] and used for the recommendation of Web pages[230]. To the best of our knowledge, this thesis has been the first one to apply the notion of citation proximity for the recommendation of scientific publications (see Chapter 4).

- **Citation context:** Citation context refers to the text surrounding the citation of a paper. The hypothesis behind the use of citation context is that this text provides an indication of the semantics with which the citation is mentioned [102, 253]. Multiple works in the literature have used the notion of citation context with a variety of purposes. When addressing the task of recommending papers for an ongoing manuscript, [102] used citation context to enrich the profile of the target manuscript. In a similar fashion, [253] used citation context to enrich user preferences, when addressing the task of recommending scientific publications for a user.

- **Citation section:** Citation section captures the particular section of a paper where the citation appears, e.g., introduction, related work, conclusions, etc. [45]. References cited in the *introduction*

section may have different importance and purpose to those cited in other sections. The assumption behind using this type of citation knowledge is that citations in different sections may have different relevance. This notion of citation knowledge was explored by [45] in combination with the citation graph. In their work, they generated a directed graph linking papers based on citations and enriched this graph with information about the positions of such citations. An edge representing a citation from paper p_1 to paper p_2 may be tagged with multiple labels, if p_1 cites p_2 in several sections. This graph is then used to provide recommendations for a given paper. They claimed to improve the performance of their citation section based recommender systems, especially the maximum improvement achieve by the section presenting *background information* (i.e. Introduction), followed by the *method section* [45].

Prior works based on citation section categorisation

- **Citation Intention:** Citation intention captures the objective of the citation. Citations can be used to provide background, compare the current work, etc. While the notion of citation intention has been extensively used in the area of scientometrics, which measures and analyses the impact of scientific literature, to the best of our knowledge, only a few works have used this notion of citation knowledge for the recommendation of scientific publications when addressing the task of recommending scientific publications for a given paper [160] and for a given snapshot of textual content [69]. The hypothesis of using citation intention is that not all intentions may reflect the same relevance. For example, papers that are cited to provide background information may be more relevant to the users than those papers cited to point to future lines of work. Since citation intention

Table 3.2: Reviewed papers categorised into recommendation task and their utilisation of different types of citation knowledge.

Citation Knowledge	Recommendation task	
	A piece of Work	A user
Citation graph	[15, 16, 34, 42, 44, 45, 50, 64, 71, 72, 80, 88, 94, 97, 98, 102, 113, 117, 127–131, 144, 151, 152, 159, 164–168, 182, 183, 192, 199, 202, 217, 232–234, 244, 247, 248, 259, 274, 279, 288, 291, 297, 298]	[7, 23, 48, 49, 58, 93, 95, 150, 163, 250–254, 267, 283, 286, 287]
Citation Context	[34, 50, 69, 71, 74, 79, 102, 113, 116, 127, 140, 144, 160, 165, 166, 168, 284, 289]	[251, 253, 254]
Citation Section	[45, 144]	[251, 253, 254]
Citation Proximity	[140, 143][4],[83–85][5]	
Citation Intention	[69, 160]	

is rarely explored in the RS domain, we studied various types of citation intention from other domains (e.g. citation analysis, scientometric etc.) and collected different categorisations that can be applied in the scientific publication RS. Table 3.8 details the gathered citation intentions from the prior works.

Table 3.2 provides a summary of the different types of citation knowledge that have been used across works in RS considering the recommendation tasks described in Section 3.2. As we can see in this table, while most reported approaches have focused on the use of the citation graph and then citation context, very few works up to date have exploited the notion of citation proximity, citation intention and citation section either individually or in combination.

In the following sections, we analyse the studied works based on different dimensions of the recommendation problem, including how target (e.g., user) preferences are modelled, how research papers (i.e., item features) are captured, and which methods are applied widely to

---

4 This is our proposition
5 Only conceptualised, no experimental results provided

recommend items to a target (user, paper, etc.) by considering both the target's preferences and the items' features.

3.4 ITEM AND TARGET MODELLING

In this section, we focus on describing how items (i.e.scientific publications) and targets (users, papers, etc.) are modelled. How items and targets are modelled is a key aspect of the recommendation process. In this section, we also describe how different notions of citation knowledge have been incorporated in the modelling process.

3.4.1 *Item Modelling*

In this thesis, the items to be recommended are scientific publications. To model scientific publications, previous works have considered two main types of information: (i) **metadata about the publications** (i.e. title [195, 282], abstract [2, 203], keywords [37, 137], authors [252, 253], publication date [159, 160], publication venue [30, 80, 168], list of references [72, 88, 101, 102, 183], etc.) and (ii) the **full textual content of the publications [16, 165, 253]**.

Typically, terms from metadata information such as, title, abstract, keywords etc. are extracted in the form of n-grams [75, 195] or topics [131] and used to model items. In terms of full content, works have also focused on the extraction of topics [184, 201] and on the extraction of citation knowledge [74, 248, 251] in order to model items.

While metadata tends to be widely available, the full textual content of scientific publications has been generally (and it is still in many

cases) hidden behind paywalls [279]. Hence, a wider range of works have been focused on the use of metadata to model such items.

Figures 3.3 and 3.4 show the distribution of features used on the studied works to model items (i.e., scientific publications). In Figure 3.3[6], we group all aspects of citation knowledge (i.e. citation graph, citation context, citation section, citation proximity and citation intention) as one feature and present it as **citation knowledge** along with other features such as title, abstract, bibliography (reference) list etc. We detail the different notions of citation knowledge employed by the reviewed papers in Figure 3.4 and in Table 3.3. It is essential to mention that some works use more than one of the listed features for modelling items and this has been accounted in the presented distribution.

According to Figure 3.3, citation knowledge and reference list are the most common features used to model items, with citation knowledge used to model items in 45% of the studied works, and the reference list used in 42% of the studied works. Other relevant features to model items include: title (33%) and abstract (32%) etc. If we look in detail into the works that use citation knowledge to model items (see Figure 3.4 and Table 3.3), we however can observe that 89% of those works are mainly based on the use of the citation graph (- which is formed using the reference list -). About 23% of the works mention the use of citation context, only 6% mention the use of citation section, 2 (3%) prior works have use some form of citation intention, and non of the prior works have used the notion of citation proximity in RS to model items. In terms of combining multiple notions of citation knowledge to model items, only the works of Sugiyama et

---

6 We have also listed the prior works and their use of different features to model item in Table A.1

al. [251, 253, 254] have explored the combinations of various types of citation knowledge, namely, citation graph, citation section and citation context to model items.

Once the features are collated, different item representations – e.g., vectors, matrices, or knowledge bases – are built to gather and exploit the above mentioned knowledge.
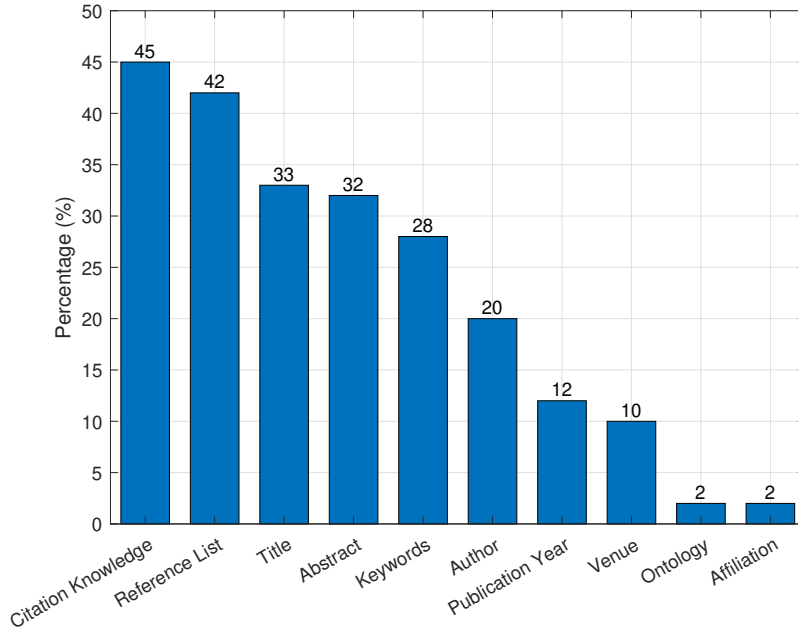
Figure 3.3: Features that are employed in the literature for modelling items

3.4.2 *Target Modelling*

As previously mentioned, the field of RS for scientific publications considers two types of targets: (i) a piece of work (represented as a paper, a set of papers, an ongoing manuscript or a snapshot of text), and (ii) a user. How these targets are modelled is different since they have different nature. In this section, we will briefly present the different techniques and features used to model these types of targets.
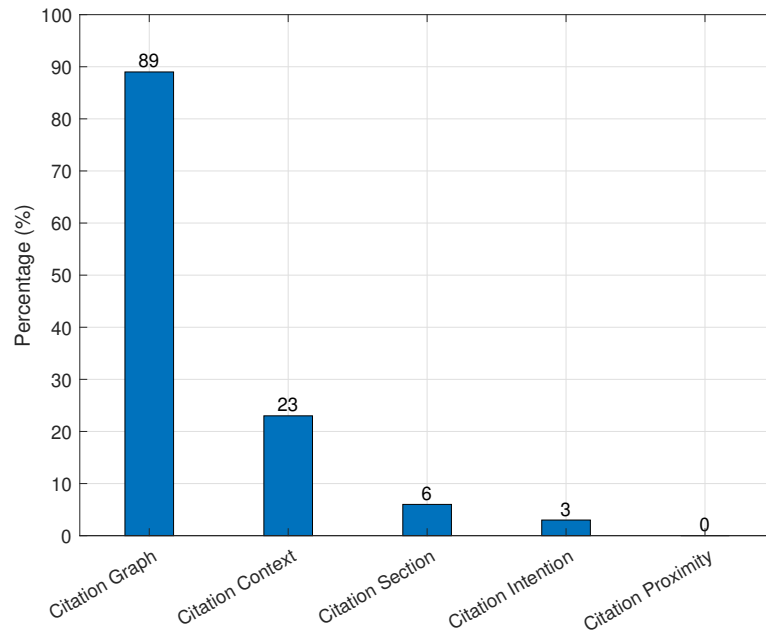
Figure 3.4: Different types of citation knowledge that are employed in the literature for modelling items

Table 3.3: List of reviewed papers utilising different aspects of citation knowledge to model items. CG stands for Citation graph, CC stands for citation context, CS stands for citation section, CP stands for citation proximity and CI stands for citation intention

References	Citation Knowledge				
	CG	CC	CS	CP	CI
[251, 253, 254]	x	x	x		
[34, 50, 71, 102, 113, 114, 127, 165, 166, 168]	x	x			
[45, 144]	x		x		
[7, 15, 16, 23, 42, 44, 48, 49, 58, 64, 72, 80, 88, 93–95, 97, 98, 117, 119, 128–131, 150–152, 159, 164, 167, 182, 183, 192, 199, 202, 217, 232–234, 244, 247, 248, 250, 252, 259, 267, 274, 279, 283, 286–288, 291, 297, 298]	x				
[74, 116, 140, 284, 289]		x			
[69]					x
[160]	x				x

User/Target preferences can be captured by considering explicit and implicit feedback. Explicit feedback may consist of *ratings* [17, 75, 204, 257, 290], scoring [93, 275] or user account with topic of interest stated by the user [197, 242]. Approaches based on implicit feedback to model a user's profile, in contrast, generally capture information from implicit actions such as browsing sessions [188, 189, 246, 276], clicks [110, 145, 214], bookmarks [51, 210, 211] and tags [204, 290], among others. In this context, it is important to note that, when limited information exists, such as authors who have published few papers or do not have many logged activities within the system, user profiles may be incomplete and inadequate to provide accurate recommendations [252]. In addition, building user profiles from the users' browsing/downloading history may lead to the development of inaccurate profiles if the user inadvertently browses or downloads irrelevant items [237]. In these scenarios, the use of citation information may be helpful to create more complete and accurate user profiles for recommendation. Sugiyama et al. [252], for example, proposed the use of the cited papers to enrich user profiles, capturing users' research interests by considering not only their past publications, but also the citations of such publications. Hence, the citing action has been adopted to capture users' preferences widely [95, 155, 250–254, 286].

A summary of the features used to model user preferences when the target of the recommendation is a user can be seen in Table 3.10. In terms of the statistics extracted from the studied papers, we can observe that to collect user preferences, implicit feedback is applied by 89% of the reviewed papers while explicit feedback is applied by only 2%. Note that, some papers have used both implicit and explicit feedback which is about 9%.

Work/Target preferences As with items, when the target of the recommendation is a particular work, 'preferences' for such target can be captured by considering metadata such as title [79, 195], abstract [120, 291], keywords [92, 128, 168], authors [15, 37], publication date [180, 289], publication venue [128, 168], and bibliography (i.e., the list of publications that are referenced in a paper) [72, 183, 248, 267]. Preferences can also be captured by considering the textual content of (the paper, set of papers, ongoing manuscript, - that constitutes the target) [79] including text around citations - or *citation-context* [252, 253, 267]. In this case, due to the inaccessibility of the full content of papers, fewer works up to date have exploited the notion of citation context to model target preferences in comparison to the citation graph. Similarly, citation knowledge in terms of citation section, citation proximity, and citation intention have been under explored in the current literature [102, 160]. The citation graph, on the other hand is a popular type of citation knowledge used to model target preferences. A summary of the different works that have used citation knowledge to capture preferences of a recommendation target, when the recommendation target is a given piece of work can be seen in Table 3.4. The distribution of features to model such target across works is summarised in Table 3.9.

3.5 RECOMMENDATION METHODS

Different recommendation methods have been designed based on the above described representations. Among the most popular recommendation methods, we can highlight Content Based Filtering (CBF),

Table 3.4: List of reviewed papers utilising different notions of citation knowledge for modelling as a target (**a piece of work**). CG stands for citation graph, CC stands for citation context, CS stands for citation section, CP stands for citation proximity and CI stands for citation intention.

References	CG	CC	CS	CP	CI
[144]	x	x	x		
[34, 50, 71, 102, 113, 127, 165, 166]	x	x			
[45]	x		x		
[15, 16, 42, 44, 64, 72, 80, 88, 94, 97, 98, 117, 129–131, 138, 151, 152, 159, 164, 167, 168, 182, 183, 192, 199, 202, 217, 232–234, 236, 244, 247, 248, 259, 274, 279, 288, 291, 297, 298]	x				
[74, 79, 116, 140, 284, 289]		x			
[69]		x			x
[160]	x	x			x

CF, and hybrid approaches.[7] A categorisation of the studied works with regards to the recommendation method that they utilised can be seen in Table 3.5. Among the studied works, the most popular method is hybrid which is used by 45% of the works, followed by CBF - 34% and CF - 22%. Note that few papers used and/or compared different approaches. For example, [182] proposed both, the use of CBF and CF.

In the following sections, we briefly describe how these methods have been applied for the recommendation of scientific publications. In addition, we categorised the studied works based on their use of these recommendation methods and their application of the various types of citation knowledge (see Table 3.6).

---

7 Note that, we have categorised works in terms of broad methods only, i.e. CBF, CF, hybrid. For example, if a paper uses multi-criteria collaborative filtering then it is categorised as CF, if a paper uses an ontology to get concepts and find similarity based on concepts then it is categorised as CBF etc.

Table 3.5: List of reviewed papers in different recommendation approaches

Recommendation approach	References
CBF	[2, 6, 10, 11, 16, 18, 22, 23, 29, 30, 37, 46–48, 51, 58–61, 69, 74, 75, 79, 80, 90, 94, 110, 116, 131, 132, 134, 138, 145, 154, 168, 177, 181, 182, 191, 195, 198, 203, 214, 215, 217, 231, 235, 240, 242, 245, 248, 252, 255, 275–277, 282, 289, 294, 295]
CF	[15, 21, 33, 42, 44, 49, 72, 83, 85, 88, 95–98, 117, 129, 130, 143, 150–152, 155, 163, 164, 169, 182–184, 199, 201, 204, 210, 244, 259, 279, 281, 297, 298]
Hybrid	[5, 7–9, 12–14, 17, 34, 45, 50, 63, 64, 66, 68, 71, 84, 91–93, 101, 102, 112–114, 118–121, 127, 128, 140, 144, 159, 160, 165–167, 173, 175, 176, 180, 188, 189, 192, 196, 197, 202, 211, 232–234, 236, 247, 250, 251, 253, 254, 256–258, 260, 261, 265, 267, 268, 271, 272, 274, 278, 283–288, 290–292]

3.5.1 *Content-based filtering (CBF)*

Content Based Filtering (CBF) approaches recommend items to a target that are similar to the items with which the target (e.g., the user) has previously interacted with. These interactions are typically citing [72], browsing [188, 246, 276], downloading [58, 93, 258, 267], reading [13], saving [272], rating [17, 75, 134, 204, 257], scoring [61, 150, 163, 275] etc.

Target profiles are created by extracting features from the interacted items. Typically, these features can be terms from title [195], abstract [2], keywords [37], citation context [69, 253], or even classification taxonomies, such as the ACM classification taxonomy[8] [6, 46, 145, 169], or DMOZ[9] [13, 189].

Extracted terms and concepts are then used to represent target profiles by converting them into machine recognisable forms (e.g. vector representation) using weighing schemes (e.g., Term Frequency Inverse Document Frequency (TF-IDF) [135]). Once the targets and

---

8 https://dl.acm.org/ccs
9 https://web.archive.org/web/20180119172207/http://www.dmoz.org/

items in the database are represented into a machine recognisable form then a similarity metric (e.g., cosine similarity [140, 144, 252, 291], Jaccard coefficient [97, 169, 177]) is applied to compute the similarity between the target and the items. The higher this similarity is, the strongly relevant items are considered to be recommended [138, 252, 254]. In addition to the well-known similarity measuring metric, some reviewed works have used citation knowledge to procure recommendation, for example, [140, 143][4] used citation proximity to compute relatedness between relevant papers to a target paper.

3.5.2 *Collaborative filtering (CF)*

Collaborative filtering (CF) approaches recommend items to a target (e.g., a user) that are preferred by like-minded users. Generally, users who rate the same items with similar score are considered like-minded. Explicit feedback given by the targets to items is captured by ratings [17, 173, 201] votes [283] or scores [59–61, 93]. However, as mentioned in Chapter 2 capturing such feedback explicitly is high-cost process, and most works rely on implicit feedback (browsing [58, 258], downloading [93]) that can be noisy and incomplete at times.

An important issue to consider with this recommendation method is that the number of targets (users) in comparison with the number of items (scientific papers papers) is significantly lower [4]. Moreover, few users are likely to rate the same papers. This creates a high degree of sparsity. Hence, finding like-minded users can be challenging. Also, CF suffers from the cold-start problem, i.e., scientific papers that are not previously rated can not be recommended (even if they are relevant) or a new user who has not provided ratings (e.g. cited, authored, scored etc.) to any publications can not get recommendations.

When the target of the recommendation is not a user, but a piece of work (e.g., a paper), preferences from the target to items are generally captured based on citations. The assumption being that a paper shows a preference for those papers it cites. In this case, instead of a user-item rating matrix, a paper-citation matrix is created for the recommendation [72, 182, 183]. While citations have been commonly considered to gather preferences this approach has also gathered criticism [26, 160]. Papers are cited for different reasons (providing background, comparing approaches, highlighting the limitations of previous works), and not all citations really express a preference. Similarly, citations may be an incomplete form of preference gathering, since sometimes, due to space limitations, important references are not included [160].

Once the rating matrix is created (capturing preferences from targets towards items) the recommendation of scientific publications is generally performed by applying traditional heuristic based neighbourhood methods (i.e. user-based collaborative filtering and item-based collaborative filtering) [72, 183]. Some works have enhanced the performance of traditional heuristic based methods by applying graph ranking algorithms such as PageRank [200], Hyperlink-Induced Topic Search (HITS) [141] and SALSA [156] [72]. These algorithms are used to obtain a score for each publication that is then used to modify the scores of the rating matrix before applying heuristic based methods for recommendation.

Some works have focused on the application of model based recommendation methods. Among those works, we can highlight the use of Matrix Factorisation (MF) [148], representing targets and items in a shared low-dimensional space, and the use of probabilistic classifiers [183], which focus on sorting items based on probability. More inform-

Table 3.6: List of reviewed papers categorised based on recommendation approaches and their utilisation of different notions of citation knowledge. Here, CG stands for Citation graph, CC stands for citation context, CS stands for Citation section, CP stands for citation proximity, CI stands for Citation intention

References	Approaches	Citation Knowledge				
		CG	CC	CS	CP	CI
[251, 253, 254]	Hybrid	x	x	x		
[34, 50, 71, 102, 113, 114, 127, 165, 166]	Hybrid	x	x			
[7, 64, 93, 119, 128, 159, 167, 192, 202, 232–234, 247, 250, 267, 274, 283, 286–288, 291]	Hybrid	x				
[45, 144]	Hybrid	x		x		
[160]	Hybrid	x				x
[140, 284]	Hybrid		x			
[168]	CBF	x	x			
[16, 23, 48, 58, 80, 94, 131, 217, 248, 252]	CBF	x				
[74, 116, 289]	CBF		x			
[69]	CBF					x
[15, 42, 44, 49, 72, 88, 95, 97, 98, 117, 129, 130, 150–152, 164, 183, 199, 244, 259, 279, 297, 298]	CF	x				

ation about model based recommendation methods can bee seen in Chapter 2.

3.5.3 *Hybrid*

Hybrid approaches jointly exploit multiple recommendation methods, commonly CBF and CF methods [39]. Hybrid approaches aim to overcome the disadvantages of individual recommendation methods by combining them together. For example, the item cold start problem present in the CF approach can be mitigated by combining CBF approach.

Various works in the literature have combined multiple recommendation methods proposing hybrid approaches. Examples include the work of [272], who combined CF and probabilistic topic modelling Latent Dirichlet Allocation (LDA)) to mitigate the item cold start problem, or the works of [253, 293], which proposed hybrid methods for alleviating the sparsity problem.

Interesting works in this direction include the work of Liu et.al [162], who combined social network data with CF to reduce the sparsity of user-item matrices, and the work of Alotaibi et.al [9] explored information about readership (reading a paper), co-readership (co-reading papers) and tags (tagging a paper) with CF, also to alleviate the sparsity problem. To uncover 'like-minded' individuals, also reducing the sparsity problem, Pera et al. [210, 211] explored the items shared across users' libraries.

Over the years, citation knowledge has been incorporated into hybrid methods, more specifically citation graph and lately citation context. Ritchie [224] showed that indexing cited articles with the terms appearing in citation context can improve the effectiveness of the document retrieval. Based on this work, [102] proposed a probabilistic hybrid model which uses a paper-citation matrix and citation context to improve the recommendation for a target manuscript. [160] created paper-citation matrix using citation graph and incorporated citation intention to find the relevance between targets and items. Likewise, [253] incorporated citation section with citation context and citation graph to recommend items to a researcher.

3.6 EVALUATION

This section summarises the different evaluation methods used in the domain in Section 3.6.1, datasets in Section 3.6.2 and metrics in Section 3.6.3 that have been applied for the validation of RS for scientific publications.

3.6.1 *Evaluation Methods*

As pointed out in Section 2.5.1, there are two types of evaluation methods to validate recommender systems, offline and online. Statistics about the use of these methods in the reviewed literature can be seen in Figure 3.5. 49% of the works follow an offline evaluation, 31% applied online evaluation, 5% employed both online and offline evaluations and 15% of the studied papers have not evaluated their systems.

Offline methods use evaluation benchmarks containing rating data that is separated into a training set that used to train the model, and a test set that used for validating it. Such evaluation benchmarks are generally gathered from real-world recommender systems from where interactions between live users and items are collected [24, 26].

Offline evaluation benchmarks for recommender systems exist in a variety of domains including: movie recommendation[10], music recommendation[11], etc. However, these benchmarks are uncommon in the field of recommender systems for scientific publications (see Table 3.7).

---

10 https://grouplens.org/datasets/movielens/latest/
11 http://files.grouplens.org/datasets/hetrec2011/hetrec2011-lastfm-2k.zip
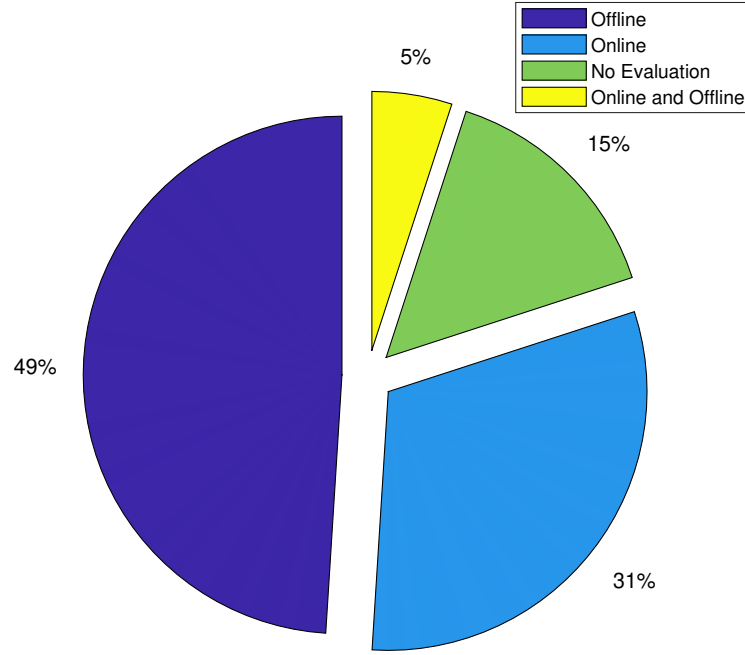
Figure 3.5: The statistics of different evaluation methods employed by the reviewed works

To tackle the lack of evaluation benchmarks researchers have considered using the reference list of each publication as ground-truth data [102, 166, 240, 248]. A cited paper is relevant to both, the paper that cites it and the users who author the paper that cites it. Note that, both can be recommendation targets. However, as previously mentioned, citations are referenced with different purposes, and a paper may be cited because it is criticised, or simply because it is popular or seminal (This is known as the Matthew effect [186]) [26, 251]. In such cases, the cited papers may not be relevant neither for the paper that cites them, nor for the users that authored such paper.

Another issue to consider on the generation of these benchmarks is the use of time-aware data split (training and test data split). Even though time-based split have been considered by some works such as [102, 192, 217]. However, they do not use it for recommending new items where new items do not have any interactions (e.g. cited). Only

the works of Ha et al. [95] and Wang et al. [272] have considered time within their evaluations for new item recommendations for users.

The **online evaluation** method is generally used by industries that already have deployed production level recommendation engines (e.g., Mendeley's suggest.[12]) In such cases A/B testing methods are used for evaluation [26].

User studies is a type of online evaluation, as pointed out in Section 2.5.1.2, have also been applied by some works in the literature. For example, the work of McNee et.al [183] made use of questionnaires to assess the familiarity and relevance of recommended papers. 120 participants were engaged in this experiment. [72] developed a web prototype and engaged 19 students on its evaluation. Participants provided a list of 5-10 query papers to the prototype, which generated a 5-item recommendation list for each target paper. Participants were asked to evaluate each list based on coverage and relevancy to the topic of interest. Similarly, [267] created a web prototype named *Tech-Lens* and engaged 110 participants in its evaluation. Each participant was asked to select a target paper. For each target paper the system provided a 5-item recommendation list. The participants were asked to evaluate each list based on the quality, familiarity and authority of the recommendation.

In our work, we have selected to use two types of evaluation, user studies in Chapter 4 and offline evaluations in Chapter 5.

3.6.2 *Evaluation Datasets*

Various datasets have been used in the past to investigate academic recommender systems. A comprehensive list of publicly available

---

12 https://www.mendeley.com/guides/web/05-mendeley-suggest

datasets is given in Table 3.7. For each dataset, the table shows a brief description of the type of data, time when accessed data, the number of users, items and ratings in the dataset, and whether the full text PDF_{av} and publication history of the users UPH_{av} are available.

Table 3.7: Publicly available datasets for academic recommender systems. Where, PDF_{av} stands for Portable Document Format (PDF) document available and UPH_{av} represents the availability of authors' publications history

Dataset	Description	Accessed /Published	Users	Items	Ratings	PDF_{av}	UPH_{av}
AMiner[13]	AMiner contains a series of datasets capturing relations among citations, academic social networks, topics, etc. We report data here about the citations dataset V11	2019	Not specified	4M	No	No	No
Open Citations[14]	Open repository of scholarly citation data	2019	Not specified	7.5M	No	No	No
Open Academic Graph[15]	Large knowledge graph combining Microsoft Academic Graph and AMiner	2019	253M	381M	No	No	No

13 https://www.aminer.cn/aminer_data
14 https://opencitations.net/corpus
15 https://www.openacademic.ai/oag/

ArXiv[16]	Open access e-prints publications in different fields such as physics, mathematics etc.	2019	Not specified	1,5M	No	Yes[17]	No
CORE[18]	Dataset of open access research publications published up to 2018	2019	No	9.8M	No	Yes[19]	No
CiteULike[20]	Dataset of users' selected bookmarks to academic papers	2019	5,551	16,980	No	No	No
Mendeley [122]	Dataset shared by Mendeley for a recommender system challenge	2010[21]	50,000	4.8M	Yes[22]	No	No
ACL anthology[23]	Corpus of scholarly publications about Computational Linguistics	2019	Not specified	22,878	No	Yes	No

16 https://arxiv.org/help/bulk_data
17 Download requester pays Amazon S3 bucket https://arxiv.org/help/bulk_data_s3
18 https://core.ac.uk/services/dataset/
19 Data needs to be requested
20 https://old.datahub.io/dataset/citeulike
21 published date
22 Anonymised data and needs to be requested
23 https://acl-arc.comp.nus.edu.sg/

SPD 1[24]	ACL anthology based papers published between 2000-2006	2019	28	597	Yes	Yes	No
SPD 2[25]	ACM proceedings based papers published between 2000-2010	2019	50	100,531	Yes[26]	No	No

24 https://acl-arc.comp.nus.edu.sg/
25 ttps://www.comp.nus.edu.sg/~sugiyama/Dataset2.html
26 Anonymised data

As we can see, most of these datasets do not provide ratings, limiting the application of a wide range of recommendation methods. Those datasets providing rating information have anonymised information, the details of users and publications, which can restrict the exploration of other features such as features from publication e.g. title, citation context etc. Moreover, none of the listed datasets captures the authors' publication history, and thus knowledge about the users, particularly their preferences (i.e., publications and references), cannot be easily captured. It is also important to observe that most of the existing datasets do not provide the full texts of publications, and hence, a wide variety of citation knowledge cannot be extracted from them.

Given the limitations of the existing datasets used for recommendation, in this thesis, we have proposed the creation of novel datasets both in Chapter 4, when targeting the recommendation of scientific publications for a given target paper, and in Chapter 5, when targeting the recommendation of scientific publications for a given target user.

3.6.3 *Evaluation Metrics*

A variety of evaluation metrics have been applied to measure the performance of recommendation approaches for scientific publications including: error-based, ranking-based and user-centric metrics. 63% of the reviewed papers evaluated their systems through the use of ranking metrics including precision, recall, nDCG, Mean Average Precision (MAP) etc. For details about these evaluation metrics the reader is referred to Chapter 2.

3.7 DISCUSSION

Through the conducted systematic literature review, we have attempted to answer two important research questions in this thesis: *RQ1: Which types of citation knowledge have been used in RS for scientific publications?* and *RQ2: Which are the different recommendation task that have been proposed in the literature of RS for scientific publications and how citation knowledge has been applied for each of these tasks?*

As we have observed, while citation graph is one of the most popular notions of citation knowledge used for recommendation, and citation context is also commonly used in those approaches that have access to the full text of scientific publications, other notions of citation knowledge, such as citation section, citation proximity, or citation intention have been barely considered in the literature of RS for scientific publications. Citation intention, in particular, has been applied to other fields like scientometrics[269], but, to the best of our knowledge, this thesis is the first work to apply citation proximity knowledge and also the combinations of different notions of citation knowledge (citation graph, citation context, citation section and citation intention) for the recommendation of scientific publications.

We have also observed that these notions of citation knowledge have been applied for two main tasks: (i) the recommendation of scientific publications for a given piece of work and (ii) the recommendation of scientific publications for a user. In different tasks, different notions of citation knowledge have been applied to model both, items and recommendation targets, as well as to enhance the performance of recommendation methods. We have observed that:

- Many of the studied works use shallow notions of citation knowledge due to the lack of access to the full-text of scientific articles. In particular, the use of citation proximity has not been applied before for the recommendation of scientific publications, and citation intention has not been applied yet for recommending scientific publications to users.

- Many of the studied works explore different types of citation knowledge in isolation, but not in combination with one another.

- Existing works are focused on providing recommendations independently on when papers are published. To the best of our knowledge, only few works ([95, 272]) have previously focused on recommending the most recent scientific publications to a user.

- Only industry-based labs with RS in production can conduct online evaluations. The rest of the studied works either rely on offline evaluations using 'incomplete' benchmarks (lack of preferences/ratings, lack of full textual content of papers, small-scale datasets, etc.) or on high-cost user studies where a small number of users is engaged in the evaluation.

In this thesis, we aim to target the above mentioned limitations of existing works. Chapter 4 addresses the task of recommending scientific publications for a given piece of work (understood in our thesis as a paper, following the majority of studied approaches). We propose in this chapter, novel recommendation methods that incorporate an unexplored notion of citation knowledge in the domain of RS for scientific publications (citation proximity). We also explore the combination of citation proximity and citation context, proposing

recommendation methods that explore simultaneously multiple types of citation knowledge.

In Chapter 5, we focus on the recommendation of scientific publications to a user, and in particular, we aim to target the real-world problem of recommending users the most recent scientific articles. As previously mentioned, to the best of our knowledge this problem has not been widely addressed in the literature, and it is a particularly challenging one. Note that, traditional recommendation methods, such as collaborative filtering, do not work in this particular scenario, since they are not able to recommend items for which ratings have not been previously provided (item cold start problem). We address this complex scenario by means of the use of deep notions of citation knowledge including the citation graph, citation section, citation context, and more importantly, citation intention. To the best of our knowledge, this thesis is the first one that explores the notion of citation intention for the recommendation of scientific publications to users. As in Chapter 4, we also explore the effectiveness of these notions of citation knowledge in combination.

In addition, this thesis contributes to the scientific field with two important evaluation benchmarks focused on the two above mentioned tasks. These benchmarks are detailed in Chapter 4 and Chapter 5.

Other challenges/gaps, have been identified while conducting this literature review. While we do not address these challenges in this thesis we list them below. They are a reflection of the possible opportunities and open research lines within this field.

- **Situational awareness:** A PhD student starting on a research field may have completely different needs that a senior researcher who have been part of the field for many years. Similarly, the types of papers that a researcher may be looking for while

working on a survey paper may be different than the ones she
is looking for while working on a concrete problem. Under-
standing all these different situations, and adapting to them, is
something that has not yet been targeted in the domain of RS
for scientific publications.

- **Sparsity:** As we previously mentioned, while preferences are
  gathered either implicitly (by downloading, reading, citing pa-
  pers, etc.) or explicitly [4] there is a much higher number of
  publications than users expressing their preferences. Hence
  the user-item rating matrix is generally really sparse. The de-
  velopment of techniques addressing this problem could help
  improving the performance of recommendation methods, such
  as collaborative filtering.

- **Heterogeneity of research fields:** Interdisciplinary is key in
  research. However, finding relevant papers belonging to differ-
  ent disciplines is a problem not yet explicitly addressed in the
  current literature. While some recommendations are provided
  based on citations, and co-citations, unless the interdisciplinary
  between fields is already established, relevant papers from differ-
  ent disciplines will not be recommended. We believe this is an
  important research line that should be explored in future work.

- **Explainability:** To the best of our knowledge, very few of the
  proposed RS do provide explanations of why those scientific
  publications, and no others, are the ones being recommended.
  Providing explanations could help users to better assess whether
  the recommended papers are valuable to them in accordance to
  their information needs [271].

- **Reproducibility:** Some analysed works do not provide concrete details of their implementations, and do not share code and/or datasets, making it difficult and sometimes impossible, to reproduce those approaches and provide appropriate comparisons. For example, prior works such as [9, 214, 240] mentioned the use of terms from research publications for building item profiles but do not indicate from where such terms were extracted (e.g., from the title, the abstract, etc.). This slows down the research and development of the field, since approaches can not be reproduce and formally compared against each other.

Table 3.8: Types of citation intention proposed by prior works grouped into top-level classes

Work	Background	Description	Discuss, Compare, Control		Use	Extension	Motivation	Future
			Criticise	Support				
[136]	Background		Compare or Contrast		Use	Extension	Motivation	Future
[185, 263]		Other, Own, Textual	Contrast			Basis	Aim	
[53]	Background Information		Result Comparison		Method			
[67]			Comparison		Technical Basis		Fundamental Idea	
[45]	Background		Alternative Approaches, Comparison		Methods			
[133]	CONC-OP (Conceptual - Operational)	ORG, PERF (Organic - Perfunctory)	NEG, JUX (Negational, Juxtapositional)	CONF (Confirmative)		EVOL (Evolutionary)		

Ref	Type	Neutral	Compare and Contrast	Weakness	Positive	Use	Based on	Significant	Future Research
[264], [249]		Neutral	Compare and Contrast	Weakness	Positive				
[158]	Standard	Neutral, Co-Citation	Corroboration, Contrast	Negative	Positive	Practical, Supply, Discover	Based on	Significant	
[160]	General		Comparable				Based on		
[81]	Assumptive	Tentative, Reader Alert	Contrastive	Negational	Affirmational	Methodological, Use	Interpretational		Future Research
[193, 194]	Type O (Other)		Type C (Compare)				Type B (Basis)		
[187]				Contrast	Corroborate		Expand		
[20]		Neutral	Mathematical	Contrast	Correct	Useful			
[1]		Neutral	Comparison	Criticising	Substantiating	Use	Basis		

	Related Work	Comparison	Using	Extending		
[99, 100, 269]						
[212]		Comparison / Limitation / Support		Basis		

Table 3.9: List of reviewed papers categorised based on target preferences when the target is **a piece of work**. Here, Ci stands for Citing, Ti stands for Title, Ab stands for Abstract, Ke stands for Keywords, Au stands for Author, Ve stands for Venue, Py stands for Publication year, Ft stands for Terms from free text, Tx stands for Taxonomy, Ck stands for Citation knowledge

References	Ci	Ti	Ab	Ke	Au	Ve	Py	Ft	Tx	Ck
[168]	x	x	x	x	x	x				x
[182]	x	x	x	x	x					x
[98, 117]	x									x
[284]	x	x	x		x	x				x
[102]	x	x	x							x
[232]	x	x		x			x			x
[236]	x	x								x
[128]	x		x	x		x				x
[131]	x		x							x
[92][27]	x			x	x					x
[297]	x				x	x				x
[50, 288]	x				x					x
[64]	x						x			x
[101, 202, 248][27]	x									x
[42, 45, 72, 88, 94, 97, 113, 127, 130, 144, 151, 152, 159, 160, 164, 183, 199, 244, 247, 259, 279]	x									x
[129]	x									
[5]		x		x	x					
[140, 143]										x
[83–85]										
[217]		x	x		x	x				

---

27 No mention of entities to extract terms

[195, 282]		x	x						
[177, 234]		x		x					
[196]		x							
[2, 138]			x						
[15]				x					
[80]					x	x			
[16][27]									x
[29, 240][27]									
[18, 30, 34, 37, 44, 69, 71, 74, 79, 96, 114, 116, 132, 154, 165–167, 180, 181, 184, 192, 203, 215, 233, 235, 274, 289, 291, 298]								x	
[196]									x	

Table 3.10: List of reviewed papers categorised based on target preferences when the target is a user. Here, A stands for authoring, B stands for browsing, T stands for tagging, Bm stands for bookmarking, Sc stands for scoring, Rd stands for reading, Cl stands for clicking, R stands for rating, V stands for viewing, D stands for downloading, P stands for has profile, Sr stands for searching, Ac stands for accessing, Sh stands for sharing, Vo stands for voting, Cm stands for commenting, An stands for annotating, Ci stands for citing

References	Implicit Feedback															Explicit Feedback			
	A	B	T	Sv	Bm	Rd	Cl	V	D	Sr	Ac	Sh	Cm	An	Ci	Sc	R	P	Vo
[58, 258]		×							×				×						
[119]		×									×								
[118, 188, 189, 198, 245, 276, 278, 294, 295]		×																	
[210]			×	×	×	×													
[275]	×		×	×									×			×			
[21]			×	×															
[51, 91, 211]			×		×														
[17, 75, 134, 204, 257, 287, 290]			×																
[214]				×	×	×	×												
[22]				×		×								×			×		

[112]	[33, 265, 272, 281]	[283]	[8, 9, 66]	[7, 256]	[13]	[6, 14, 68]	[93]	[163]	[59–61, 150]	[173]	[110, 145]	[11, 201, 231]	[191, 271, 277, 292]	[267]
		×												
					×									
										×		×		
							×	×	×					
		×					×			×				
								×						
		×												
							×							×
		×							×			×		
		×					×		×	×				
×		×			×	×								
		×	×	×										
×	×													
				×	×									

[197, 242]																		x	
[285]										x			x						
[121]											x								
[23]													x						
[95, 155, 250–254, 286]	x														x				
[63]															x				
[10, 12, 46–49, 169, 175, 255, 260, 261, 268]	x																		
[90]																			
[176]																			

Part II

PROPOSED SOLUTIONS

4

RECOMMENDING SCIENTIFIC PUBLICATIONS FOR
A GIVEN PIECE OF WORK USING CITATION
CONTEXT AND CITATION PROXIMITY

As pointed out in the previous chapters, one of the main tasks addressed by academic Recommender Systems (RS) is the recommendation of scientific publications for a given piece of work. In the literature, that piece of work is captured by a paper, an on-going manuscript, or components of a paper such as its title and abstract. In this chapter, we focus on addressing this particular task, i.e., recommending scientific publications for a given piece of work, and more specifically for an input paper. Our goal is to address RQ3, investigating whether the use of citation knowledge can help improving existing recommendation methods in the above task. For such purpose, we define two types of citation knowledge –citation proximity and citation context– and propose novel recommendation methods that exploit such knowledge. An evaluation of our methods is conducted by means of user studies.

> **RQ3 : When addressing the task of recommending scientific publications for a particular piece of work, can citation knowledge help improving existing RS?**

The remainder of the chapter is structured as follows: Section 4.1 motivates the work, Section 4.2 describes the data used to conduct this research, Section 4.3 presents the proposed recommendation methods, Section 4.4.1 describes the conducted experiments, and Section 4.5 provides some conclusions.

4.1 INTRODUCTION

Recommending scientific publications for a given piece of work

Recommending scientific publications for a given piece of work is a relevant task. It helps researchers to find relevant literature with respect to their ongoing work: writing a paper for which relevant references need to be cited, finding relevant papers within a particular research field, etc.

As observed in Chapter 3, multiple recommendation methods have emerged in the literature of RS attempting to address the above problem [45, 140, 160, 183]. Existing methods tend to suggest relevant publications based on metadata, such as title [195], abstract [172] and references [183, 267]. However, the use of metadata only may not be sufficient or entirely reliable. Titles and abstracts are sometimes written in a style to draw attention rather than to comprehensively describe a piece of work [26]. Moreover, these methods do not explore the use of fine-grained citation knowledge that can be extracted from the full texts of scientific publications. It is important to highlight that until very recently the full textual content of scientific publications was hidden behind pay walls. Thanks to the open access movement, more publications than ever are now freely available, forming an important resource, from which fine-grained knowledge has yet to be explored.

In this chapter, we explore the full textual content of publications and introduce two notions of citation knowledge to address the above mentioned recommendation task: (i) citation proximity and (ii) citation context.

Citation proximity

The main hypothesis behind the idea of citation proximity [85] is that "the closer the documents are co-cited, the strongly related they are", i.e., publications whose citations co-occur in a close vicinity tend to have a certain relation (convey similar ideas, cover similar topics

or methods, etc.); when authors write research papers, they tend to strengthen their arguments by citing other scientific publications in their papers [252, 253].

An example of the notion of citation proximity is displayed in Figure 4.1. This image captures an excerpt from [147], where the citations (*Billsus and Pazzani 1998*) and (*Sarwar et al. 2002*), cited in a close vicinity, are associated with the same topic –*dimensionality reduction*. Similarly, the citations (*Deerwester et al. 1990*) and (*Zha et al. 1990*) also refer to a similar topic but specific to the challenges.

When introducing the idea of citation proximity [85], the authors proposed a method to measure citation proximity: Citation Proximity Index (CPI). This method computes a proximity value based on whether documents are co-cited within the same sentence, paragraph, chapter, journal, etc. While this method was used to measure proximity for the recommendation of web pages [84], this idea has rarely been applied to the recommendation of scientific publications. In this chapter, our first aim is to assess whether the notion of citation proximity can be (i) extracted, considering that we have access to the full textual content of scientific publications, and (ii) applied to enhance existing recommendation methods.

One of the issues identified when doing this first assessment is that recommendations based solely on citation without any content *Citation context* may suffer from topic drifting [26, 113]. Topic drifting can be defined as moving away from the main topical concept of the target paper. Not considering the topical focus of the citations could derive on treating all citations equally and recommend, for example, scientific publications that define mathematical concepts when looking for recommendations about image classification; as machine learning based research papers tend to use mathematics and acknowledge them ac-

cordingly. To address this problem, we propose to combine citation position and citation context (i.e., text around citations) to generate paper recommendations.

The hypothesis is that by knowing the context behind citing a particular reference, we can improve the performance of existing RS. Note that, these two notions of citation knowledge have not been previously combined in the literature, which constitutes one of the key innovations proposed in this chapter.

We propose an evaluation based on user studies to assess whether citation proximity and citation context knowledge do indeed help enhancing existing recommendation methods. In a first user study, we evaluate our proposed recommendation model, which applies only citation-proximity knowledge, and we compare it against existing baselines. A second user study is conducted to assess combination of citation-proximity and citation context, also comparing this model against existing baselines. Results show that our proposed model, utilising both citation proximity and citation context, outperforms the baselines by around 25% at different Normalised Discounted Cumulative Graph (nDCG) cutoffs.

4.2 DATASET BUILDING

Collecting Data

To conduct our research we needed to compile appropriate datasets. As shown in Table 3.7, existing datasets used in RS for scientific publications do generally provide a wide range of metadata, but do not provide the full textual content of scientific articles. We have therefore compiled and built appropriate datasets from which our previously described notions of citation knowledge (citation proximity and citation context) could be extracted.

Alternative algorithms based on dimensionality reduction (Billsus and Pazzani 1998, Sarwar et al. 2002) showed early promise for commercial application and have been adapted in many ways to deliver high performance and high quality recommendations. These methods, commonly based on singular value decomposition, start with the recognition that a user–item ratings matrix actually has too many independent dimensions and thus loses some of the underlying relationships between user tastes. The algorithms reduce the dimensionality to an underlying set of latent taste dimensions, expressing both user preferences and item characteristics in terms of these latent dimensions. Once this dimensionalization (which is costly to compute) is established, prediction and recommendation are quite efficient, even for very large datasets. One challenge is that the singular value decomposition algorithm is too expensive to re-compute for each new rating that arrives. In the SVD literature a technique called *folding-in* is used to incorporate new data into an existing decomposition (Deerwester et al. 1990). After a significant amount of folding-in the decomposition loses its accuracy, and must be updated. More recent algorithms enable the SVD to be updated "in place" (Zha et al. 1999), but these algorithms are themselves complicated and computationally intensive.

Generic Description

Challenges

Figure 4.1: A toy example for our proximity based hypothesis.

To gather data, we relied on CORE[1], the world's largest collection of open access research papers. CORE provides a large collection of papers, including their full text, making this data suitable for our research purposes. We originally downloaded from CORE a dataset of two million scientific publications. This set of scientific publications has been used to generate two datasets for the research presented in this chapter. These datasets are described in detail in Sections 4.4.1.1 and 4.4.2.1. The first dataset contains 368,385 publications with 142,157,561 co-citations among them. The second dataset is an extension of the first one containing about two million publications with 665,330,651 co-citations among them.

To parse the content from the collected papers in Portable Document Format (PDF), we made use of the GeneRation Of Bibliographic Data (GROBID) parser.[2] We implemented a software library that receives as input the text of a scientific publication parsed by GROBID, identifies where the citations are in the text, and extracts for each citation its position, and its context, captured as the sentence where the reference has

Parsing Data

---

[1] https://core.ac.uk/services/dataset/
[2] https://github.com/kermitt2/grobid

been cited as well as the preceding, and the following sentences. Note that, if the citation is located at the start or end of a paragraph, the preceding or following sentences, respectively, are not extracted. During the parsing process, the reference list of the scientific publication is also extracted and process to provide additional information about the citations (the title of the paper, its list of authors, published date, etc.). More information about how citation information is extracted and applied for recommendation is explained in Section 4.3.

4.3   PROPOSED APPROACH

This section presents the types of citation knowledge explored to address the task of recommending scientific publications for a given paper in Section 4.3.1, and our proposed recommendation methods, which exploit such knowledge in Section 4.3.2.

4.3.1   *Citation Knowledge*

We have selected two types of citation knowledge to address this task: citation proximity and citation context.

Citation proximity analysis

Citation Proximity represents a distance between co-cited publications [85]. As mentioned before, the main hypothesis of this approach is "the closer two documents are co-cited, the strongly related they are." In [85], CPI is computed as follows: if two documents are co-cited in a same sentence level, then $CPI = 1$, if they are cited in same paragraphs, then $CPI = 1/2$. If documents are co-cited in same chapters of the paper, $CPI = 1/4$, if they are co-cited in the same edition of the journal, $CPI = 1/8$, and if they are co-cited in the same
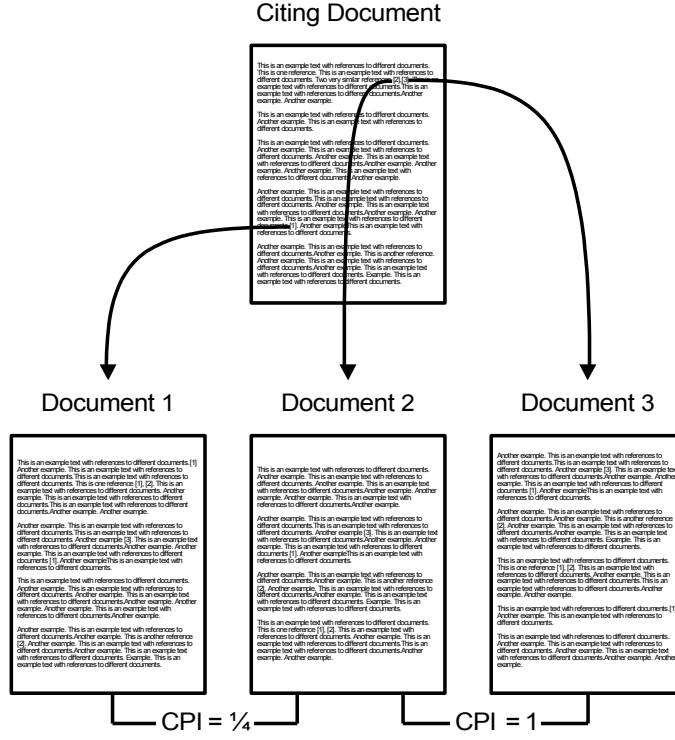
Figure 4.2: Illustration of Citation Proximity [85]

journal but in different edition, $CPI = 1/16$. An illustration of this computation is shown in Figure 4.2. The final citation proximity value is the average of all the proximities among the co-cited documents. While in their work, Gipp et al. [85] presented a design of this method, the method was not applied for the recommendation for scientific publications.

In this chapter, we propose three new citation proximity functions, which use character counts between co-cited documents rather than arbitrary CPI values as proposed by [85], to compute a final citation proximity score. We implement and evaluate a RS for scientific publications based on this notion of citation proximity. More information about how citation information is extracted and applied for recommendation is explained in Section 4.3.2.

Proposed proximity methods

Citation Context represents the semantic context in which the article is cited and it is captured by the textual content around the citation. It is computed as the sentence where the reference has been cited as well as the preceding, and the following sentence. Note that, if the citation is located at the start or end of a paragraph, the preceding or following sentences, respectively, are not extracted. This adoption of citation context has been inspired by the work of Ritchie et al. [224]. While other works [102, 127] have used a slightly different approach to extract citation-contexts, by adopting a fixed window of terms around a citation, we focus on a sentence-level definition of citation context. Note that a fixed window may cut short relevant sentences, hence not capturing the full semantic meaning of the citation context. An example of citation context is provided in Figure 4.3 where three sentences are highlighted with yellow, green and red colours. The green highlighted sentence is the sentence where a citation i.e. **Deerwester et al. 1990** is mentioned for which citation context is getting created, the yellow and the red highlighted sentences are respectively preceding and succeeding sentences of the citation.

Figure 4.3: A toy example of a citation context formed from citing, preceding and succeeding sentences

As described in Chapter 3, the notion of citation proximity has been unexplored in the area of RS for scientific publications. A key innovation of our work, therefore, includes the provided definition and application of citation proximity knowledge. The notion of citation context, on the other hand, has been previously explored in the literature to address this task, but not in combination with citation proximity.

Citation proximity and citation context

4.3.2 *Proposed Recommendation Methods*

In this section, we describe our proposed recommendation methods. These methods incorporate the above notions of citation knowledge. The first set of methods focus on citation proximity, the second set of methods focus on incorporating citation proximity and context in combination. These methods are described in Sections 4.3.2.1 and 4.3.2.2 respectively.

4.3.2.1 *Citation Proximity (CP) Recommendation Method*

Our Citation Proximity (CP) recommendation method consists of four main components: Citation Knowledge Extraction, Citation Knowledge Normalisation, Citation Proximity Analysis, and Recommendation. These steps are depicted in Figure 4.4.

- **Citation Knowledge Extraction**: As described in Section 4.3.1, in this step we extract and parse the set of references for papers in the test collection, as well as all the citations of these papers within the text. For each scientific publication that is introduced
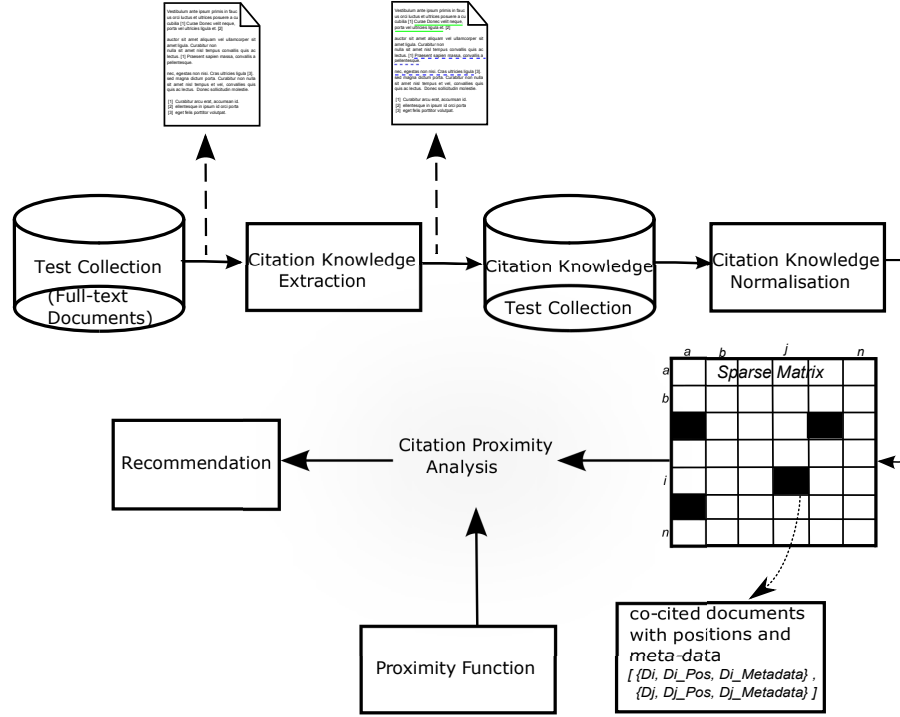
Figure 4.4: Model diagram of the proposed citation proximity (CP) recommendation method. Solid arrows indicate data flow and dashed arrows exhibit data details at that particular steps.

as input, this component outputs a set of tuples, one for each paper referenced and cited in the text of the publication.

$$(\text{referenceId, title, authors [], characterOffsets [], yearPublished, sourceId})$$

$$(4.1)$$

where *referenceId* is the identifier of the referenced paper, *title* is the title of the referenced paper, *authors* is the list of authors, *characterOffsets* is the array of positions within the body of the input scientific publication where the *referenceId* paper is cited, *yearPublished* is the year of publication and *sourceID* is a unique identifier of the input scientific publication. The list of paramet-

ers extracted from each parsed scientific publication is listed in Table 4.1.

Table 4.1: List of parameters extracted from a parsed scientific publication

Identifier	Description
referenceId	Identifier of the publication that is being referenced/cited
sourceId	Identifier of the publication that cites referenceId
title	Title of the referenced/cited publication (referenceId)
authors	Authors of the referenced/cited publication (referenceId)
characterOffsets	A collection of position(s) of the referenced/cited (referenceId) paper within the citing paper (SourceID)
yearPublished	The publication year of the referenced/cited publication (referenceId)

- **Citation Knowledge Normalisation**: Once the full collection of scientific publications has been parsed using the Citation Extraction component, the Citation Normalisation component takes as an input the full set of reference tuples and deduplicates them. The goal of this step is to avoid having duplicate entries for the same research article. We use a naive deduplication method that targets precision at the expense of recall. This method takes into consideration the title of the publication, the publication date and at least one author from the list of authors to identify duplicates. The output of this component is a matrix of co-cited documents where each cell, for example, $V_{i,j}$ contains citation position information where a paper i co-occurs with a paper j in a given source document.

- **Citation Proximity Analysis**: Two papers can be co-cited several times within the same scientific publication, as well as in different scientific publications. This component computes a co-

citation proximity score *prox* for each two given papers taking into consideration all the times that they are co-cited together as well as the proximity in which they are co-cited. The intuition behind the proposed metric is that, the higher the number of times two papers are co-cited, as well as the closer their proximity, the more strongly relevant they are to one another. Our proposed method to compute co-citation scores is defined in Equation (4.2).

$$prox^{ab} = \frac{|Docs|_{a \in Docs \wedge b \in Docs}}{selected\ distance\ metric\ applied\ on\ (\{d_1^{ab}, ..., d_n^{ab}\})} \quad (4.2)$$

where *a* and *b* are two papers co-cited within the scientific publications of the input dataset *Docs*. *n* is the number of times *a* and *b* are cited together in *Docs*. d_1 denotes the first distance between the co-cited pair *a* and *b*. d_n denotes the last distance between them. The numerator captures the number of times the co-cited pair *a* and *b* are cited together in documents in the dataset *Docs* and the denominator captures the distance metrics applied to a set of distances (proximity). Three different metrics (proximity functions) are proposed to compute this proximity: *MinProx*, *SumProx* and *MeanProx*. Each of these measures is introduced below:

MinProx   This method selects the closest distance between a co-cited pair. It chooses the distance that has minimum value

Minimum distance between co-cited pair

from the set of distances as depicted in Equation (4.2). The formulation of *MinProx* is given in Equation (4.3).

$$prox_{Min}^{ab} = \frac{|Docs|_{a \in Docs \wedge b \in Docs}}{\log(min\{d_1^{ab}, ..., d_n^{ab}\})} \tag{4.3}$$

where logarithm (log) is applied to smooth the larger distance. It is essential to mention that during distance computation, one of the hypothesis is the distance between co-cited documents are never Zero or One. Hence, log of zero or one will not occur. For example, a case of citations being cited together depicting as $[X, Y]$ in a citing document in the dataset (Doc) and where a separator character between them is present. If the reference "X" has character offset 102 and reference "Y" will have character offset 104 then the distance between them will be $104 - 102 = 2$.

SumProx   This method adds all the co-cited distances between the co-cited pair as depicted in Equation (4.2). The formulation of *SumProx* is given in Equation (4.4).

Summed distance between co-cited pair

$$prox_{Sum}^{ab} = \frac{|Docs|_{a \in Docs \wedge b \in Docs}}{\sum_{i=1}^{n} \log(d_i^{ab})} \tag{4.4}$$

where d_i is the ith distance between the co-cited documents a and b, and logarithm (log) is applied to smooth the larger distance.

MeanProx   This method computes the mean from all the co-cited distances between the co-cited pair as depicted in Equa-

Averaged distance between co-cited pair

tion (4.2). The formulation of *MeanProx* is given in Equation (4.5).

$$prox^{ab}_{Mean} = \frac{|Docs|_{a \in Docs \wedge b \in Docs}}{\log(mean\{d^{ab}_1, ..., d^{ab}_n\})} \qquad (4.5)$$

where logarithm (log) is applied to smooth the larger distance.

- **Recommendation of Scientific Publications**: When recommending scientific publications for a given paper a, we take into consideration all the papers that are co-cited with a in the corpus. The recommendation module then provides a ranked list of recommendations based on the previously computed *prox* score. The list of publications with higher score in decreasing order is provided as recommendation. It is important to highlight two key limitations of this recommendation method. First, recommendations can not be provided for papers that are not in the original corpus. Second, recommendations can not be provided for papers that are not co-cited with other scientific publications.

The evaluation of our proposed method and the results obtained by using citation proximity are presented in Section 4.4.

4.3.2.2 *Citation Proximity-Context (CPC) Recommendation Method*

In this section, we introduce our Citation Proximity-Context (CPC) recommendation method. This method explores two types of citation knowledge in combination, citation proximity and citation context. Our previously proposed method, solely based on citation proximity, determines the relevancy of scientific publications based on a numeric one dimensional connection (i.e., whether the scientific publications

are co-cited). However, the context of such connections (i.e., how the papers are co-cited together) is lost in our previous model. This can lead to topic-drifting, i.e., recommending papers that although, co-cited together do not cover similar topics. For instance, papers focused on mathematical methods may be recommended when searching for scientific publications about image classification.

Our hypothesis is that, by analysing and exploiting citation context, we can enhance the performance of the recommendation method solely based on proximity. We, therefore, propose to combine citation proximity and citation context to provide recommendations that are topically relevant. As explained in Section 4.3.1, citation context is computed as the sentence where the reference has been cited as well as the preceding, and the following sentences. Note that, if the citation is located at the start or end of a paragraph, the preceding or following sentences, respectively, are not extracted. This adoption of citation context has been inspired by the work of Ritchie et al. [224].

Our method delivers recommendations for a scientific publication in a two-stage process. First, we employ citation proximity method to select an initial set of relevant publications. We then infer the topics from each recommendation generated in the first stage and compare these topics to the ones of the target paper. The pseudo-code of the proposed method is presented in Algorithm 1 where **D** is a corpus containing full text documents, **X** is a set containing citation knowledge of all the documents in **D**, **L** is a Latent Dirichlet Allocation (LDA) model, **T** is a set containing learned topics, and **R** is a list containing recommended items. We used 20 documents as an initial list of recommended items in **R**.

Our recommendation method can be summarised in five main components:

Algorithm 1: Pseudocode for generating recommendations from Citation Proximity-Context (CPC) method

---

Input: Corpus **D** containing full text documents, Set **X** containing citation features of all the documents in **D** and query (target) document $\vec{q}$

Output: n number of item in the recommendation list

1 Run LDA on **D** to generate model L and topic set T

2 Create $\mathbf{R} \subseteq \mathbf{X}$ using proximity-based method

3 for *each* $\vec{x} \in \mathbf{R}$ **do**

4 | Assign topic t to $\vec{x}$ using L.

5 | Find $\{sw\}$ semantically similar top word from the topic t using GloVe's Wikipedia corpus.

6 | Find vector representations $\vec{v}$ of $\{sw\}$

7 | Calculate Cosine similarity between $\vec{q}$ and $\vec{v}$

8 end

9 Reorder **R** into descending order of cosine similarity

10 Output top n items in **R**

- **Citation Knowledge Extraction**: This component is used to extract the citation context for all citations in the corpus. As previously mentioned, citation context is defined in our work as the sentence where the reference has been cited as well as the preceding, and the following sentences. If the citation is located at the start or end of a paragraph, the preceding or following sentences, respectively, are not extracted.

- **Citation Proximity Recommendations**: We use Gipp's citation proximity method [84] to provide an initial list of relevant scientific publications based on citation proximity. We define this list as **R**.

- **Topic Inference from Citation Contexts**: This component extract a list of topics from the previously extracted citation contexts. The idea of applying topic modelling on citation context is to cluster the documents which are focused on the same concept but portrayed in different ways by different authors. Finding

different mentions for the same idea can help providing a meaningful understanding of the research domain. Topic models are widely used to infer latent topics from a corpus of documents. According to [32], documents can be considered as random mixtures over latent topics where each topic is characterised by a distribution over all the words. In our work, we follow the generative process proposed by [32] to discover latent topics from the citation contexts in the corpus D. Inferring latent topics from short-texts (citation contexts) is, however, a difficult task. Inspired by the work of Hong et al. [111], who conducted topic modelling over Twitter data by aggregating tweets (texts of a maximum of 140 characters), we conducted topic modelling over our data by aggregating all the extracted citation contexts for each paper and treating this as a document. For example, a scientific publication d is cited ten times in a corpus, we aggregated the 10 different citation contexts in which d is cited and treated this aggregation as one unique document for topic inference. Once the topics are extracted from these generated documents, we assign the most prominent topic to each scientific publication. The outputs of this component are (i) a list of topics T, each topic described by a list of words $\{w_1, w_2, ..., w_m\}$ where m is the number of words assigned for each topic, and (ii) a list of scientific publications, each of them with one topic T_i assigned.

- **Topic Mapping to Word Embeddings**: The goal of this component is to refine the list of words that define each of the topics extracted in the previous step by means of word embeddings. Word embeddings can capture the subtle semantic relationships between terms in a corpus. For example, the sum of the embedding vectors for *Capital* and *France* is very similar to the embed-

ding vector that describes *Paris* (i.e. $\vec{France} + \vec{Capital} \approx \vec{Paris}$) [190]. Taking this idea as an inspiration, this component projects the words of each topic, T_i, to an vector space. To conduct this step, we use 'GloVe'[3], a statistical model introduced by [209], and its publicly available vector representation of the Wikipedia corpus[4]. We chose this Glove vectors because of its immense and diverse range of enriched topics embedded, and also because this model [209] has been shown to perform better than Word2Vec embedding [190]. The output of this component is a single vector representation v_i extracted for each topic T_i. This vector is computed as mean of the projection weight vectors of each of the words associated to T_i.

We applied a topic modelling method [32] to cluster the documents which are focused on the same concept but portrayed in different ways by different authors. This can help provide a meaningful understanding of the research domain. In addition, we wanted to capture the subtle semantic relationships between concepts. We therefore opted to use the word embeddings in addition to the topic modelling. We adopted the Glove vectors [209] trained on the corpus of Wikipedia as Wikipedia encapsulates a diverse range of topics embedded. Alternatively, another option would be to use the GloVe learning algorithm alone (i.e. without topic modelling) for obtaining vector representations for terms/words by training on aggregated global word-word co-occurrence statistics from our corpus containing two million documents. Then the resulting representations can showcase interesting linear substructures of the word vector space and

---

3 https://github.com/thomasjungblut/glove
4 https://nlp.stanford.edu/projects/glove/

compute the Euclidean distance of the mean GloVe vectors of documents. However, we did not choose the option due to the lack of time and computing resources. We believe comparing these two methodologies in future can be an interesting task.

- **Recommendation of Scientific Publications**: In the final step, a cosine similarity metric is used to measure the similarity between the target publication and the initial list of recommendations, **R**; see Equation (4.6). The recommended publications are re-ranked based on the decreasing cosine similarity value between q and x; where $x \in \mathbf{R}$.

$$d(v_q, v_x) = \frac{v_q.v_x}{\|v_q\|\|v_x\|} \tag{4.6}$$

where v_q and v_x are the vector representations of the target paper and the recommended scientific publication, respectively. Finally, we choose the top five documents as recommendations for the target document. The decision to show only five recommendations is based on not to overwhelm the user with loads of choices.

4.4 EVALUATION

This section reports the experiments conducted to assess our recommendation methods. The evaluation of Citation Proximity (CP) recommendation method is described in Section 4.4.1, and the evaluation of Citation Proximity-Context (CPC) recommendation method is described in Section 4.4.2. For each section, we present the proposed evaluation set-up as well as the obtained results.

4.4.1   *Citation Proximity (CP) Recommendation Method*

This section reports the experiment conducted to evaluate our CP recommendation method. The proposed evaluation set up is described in Section 4.4.1.1 including the evaluation method, metrics and baselines. Results of this evaluation are described in Section 4.4.1.2. Note that, the use of citation proximity for recommendation is still in its infancy. The key objective of this first small-scale evaluation has therefore been to assess whether citation proximity could be a key aspect of citation knowledge to help improving existing recommendation methods for scientific publications.

4.4.1.1   *Evaluation Set-up*

In this section, we report the evaluation set-up for our proposed CP recommendation method, including the dataset used for evaluation, the proposed evaluation method, the baseline used for comparison, and the selected evaluation metrics.

- **Dataset**: For our experiments, we depart from the CORE dataset described in Section 4.2. We used a subset of 368,385 open source scientific publications, all in PDF format and applied GROBID [170] to convert these PDF files into the Text Encoding Initiatives (TEI) format in order to extract the required citation knowledge. Citations and their positions were extracted for a subset of 368,385 documents. The rest of the PDFs were scans or erroneously encoded and hence, no citation knowledge could be extracted from them. A total of 6,609,147 papers were referenced in the 368,385 original scientific publications. On average, we are able to extract 18 references for each publication. In addition,

from this dataset we extracted a matrix of 142,157,561 co-cited pairs with their proximity information.

- **Evaluation Method**: We opted to conduct a user evaluation to assess whether citation proximity can help to enhance existing recommendation methods. Ten members of a Computer Science department (all of them working on the areas of "Data mining" and "Information Retrieval") volunteer for this experimentation. Following this sample of human evaluators, we randomly selected six scientific publications on the above areas of expertise. For each of the six papers, we presented the human evaluators four sets of recommendations, each of them containing five papers. Three of the recommendations were generated using the MinProx, SumProx, and MeanProx metrics presented above, and the fourth recommendation was generated using the baseline (see below). Each evaluator provided 120 relevant judgements (6 query papers - 4 sets of recommendations per target paper - 5 suggested papers per recommendation). The relevant judgements where binary ratings, i.e., the evaluators assigned the recommended scientific publications a value of 1 if they considered the recommended paper relevant with respect to the query paper and 0 otherwise. A total of 1,200 binary judgements were generated using this method.

- **Evaluation Baseline**: We used as a baseline a method that does not include citation proximity information. The suggested baseline method, proposed by [243], is defined in Equation (4.7). This baseline method only takes into consideration the number

of documents where a and b are co-cited, but not the proximity of these co-citations.

$$cocit^{ab}_{baseline} = |Docs|_{a \in Docs \wedge b \in Docs}, \tag{4.7}$$

- **Evaluation Metrics**: We selected Precision @1, @3 and @5, as key evaluation metrics, considering a scenario where the key objective of the evaluation is to provide top relevant items first.

4.4.1.2 *Evaluation Results*

As explained in the previous section, we have calculated precision at three different levels. The results of this experimentation are displayed in Table 4.2. Our experimental results indicate that the use of citation proximity helps generating more relevant recommendations in comparison with the baseline method. More specifically, two out of the three proposed proximity functions, *SumProx* and *MeanProx*, outperform the *Baseline*, indicating that averaging or summing the proximity distances among the cited papers seems to yield more precise results. The improvement over the baseline for P@5 ranges from 0.27 to 0.34, i.e., more than 25% improvement.

Table 4.2: Precision at three different levels. For each metric, a gray scale is used to highlight higher (dark gray) and lower (white) precision values. Higher precision values are represented in bold font.

Method	p@1	p@3	p@5
Co-Citation	0.29	0.27	0.27
MinProx	0.2	0.25	0.25
SumProx	**0.32**	0.33	**0.34**
MeanProx	0.32	**0.34**	0.3

To assess the subjectivity of the task, we have also calculated an inter-rater reliability statistic to weight the agreement between the evaluators. To do so, we have used Fleiss's κ as follows:

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \tag{4.8}$$

where, $\bar{P}_e$ denotes the observed agreement and $\bar{P}$ denotes the probability of agreement. Hence, $(1 - \bar{P})$ is the degree of agreement which is obtainable by chance and $\bar{P} - \bar{P}_e$ gives the degree of agreement which is actually obtained. For all the sample data and its recommendations, we observed $\kappa = 0.25$ suggesting a fair agreement. The interpretation of the level of agreement is shown in Table 4.3 provided by [153].

Table 4.3: κ values interpretation [153]

κ	Interpretation
< 0	Poor agreement
0.01 − 0.20	Slight agreement
0.21 − 0.40	Fair agreement
0.41 − 0.60	Moderate agreement
0.61 − 0.80	Substantial agreement
0.81 − 1.00	Almost perfect agreement

To quantify the statistical significance, we conducted the chi-square test which gives an indication of the extent to which judgements (in our case, relevance/irrelevance) differ across the methods. We obtained a chi-square value of 9.64 and a p-value of 0.0219 with 3 degrees of freedom and an alpha value of 0.05. Since the p-value is less than the alpha value and the chi square value is higher than the critical value 7.815 (– the critical value is obtained from the chi-square test table [77]), the methods and observed data are statistically significant.

4.4.2   *Citation Proximity-Context (CPC) Recommendation Method*

This section presents the experiment conducted to evaluate our proposed CPC recommendation method. This method uses two types of citation knowledge in combination, citation proximity and citation context. The proposed evaluation set up is described in Section 4.4.2.1 including the evaluation method, metrics and baselines. Results of this evaluation are described in Section 4.4.2.2.

4.4.2.1   *Evaluation Set-up*

In this section, we report the evaluation set-up for the proposed CPC recommendation method, including the dataset used for evaluation, the proposed evaluation method, the baseline used for comparison, and the selected evaluation metrics.

- **Dataset**: For the experiment, We extended the dataset in our previous experiment to two million open source scientific publications, all in PDF format. We used GROBID [170] to convert these PDF files into the TEI format in order to extract the required citation knowledge. A total of 30,721,863 papers were referenced in the 2 million scientific publications. On average, we were able to extract 15 references for each publication. Then, citations, their positions and their context were extracted for two millions documents. In addition, from this dataset we extracted a matrix of 665,330,651 co-cited pairs with their proximity information. To the best of our knowledge, this is the largest scientific dataset up to date containing citation-context information. This dataset has been made available for the scientific community under [139], and it is one of the key contributions of this thesis.

Moreover, we inferred 200 topics using LDA topic modelling method [32] from the scientific publications in the dataset. These topics were used in our CPC method to detect semantic relationships between papers. We selected 200 topics inspired by the literature [272].

- **Evaluation Method**: We opted to conduct a user evaluation to assess whether citation proximity, in combination with citation context, can help to enhance existing recommendation methods. Fourteen members of a Computer Science department (all of them working on the areas of "Data mining" and "Machine Learning") volunteer for this experimentation. Following this sample of human evaluators, we randomly selected five scientific publications on the above areas of expertise. For each of the five target papers, we presented the human evaluators five sets of recommendations, including the proposed method and four baselines (see below). Each evaluator provided 125 relevant judgements (5 query papers - 5 sets of recommendations per paper - 5 suggested papers per recommendation). The relevant judgements where on a Likert scale [161] (extremely relevant, very relevant, somewhat relevant, and not relevant).

- **Evaluation Baselines**: We selected four different baselines for this evaluation.

  - *Co-Citation:* Our first baseline is the one used in our previous evaluation, [243], which is defined in Equation (4.7). As explained earlier this baseline method only takes into consideration the number of documents where *a* and *b* are co-cited, but not the proximity of these co-citations, or the citation context, to provide recommendations.

- $CP_{MeanProx}$: Our second baseline is our CP method (presented in Section 4.3.2.1), using *MeanProx* to compute the mean citation proximity.

- *CPA:* Our third baseline is the Citation Proximity Analysis (CPA) method proposed by Gipp et al. [85] and also described in Section 4.3.1. This method has been replicated for the purpose of this study using the exact values of citation proximity index described in the paper.

- *CBF:* The last implemented baseline is a content-based filtering recommendation method. Scientific publications are represented as vectors of terms weighted based on $tf * idf$ [135]. The cosine similarity metric is used to compute the similarity between the target paper and the papers to be recommended. Scientific publications with a higher similarity score to the target paper are the ones provided as recommendation.

4.4.2.2   *Evaluation Results*

The obtained nDCG results at both 3rd and 5th are displayed in Table 4.4. Our results show how our method, CPC, performs better than all baseline methods at both nDCG levels. However, results from both proximity-based citation analyses (*CPA* and $CP_{MeanProx}$) are surprising; with nDCG@5 values of 69% and 79% respectively in comparison to 86% in content-based and co-citation methods. According to [85], the performance of CPA should have been better than $Co - Citation$ and our first experiment for CP also corroborated the concept. We investigated our evaluation dataset and we believe that the length of documents has a higher impact on the proximity-based approach than the originally envisioned. This suggests that looking

into the ways of normalising documents' length might be a plausible next step for future research.

Table 4.4: nDCG results at 3rd and 5th for the proposed method and baselines. For each metric, a gray scale is used to highlight high (dark gray) and low (white) values.

Method	nDCG@3	nDCG@5
$Co - Citation$	0.717	0.864
CPA	0.575	0.688
$CP_{MeanProx}$	0.659	0.782
CBF	0.764	0.865
CPC	**0.838**	**0.902**

To quantify the statistical significance of the obtained result, we conducted the Kruskal-Wallis test [149] which gives an indication of the extent to which the scores (judgements) given by the participants to the documents differ across methods. We obtained a statistic value of 155.477 and a p-value of 1.288×10^{-08} with 4 degrees of freedom and an alpha value of 0.05. Since the p-value is less than the alpha value and the statistic value is higher than the critical value 9.488 (- the critical value is obtained from the chi-square test table [77]), the methods and observed scores are statistically significant.

To check the homogeneity in the ratings of participants, an inter-rater reliability check is performed using Cronbach's alpha [57] which resulted in the value of 0.904, which indicates that participants are in agreement with excellent internal consistency. The interpretation of the level of agreement is obtained from Table 4.5 provided by [86] which shows that a score of more than 0.7 value is acceptable [55, 86].

Table 4.5: Cronbach's alpha interpretation table [86]

Cronbach's alpha	Internal consistency
$\alpha \geq 0.9$	Excellent
$0.9 > \alpha \geq 0.8$	Good
$0.8 > \alpha \geq 0.7$	Acceptable
$0.7 > \alpha \geq 0.6$	Questionable
$0.6 > \alpha \geq 0.5$	Poor
$0.5 > \alpha$	Unacceptable

4.5 DISCUSSION AND CONCLUSIONS

In this chapter, we have investigated whether the use of citation knowledge can help improving existing RS for scientific publications when addressing task of recommending scientific publications for a given piece of work. In the literature, this work is captured by a paper, an on-going manuscript, or a snapshot of text (e.g. components of a paper such as title or abstract). We have captured it by using a paper.

Two types of citation knowledge have been explored in this task: **citation proximity** and **citation context**. We have extracted such knowledge from the content of scientific publications and use it to recommend scientific publications. Upon doing so, we have proposed **novel recommendation methods** encapsulating this knowledge. Our hypothesis for using citation proximity for this particular task is that papers that are co-cited together, and in close proximity, tend to be relevant to one another. Similarly, papers that are cited based on similar terminologies and expressions (i.e., whose citation contexts are similar), tend to refer to similar topics.

Our experiments have shown how the proposed **recommendation methods**, which use these two types of citation knowledge, **perform better than the baselines**. This shows how citation proximity and

citation context, particularly in combination can indeed help to enhance the performance of exiting RS for scientific publications when addressing the task of recommending scientific publications for a given target paper.

In conducting these experiments, we have also provided a **novel dataset** containing the extracted citation knowledge from the publications. This includes not only metadata of the referenced papers but also the citation context and citation position (in terms of sentence, paragraph and chapter level) while citing them. Note that, as pointed out in Chapter 3, public datasets with the full-text of scientific publications are limited, and their size is generally small, typically in the thousands of documents. Our provided large-scale dataset, containing 53 million unique citation-based records extracted from 2 million scientific publications, is a key contribution of this PhD to the scientific community.

Multiple challenges do however remain open, and we acknowledge several limitations of the proposed methods and conducted research. First of all, while our proposed citation-proximity recommendation method is a pioneer method on exploring the use of citation proximity to provide recommendations of scientific publications, it suffers from an important limitation. Papers that are not cited, which will generally happen with the most recent papers, will not receive any recommendation. Since we are in a task of recommending papers for a particular paper this encapsulates both, the **target and item cold start problems**.

A second limitation of the presented research is the lack of ratings, or ground truth. Note that our dataset does not contain ratings (or associated preferences). Hence, we have conducted user-studies in our evaluation. With the lack of ratings, it is not possible to train

other potential baselines (such as Collaborative Filtering (CF)) and conduct a more **comprehensive evaluation** against a wider variety of methods. Obtaining reliable ground truth, particularly for this recommendation task, is not trivial. Note that, the recommendation is not made for users, but for a particular piece of work (in this case, represented by a paper). Subjectivity therefore comes into place, since different users, with the same query paper, may be looking for different recommendations.

To address this task, we have focused on two types of citation knowledge, citation proximity and citation context. We selected these types of knowledge since we hypothesise they could be the useful ones for the given recommendation task. However, other types of citation knowledge, such as the ones that will be presented in Chapter 5, could have also been applied and evaluated for this task. This would have required to propose additional recommendation methods, or modifications of existing methods, to incorporate this knowledge. This is part of our future work.

Finally, it is also important to highlight that our evaluations were based on a maximum of fourteen users. It would have been desirable to recruit more users for the conducted evaluations. However, as specified in Chapter 2, user studies are costly since users' time is expensive. Also, recruiting users with the appropriate expertise to assess RS for scientific publications is not trivial. Note that, these users need to be researchers with sufficient knowledge of domain. We used computer science domain in our user study during evaluation.

Despite the above-mentioned limitations, this chapter presents multiple important contributions to the field including (i) novel conceptualisations of citation knowledge, including citation proximity and citation context (ii) a large dataset containing citation knowledge extracted

from scientific publications, (iii) novel recommendation methods that incorporate, by means of a wide range of metrics, various types of citation knowledge, and (iv) an evaluation that shows how, for the particular task of recommending papers to a given paper, the use of citation knowledge can help enhancing existing recommendation methods.

RECOMMENDING SCIENTIFIC PUBLICATIONS FOR A USER USING CITATION GRAPH, CITATION CONTEXT, CITATION SECTION AND CITATION INTENTION

As pointed out in Chapter 3, the second main task addressed by Recommender Systems (RS) for scientific publications is the recommendation of relevant scientific publications for a user. As opposed to most of the existing solutions for this task, which have focused on recommending relevant papers to a user independently of the time when such papers were published, we address the real-world problem of recommending recently published papers [95]. This is a particularly challenging problem where existing RS methods, such as Collaborative Filtering (CF), are ineffective to provide recommendations. Note that, recent scientific publications are not previously seen, rated or cited and hence, they have no ratings associated to them.

Our purpose in this chapter is, therefore, to address RQ4, investigating whether the use of wider granularity of citation knowledge can help improving existing recommendation methods when addressing the task of providing personalised recommendations of recent scientific publications. With this purpose, we define four types of citation knowledge –citation section, citation graph, citation context and citation intention–, and propose novel recommendation methods that exploit such knowledge. An evaluation of the proposed methods

is conducted by means of offline evaluations. A novel benchmark has been also developed to conduct such evaluations.

RQ4 : When addressing the task of recommending recent scientific publications to a user, can citation knowledge help improving existing RS?

The remainder of the chapter is structured as follows: Section 5.1 motivates the work, Section 5.2 describes the data used to conduct this research, Section 5.3 presents our recommendation methods, Section 5.4 reports the conducted experiments, and Section 5.5 concludes the work presented in the chapter.

5.1   INTRODUCTION

With the continuous growth of scientific literature, discovering relevant academic papers for a researcher has become a critical task. This task is particularly important when considering that researchers need to be up to date with the latest developments in their scientific fields. Hence, the need of RS that can recommend them the latest, more recent relevant papers.

While multiple works in the literature [11, 245, 267] (more works in Table 3.1) have aimed to target the problem of recommending scientific publications to a user, they have done this independently on when those papers were published. To the best of our knowledge, only the works by Ha et al. [95] and Wang et al. [272] have considered the time and item cold start aspects of this problem.

To address this problem, we investigate how **citation knowledge** could be captured and exploited to support users towards the discovery of recent and relevant scientific publications. On doing so, we propose **novel hybrid recommendation methods** that explore the

users' history (in terms of their publications and cited papers) to build their profiles, and various notions of citation knowledge to provide personalised recommendations. In particular, we introduce four notions of citation knowledge to address the above mentioned recommendation task: the **citation graph** (or relations between papers based on citations), **citation context** (or texts that surround citations within the papers), **citation section** (or sections of the papers where citations appear) and **citation intention**, or intention behind citing the paper (providing background knowledge, comparing against other approaches, etc.).

The idea of using citation knowledge to recommend scientific publications to users has been previously applied in the literature. Sugiyama et al. [252], for example, explored the use of citations to enrich user profiles, capturing a user's research interests by considering not only their past publications, but also the citations of such publications, and the publications citing the user's work. In an extension of this work [253], the authors explored the use of the content of the publications citing the user's work, including: (i) citation context, or text around the citation, since it may be viewed as an endorsement of the work, and (ii) textual content from other sections (abstract, introduction, and conclusions) to complement citation context. Torres et al. [267] explored the use of the citation graph of CiteSeer to provide recommendations. Their proposed algorithm exploits the content of papers as well as their citations, when available.

While these works show how citation knowledge can help enhancing recommendation accuracy, they do not explore the use of citation knowledge on the particular real-word scenario of recommending the latest scientific publications to users. In addition to exploring the use of citation knowledge in this scenario, our work proposes a

more comprehensive view of citation knowledge. While knowledge in terms of citation context, citation section, or citation graph (relations between papers based on citations) has been previously used in RS for scientific publications [45, 253], to the best of our knowledge, the concept of **citation intention** has rarely been applied to this domain (see Section 3.3). Similarly, **combinations of these different aspects of citation knowledge** have also not been previously used for the recommendation of scientific publications. It is important to highlight that to capture the above knowledge, access to the publications of users and the textual content of papers is needed. Existing datasets used for training and testing academic RS do not provide either the entire user publication history or scientific publication texts, but just their metadata, e.g., title, abstract and keywords. As part of this work, we have also built and made publicly available a **novel evaluation benchmark** to enable the implementation and evaluation of recommender systems for the particular setting of recommending recently published papers to a user.

Building this benchmark has allowed us to conduct offline evaluations to assess whether the above notions of citation knowledge do indeed help enhancing the performance of existing recommendation methods for the task at hand. Experimental results show improvements over baseline methods, evidencing the potential of using citation knowledge to recommend recently published papers in a personalised way.

In the next section, we introduce the above benchmark, followed by our proposed recommendation approaches in Section 5.3 and their evaluation in Section 5.4.

5.2 DATASET BUILDING

Multiple datasets have been used in the past to evaluate recommender systems for academic publications. A comprehensive list of publicly available datasets is given in Chapter 3, and more particularly in Table 3.7. For each dataset, the table shows a brief description of the type of data and time period when its data were collected, the number of users, items and ratings in the dataset, and whether the full text PDF_{av} and publication history of the users UPH_{av} are available.

As we can observe from the table, most of the existing datasets do not provide the full texts of publications, and hence, a wide range of citation knowledge (e.g., citation section, context, intention, etc.) cannot be extracted from them. Similarly, many datasets do not provide the authors' publication history, and thus knowledge about the users, particularly their preferences (publications and citations), cannot be easily captured. To address these limitations, in this thesis, we have built a new dataset that includes both the textual content of papers, as well as the authors' publication history. Our dataset building process is described next.

5.2.1 *Collecting Data*

As previously explained, we aimed to build a new dataset that (i) could serve to investigate the particular recommendation scenario of discovering the most recent academic papers relevant for a target user, and (ii) provides the textual content of papers in addition to their metadata, so that fine grained citation knowledge could be extracted and used as by recommendation methods.

Since we were interested in exploring the usage of citation knowledge for recommendation purposes, we needed to ensure that there are sufficient items (papers) cited by other items within the dataset. Following this requirement, we gathered the publication history of authors working on the same field (e.g., publishing in the same conference), since they are likely to cite each other's publications.

Specifically, we selected the ACM Conference Series on Recommender Systems (RecSys)[1] and collected data for 1,931 authors who have published in the conference from the first RecSys conference in 2007 until the twelfth RecSys conference in 2018. The complete publication histories of these authors were collected from the well-known computer science bibliography data provider DBLP[2]. Collectively 1,931 authors published 80,808 research publications. Note that, the publication history of an author contains not only their RecSys papers but also papers published in other venues (journals, conferences, etc.).

As shown in Figure 5.1, an author's publication history contains metadata of each of the author's papers, including their titles, abstracts, publication dates and venues. The metadata also includes a URL to the corresponding Google Scholar page of each paper, which we used to gather the Portable Document Format (PDF) file of the paper.

Out of the 80,808 papers crawled from DBLP, we were able to obtain full content for 35,473 of them (about 44%). We note that, while initiatives like open access and pre-prints enabled full access to many scientific publications, many of them are still hidden behind pay-walls and thus are not publicly accessible.[3] Then, to ensure that we had sufficient historical data to capture user preferences, we discarded all

---

1 https://recsys.acm.org/
2 https://dblp.uni-trier.de/
3 https://www.theguardian.com/higher-education-network/2018/may/21/
  scientists-access-journals-researcher-article

the authors for which we obtained less than 4 publications, keeping a total of 1,336 authors.

To extract **citation knowledge**, we parsed the available PDF files using the GROBID parser[4] and the classifier for citation intention generation provided by [136]. Among the 35,473 papers for which we have textual content, we could obtain citation intention for 21,924 of them using the classifier of citation intentions provided by [136]. The parser used as a part of the approach of [136] is ParsCit[5] which could only extract citation information for 61% of the collected papers, reducing the completeness of the data. Then, to ensure that we had sufficient historical data to capture user preferences, we discarded all the authors for which we obtained less than four publications, keeping a total of 1,102 authors.

The parser problem led to the creation of two different datasets: (i) the first one containing data for 1,336 authors, and citation knowledge in the form of citation section, citation graph and citation context, we named it as **Dataset_sgc**; (ii) the second one containing data for 1,102 authors and citation knowledge in the form of citation section, citation graph, citation context and citation intention, we named it as **Dataset_sgci**. Both datasets were divided into training and test sets by observing the publication time distribution, and selecting as breaking date the 1st of January 2018 (see Figure 5.2). All papers published before that date were considered part of the training set and all papers published after that date were considered part of the test set. Lastly, we kept those authors having at least 60% of the data in the training set, and at least 10% of the data in the test set.

The final two datasets consists of:

---

4 https://github.com/kermitt2/grobid
5 https://parscit.comp.nus.edu.sg/

- **Dataset_sgc**: This dataset contains 547 authors and 15,174 academic papers, from which 12,641 belong to the training set and 2,533 represent the test set. Citation knowledge in the form of citation section, citation graph and citation context is available for this dataset.

- **Dataset_sgci**: This dataset consists of 446 authors and 9,399 academic papers from which 7,786 belong to the training set and 1,613 represent the test set. Citation knowledge in the form of citation section, citation graph, citation context and citation intention is available for this dataset.

5.2.2 *Modelling Authors, Papers and Citations*

Figure 5.1 shows the different features that are captured for authors (users), publications and citations, as well as different relations among them. For each user, we consider four different identifiers, including: the internal identifier within the dataset, the ORCID identifier[6], and both, the DBLP and the Google Scholar URLs. In addition, we also capture name, last name, website and affiliation.

For each publication, we also capture multiple identifiers, including the internal identifier within the dataset, the DBLP URL (from where meta-data about the paper has been extracted) and the Google Scholar URL (from where the PDF file of the paper has been downloaded –if available–). Metadata about the publication includes: title, abstract, publication date, and publication venue.

To extract citation knowledge, we have parsed the available PDF files using the GROBID parser[4]. From the PDF of each publication,

---

6 https://orcid.org/
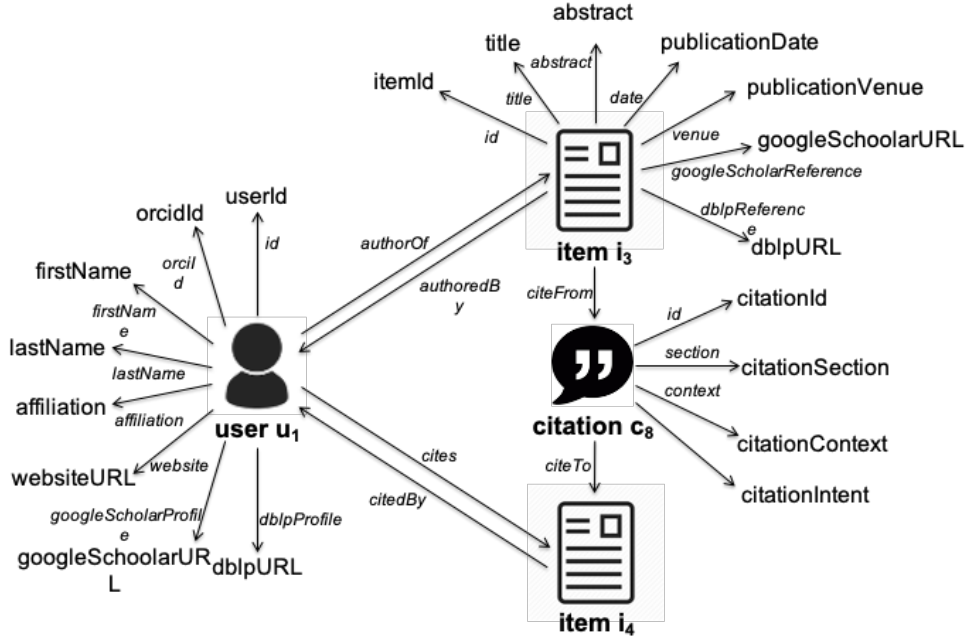
Figure 5.1: Capturing citation knowledge

we have extracted: (i) the reference list (i.e., all the papers that are cited within the publication), (ii) the sections within the publication where those citations appear (introduction, related work, etc.), (iii) the citation context (the text that surrounds the citation in the publication). We consider as citation context three sentences: the one where the citation appears, and the ones before and after, when available), and (iv) the citation intention (using the approach proposed by [136]).[7] More information about how citation knowledge has been extracted is described in Section 5.3.1.

5.2.3 *Modelling User Preferences*

As mentioned in Chapter 3, when the publication history of a user is sparse, its data may be insufficient to build a reliable profile for

---

7 Relation between authors, their publications and citations are captured in a database and its entity relationship is provided in Appendix A.2.

personalised recommendation. In such a case, relying on citation information could help enriching the user's profile.

Hence, we distinguish between two main ways of capturing user preferences. On the one hand, we consider that a user has a preference (positive rating) for all the papers they have authored. Authored papers encapsulate the user's interest in terms of research areas, topics, methods, etc., and constitute a relevant source of information to build their profile. On the other hand, we consider that a user has a preference for all the papers they have authored as well as the papers they have cited, since cited papers also encapsulate research that the user considers relevant in relation to their work. By doing so, we explore the use of the citation-graph to enrich a user-item rating matrix.

Figure 5.2 illustrates these two preference models. In the left part of the figure, we show a rating matrix R_P relating authors (rows) and papers (columns) where a cell has a value 1 if the corresponding user authored the associated paper, and 0 otherwise. In the right part of the figure, we show a rating matrix R_{PC} where a cell has a value 1 if the user authored or cited the paper, and 0 otherwise. In addition to these rating matrices, we also considered four enriched versions of R_{PC}: (i) R_{PCX}, where X stands for context, R_{PCXS}, where S stands for section, R_{PCXI}, where I stands for intention and R_{PCXSI}, where SI stands for section and intention. A more detailed description of these matrices is provided in Section 5.4.1.

5.3 PROPOSED APPROACH

This section presents the various types of citation knowledge explored to address the task of recommending recent scientific publications
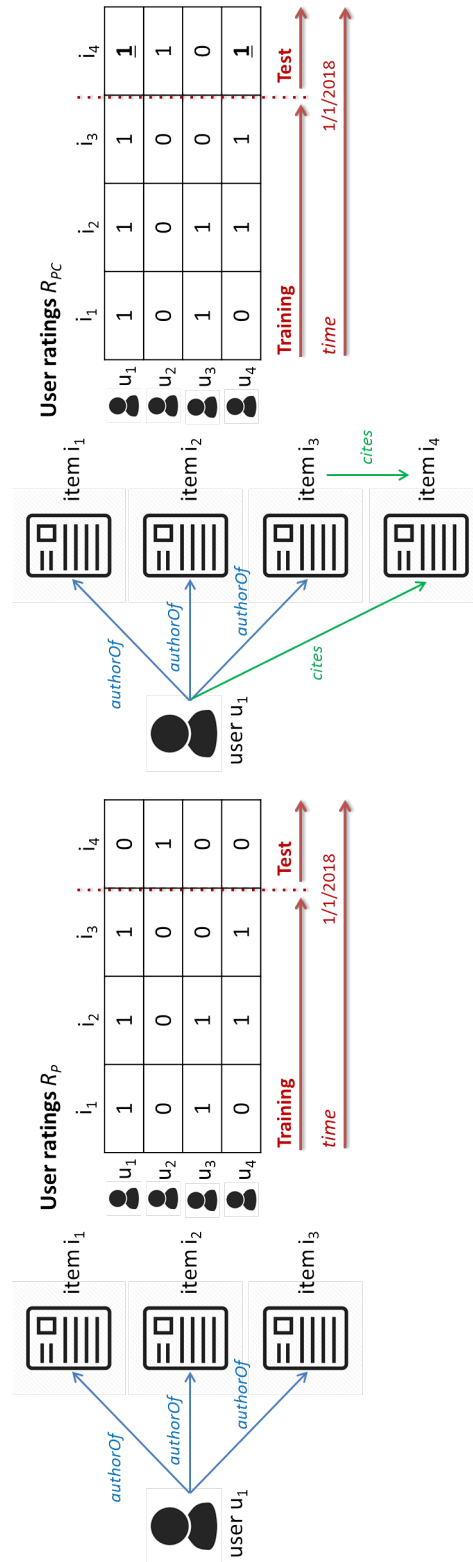
Figure 5.2: Modelling user preferences

for a given user (Section 5.3.1), and our proposed recommendation methods, which incorporate such knowledge (Section 5.3.2).

5.3.1   *Citation Knowledge*

We have selected four types of citation knowledge to address this task: citation section, citation graph, citation context, and citation intention. This section explains the rationale of using each of these types of citation knowledge, as well as the process of extracting this knowledge from the collected data.

- **Citation graph**: The citation graph represents relations between papers based on citations. In this graph, nodes represent papers and edges represent relations between such papers based on their citations. Relations in the citation graph can be either directed –i.e., they capture the explicit relation between source (citing) paper and target (cited) papers [183]–, or undirected –i.e., they do not consider which paper is the one citing and which paper is the one being cited [95]. The following process was followed to extract this knowledge from the collected data. From the PDF of each publication, we extracted the reference list (i.e., all the papers that are cited within the publication). The reference lists are then matched against all the publications of the dataset to identify the citation-based relations and generate the citation graph. Paper titles, authors, and publication years were matched considering a series of heuristics to minimise errors including: applying lower case, matching at least one author, and computing the Levenshtein distance[8] between the title of

---

8 https://en.wikipedia.org/wiki/Levenshtein_distance

the publication and the title of the reference. An 85% minimum threshold was empirically selected for this distance. These heuristics were needed since references sometimes contained errors or incomplete information.

- **Citation Section**: The citation section refers to the particular section of the paper where the citation appears, e.g., introduction, literature review, and conclusions. Citations in different sections of a paper generally serve different purposes. For example, citations appearing in the introduction section tend to help motivating and describing the addressed problem, while citations appearing in the literature review section are generally used for comparison [45]. To extract this knowledge, we parsed the PDF of each publication and extracted the sections within the publication where citations appear. We considered four sections in this work: *introduction*, *related work*, *conclusions*, and *other*.

- **Citation context**: We use the same notion of citation context as the one described in Chapter 4, the sentence-based text surrounding the citation, i.e., the sentence where the citation appear, as well as the sentences before and after, when available. Citation context provides an indication of the semantics with a citation is mentioned, which could provide relevant information when it comes to determine relevancy for a recommendation [253]. We use the same process described in Chapter 4 to extract this knowledge.

- **Citation intention**: The citation intention mainly reflects the objective of the citation: provide background, criticise, support, etc. [160]. Understanding the intention behind a citation, could also be a relevant indicator to enhance the recommendation

process. It is important to note that, while the definition and use of citation intention has been applied in the area of scientometrics to study the evolution of scientific fields [136], or to quantify the influence of research works [99, 269], to the best of our knowledge, it has not been explored before for the recommendation of scientific publications to users. To assign citation intentions to the referenced papers in our dataset, we used the classifier provided by [136], which categorises citations into six main types: *background, compare or contrast, use, extension, motivation and future*. More information about the different categorisations of citation intentions can be found in Section 3.3. This classifier was trained with papers from the ACL Anthology corpus[9], a corpus of scholarly publications about Computational Linguistics. To evaluate the performance of this classifier when identifying citation intentions from the RS field we conducted a manual assessment of 100 randomly selected citations. Based on 3 annotators (all of them with computer science background), displaying a moderate agreement (0.62 Fleiss kappa [76]), we identified a 69% average correct classification, which is in line to the one reported by the authors [136].

5.3.2   *Proposed Methods*

In this section, we describe the proposed hybrid citation knowledge-based recommendation methods. These methods, aside from incorporating citation knowledge, jointly exploit the content of the papers and the user-item ratings (see Section 5.2.3) to provide personalised recommendations.

---

9 https://acl-arc.comp.nus.edu.sg/

Hybrid methods [38] aim to mitigate the disadvantages of individual approaches by combining the strengths of each type of approach, in general Content Based Filtering (CBF) and CF. In our case, we aim to mitigate the ineffectiveness of CF when recommending the latest scientific publications by combining it with CBF and exploiting the captured citation knowledge.

We propose four different hybrid methods that explore different combinations of citation knowledge presented in this chapter. These methods, and the citation knowledge they use are summarised in Table 5.1, and described in the following subsections.

Table 5.1: List of hybrid methods and various citation knowledge they utilised

Method	Exploited citation knowledge
hyb	citation graph and citation context
hybSec	citation graph, citation context and citation section
hybIntent	citation graph, citation context and citation intention
hybIntentSec	citation graph, citation context, citation section and citation intention

5.3.2.1 *Hybrid Recommendation Method: hyb*

The first proposed recommendation method, **hyb**, is based on the item-based CF nearest neighbour heuristic[10] where content-based features are used to compute item similarities. In item-based CF algorithms [228], similarities between items are used to estimate scores for a (user, item) pair. In our case, item (paper) profiles are generated based on textual features, where textual features vary with respect to the available citation knowledge (see Section 5.2.3): for R_P, item profiles are build from the title of the item; for R_{PC}, item profiles are build from

---

10 We also tested the user-based CF heuristic, but discarded it due to its non competitive performance results.

the title of the item and the titles of the cited papers; and for R_{PCX}, item profiles are built from the title of the item, the titles of the cited papers, and the citations context. To create a profile (– either a user profile or an item profile), texts are represented as vectors in a vector space model using Term Frequency Inverse Document Frequency (TF-IDF) method. We formulate our **hyb** method in Equation (5.1):

$$\hat{r}_{u,i} = \frac{\sum_{i \epsilon N(i')} Sim(i,i').r_{u,i'}}{\sum_{i \epsilon N(i')} |Sim(i,i')|} \tag{5.1}$$

where $\hat{r}_{u,i}$ is the preference score to be predicted for the target user u and item i, $Sim(i',i)$ is the similarity between the interacted item i' and an item i from the neighbourhood $N(i')$ of the item i'. Cosine similarity is used to measure the similarity between items (i.e. item profiles). Lastly, $r_{u,i'}$ is the preference (rating) given by user u to an item i'. We also use different sizes of neighbours, specifically 5, 10, 15 and 20.

5.3.2.2 *Hybrid Recommendation Method: hybSec*

To investigate the relevance of citation sections, we further modified our hybrid method Equation (5.1) by incorporating a weight, $w_{u,i}$, that reflects the strength of an item i for a user u based on the different sections where u cites i in their publications. Four different sections are considered in this method: *Introduction, Related Work, Conclusions and Other*.

We formulate this method, **hybSec**, in Equations (5.2) and (5.3).

$$\hat{r}_{u,i} = \frac{\sum_{i \epsilon N(i')} Sim(i,i').r_{u,i'}.w_{u,i'}}{\sum_{i \epsilon N(i')} |Sim(i,i')|} \tag{5.2}$$

where the strength (weight) $w_{u,i'}$ is computed by considering all the instances where i' is cited by u; Note that, an item i' may be cited

by u in several publications, and in different sections of the same publication. Then, the weight is normalised by the total number of instances. More formally, the strength $w_{u,i'}$ of item i' for user u is calculated as:

$$w_{u,i'} = \frac{\sum_{j=1}^{n}(w_{j_{int}} + w_{j_{relWork}} + w_{j_{concl}} + w_{j_{others}})}{n_{u,i'}} \tag{5.3}$$

where $n_{u,i'}$ is the number of times i' is cited by u in their papers, and $w_{j_{int}}$, $w_{j_{relWork}}$, $w_{j_{concl}}$ and $w_{j_{others}}$ reflect the number of times i' is cited in the introduction, related work, conclusion or other sections, respectively.

5.3.2.3 *Hybrid Recommendation Method: hybIntent*

To investigate the relevance of the different intents of citations, we further modified our hybrid method by incorporating a weight, $w_{u,i'}$ in the Equation (5.1) heuristic, that reflects the strength of an item i' for a user u based on the different citation intents, named as **hybIntent**. Five different intentions are considered [136]: *Background, Compare or Contrast, Motivation, Use, Extension.*

Equation (5.4) represents **hybIntent** hybrid method. We formulate this method **hybIntent** in Equations (5.4) and (5.5).

$$\hat{r}_{u,i} = \frac{\sum_{i \in N(i')} Sim(i,i').r_{u,i'}.w_{u,i'}}{\sum_{i \in N(i')} |Sim(i,i')|} \tag{5.4}$$

where the strength (weight) $w_{u,i'}$ is computed by considering all the instances where i' is cited by u; Note that, an item i' may be cited by u in several publications, and with different citation intentions. Then,

the weight is normalised by the total number of instances. Formally, the strength $w_{u,i'}$ of item i' for user u is calculated as:

$$w_{u,i'} = \frac{\sum_{j=1}^{n}\left(w_{j_{bkg}} + w_{j_{com}} + w_{j_{mot}} + w_{j_{use}} + w_{j_{ext}} + w_{j_{fut}}\right)}{n_{u,i'}} \tag{5.5}$$

where $n_{u,i'}$ is the number of times i' is cited by u in their papers, and $w_{j_{bkg}}$, $w_{j_{com}}$, $w_{j_{mot}}$, $w_{j_{use}}$, $w_{j_{ext}}$ and $w_{j_{fut}}$ reflect the weight when i' is cited as background, compareOrContrast, motivation, use, extension, or future respectively.

5.3.2.4 *Hybrid Recommendation Method: hybIntentSec*

Our last hybrid method that combines all the citation knowledge previously mentioned, i.e., citation graph, citation context, citation section and citation intention. To investigate the relevance of different citation intentions within different sections, we modified our **hyb** method and incorporated a weight reflecting the strength of an item cited with a specific intention on a particular section.

We assume that a paper being cited with one intention in a particular section (e.g. intention:motivation / section:introduction) may carry different intention in the same or other section (e.g. intention:compare or Contrast / section:related work). Therefore, we incorporate a weight, $w_{u,i'}$ in **hyb**, that reflects the strength of an item i' for a user u based on different intentions and sections. This method, **hybIntentSec**, is presented in Equations (5.6) and (5.7).

$$\hat{r}_{u,i} = \frac{\sum_{i \epsilon N(i')} Sim(i,i').r_{u,i'}.w_{u,i'}}{\sum_{i \epsilon N(i')} |Sim(i,i')|} \tag{5.6}$$

where the strength (weight) $w_{u,i'}$ is computed by considering all the instances where i' is cited by u; Note that, a publication i' may be cited by u in several publications, and with different citation intention and section of the publications. Then, the weight is normalised by the total number of instances. More formally, the strength $w_{u,i'}$ of item i' for user u is calculated as below:

$$w_{u,i'} = \frac{\sum_{s=1,t=1}^{s=S,t=T}(w_{s,t})}{n_{u,i'}} \tag{5.7}$$

where $n_{u,i'}$ is the number of times, the publication i' is cited by the user u in the author's (u's) papers, and s and t are citation section and citation intention respectively. s ranges from 1 to the total number of chosen citation sections, that is, S. t ranges from 1 to the total number of chosen citation intention, that is, T.

$w_{s,t}$ reflects the number of times a publication is cited in a citation section chosen from s as a citation intention from t. There are numerous combinations how the weights can be assigned. An example is shown in Table 5.2, where a publication is cited with *Background* intention in the *Introduction* section, hence, this cell is assigned as weight of 1. Heuristically, various combinations of citation intention and section can be taken into account by shifting the weight (i.e. 1).

Table 5.2: An example of the combination of citation intention and citation section that are applied in Equation (5.7). Here, *BKG* and *Introduction* are chosen as citation intention and citation section respectively.

	BKG	COM	MOT	USE	EXT	FUT
Introduction	1	0	0	0	0	0
Related work	0	0	0	0	0	0
Conclusion	0	0	0	0	0	0
Other	0	0	0	0	0	0

5.4   EVALUATION

This section reports the experiments conducted to assess the proposed recommendation methods. Two evaluations are presented here. The first evaluation (see Section 5.4.2) has been conducted with the complete dataset and three types of citation knowledge: citation graph, citation section and citation context. The second evaluation (see Section 5.4.3), has been conducted with a subset of the original dataset and with all types of citation knowledge, including citation intention. Note that, as described in Section 5.2.1, the classifier used to extract citation intention [136] could not parse all PDF documents. Hence, our second evaluation has been conducted with a reduced subset of the data, but with the complete set of citation knowledge. We present below the proposed evaluation set-up as well as the obtained results.

5.4.1   *Evaluation Set-up*

In this section, we present the followed evaluation set-up, including the datasets used for evaluation, the proposed evaluation methodology, the baselines used for comparison, and the selected evaluation metrics.

- **Datasets**: As described in Section 5.2, we have two datasets for evaluation, **Dataset_sgc**, for which we have citation knowledge in the form of citation section, citation context and citation graph and **Dataset_sgci**, a subset of Dataset_sgc for which we also have citation intention.

  The knowledge of the citation graph in these datasets is used to capture user preferences based on two different preference models. On the first model, we consider that a user has a

preference (positive rating) for all the papers they have authored. On the second model, we consider that a user has a preference for all the papers they have authored as well as the papers they have cited. Figure 5.2 illustrates these two preference models. In the rating matrix R_P, a cell has a value 1 if the corresponding user authored the associated paper, and 0 otherwise. In the rating matrix R_{PC} a cell has a value 1 if the user authored or cited the paper, and 0 otherwise.

In addition to these rating matrices, we also considered four enriched versions of R_{PC}: (i) R_{PCX}, where X stands for context, ii) R_{PCXS}, where S stands for section, iii) R_{PCXI}, where I stands for intention, and iv) R_{PCXSI}, where SI stands for section and intention.

In case of R_{PCXS} for every rating based on citations, it captures the section in the paper and the text around a citation. In case of R_{PCXI} for every rating based on citations, it captures the intention of citing and the text around a citation. In the case of R_{PCXSI}, for every rating based on citations, it captures the citation intention, the citation section, and the citation context. In our experiments, the above matrices are split into training and test sets according to a target time, in particular, the 1st of January 2018. The final dataset splits for the two datasets are as follows:

- **Dataset_sgc**: This dataset contains 547 authors and 15,174 academic papers, from which 12,641 belong to the training *Dataset for Hyb,* set and 2,533 represent the test set. These publications were *HybSec* cited 4,358 times in the introduction sections, 3,999 in the related work sections, 82 times in the conclusion sections,

and 12,213 times in other sections. Citation intention knowledge is not available for this dataset, hence, only the first four matrices are available. The splits are:

* $R_P^{training}$: 547 users, $12,641$ items and $14,555$ ratings

* R_P^{test} : 547 users, $2,533$ items and $3,082$ ratings

* $R_{PC}^{training}$: 547 users, $12,641$ items and $20,756$ ratings

* R_{PC}^{test} : 547 users, $2,533$ items and $3,233$ ratings

* $R_{PCX}^{training}$: 547 users, $12,641$ items, $20,756$ ratings, and citation context and section information

* R_{PCX}^{test} : 547 users, $2,533$ items and $3,233$ ratings, and citation context and section information

* $R_{PCXS}^{training}$: 547 users, $12,641$ items, $20,756$ ratings, and citation context and section information

* R_{PCXS}^{test} : 547 users, $2,533$ items and $3,233$ ratings, and citation context and section information

– **Dataset_sgci**: This dataset consists of 446 authors and 9,399 academic papers, from which 7,786 belong to the training set and 1,613 represent the test set. These publications were cited 2,820 times in the introduction sections, 2,784 in the related work sections, 44 in the conclusion sections, and 8,113 in other sections. Citations were classified in terms of their intention 9,933 times as background, 3,032 as Compare Or Contrast, 225 times as Extend, 449 times as Use, 83 times as Motivation and 39 times as Future. Since this dataset contains citation intention, the six matrices are available for it. The splits are:

Dataset for Hyb, HybIntent, HybIntentSec

* $R_P^{training}$: 446 users, $7,786$ items and $9,348$ ratings

* R_P^{test} : 446 users, 1,613 items and 2,033 ratings

* $R_{PC}^{training}$: 446 users, 7,786 items and 13,104 ratings

* R_{PC}^{test} : 446 users, 1,613 items and 2,126 ratings

* $R_{PCX}^{training}$: 446 users, 7,786 items and 13,104 ratings, and citation context and section information

* R_{PCX}^{test} : 446 users, 1,613 items and 2,126 ratings, and citation context and section information

* $R_{PCXS}^{training}$: 446 users, 7,786 items and 13,104 ratings, and citation context and type information

* R_{PCXS}^{test} : 446 users, 1,613 items and 2,126 ratings, and citation context and type information

* $R_{PCXI}^{training}$: 446 users, 7,786 items and 13,104 ratings, and citation context and type information

* R_{PCXI}^{test} : 446 users, 1,613 items and 2,126 ratings, and citation context and type information

* $R_{PCXSI}^{training}$: 446 users, 7,786 items and 13,104 ratings, and citation context and type information

* R_{PCXSI}^{test} : 446 users, 1,613 items and 2,126 ratings, and citation context and type information

- **Evaluation Method**: Based on the above generated benchmarks, containing ground truth data, we are able to conduct an offline evaluation that allow us to compare various baselines systematically. More information about the offline evaluation methodology is described in Chapter 2.

- **Evaluation Baselines**: We tested our recommendation methods against a variety of baselines. Note that, CF is not among these

baselines. As stated at the beginning of the chapter, this recommendation method is not able to provide recommendations for the task we are addressing, i.e, the recommendation of recent scientific publications. A similar problem can be observed with the methods like PageRank [36] and ItemRank [89] where there are no connections between training and testing nodes (i.e. papers). The rest of the baselines are described below:

– **Content-based recommendation**:

As described in Chapter 2, Content-based filtering methods recommend items (academic papers) to a user (author) that are 'similar' to those they positively rated (i.e., authored or cited). The similarity between users and items is computed based on profiles built from textual information. User preferences and item attributes correspond to text features; in our case, keywords extracted from the titles of the papers, title of referenced papers and citation context. Recommendations are generated by means of user and item similarities in the text feature space.

More formally, an item i_n's profile consists of a vector $i_n = w_{n,1}, w_{n,2}, ..., w_{n,l} \in \mathbb{R}^L$ where $w_{n,l}$ denotes the relative relevance (weight) of feature f_l for i_n, and L is the number of existing features. To compute the weights $w_{m,l}$ we use TF-IDF [135].

Similarly, a user u_m's profile is represented as a vector $u_m = w_{m,1}, w_{m,2}, ..., w_{m,l} \in \mathbb{R}^L$, where $w_{m,l}$ denotes the relative relevance (weight) of feature f_l for u_m, and L is computed by aggregating the contents of all the papers that have a rating associated to the user, i.e., all the papers for which

the user has expressed an interest. The recommendation score of an item i for a target user u is then computed as the cosine similarity $score(u,i) = cos(u,i)$. We refer to this method as **cb**.

The textual features used to model user profiles for **cb** varies according to the available citation knowledge. For R_P, a user's profile is built by considering the titles of the papers that user has authored. For R_{PC}, a user's profile is built by considering the titles of the papers they authored and the papers they cited. Lastly, for R_{PCX}, R_{PCXS}, R_{PCXI} and R_{PCXSI}, a user's profile is built by considering the titles of the papers they authored, the tiles of the papers they cited, and the citation contexts, i.e., the texts around citations within the papers.

– **Collaborative Topic Modelling (ctr)**:

Wang et al. [272] built a hybrid method (–named as ctr) by combining the well-known Matrix Factorisation (MF) and topic modelling (Latent Dirichlet Allocation (LDA)) methods to recommend scientific publications to users. It represents users with their topics of interests and assumes that documents are generated by a topic model [272]. It is a generative model where for each user u, a latent vector is drawn based on their interaction with items. For each item, latent topics are inferred following a multi-nominal distribution of words from the textual contents of items using a probabilistic topic modelling model, LDA. We adapted the work of [272] considering users' authored as well as cited publications to generate the rating matrices R_P, R_{PC} and R_{PCX}. To generate the latent vectors for users and items, we

consider the titles of authored publications R_P, and titles of both authored and cited publications for R_{PC}, R_{PCX}. For the last matrix, we also consider citation context when generating the latent vectors. Note that, this baselines does not use citation section and citation intention. We refer it as **ctr**.

– **Factorisation Machine (fm)**:

As explained in Chapter 2, factorisation Machines have the benefit of being able to work with feature-rich datasets. We have selected Rendle's model [218] available as part of the RankSys recommendation framework.[11] for our work In addition to the user-rating matrix, we have included textual features in this model. We have also adjusted the dimensionality hyperparameter, k, and the global bias w_0 in our matrices as follows:

* R_P: textual features are extracted from the titles of users' authored publications. The number of textual features = 8,000. k=100, w_0 = 0.01.

* R_{PC} textual features are extracted from the titles of users' authored and cited publications. The number of textual features = 25,000. k=100, w_0 = 0.01.

* R_{PCX}, R_{PCXS}, R_{PCXI} and R_{PCXSI} textual features are extracted from the titles of users' authored and cited publications, as well as from the citation contexts The number of textual features = 60,000. k=100, w_0 = 0.01.

Note that, this baselines does not use citation section and citation intention. We refer it as **fm**.

---

11 https://github.com/RankSys/RankSys

- **Random (random**): We also consider as baseline the ran-
  dom recommendation method, where a recommendation
  list for a user is generated randomly. We use the random re-
  commendation method from the Ranksys recommendation
  framework[11]. We refer it as **random**.

- **Evaluation Metrics**: We have selected a variety of evaluation
  metrics to assess the different aspects of the recommendation. In
  particular, we consider precision (P), recall (R), F1 measure ($F1$),
  Mean Average Precision (MAP) (M), and Normalised Discounted
  Cumulative Graph (nDCG) (N). All of them at different levels
  of 5 and 10. More information about these metrics can be found
  in Chapter 2. Metrics were computed through the RiVal open
  source evaluation framework [225].

5.4.2 *Evaluation Results: Hyb & HybSec Recommendation Methods*

This section presents the results of our first evaluation using **Data-
set_sgc** and assessing our proposed Hyb and HybSec recommenda-
tion methods. The key objective of this evaluation has been to assess
whether citation graph, section and context are aspects of citation
knowledge that could help improving existing recommendation meth-
ods for scientific publications when addressing the task of providing
personalised recommendations of recent scientific publications to a
user.

The results of this offline evaluation, using *Dataset_sgc* for Hyb and
HybSec methods, against the baselines described above are presented
in Tables 5.3 and 5.4. Note that, the *hyb*5, *hyb*10, *hyb*15 and *hyb*20 refer
to our proposed hybrid methods with 5, 10, 15 and 20 neighbourhoods

respectively. Likewise, hybSecN (.,.,.,.) refers to our proposed hybrid method with N neighbourhoods containing weights for four selected sections; introduction, related work, conclusion and other sections. For example, the hybSec5 (1,0,0,0) method is a hybrid method with five neighbourhoods which indicates the citation is in the introduction section and not in any of the other sections, namely related work (second position), conclusion (third position) and other (fourth position) sections. The weight selected in this case is 1 for the introduction section and 0 for the rest of the sections.

The first conclusion that we derive from them is the fact that incorporating knowledge from the *citation graph* into the original author-publication matrix R_P entails an improvement of the generated recommendations for hybrid methods for all metrics, but only when the number of neighbours is higher than 10. The less rating sparsity of the R_{PC} matrix allows finding valuable item similarities and relations that are effectively exploited when more than 10 neighbours are considered.

When incorporating knowledge about *citation context*, captured in the R_{PCX} matrix, we achieve further improvements on our **hyb** approach with all neighbourhood sizes (i.e., 5, 10, 15 and 20 neighbours), for all metrics over both the matrices R_P and the R_{PC}. The best results for all metrics is obtained when considering a size of five neighbours. This indicates that citation context, is a prominent feature to enhance recommendations in our given setting. Even though there is an increment in the performance of the hybrid methods, but a decrease in the performance of the **cb** method. The saturation in the textual features between R_{PC} and R_{PCX} matrices may be the cause of this drop in performance.

Table 5.3: Experimental results of the baselines and proposed hybrid recommendation methods. A gray scale is used to highlight better (dark gray) and worst (white) values for each each metric (column). For every metric, the best values are in bold. Here, P stands for precision, R stands for recall, M stands for MAP and N stands for nDCG.

Matrix	Method	P@5	P@10	R@5	R@10	F1@5	F1@10	M@5	M@10	N@5	N@10
R_P	hyb5	0.056	0.04	0.071	0.09	0.062	0.055	0.044	0.047	0.08	0.083
	hyb10	0.059	0.039	0.072	0.093	0.065	0.055	0.044	0.048	0.084	0.084
	hyb15	0.055	0.039	0.068	0.088	0.061	0.054	0.041	0.045	0.078	0.081
	hyb20	0.05	0.037	0.064	0.086	0.056	0.052	0.041	0.046	0.076	0.081
	cb	0.054	0.039	0.063	0.088	0.058	0.054	0.044	0.048	0.081	0.084
	ctr	0	0	0.001	0.001	0	0	0	0	0.001	0.001
	fm	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002
	random	0	0	0.002	0.002	0	0	0.001	0.001	0.001	0.001
R_{PC}	hyb5	0.052	0.04	0.066	0.095	0.058	0.056	0.043	0.049	0.078	0.085
	hyb10	0.056	0.037	0.068	0.091	0.062	0.053	0.044	0.048	0.08	0.083
	hyb15	0.055	0.04	0.068	0.09	0.061	0.055	0.044	0.048	0.081	0.085
	hyb20	0.054	0.041	0.067	0.092	0.06	0.057	0.044	0.048	0.081	0.086
	cb	0.052	0.041	0.062	0.091	0.056	0.056	0.041	0.047	0.077	0.084
	ctr	0	0	0.001	0.001	0	0	0	0	0.001	0.001
	fm	0.002	0.002	0.001	0.002	0.002	0.002	0.001	0.001	0.002	0.002
	random	0.001	0.001	0.001	0.001	0.001	0.001	0	0	0.001	0.001
R_{PCX}	hyb5	0.065	0.048	0.076	0.108	0.07	0.066	0.053	0.06	0.095	0.102
	hyb10	0.062	0.047	0.077	0.106	0.069	0.065	0.052	0.058	0.092	0.098
	hyb15	0.062	0.045	0.076	0.105	0.068	0.063	0.051	0.056	0.091	0.096
	hyb20	0.06	0.044	0.073	0.104	0.066	0.062	0.049	0.055	0.088	0.095
	cb	0.029	0.023	0.04	0.061	0.034	0.033	0.025	0.029	0.043	0.05
	ctr	0	0	0.001	0.001	0	0	0.001	0.001	0.001	0.001
	fm	0.002	0.002	0.001	0.003	0.002	0.003	0.001	0.001	0.002	0.003
	random	0	0	0	0	0	0	0	0	0	0
R_{PCXS}	hybSec5 (0, 0, 0, 1)	0.059	0.045	0.073	0.103	0.065	0.063	0.05	0.056	0.086	0.095
	hybSec5 (0, 0, 1, 0)	0.058	0.043	0.07	0.098	0.063	0.06	0.049	0.054	0.085	0.091
	hybSec5 (0, 1, 0, 0)	0.058	0.044	0.072	0.103	0.064	0.062	0.05	0.056	0.086	0.093
	hybSec5 (1, 0, 0, 0)	0.059	0.044	0.074	0.101	0.065	0.061	0.05	0.055	0.086	0.093
	hybSec5 (0, 0, 0.5, 0.5)	0.056	0.044	0.068	0.1	0.061	0.061	0.048	0.054	0.082	0.091
	hybSec5 (0, 0.5, 0.5, 0)	0.057	0.043	0.069	0.097	0.062	0.059	0.048	0.053	0.083	0.09
	hybSec5 (0.5, 0.5, 0, 0)	0.057	0.043	0.069	0.096	0.062	0.059	0.048	0.054	0.083	0.09
	hybSec5 (0.5, 0, 0, 0.5)	0.055	0.043	0.069	0.1	0.061	0.06	0.048	0.054	0.082	0.091
	hybSec5 (0.3, 0.3, 0.3, 0.1)	0.054	0.041	0.067	0.094	0.06	0.057	0.047	0.052	0.08	0.086
	hybSec5 (0.3, 0.3, 0.1, 0.3)	0.054	0.041	0.066	0.095	0.06	0.058	0.047	0.052	0.08	0.087
	hybSec5 (0.3, 0.1, 0.3, 0.3)	0.054	0.041	0.066	0.095	0.059	0.058	0.047	0.052	0.079	0.087
	hybSec5 (0.1, 0.3, 0.3, 0.3)	0.054	0.041	0.066	0.094	0.059	0.057	0.047	0.052	0.079	0.086
	hybSec5 (0.25, 0.25, 0.25, 0.25)	0.054	0.041	0.066	0.094	0.059	0.057	0.047	0.052	0.079	0.086

Table 5.4: Experimental results of the baselines and proposed hybrid recommendation methods. A gray scale is used to highlight better (dark gray) and worst (white) values for each each metric (column). For every metric, the best values are in bold. Here, P stands for precision, R stands for recall, M stands for MAP and N stands for nDCG.

Matrix	Method	P@5	P@10	R@5	R@10	F1@5	F1@10	M@5	M@10	N@5	N@10
	hybSec5 (0, 0, 0, 1)	0.059	0.045	0.073	0.103	0.065	0.063	0.05	0.056	0.086	0.095
	hybSec5 (0, 0, 1, 0)	0.058	0.043	0.07	0.098	0.063	0.06	0.049	0.054	0.085	0.091
	hybSec5 (0, 1, 0, 0)	0.058	0.044	0.072	0.103	0.064	0.062	0.05	0.056	0.086	0.093
	hybSec5 (1, 0, 0, 0)	0.059	0.044	0.074	0.101	0.065	0.061	0.05	0.055	0.086	0.093
	hybSec5 (0, 0, 0.5, 0.5)	0.056	0.044	0.068	0.1	0.061	0.061	0.048	0.054	0.082	0.091
	hybSec5 (0, 0.5, 0.5, 0)	0.057	0.043	0.069	0.097	0.062	0.059	0.048	0.053	0.083	0.09
	hybSec5 (0.5, 0.5, 0, 0)	0.057	0.043	0.069	0.096	0.062	0.059	0.048	0.054	0.083	0.09
	hybSec5 (0.5, 0, 0, 0.5)	0.055	0.043	0.069	0.1	0.061	0.06	0.048	0.054	0.082	0.091
	hybSec5 (0.3, 0.3, 0.3, 0.1)	0.054	0.041	0.067	0.094	0.06	0.057	0.047	0.052	0.08	0.086
	hybSec5 (0.3, 0.3, 0.1, 0.3)	0.054	0.041	0.066	0.095	0.06	0.058	0.047	0.052	0.08	0.087
	hybSec5 (0.3, 0.1, 0.3, 0.3)	0.054	0.041	0.066	0.095	0.059	0.058	0.047	0.052	0.079	0.087
	hybSec5 (0.1, 0.3, 0.3, 0.3)	0.054	0.041	0.066	0.094	0.059	0.057	0.047	0.052	0.079	0.086
	hybSec5 (0.25, 0.25, 0.25, 0.25)	0.054	0.041	0.066	0.094	0.059	0.057	0.047	0.052	0.079	0.086
	hybSec10 (0, 0, 0, 1)	0.053	0.041	0.07	0.098	0.061	0.058	0.048	0.053	0.082	0.089
	hybSec10 (0, 0, 1, 0)	0.05	0.038	0.066	0.089	0.057	0.054	0.044	0.049	0.076	0.082
	hybSec10 (0, 1, 0, 0)	0.051	0.039	0.068	0.092	0.058	0.055	0.047	0.051	0.079	0.085
	hybSec10 (1, 0, 0, 0)	0.05	0.039	0.064	0.092	0.056	0.055	0.045	0.05	0.077	0.084
	hybSec10 (0, 0, 0.5, 0.5)	0.049	0.038	0.064	0.089	0.055	0.053	0.044	0.048	0.075	0.081
	hybSec10 (0, 0.5, 0.5, 0)	0.05	0.037	0.065	0.088	0.057	0.053	0.045	0.049	0.077	0.082
	hybSec10 (0.5, 0.5, 0, 0)	0.05	0.038	0.065	0.088	0.056	0.053	0.045	0.049	0.077	0.082
	hybSec10 (0.5, 0, 0, 0.5)	0.05	0.038	0.065	0.09	0.056	0.054	0.045	0.05	0.077	0.083
	hybSec10 (0.3, 0.3, 0.3, 0.1)	0.047	0.036	0.063	0.085	0.054	0.05	0.043	0.047	0.072	0.078
	hybSec10 (0.3, 0.3, 0.1, 0.3)	0.048	0.037	0.063	0.086	0.055	0.051	0.043	0.047	0.074	0.079
	hybSec10 (0.3, 0.1, 0.3, 0.3)	0.048	0.036	0.063	0.085	0.055	0.051	0.043	0.047	0.074	0.079
	hybSec10 (0.1, 0.3, 0.3, 0.3)	0.048	0.036	0.063	0.086	0.054	0.051	0.043	0.047	0.073	0.079
	hybSec10 (0.25, 0.25, 0.25, 0.25)	0.048	0.036	0.063	0.085	0.055	0.051	0.043	0.047	0.073	0.078
R_{PCXS}	hybSec15 (0, 0, 0, 1)	0.054	0.041	0.07	0.099	0.061	0.058	0.047	0.052	0.082	0.089
	hybSec15 (0, 0, 1, 0)	0.05	0.038	0.066	0.089	0.057	0.053	0.045	0.049	0.077	0.082
	hybSec15 (0, 1, 0, 0)	0.052	0.038	0.069	0.089	0.059	0.053	0.047	0.05	0.08	0.084
	hybSec15 (0, 0.5, 0.5, 0)	0.05	0.037	0.065	0.085	0.057	0.051	0.046	0.05	0.078	0.082
	hybSec15 (1, 0, 0, 0)	0.052	0.038	0.067	0.091	0.058	0.054	0.046	0.05	0.079	0.084
	hybSec15 (0, 0, 0.5, 0.5)	0.05	0.037	0.066	0.085	0.057	0.051	0.046	0.049	0.078	0.082
	hybSec15 (0.5, 0.5, 0, 0)	0.051	0.037	0.066	0.086	0.057	0.051	0.046	0.049	0.078	0.082
	hybSec15 (0.5, 0, 0, 0.5)	0.051	0.037	0.067	0.088	0.058	0.052	0.047	0.051	0.08	0.084
	hybSec15 (0.3, 0.3, 0.3, 0.1)	0.047	0.035	0.063	0.084	0.054	0.05	0.043	0.047	0.074	0.078
	hybSec15 (0.3, 0.3, 0.1, 0.3)	0.049	0.036	0.065	0.084	0.056	0.05	0.045	0.048	0.076	0.079
	hybSec15 (0.3, 0.1, 0.3, 0.3)	0.049	0.036	0.065	0.086	0.056	0.051	0.044	0.048	0.076	0.08
	hybSec15 (0.1, 0.3, 0.3, 0.3)	0.048	0.035	0.064	0.085	0.055	0.05	0.044	0.048	0.075	0.079
	hybSec15 (0.25, 0.25, 0.25, 0.25)	0.049	0.035	0.064	0.085	0.055	0.05	0.044	0.048	0.076	0.079
	hybSec20 (0, 0, 0, 1)	0.054	0.04	0.07	0.098	0.061	0.057	0.047	0.052	0.082	0.088
	hybSec20 (0, 0, 1, 0)	0.05	0.036	0.066	0.088	0.057	0.051	0.044	0.048	0.077	0.081
	hybSec20 (0, 1, 0, 0)	0.052	0.037	0.069	0.089	0.059	0.052	0.046	0.049	0.08	0.082
	hybSec20 (1, 0, 0, 0)	0.051	0.037	0.066	0.092	0.058	0.053	0.045	0.049	0.078	0.083
	hybSec20 (0, 0, 0.5, 0.5)	0.051	0.037	0.066	0.087	0.058	0.051	0.045	0.048	0.078	0.081
	hybSec20 (0, 0.5, 0.5, 0)	0.052	0.036	0.066	0.085	0.058	0.05	0.045	0.049	0.078	0.081
	hybSec20 (0.5, 0.5, 0, 0)	0.051	0.036	0.065	0.085	0.057	0.05	0.045	0.048	0.077	0.08
	hybSec20 (0.5, 0, 0, 0.5)	0.051	0.037	0.067	0.088	0.058	0.052	0.046	0.05	0.079	0.083
	hybSec20 (0.3, 0.3, 0.3, 0.1)	0.048	0.035	0.063	0.086	0.054	0.05	0.042	0.046	0.074	0.078
	hybSec20 (0.3, 0.3, 0.1, 0.3)	0.05	0.035	0.065	0.083	0.056	0.049	0.044	0.047	0.076	0.078
	hybSec20 (0.3, 0.1, 0.3, 0.3)	0.05	0.035	0.065	0.086	0.056	0.05	0.044	0.048	0.076	0.08
	hybSec20 (0.1, 0.3, 0.3, 0.3)	0.049	0.035	0.064	0.083	0.056	0.049	0.043	0.047	0.075	0.078
	hybSec20 (0.25, 0.25, 0.25, 0.25)	0.049	0.035	0.064	0.083	0.056	0.049	0.043	0.046	0.075	0.078

When adding information about *citation section*, on the R_{PCX} matrix, our **hybSec** approach does not outperform its **hyb** counterpart. This indicates that the section of the citation within the paper may not be a relevant feature to enhance recommendations in the studied scenario.

The proposed baselines also performed poorly in this scenario, with content based methods being the most competitive. We hypothesise that the low performance from **ctr** may be due to the model dependency on the LDA topic model. LDA is an unsupervised learning model. To generates topics it requires a large amount of textual contents [32], otherwise the model may be unable to effectively learn available features. Regarding the **fm** baseline we hypothesise that, with the incorporated features were insufficient, or the dataset is too sparse to detect pairwise feature interactions.

It is important to highlight that we are targeting a particular difficult scenario, recommending the most recent publications to users. In this scenario, items in the test set do not have any connections to items in the training set, hence CF methods do not work, and some of the studied baselines (see Section 5.4.1) performed very poorly. Even our proposed hybrid method, which has proven to outperform the best performing baseline, in this case content-based, also shows moderate performance. It is, however, promising to observe how, in this scenario, the use of citation knowledge, and more particularly the use of the citation-graph and citation context, can help providing more accurate recommendations to users.

In Tables 5.3 and 5.4, we have shown the results in a grayscale to differentiate among the highest and the lowest values of each metric across methods (– both proposed and baselines). The darker shade represents the highest values obtained for the metric across methods while the lighter shade encapsulates the lowest values in comparison.

To obtain this, we have used the 'conditional formatting' functionality of 'Microsoft Excel'. This functionality is often used as colour-based formatting to highlight, emphasise, or differentiate among data.

To be consistent with the previous technical chapter, we conducted a statistical test in this chapter as well. However, the choice of a statistical hypothesis test is a challenging open problem for interpreting machine learning results. Given the nature of the problem (recommending recent scientific publications to a user) we could not use a standard cross-validation approach for evaluation [65] but we used a time-split based partition of train and test sets of the data.

According to [65], the McNemar's test is a suitable test for those cases where the algorithms that are being compared are evaluated on a single test set, as opposed to repeated evaluations via a re-sampling technique, such as k-fold cross-validation. Therefore, we conducted the McNemar' test which is a paired distribution-free statistical hypothesis test.

Due to a large number of conducted experiments, for simplicity, we show the paired comparison between the competitive baselines and the highest performing methods. The results are shown in Table 5.5. The alpha value is higher than the obtained p-values, showing that the results are statistically significant.

Table 5.5: McNemar hypothesis significance test results between the baseline and the proposed methods.

Matrix	Baseline	Methods	Statistics	p-value	alpha
R_P	cb	hyb5	139	0.025	
R_{PC}	cb	hyb5	134	0.0011	
R_{PCX}	cb	hyb5	103	1.767×10^{-14}	
		hybSec5 (0, 0, 0, 1)	104	3.194×10^{-13}	
		hybSec5 (0, 0, 1, 0)	112	6.705×10^{-11}	0.05
R_{PCXS}	cb	hybSec5 (0, 1_0, 0)	109	6.106×10^{-11}	
		hybSec5 (1, 0, 0, 0)	111	1.464×10^{-10}	

5.4.3   *Evaluation Results: HybIntent & HybIntentSec Recommendation Methods*

This section presents the results of our second evaluation using *Dataset_sgci* and assessing our proposed Hyb, HybIntent and HybIntentSec recommendation methods. The key objective of this evaluation has been to assess whether citation graph, section, context and intention are aspects of citation knowledge that could help improving existing recommendation methods for scientific publications when addressing the task of recommending the most recent scientific publications to a user.

The results of this offline evaluation, using Dataset_sgci, against the baselines are presented in Tables 5.6 and 5.7. Note that, the hybIntentN (.,.,.,.,.,.) refers to our proposed hybrid method with *N* neighbourhoods, where intention is represented with six weights that define whether the citation is used as background, compareOrcontrast, motivation, use, extend and future work. For example, the hybIntent5 (1,0,0,0,0,0) method is a hybrid method with five neighbourhoods which indicates the citation has been used with the intention of providing background as the weight at the first position is 1. The rest of the weights, compareorcontrast (second position), motivation (third position), use (fourth position), extend (fifth position) and future (sixth position) are 0. Likewise, in the hybrid methods leveraging both citation section and citation intention, for example, the *hybIntentSec5_concl − use* method is a hybrid method with five neighbourhoods where we select a weight of 1 when a paper is cited with the *use* intention in the *conclusion* (*concl*) section, for everything else we select a weight of 0.

Table 5.6: Experimental results of the baselines and proposed hybrid recommendation methods including **hybIntent**. A gray scale is used to highlight better (dark gray) and worst (white) values for each ranking metric. For every metric, the best values are highlighted in bold. Here, P stands for precision, R stands for recall, M stands for MAP and N stands for nDCG.

Matrix	Method	P@5	P@10	R@5	R@10	F1@5	F1@10	M@5	M@10	N@5	N@10
R_P	hyb5	0.05	0.037	0.063	0.094	0.056	0.053	0.039	0.045	0.073	0.08
	hyb10	0.05	0.039	0.063	0.099	0.056	0.056	0.039	0.045	0.073	0.081
	hyb15	0.049	0.039	0.061	0.099	0.054	0.056	0.039	0.045	0.072	0.081
	hyb20	0.048	0.039	0.057	0.096	0.052	0.055	0.037	0.043	0.069	0.079
	cb	0.053	0.043	0.065	0.105	0.058	0.061	0.044	0.052	0.08	0.091
	ctr	0	0	0.001	0.001	0	0	0	0	0.001	0.001
	fm	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002
	random	0	0	0	0	0	0	0	0	0	0
R_{PC}	hyb5	0.05	0.035	0.065	0.095	0.057	0.051	0.042	0.047	0.075	0.081
	hyb10	0.055	0.041	0.074	0.106	0.063	0.059	0.046	0.052	0.082	0.09
	hyb15	0.057	0.042	0.077	0.108	0.066	0.06	0.048	0.053	0.085	0.092
	hyb20	0.056	0.041	0.077	0.108	0.065	0.059	0.047	0.052	0.083	0.09
	cb	0.056	0.046	0.071	0.112	0.063	0.065	0.047	0.052	0.083	0.095
	ctr	0	0.001	0.001	0.001	0	0.001	0	0.001	0.001	0.001
	fm	0.002	0.002	0.001	0.002	0.002	0.002	0.001	0.001	0.002	0.002
	random	0	0	0	0	0	0	0.001	0	0	0
R_{PCX}	hyb5	0.054	0.039	0.07	0.098	0.061	0.056	0.043	0.048	0.08	0.085
	hyb10	0.058	0.043	0.074	0.112	0.065	0.062	0.049	0.055	0.086	0.095
	hyb15	0.059	0.043	0.074	0.108	0.066	0.062	0.051	0.057	0.089	0.096
	hyb20	0.057	0.044	0.073	0.108	0.064	0.063	0.052	0.058	0.088	0.097
	cb	0.036	0.027	0.048	0.07	0.041	0.039	0.032	0.036	0.054	0.059
	ctr	0.002	0.002	0.003	0.003	0.002	0.001	0.005	0.005	0.003	0.002
	fm	0.002	0.002	0.001	0.003	0.002	0.003	0.001	0.001	0.002	0.003
	random	0	0	0.001	0.001	0	0.001	0.001	0.001	0.001	0.001
R_{PCXI}	hybIntent5(0, 0, 0, 0, 0, 1)	0.067	0.062	0.111	0.172	0.084	0.091	0.066	0.078	0.099	0.126
	hybIntent5(0, 0, 0, 0, 1, 0)	0.067	0.062	0.11	0.168	0.083	0.091	0.066	0.079	0.101	0.126
	hybIntent5(0, 0, 0, 1, 0, 0)	0.06	0.058	0.106	0.161	0.077	0.085	0.064	0.075	0.096	0.121
	hybIntent5(0, 0, 1, 0, 0, 0)	0.067	0.061	0.111	0.167	0.084	0.089	0.067	0.079	0.101	0.126
	hybIntent5(0, 1, 0, 0, 0, 0)	0.04	0.039	0.074	0.125	0.052	0.059	0.046	0.055	0.069	0.089
	hybIntent5(1, 0, 0, 0, 0, 0)	0.037	0.034	0.063	0.104	0.047	0.051	0.041	0.048	0.063	0.078
	hybIntent5 (0.15, 0.15, 0.15, 0.15, 0.15, 0.25)	0.037	0.03	0.06	0.093	0.046	0.045	0.041	0.047	0.062	0.073
	hybIntent5 (0.15, 0.15, 0.15, 0.15, 0.25, 0.15)	0.037	0.03	0.061	0.094	0.046	0.045	0.041	0.047	0.062	0.074
	hybIntent5 (0.15, 0.15, 0.15, 0.25, 0.15, 0.15)	0.037	0.031	0.06	0.096	0.046	0.047	0.041	0.047	0.061	0.074
	hybIntent5 (0.15, 0.15, 0.25, 0.15, 0.15, 0.15)	0.037	0.03	0.06	0.094	0.046	0.045	0.041	0.046	0.061	0.073
	hybIntent5 (0.15, 0.25, 0.15, 0.15, 0.15, 0.15)	0.035	0.03	0.058	0.093	0.044	0.045	0.04	0.046	0.059	0.072
	hybIntent5 (0.25, 0.15, 0.15, 0.15, 0.15, 0.15)	0.037	0.03	0.06	0.091	0.046	0.045	0.041	0.047	0.062	0.073
	hybIntent10(0, 0, 0, 0, 0, 1)	0.048	0.042	0.094	0.144	0.064	0.065	0.051	0.061	0.075	0.096
	hybIntent10(0, 0, 0, 0, 1, 0)	0.046	0.041	0.087	0.143	0.06	0.064	0.048	0.059	0.073	0.095
	hybIntent10(0, 0, 0, 1, 0, 0)	0.049	0.04	0.097	0.14	0.065	0.062	0.054	0.063	0.08	0.097
	hybIntent10(0, 0, 1, 0, 0, 0)	0.049	0.043	0.096	0.147	0.065	0.067	0.052	0.062	0.077	0.098
	hybIntent10(0, 1, 0, 0, 0, 0)	0.035	0.027	0.067	0.095	0.046	0.042	0.039	0.044	0.06	0.07
	hybIntent10(1, 0, 0, 0, 0, 0)	0.041	0.029	0.068	0.092	0.051	0.044	0.04	0.044	0.063	0.07
	hybIntent10 (0.15, 0.15, 0.15, 0.15, 0.15, 0.25)	0.039	0.028	0.062	0.087	0.048	0.042	0.036	0.041	0.06	0.066
	hybIntent10 (0.15, 0.15, 0.15, 0.15, 0.25, 0.15)	0.039	0.028	0.062	0.088	0.048	0.042	0.037	0.041	0.06	0.067
	hybIntent10 (0.15, 0.15, 0.15, 0.25, 0.15, 0.15)	0.039	0.027	0.062	0.087	0.048	0.041	0.036	0.04	0.059	0.066
	hybIntent10 (0.15, 0.15, 0.25, 0.15, 0.15, 0.15)	0.039	0.028	0.062	0.087	0.048	0.042	0.036	0.041	0.06	0.066
	hybIntent10 (0.15, 0.25, 0.15, 0.15, 0.15, 0.15)	0.036	0.027	0.057	0.084	0.044	0.041	0.034	0.039	0.056	0.064
	hybIntent10 (0.25, 0.15, 0.15, 0.15, 0.15, 0.15)	0.038	0.026	0.062	0.086	0.047	0.04	0.036	0.041	0.058	0.065
	hybIntent15(0, 0, 0, 0, 0, 1)	0.034	0.033	0.068	0.125	0.045	0.052	0.035	0.045	0.053	0.075
	hybIntent15(0, 0, 0, 0, 1, 0)	0.035	0.031	0.063	0.114	0.045	0.049	0.034	0.043	0.054	0.072
	hybIntent15(0, 0, 0, 1, 0, 0)	0.037	0.033	0.074	0.123	0.049	0.052	0.039	0.048	0.059	0.078
	hybIntent15(0, 0, 1, 0, 0, 0)	0.034	0.033	0.068	0.127	0.045	0.052	0.035	0.045	0.053	0.076
	hybIntent15(0, 1, 0, 0, 0, 0)	0.03	0.024	0.053	0.085	0.038	0.037	0.03	0.035	0.049	0.06
	hybIntent15(1, 0, 0, 0, 0, 0)	0.04	0.027	0.063	0.086	0.049	0.041	0.036	0.04	0.059	0.065
	hybIntent15 (0.15, 0.15, 0.15, 0.15, 0.15, 0.25)	0.036	0.026	0.055	0.079	0.044	0.039	0.033	0.038	0.054	0.061
	hybIntent15 (0.15, 0.15, 0.15, 0.15, 0.25, 0.15)	0.037	0.026	0.056	0.079	0.045	0.039	0.033	0.038	0.055	0.061
	hybIntent15 (0.15, 0.15, 0.15, 0.25, 0.15, 0.15)	0.035	0.026	0.054	0.079	0.042	0.039	0.033	0.037	0.053	0.061
	hybIntent15 (0.15, 0.15, 0.25, 0.15, 0.15, 0.15)	0.036	0.026	0.055	0.079	0.044	0.039	0.033	0.038	0.054	0.061
	hybIntent15 (0.15, 0.25, 0.15, 0.15, 0.15, 0.15)	0.036	0.025	0.053	0.075	0.043	0.038	0.032	0.036	0.053	0.059
	hybIntent15 (0.25, 0.15, 0.15, 0.15, 0.15, 0.15)	0.036	0.027	0.058	0.084	0.044	0.041	0.035	0.039	0.055	0.063
	hybIntent20(0, 0, 0, 0, 0, 1)	0.027	0.026	0.05	0.101	0.035	0.041	0.028	0.037	0.042	0.061
	hybIntent20(0, 0, 0, 0, 1, 0)	0.028	0.025	0.046	0.091	0.035	0.039	0.028	0.036	0.043	0.059
	hybIntent20(0, 0, 0, 1, 0, 0)	0.03	0.027	0.055	0.101	0.039	0.043	0.033	0.041	0.049	0.067
	hybIntent20(0, 0, 1, 0, 0, 0)	0.026	0.026	0.05	0.101	0.034	0.041	0.028	0.037	0.042	0.061
	hybIntent20(0, 1, 0, 0, 0, 0)	0.027	0.022	0.041	0.074	0.033	0.034	0.026	0.031	0.042	0.053
	hybIntent20(1, 0, 0, 0, 0, 0)	0.039	0.026	0.06	0.08	0.047	0.039	0.035	0.038	0.057	0.062
	hybIntent20 (0.15, 0.15, 0.15, 0.15, 0.15, 0.25)	0.033	0.025	0.051	0.076	0.04	0.038	0.031	0.036	0.05	0.058
	hybIntent20 (0.15, 0.15, 0.15, 0.15, 0.25, 0.15)	0.033	0.025	0.051	0.076	0.04	0.038	0.031	0.036	0.05	0.058
	hybIntent20 (0.15, 0.15, 0.15, 0.25, 0.15, 0.15)	0.033	0.024	0.051	0.075	0.04	0.036	0.03	0.035	0.049	0.057
	hybIntent20 (0.15, 0.15, 0.25, 0.15, 0.15, 0.15)	0.033	0.025	0.051	0.076	0.04	0.038	0.031	0.036	0.05	0.058
	hybIntent20 (0.15, 0.25, 0.15, 0.15, 0.15, 0.15)	0.033	0.024	0.048	0.072	0.039	0.036	0.029	0.034	0.049	0.056
	hybIntent20 (0.25, 0.15, 0.15, 0.15, 0.15, 0.15)	0.036	0.026	0.056	0.081	0.044	0.039	0.033	0.038	0.054	0.061

Table 5.7: Experimental results of the baselines and proposed hybrid recommendation methods including competitive part of **hybIntent** and **hybIntentSec**. A gray scale is used to highlight better (dark gray) and worst (white) values for each ranking metric. For every metric, the best values are highlighted in bold. Here, P stands for precision, R stands for recall, M stands for MAP and N stands for nDCG.

Matrix	Method	P@5	P@10	R@5	R@10	F1@5	F1@10	M@5	M@10	N@5	N@10
R_{PCXI}	hybIntent5 (0, 0, 0, 0, 0, 1)	0.067	0.062	0.111	0.172	0.084	0.091	0.066	0.078	0.099	0.126
	hybIntent5 (0, 0, 0, 0, 1, 0)	0.067	0.062	0.11	0.168	0.083	0.091	0.066	0.079	0.101	0.126
	hybIntent5 (0, 0, 0, 1, 0, 0)	0.06	0.058	0.106	0.161	0.077	0.085	0.064	0.075	0.096	0.121
	hybIntent5 (0, 0, 1, 0, 0, 0)	0.067	0.061	0.111	0.167	0.084	0.089	0.067	0.079	0.101	0.126
	hybIntent5 (0, 1, 0, 0, 0, 0)	0.04	0.039	0.074	0.125	0.052	0.059	0.046	0.055	0.069	0.089
	hybIntent5 (1, 0, 0, 0, 0, 0)	0.037	0.034	0.063	0.104	0.047	0.051	0.041	0.048	0.063	0.078
R_{PCXSI}	hybIntentSec5_concl-use	0.068	0.062	0.112	0.169	0.085	0.09	0.067	0.079	0.102	0.127
	hybIntentSec5_intro-bkg	0.041	0.045	0.071	0.138	0.052	0.067	0.048	0.059	0.07	0.097
	hybIntentSec5_intro-com	0.069	0.062	0.114	0.171	0.086	0.091	0.07	0.083	0.105	0.13
	hybIntentSec5_intro-ext	0.07	0.063	0.114	0.172	0.087	0.092	0.069	0.081	0.105	0.129
	hybIntentSec5_other-bkg	0.043	0.036	0.075	0.115	0.054	0.055	0.041	0.048	0.066	0.081
	hybIntentSec5_other-com	0.061	0.062	0.106	0.172	0.078	0.091	0.066	0.08	0.098	0.128
	hybIntentSec5_other-ext	0.065	0.061	0.109	0.167	0.081	0.089	0.065	0.078	0.099	0.125
	hybIntentSec5_other-fut	0.067	0.062	0.111	0.172	0.083	0.091	0.066	0.078	0.099	0.126
	hybIntentSec5_other-mot	0.067	0.061	0.111	0.167	0.083	0.09	0.067	0.079	0.101	0.126
	hybIntentSec5_other-use	0.059	0.058	0.105	0.163	0.076	0.085	0.063	0.074	0.095	0.121
	hybIntentSec5_relW-bkg	0.068	0.063	0.113	0.172	0.085	0.092	0.068	0.08	0.103	0.129
	hybIntentSec5_relW-com	0.039	0.039	0.073	0.123	0.051	0.059	0.044	0.053	0.067	0.087
	hybIntentSec10_concl-use	0.048	0.042	0.096	0.146	0.064	0.065	0.052	0.062	0.077	0.098
	hybIntentSec10_intro-bkg	0.042	0.03	0.07	0.095	0.052	0.045	0.04	0.045	0.063	0.072
	hybIntentSec10_intro-com	0.051	0.044	0.098	0.149	0.067	0.068	0.055	0.066	0.08	0.102
	hybIntentSec10_intro-ext	0.052	0.045	0.098	0.15	0.068	0.069	0.054	0.064	0.08	0.101
	hybIntentSec10_other-bkg	0.043	0.03	0.068	0.105	0.053	0.047	0.037	0.043	0.062	0.073
	hybIntentSec10_other-com	0.047	0.04	0.094	0.141	0.062	0.062	0.056	0.065	0.079	0.099
	hybIntentSec10_other-ext	0.044	0.04	0.086	0.142	0.059	0.063	0.048	0.059	0.071	0.094
	hybIntentSec10_other-fut	0.048	0.042	0.094	0.144	0.063	0.065	0.051	0.061	0.075	0.096
	hybIntentSec10_other-mot	0.049	0.043	0.096	0.147	0.065	0.066	0.052	0.062	0.077	0.098
	hybIntentSec10_other-use	0.047	0.04	0.095	0.14	0.063	0.062	0.052	0.061	0.076	0.095
	hybIntentSec10_relW-bkg	0.049	0.043	0.096	0.148	0.065	0.067	0.052	0.063	0.077	0.099
	hybIntentSec10_relW-com	0.032	0.026	0.064	0.095	0.043	0.04	0.035	0.04	0.055	0.066
	hybIntentSec15_concl-use	0.033	0.033	0.068	0.126	0.045	0.053	0.035	0.045	0.053	0.076
	hybIntentSec15_intro-bkg	0.034	0.027	0.055	0.086	0.042	0.041	0.033	0.038	0.052	0.063
	hybIntentSec15_intro-com	0.037	0.034	0.072	0.128	0.049	0.054	0.039	0.049	0.058	0.079
	hybIntentSec15_intro-ext	0.037	0.035	0.07	0.129	0.049	0.055	0.037	0.047	0.057	0.078
	hybIntentSec15_other-bkg	0.04	0.028	0.062	0.091	0.049	0.043	0.033	0.038	0.057	0.065
	hybIntentSec15_other-com	0.036	0.033	0.072	0.124	0.048	0.052	0.041	0.049	0.06	0.08
	hybIntentSec15_other-ext	0.034	0.03	0.063	0.114	0.044	0.047	0.034	0.043	0.053	0.071
	hybIntentSec15_other-fut	0.034	0.033	0.068	0.125	0.045	0.052	0.035	0.045	0.053	0.075
	hybIntentSec15_other-mot	0.034	0.033	0.068	0.127	0.045	0.053	0.035	0.045	0.053	0.076
	hybIntentSec15_other-use	0.035	0.032	0.071	0.12	0.047	0.05	0.036	0.045	0.055	0.074
	hybIntentSec15_relW-bkg	0.034	0.033	0.069	0.127	0.046	0.053	0.036	0.046	0.054	0.076
	hybIntentSec15_relW-com	0.027	0.022	0.049	0.082	0.035	0.035	0.026	0.032	0.044	0.055
	hybIntentSec20_concl-use	0.026	0.026	0.05	0.101	0.034	0.041	0.028	0.037	0.042	0.062
	hybIntentSec20_intro-bkg	0.03	0.024	0.046	0.077	0.037	0.036	0.03	0.035	0.047	0.057
	hybIntentSec20_intro-com	0.03	0.027	0.055	0.105	0.039	0.043	0.033	0.041	0.047	0.066
	hybIntentSec20_intro-ext	0.03	0.028	0.052	0.103	0.038	0.044	0.03	0.039	0.046	0.064
	hybIntentSec20_other-bkg	0.039	0.026	0.056	0.083	0.046	0.04	0.031	0.035	0.054	0.061
	hybIntentSec20_other-com	0.03	0.026	0.056	0.1	0.039	0.041	0.035	0.042	0.05	0.067
	hybIntentSec20_other-ext	0.026	0.024	0.046	0.091	0.034	0.038	0.028	0.036	0.042	0.059
	hybIntentSec20_other-fut	0.027	0.026	0.05	0.101	0.035	0.041	0.028	0.037	0.042	0.061
	hybIntentSec20_other-mot	0.026	0.026	0.05	0.101	0.035	0.041	0.028	0.037	0.042	0.061
	hybIntentSec20_other-use	0.029	0.026	0.053	0.098	0.038	0.041	0.03	0.039	0.046	0.063
	hybIntentSec20_relW-bkg	0.027	0.026	0.051	0.102	0.035	0.041	0.029	0.038	0.043	0.062
	hybIntentSec20_relW-com	0.023	0.02	0.037	0.07	0.029	0.031	0.022	0.027	0.037	0.048

The foremost conclusion we draw from the experiment is that the incorporation of citation knowledge helps improving the performance of recommendation methods. Between the rating matrices R_P and R_{PC} (where R_{PC} reduces the sparsity because of the additional citation information added to it), **cb** and our **hyb** method show improvements with all the evaluation metrics at the cut-off points of 5 and 10. The improvement continues in all the sizes of neighbourhoods, i.e., 5, 10, 15 and 20. This shows that the decrease on the sparsity of the matrix, thanks to the incorporation of user preferences based on the citation graph, allows finding valuable relationships between items and users that are productively exploited.

When incorporating citation knowledge in the form of citation context, captured in the R_{PCX} matrix, we notice further increment in the performance of the hybrid methods, but a decrease in the performance of the **cb** method. The saturation in the textual features between R_{PC} and R_{PCX} may be the cause of this drop in performance.

Exploiting citation knowledge in the form of citation intentions (**hybIntent**), captured in the R_{PCXI} matrix, outperformed all other methods (i.e., **hyb** and **cb**) for all the evaluation metrics. This shows that citation intention is a useful feature for recommending recent papers to a user. In particular, the performance of our **hybIntent** method is higher when the citation belongs to a **Future** or **Extension** categories. This corroborates the intuition that citations that are related to future research and to work that is being extended or enhanced, represent relevant pointers for new directions in a research field. As can be seen in Table 5.6, precision, recall and nDCG@5 values are higher when citations belong to *Future* category, while nDCG, precision and recall@5 values are higher when the citation intention is *Extension*. In this context, while higher precision and recall values

show the methods were able to find relevant and new items, the higher nDCG value shows that novel papers are appearing earlier in the recommendation lists. In addition, the **Motivation** intention category also has positive impact on the recommendation. This can imply, citations that motivate the work are helpful on reflecting problems that may be targeted not only in the present, but also in future scientific publications.

Our final recommendation method, combining all the different types of citation knowledge **hybIntentSec**, outperforms *hybIntent*. As can be seen in Table 5.7, in particular, citations intended to provide **Extensions** or to **compare or contrast** in the **Introduction (Intro)** section proved to be more effective than other combinations. This corroborates the intuition that a paper cited in the introduction referring to a work that is being extended, represent an important pointer for new directions in a research field.

As in our previous experiments, our baselines did not perform adequately. The item cold-start problem of the addressed recommendation scenario, the sparsity of the rating matrices, and amount of text from which topics are being generated (particularly for the *ctr* and Factorisation Machine (FM) methods) are some of the factors influencing their performance.

In order to quantify the statistical significance, similar to Section 5.4.2, we conducted the McNemar's test. Due to a large number of conducted experiments, for simplicity, we show the paired comparison between the competitive baselines and the highest performing methods. The results are shown in Table 5.8. The alpha value is higher than the obtained p-values show that the results are statistically significant.

Table 5.8: McNemar hypothesis significance test results between the baseline and the proposed methods.

Matrix	Baseline	Methods	Statistics	p-value	alpha
R_P	cb	hyb5	82	0.0018	
R_{PC}	cb	hyb5	84	0.0025	
R_{PCX}	cb	hyb5	60	7.125×10^{-11}	
R_{PCXI}	cb	hybIntent5 (0, 0, 0, 0, 0, 1)	49	2.755×10^{-52}	
		hybIntent5_0, 0, 0, 0, 1, 0)	48	3.496×10^{-51}	
		hybIntent5 (0, 0, 0, 1, 0, 0)	46	1.145×10^{-50}	
		hybIntent5 (0, 0, 1, 0, 0,0)	48	7.593×10^{-54}	
		hybIntent5 (0, 1, 0, 0, 0, 0)	53	6.651×10^{-24}	
		hybIntent5 (1, 0, 0, 0, 0, 0)	55	7.435×10^{-15}	
R_{PCXSI}	cb	hybIntentSec5_concl-use	48	0	0.05
		hybIntentSec5_intro-bkg	55	3.069×10^{-27}	
		hybIntentSec5_intro-com	49	0	
		hybIntentSec5_intro-ext	49	0	
		hybIntentSec5_other-bkg	50	2.112×10^{-20}	
		hybIntentSec5_other-com	48	2.661×10^{-48}	
		hybIntentSec5_other-ext	48	3.496×10^{-51}	
		hybIntentSec5_other-fut	49	2.755×10^{-52}	
		hybIntentSec5_other-mot	48	7.593×10^{-54}	
		hybIntentSec5_other-use	47	2.546×10^{-50}	
		hybIntentSec5_relW-bkg	49	5.196×10^{-53}	
		hybIntentSec5_relW-com	54	2.514×10^{-24}	

5.5 DISCUSSION AND CONCLUSION

In this chapter, we have addressed the problem of **providing personalised recommendations of recently published papers**. For this problem, traditional recommendation methods, such as CF, cannot be used since they are unable to establish rating-based similarities and patterns between new items. Motivated by this fact, in addition to content-based features, we advocate for the exploitation of paper citations knowledge as a bridge to link related papers.

Note that, while citation knowledge has been explored in the literature to provide paper recommendations in different scenarios, our research brings two key novelties with respect to previous works: (i) a real-world and challenging scenario, where recent papers are the

ones to be recommended, and (ii) an exploration of a wider notion of citation knowledge for this particular scenario, which includes the citation graph, citation context, citation section, and citation intention.

In particular, we have presented a series of hybrid approaches that make use of the citation graph to enrich the rating matrix, while exploring the use of citation context, citation section and citation intention as features for recommendations. Our experimental results have shown that incorporating citation knowledge in terms of the citation graph, citation context, and citation intention allows for effective item suggestions, while the use of citation section with citation context and citation graph decreases the performance of recommendations. Nonetheless, the combination of all the citation knowledge (i.e. citation graph, citation context, citation section, and citation intention) showed higher performance against the baselines.

It is important to note that, while we have implemented multiple baselines (including PageRank, ItemRank, Matrix Factorisation, Factorisation Machines, and the Random method) to compare against our proposed hybrid methods, the results obtained with these baselines were significantly worst than the ones achieved by content-based methods. Our hypothesis is that the scenario that we are targeting in this work poses significant challenges for these methods, since items in the test set do not have any connections to items in the training set. This problem was also highlighted in the prior work of [272], where a well-established CF method like MF was not able to provide recommendations.

It is also important to highlight that, while existing services, such as Google Scholar, do have their own recommender systems to provide paper recommendations, the methods behind these systems are not public, and hence it has not been possible for us to replicate them here

as baselines. Comparisons against these systems could be conducted by means of user studies, which is one of the future directions of this work.

We should also note that, based on the distribution of the collected dataset, and based on the requirement of recommending the most recent publications to users, we have selected as time split 01/01/2018. However, exploring various time-based data splits, and hence different data distributions, could provide further insights on the desired amount of authored and cited papers to build users profiles, and on the optimal ways of handling the time aspect to provide more relevant suggestions to users.

Our work also provides a novel evaluation benchmark which is publicly available[12], that we hope will encourage and serve the community to further investigate this interesting and challenging problem. A particular interesting form of citation knowledge captured in this benchmark is citation intention, including *background, compare or contrast, use, extension, motivation* and *future*. Citation intention are not only useful to enhance paper recommendation (as shown in our evaluation), but can also enable further refinement of the recommendation process based on the users' intention. For instance, users with an interest on how to use or apply certain algorithms or techniques, may receive recommendations for papers cited under the type 'use' or 'extension.' Refining recommendations based on the users' intent, in addition to the users' publication and citation history, is one of our future lines of work.

---

12 https://ordo.open.ac.uk/articles/Citation_Knowledge_based_Dataset/10673132/1

Despite the timeliness and potential of this evaluation benchmark we also acknowledge several limitations including issues with data sampling, data annotation and data filtering.

Regarding *data sampling*, it is important to note that our benchmark is representative of the area of RS. Complementing it with papers from different fields will help assessing whether the obtained findings are specific to the RS research field or are representative of other areas.

Regarding *data annotation*, although we tested that the classifier provided by [136] achieved comparable results when identifying citation intentions for our collected papers, the average obtained accuracy is 69%, indicating the presence of noise in the classification process. In addition, the PDF parser used as part of this classifier could only extract citation information for 61% of the collected papers, reducing the completeness of the data. Providing more effective citation intention classification methods is an interesting research question for the analysis, search and recommendation of scientific publications.

Regarding *data filtering*, in order to capture users' preferences we discarded from our dataset all authors for which we obtained less than four publications. Our dataset therefore does not capture the scenario of recommending recent papers to a user for which no preferences have been gathered, i.e., user cold start situations.

Despite the above mentioned limitations over the generated benchmarks and conducted research, we believe the work presented in this chapter opens new directions on the investigation of recommender systems for scientific publications, and on targeting the particular real-world scenario of recommending a user with the most recent papers that may be of her interests.

Part III

DISCUSSION, CONCLUSION AND FUTURE
WORK

DISCUSSIONS, CONCLUSIONS AND FUTURE WORK

In this chapter, we describe how the conducted work has helped us to answer our research questions. We highlight its strengths, acknowledge its limitations, and point directions for future work.

As stated in the introductory chapter, the key **goal** of this thesis has been to capture and exploit citation knowledge for the recommendation of scientific publications. To achieve this goal, we investigated the following **research questions**:

- RQ1: Which types of citation knowledge have been used in Recommender Systems (RS) for scientific publications?

- RQ2: Which are the different recommendation tasks that have been proposed in the literature of RS for scientific publications and how citation knowledge has been applied for each of these tasks?

- RQ3: When addressing the task of recommending scientific publications for a particular piece of work, can citation knowledge help improving existing RS?

- RQ4: When addressing the task of recommending recent scientific publications to a user, can citation knowledge help improving existing RS?

In the **first part** of the thesis, we have focused on addressing RQ1 and RQ2. We have conducted a systematic literature review where

more than 170 papers have been analysed and categorised to better understand the landscape of academic RS. We particularly focused on understanding the different notions of citation knowledge that have been used up to date for recommendation purposes, and how this citation knowledge has been incorporated into the recommendation processes, including the modelling of items and targets, as well as the proposed recommendation methods. We also observed how two main recommendation tasks have been so far investigated in the literature: (i) the recommendation of scientific publications for a given piece of work, and (ii) the recommendation of scientific publications to a user. We note that these two tasks are different in nature since, for the first one, target (e.g., paper) preferences are fixed, while for the second one, target (user) preferences are dynamic and evolve over time.

In the **second part** of the thesis, we focus on addressing RQ3 and RQ4. While RQ3 targets the task of recommending scientific publications for a given piece of work, RQ4 focuses on recommending scientific publications to a user. On targeting RQ3, we have focused on exploring existing and novel notions of citation knowledge, in particular, citation proximity and citation context. We have proposed novel recommendation methods that incorporate such knowledge for the the recommendation of scientific publications. By means of user studies, we have shown how the incorporation of fine-grained notions of citation knowledge, particularly in combination, can enhance the recommendation of scientific publications in comparison with baseline methods.

RQ4, on the other hand, focuses on the task of recommending scientific publications to a user. We have focused on the particular real-world scenario of recommending recent scientific publications to users. This is a challenging scenario scarcely addressed in the literature,

where traditional recommendation methods such as collaborative filtering, are ineffective of providing recommendations, due to the item cold-start problem. We have explored this scenario by proposing a range of hybrid recommendation methods that incorporate existing and novel notions of citation knowledge: citation graph, citation section, citation context, and citation intention. The use of citation intention is particularly novel since, to the best of our knowledge, this type of citation knowledge has not been previously used for the recommendation of scientific publications to users. We have assessed our proposed recommendation methods against a variety of baselines. We have shown how the incorporation of the various notions of citation knowledge, particularly in combination, enhances the performance of existing recommendation methods. As part of this evaluation, we have also generated a novel benchmark to conduct offline evaluations, which constitutes a key contribution of this thesis to the scientific community.

In the next sections, we discuss some of the key contributions of our work and its limitations. We organise our discussion around the topics introduced by our research questions (see Section 6.1). We point to future lines of work (see Section 6.2) and conclusions (see Section 6.3).

6.1 DISCUSSION

In this section, we discuss the key contributions and limitations of the conducted work.

6.1.1   *Recommendation of Scientific Publications*

We have analysed more than 170 scientific articles during the course
of this thesis. A key contribution of this analysis has been the identific-
ation of the different recommendation tasks proposed in the literature
of RS for scientific publications, and an exploration of how citation
knowledge has been used for each of the existing recommendation
tasks. In conducting this literature analysis, we have also identified
some of the key challenges and opportunities of the field.

We have broadly identified two main tasks: (i) **the recommendation
of scientific publications to a user** and (ii) **the recommendation of
scientific publications for a given piece of work**. The first task is
more commonly mapped to the area of RS, where the target of the
recommendation is a user, whose preferences may vary over time. The
second recommendation task can be understood as an information
retrieval problem where the user uses a specific work (e.g., a paper) to
query for relevant and related scientific publications. In this case, pref-
erences are static (i.e., do not vary over time). While we acknowledge
that there may be a fuzzy line on whether this is a recommendation or
a retrieval problem, we have treated it as a recommendation problem
in this thesis following previous works in the literature [45, 183, 195,
259]. This last task can, at the same time, be divided into different
sub-tasks depending on the target of the recommendation. Specifically,
it can be defined as: (i) a paper, (ii) a set of papers, (iii) an ongoing
manuscript (i.e., a paper under development), or (iv) a fragment of
text (this can be a title, an abstract, etc.).

We have identified multiple limitations of existing works when
studying the above mentioned recommendation tasks. For a com-
prehensive list of the identified limitations, the reader is referred to

Chapter 3. We, however, highlight here some of the main **limitations** addressed in the thesis.

- **Publication Time Awareness:** The first limitation that we noticed is that, when addressing the task of recommending scientific publications to a user, existing works are focused on providing recommendations independently on when papers are published. To the best of our knowledge, only the works of [95, 272] have previously focused on recommending the most recent scientific publications to a user. This is one of the key limitations that we have aimed to address in the thesis with the work presented in Chapter 4.

- **Citation Knowledge Awareness**: We noticed that many of the studied works use limited notions of citation knowledge when addressing the above mentioned tasks. We hypothesised that this is due to the lack of access to the full-text of scientific articles when extracting such knowledge. Fine-grained notions of citation knowledge (such as citation intention), that have been applied in other fields, such as scientometrics, but have been under explored for the recommendation of scientific publications.

- **Citation Knowledge Application**: We noticed that many of the studied works, while they explore the use of different types of citation knowledge in isolation, they rarely explore the use of different types of citation knowledge in combination with one another.

- **Evaluation Benchmarks**: We noticed multiple problems with the conducted offline evaluations, particularly in terms of the used evaluation datasets and benchmarks. These benchmarks

either do not provide preferences/ratings, or do not provide the full textual content of papers. In those cases, when benchmarks do provide the full content of papers, the datasets are relatively small - in the hundreds of articles. This can be seen in Table 3.7.

Next, we present how we have addressed some of these limitations, and the contributions provided during our investigation.

6.1.2 *Citation Knowledge*

A key contribution of our analysis has been the identification of the different types of citation knowledge used so far in the literature of recommender systems for scientific publications. We have identified those types of citation knowledge that are largely available and used, since they are extracted from the metadata of publications, (e.g., citation graph) as well as other types of citation knowledge that have been less frequently used in the literature due to the inaccessibility to the full text of scientific publications (e.g., citation context, citation section).

In addition, we have identified several notions of citation knowledge used in different fields, like scientometrics (e.g., citation intention), and adapted and incorporated such notions of citation knowledge to the recommendation of scientific publications.

Our contributions in this particular case can be summarised as follows:

- The identification of the different types of citation knowledge that have been used in the literature of RS for scientific publications.

- The identification of different types of citation knowledge that, while used in different fields, were not yet applied for the recommendation of scientific publications.

- The definition, extraction and incorporation of fine-grained characterisations of citation knowledge (extracted from the full textual content of publications) to the recommendation of scientific publications.

One of the key challenges faced during the work conducted in the thesis was the extraction of citation knowledge from the full content of publications. Most publications online are available in Portable Document Format (PDF) format. While PDF parsers do exist, its effectiveness is limited, and multiple errors emerge while using such parsers. The extraction and identification of sections, positions, references, etc., and the disambiguation and mapping of the extracted references to the corresponding articles was a complex engineering task. While we have not detailed all the complexities behind the designed and developed parsers in the thesis (since this has been an engineering task, more than a research task), we have made available for the scientific community a variety of datasets with citation knowledge already extracted from them. The details of these datasets can be found in Chapter 4 and Chapter 5.

6.1.3 *Exploitation of Citation Knowledge for the Recommendation of Scientific Publications*

Chapter 4 and Chapter 5 show how we have employed fine-grained notions of citation knowledge for the recommendation of scientific publications. While Chapter 4 addresses the task of recommending

scientific publications for a given piece of work, Chapter 5 addresses the task of recommending scientific publications to a user. The contributions of these two chapters can be summarised as follows:

- Proposal of novel notions of citation knowledge to be explored in the recommendation of scientific publications.

- Proposal of novel recommendation methods that incorporate different types of citation knowledge (in isolation and in combination) for the recommendation of scientific publications.

- Addressing a novel recommendation scenario where recent scientific publications are suggested to users in a personalised way.

- Conducting rigorous evaluations of the proposed methods against multiple baselines by means of offline evaluations and user studies.

We have highlighted specific limitations of the conducted work at the end of each chapter. We next highlight some of the high-level limitations.

- **Explored citation knowledge per recommendation tasks**: We have explored a much wider range of citation knowledge for the recommendation of scientific publications to a user than for the recommendation of scientific publications for a given piece of work. It is important to notice that not all types of citation knowledge are useful for all recommendation tasks. For example, citation proximity is not applicable for the recommendation of the most recent scientific publications to users. Note that, these publications, because they are recent, would not be co-cited with

other works. Hence, citation proximity is not applicable for this task. Other notions of citation knowledge (e.g., citation intention, citation section, etc.), however, could have been explored for the recommendation of scientific publications for a given piece of work. We have not conducted a full exploration of citation knowledge for this recommendation task due to lack of time, but we acknowledge that this is an interesting future line of work.

- **Combinations of citation knowledge**: We observed in our experiments that some combinations of different types of citation knowledge were more successful than others. In particular, the application of citation section was somehow controversial. While it decreases the performance of the recommendation when it is combined with citation graph and citation context knowledge, it helps enhancing the performance of recommendation methods when it is combined with citation intention. Further experiments and in-depth error analyses are needed to better understand why certain combinations are more successful than others.

- **Modelling user preferences**: In chapter 5, we proposed two different ways of capturing the users' preferences based on their publication history. First, we assumed that a user has a preference for all the papers they have published. Then, we assumed that a user has a preference for all the papers they have published as well as all the papers they have cited. Please note that, this may not be entirely true since a user may cite a paper to criticise it, and not necessarily because she has a preference for it. Similarly, user preferences do change over time and according to different purposes (conducting a survey, writing a paper, etc.), user types (senior researchers, junior researchers), etc. We have

not taken into consideration any of these factors when defining user preferences, and acknowledge that this is an important limitation of our work, but also a great opportunity for future work.

- **Exploration of the Time Aspect**: It is important to highlight that, when addressing the task of recommending recent scientific publications to a user, we used one unique time split. However, exploring various time-based data splits, and hence different data distributions, could provide further insights on the desired amount of authored and cited papers to build users profiles, and on the optimal ways of handling the time aspect to provide more relevant suggestions to users.

- **Scientific Fields**: The datasets used for our research contain papers from the Computer Science research field. It is possible that papers that belong to different research fields (humanities, social sciences, physics, etc.) are somehow structured in different ways, or follow different citation patterns. Our results are therefore bounded to the computer science field, and should not be generalised to different research fields without conducting appropriate experiments.

6.1.4   *Evaluation of Recommender Systems for Scientific Publications*

As we have previously discussed along the thesis, the evaluation of RS for scientific publications is still in its infancy compared with the evaluation of RS in other domains. This is rooted in multiple problems.

Firstly, only industries with RS in production, such as Mendeley, can conduct online evaluations and perform A/B testing of their

recommendation methods. In terms of offline evaluations, existing (and publicly available) datasets either do not provide user-preferences (which means that evaluations need to be conducted by means of user studies - a highly expensive process) or do not provide the full textual content of documents. Moreover, in those cases where the full textual content of documents is provided, the collections are small as mentioned in Table 3.7.

Aiming to alleviate some of the above mentioned issues, this thesis has provided two important evaluation benchmarks.

- The first evaluation benchmark is provided as part of Chapter 4. It consists of citation information content from 2 Million publications. These publications encompass 665,330,651 co-citations among them. Citation knowledge, in terms of citation context and citation proximity, has been extracted from this dataset. This dataset can be downloaded from here.[1]. It is important to notice that, although a large number of citation records is made available as part of this dataset, no ratings are provided.

- The second evaluation benchmark is provided as part of Chapter 5. It consist on 446 authors and 9,399 academic papers (including their content), from which 7,786 belong to the training set and 1,613 represent the test set. Ratings (user preferences) are provided for each of these authors. Citation knowledge in the form of citation graph, citation section, citation intention, and citation context, is also provided. This resource is published in scientific conference of Language Resources (LREC '20) - (see Chapter 1 for the list of publication associated to this thesis), addresses all the above mentioned limitations of the previously

---

1 Citation-context dataset (C2D): https://ordo.open.ac.uk/articles/Citation-Context_Dataset_C2D_/6865298

inspected evaluation benchmarks. It indeed constitutes one of the key contributions of this work to the scientific community and can be accessible from here.[2]

While we acknowledge that these benchmarks could be improved (e.g., by providing more refined user preferences, a higher number of scientific publications, etc.), considering the actual evaluation landscape of RS for scientific publications (see Table 3.7), we believe that the provided benchmarks are a key contribution of this thesis to the scientific community. Note that, these resources would enable to conduct more robust and systematic evaluations, e.g., using the provided benchmarks to compare approaches against each other.

6.2 FUTURE WORK

Following our previous discussion, as well as the discussion provided in Chapter 3 about the limitations found in the area of RS for scientific publications, we highlight here three future research directions that could constitute an important follow up of the work in this thesis:

- **Modelling user preferences**: Capturing user preferences is an important aspect of RS and a wide range of works have focused on it. However, when it comes to the particular field of recommending scientific publications, not many works have attempted to provide appropriate user models capturing the various relevant aspects of the problem. While some works have considered seniority (i.e., whether the user is a junior or a senior researcher [251]), to the best of our knowledge, important aspects, such as the intention of the user (produce a survey paper, get up to date

---

2 Citation Knowledge based dataset: https://ordo.open.ac.uk/articles/Citation_Knowledge_based_Dataset/10673132

with the latest literature of the field, find seminal papers etc.), the variability of preferences across research topics, the potential multidisciplinary of preferences, etc. have not been investigated. This constitutes an important area of future research in the field.

- **Further exploration of the time aspect**: While we have so far explored the scenario of recommending the most recent scientific publications to users, we have not assess the impact of the variability of the time aspect. In this scenario, it is important to understand which papers are considered 'recent', for a particular research field/user/situation. Some research fields, like computer science, can progress at a very fast rate, while others, like physics, may take longer periods of time to produce novel results. In a similar manner, not all users process scientific literature at a similar rate. While some users may be more up to date with the latest developments in their scientific fields, others may study those fields at a slower pace. Studying how different variations of time may affect the recommendation process could shed further light on the effectiveness of different time splits when taking into consideration variability across users, research fields, etc.

- **Cross comparison/validation across scientific fields**: Our research has explored how citations, and knowledge extracted from those citations, can help improving RS for scientific publications. We have particularly focused on the Computer Science research field (as our database in Chapter 5 contains publications from the computer science domain). However, as previously mentioned, different research fields may cite works differently. Understanding the particulates of how citations are applied in

different research fields could provide a better idea of the real impact of citations in the recommendation of scientific publications.

6.3 CONCLUSION

In this thesis, we have studied the use of citation knowledge to enhance the recommendation of scientific publications. Starting from this position we have investigated different types of citation knowledge, namely citation proximity, citation context, citation section, citation graph, and citation intention. We have proposed definitions and adaptations for such notions of citation knowledge and design and develop new recommendation methods that incorporate such knowledge, individually and in combination. By conducting offline evaluations, as well as user studies, we have shown how the use of citation knowledge does indeed help enhancing the performance of existing recommendation methods when addressing both (i) the task of recommending scientific publications for a given work, and (ii) the task of recommending scientific publications to a user. For the latter, we have particularly concentrated on the challenging scenario of recommending the latest scientific publications to a user. An additional contribution of this thesis has been the development of two evaluation benchmarks that we have made publicly available for the scientific community. Strengths and limitations of this work have also been discussed as well as the most prominent future lines of work, which include (i) further investigations on the different aspects to consider when capturing user preferences, (ii) further exploration of the time aspect and how it may influence the recommendation of scientific pub-

lications, and (iii) a comparison across scientific fields on the impact of citations for recommendation.

Part IV

APPENDICES

APPENDIX

In this appendix, we present additional information relating to Chapter 3 and Chapter 5. In A.1, we present a list of prior works using various features for building item profiles. In A.2, a database schema of an entity-relationship diagram is exhibited that has been used while preparing our benchmarks in Chapter 5.

A.1 FEATURES FOR ITEM MODELLING

In this section, we present a list of reviewed papers and their use of features from scientific publications to represent items' (publications') profiles. Table A.1 exhibits the feature information for item modelling.

Table A.1: List of reviewed papers utilising different item features for modelling item profiles. Ti stands for Title, Ab stands for Abstract, Ke stands for Keywords, Au stands for Author, Af stands from Affiliation, Pd stands for Publication date, Ve stands for Venue, Tx stands for Taxonomy, Rl stands for Reference list, Ck stands for Citation knowledge

References	Ti	Ab	Ke	Au	Af	PD	V	Tx	Rl	Ck
[166]	x	x	x	x	x		x			x
[121]	x	x	x	x		x	x			
[10]	x	x	x	x		x				
[168]	x	x	x	x			x			x
[119, 192]	x	x	x	x					x	x
[92]	x	x	x	x					x	
[155]	x	x	x	x						

[256]	x	x	x			x	x		
[286]	x	x	x			x			x	x
[255]	x	x	x				x		
[260]	x	x	x				x		
[34]	x	x	x						x	x
[48]	x	x	x						x	x
[6, 11, 21, 118, 169, 173, 175, 198, 211, 257, 275, 294, 295]	x	x	x						
[289]	x	x		x		x	x			x
[283]	x	x		x		x			x	x
[217]	x	x		x			x		x	x
[284]	x	x		x			x			x
[181]	x	x		x			x		
[102, 167]	x	x		x					x	x
[231]	x	x		x					
[132]	x	x				x			x	
[30]	x	x					x		
[251, 253]	x	x							x	x
[101, 112, 191, 195, 265, 272, 282, 285]	x	x							
[33]	x		x	x		x	x		
[291]	x		x	x					x	x
[5]	x		x	x					x	
[232]	x		x			x			x	x
[234]	x		x							x
[110, 177]	x		x						
[71]	x			x						x
[197]	x					x			

[196]	x							x		
[233]	x								x	x
[236]	x								x	
[44]		x	x						x	x
[120]		x		x						
[18]		x				x				
[182]		x							x	x
[131]		x								x
[2, 75, 91, 138, 203]		x								
[276]			x	x	x	x	x			
[128]			x	x			x		x	x
[58]			x	x					x	x
[180]			x		x	x			x	
[15, 287]			x						x	x
[154, 215, 235]			x						x	
[17, 37, 59–61, 145, 271]			x							
[90]				x		x	x			
[261]				x		x	x			
[9, 242]				x		x				
[165, 297]				x			x		x	x
[274]				x					x	x
[163, 184]				x					x	
[79, 281]				x						
[281]				x						
[64]						x			x	x
[150]						x				x
[277]						x				
[63]						x				

[47]								x		
[46]								x		
[278]								x		
[23, 80, 202, 252, 267]									x	x
[7, 16, 42, 45, 49, 72, 83–85, 93, 95, 97, 98, 113, 116, 117, 127, 130, 140, 143, 144, 151, 152, 159, 160, 164, 183, 199, 244, 250, 254, 259, 279, 288, 298]									x	x
[96]									x	
[248]										x
[50, 69, 74, 88, 94, 129, 247]										x

A.2   DATABASE SCHEMA DIAGRAM

An entity-relationship diagram for the database created in Chapter 5 is presented in Figure A.1, which shows the relationships between tables in the database. The database contains eight tables and each table is briefly introduced below.

- authors : This table contains authors' details.

- publications : This table holds records for scientific publication written by the author present in the *authors* table.
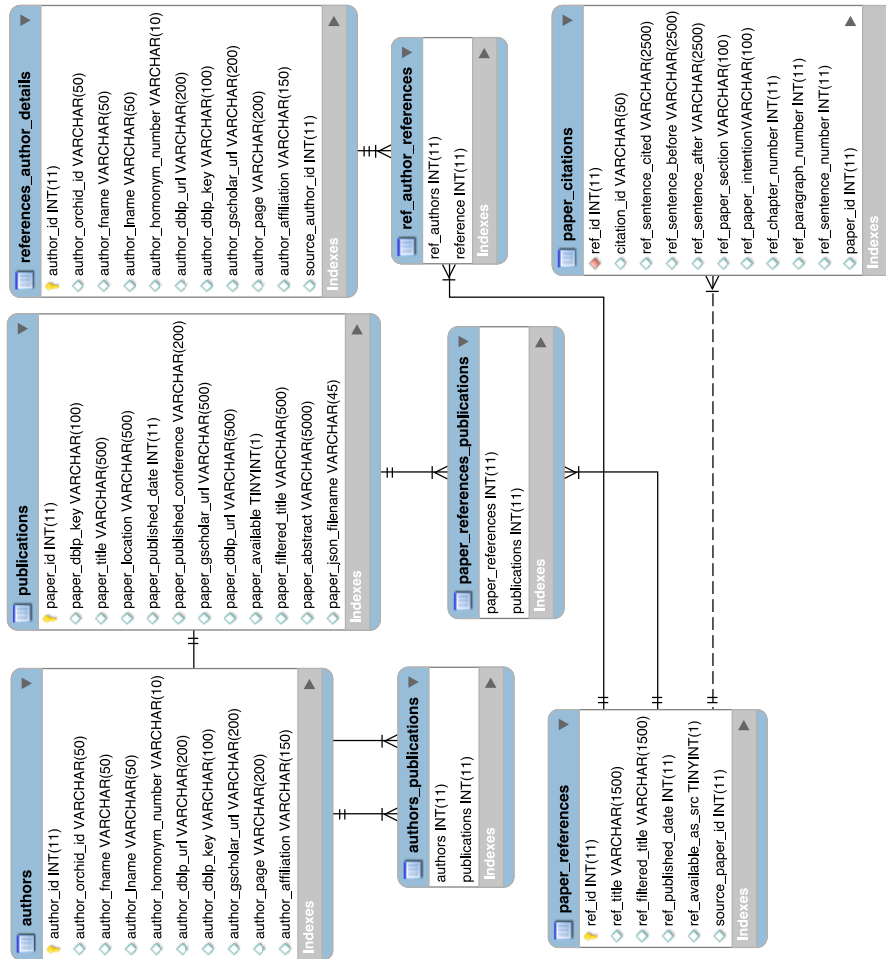
Figure A.1: A schema diagram of the database created in Chapter 5

- paper_references : This table contains records for publications that are referenced in the papers available in the *publications* table.

- paper_citation : This table holds the records of citation details obtained for those papers that are cited in the papers available in the *publications* table. This table contains different notions of citation knowledge.

- author_publications : Relationships between authors and publications.

- paper_references_publications : Relationships between publications and their references.

- ref_author_references : This table records details for authors of the referenced papers from the *paper_references* table.

BIBLIOGRAPHY

[1] Amjad Abu-Jbara, Jefferson Ezra and Dragomir R. Radev. 'Purpose and Polarity of Citation: Towards NLP-based Bibliometrics'. In: *HLT-NAACL*. 2013 (cit. on p. 103).

[2] T. Achakulvisut et al. 'Science Concierge: A fast content-based recommendation system for scientific publications'. In: *PLoS ONE* 11.7 (2016). ISSN: 19326203. DOI: 10.1371/journal.pone.0158423 (cit. on pp. 69, 77, 84, 106, 209).

[3] G. Adomavicius and A. Tuzhilin. 'Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions'. In: *IEEE Transactions on Knowledge and Data Engineering* 17.6 (2005), pp. 734–749. ISSN: 2326-3865. DOI: 10.1109/TKDE.2005.99 (cit. on pp. 26, 27, 35, 43, 45, 46, 59, 61).

[4] Nitin Agarwal et al. 'Research Paper Recommender Systems: A Subspace Clustering Approach'. In: *Advances in Web-Age Information Management*. Ed. by Wenfei Fan, Zhaohui Wu and Jun Yang. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 475–491. ISBN: 978-3-540-32087-6 (cit. on pp. 37, 46, 85, 100).

[5] S. Ahmad and M.T. Afzal. 'Combining Co-citation and Metadata for Recommending More Related Papers'. In: vol. 2017-January. Institute of Electrical and Electronics Engineers Inc., 2018, pp. 218–222. ISBN: 9781538635674. DOI: 10.1109/FIT.2017.00046 (cit. on pp. 69, 84, 105, 208).

[6] M. Al Alshaikh, G. Uchyigit and R. Evans. 'A research paper recommender system using a Dynamic Normalized Tree of Concepts model for user modelling'. In: IEEE Computer Society, 2017, pp. 200–210. ISBN: 9781509054763. DOI: 10.1109/RCIS. 2017.7956538 (cit. on pp. 69, 84, 108, 208).

[7] S. Alotaibi and J. Vassileva. 'Trust-based recommendations for scientific papers based on the researcher's current interest'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7926 LNAI (2013), pp. 717–720. ISSN: 03029743. DOI: 10.1007/ 978-3-642-39112-5-96 (cit. on pp. 69, 76, 80, 84, 87, 108, 210).

[8] S. Alotaibi and J. Vassileva. 'Effect of different implicit social networks on recommending research papers'. In: Association for Computing Machinery, Inc, 2016, pp. 217–221. ISBN: 9781450343701. DOI: 10.1145/2930238.2930293 (cit. on pp. 60, 69, 84, 108).

[9] S. Alotaibi and J. Vassileva. 'Personalized recommendation of research papers by fusing recommendations from explicit and implicit social networks'. In: vol. 1618. CEUR-WS, 2016. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84984650505&partnerID=40&md5=e957fcbe64641328466095c10ef0825a (cit. on pp. 69, 84, 88, 101, 108, 209).

[10] A. Alzoghbi et al. 'PubRec: Recommending publications based on publicly available meta-data'. In: vol. 1458. CEUR-WS, 2015, pp. 11–18. URL: https://www.scopus.com/inward/record. uri?eid=2-s2.0-84944345517&partnerID=40&md5=ff602dba85d8b93e3947f5123f5 (cit. on pp. 69, 84, 109, 207).

[11] A. Alzoghbi et al. 'Learning-to-rank in research paper CBF recommendation: Leveraging irrelevant papers'. In: vol. 1673. CEUR-WS, 2016, pp. 43–46. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84988981081&partnerID=40&md5=55c0e0a035890af6b17b55940695432c (cit. on pp. 69, 84, 108, 146, 208).

[12] M. Amami et al. 'A graph based approach to scientific paper recommendation'. In: Association for Computing Machinery, Inc, 2017, pp. 777–782. ISBN: 9781450349512. DOI: 10.1145/3106426.3106479 (cit. on pp. 69, 84, 109).

[13] B. Amini, R. Ibrahim and M.S. Othman. 'Exploiting scholar's background knowledge to improve recommender system for digital libraries'. In: *International Journal of Digital Content Technology and its Applications* 6.22 (2012), pp. 119–128. DOI: 10.4156/jdcta.vol6.issue22.12 (cit. on pp. 69, 84, 108).

[14] B. Amini et al. 'Incorporating scholar's background knowledge into recommender system for digital libraries'. In: 2011, pp. 516–523. ISBN: 9781457715310. DOI: 10.1109/MySEC.2011.6140721 (cit. on pp. 69, 84, 108).

[15] A. Anand, T. Chakraborty and A. Das. 'FairScholar: Balancing relevance and diversity for scientific paper recommendation'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10193 LNCS (2017), pp. 753–757. ISSN: 03029743. DOI: 10.1007/978-3-319-56608-5_76 (cit. on pp. 69, 76, 80, 82–84, 87, 106, 209).

[16] M.A. Arif. 'Content aware citation recommendation system'. In: Institute of Electrical and Electronics Engineers Inc., 2017.

ISBN: 9781509037513. DOI: 10.1109/ICETT.2016.7873690 (cit. on pp. 69, 76, 77, 80, 83, 84, 87, 106, 210).

[17] N.Y. Asabere et al. 'Scholarly paper recommendation based on social awareness and folksonomy'. In: *International Journal of Parallel, Emergent and Distributed Systems* 30.3 (2015), pp. 211–232. ISSN: 17445760. DOI: 10.1080/17445760.2014.904859 (cit. on pp. 69, 81, 84, 85, 107, 209).

[18] Frederick Ayala-Gomez et al. 'Global citation recommendation using knowledge graphs'. In: *Journal of Intelligent & Fuzzy Systems* 34.5 (2018), pp. 3089–3100 (cit. on pp. 69, 84, 106, 209).

[19] X. Bai et al. 'Scientific Paper Recommendation: A Survey'. In: *IEEE Access* 7 (2019), pp. 9324–9339. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2890388 (cit. on pp. 6–8).

[20] Khadidja Bakhti, Zhendong Niu and Ally S. Nyamawe. 'A New Scheme for Citation Classification based on Convolutional Neural Networks'. In: *SEKE*. 2018 (cit. on p. 103).

[21] T. Bansal, D. Belanger and A. McCallum. 'Ask the GRU: Multi-task learning for deep text recommendations'. In: Association for Computing Machinery, Inc, 2016, pp. 107–114. ISBN: 9781450340359. DOI: 10.1145/2959100.2959180 (cit. on pp. 37, 69, 84, 107, 208).

[22] J. Beel et al. 'Introducing docear's research paper recommender system'. In: Indianapolis, IN, 2013, pp. 459–460. ISBN: 9781450320764. DOI: 10.1145/2467696.2467786 (cit. on pp. 69, 84, 107).

[23] J. Beel et al. 'Exploring the potential of user modeling based on mind maps'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9146 (2015), pp. 3–17. ISSN: 03029743. DOI:

`10.1007/978-3-319-20267-9_1` (cit. on pp. 69, 76, 80, 84, 87, 109, 210).

[24] Joeran Beel et al. 'A Comparative Analysis of Offline and Online Evaluations and Discussion of Research Paper Recommender System Evaluation'. In: *Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation*. RepSys '13. Hong Kong, China: Association for Computing Machinery, 2013, pp. 7–14. ISBN: 9781450324656. DOI: `10.1145/2532508.2532511`. URL: `https://doi-org.libezproxy.open.ac.uk/10.1145/2532508.2532511` (cit. on pp. 51, 89).

[25] Joeran Beel et al. 'The Impact of Demographics (Age and Gender) and Other User-Characteristics on Evaluating Recommender Systems'. In: *Research and Advanced Technology for Digital Libraries*. Ed. by Trond Aalberg et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 396–400. ISBN: 978-3-642-40501-3 (cit. on p. 60).

[26] Joeran Beel et al. 'Research Paper Recommender Systems: A Literature Survey'. In: *International Journal on Digital Libraries* 17.4 (2016), pp. 305–338 (cit. on pp. 8, 65, 86, 89–91, 114, 115).

[27] R. M. Bell and Y. Koren. 'Scalable Collaborative Filtering with Jointly Derived Neighborhood Interpolation Weights'. In: *Seventh IEEE International Conference on Data Mining (ICDM 2007)*. 2007, pp. 43–52 (cit. on p. 45).

[28] Alejandro Bellogín, Pablo Castells and Iván Cantador. 'Statistical biases in Information Retrieval metrics for recommender systems'. In: *Information Retrieval Journal* 20 (2017), pp. 606–634 (cit. on pp. 52, 57).

[29]   M. Benard Magara, S.O. Ojo and T. Zuva. 'A comparative analysis of text similarity measures and algorithms in research paper recommender systems'. In: Institute of Electrical and Electronics Engineers Inc., 2018, pp. 1–5. ISBN: 9781538610015. DOI: 10.1109/ICTAS.2018.8368766 (cit. on pp. 69, 84, 106).

[30]   Steven Bethard and Dan Jurafsky. 'Who Should I Cite: Learning Literature Search Models from Citation Behavior'. In: *19th ACM International Conference on Information and Knowledge Management*. 2010, pp. 609–618 (cit. on pp. 7, 12, 69, 77, 84, 106, 208).

[31]   Daniel Billsus and Michael J. Pazzani. 'Learning Collaborative Information Filters'. In: *Proceedings of the Fifteenth International Conference on Machine Learning*. ICML '98. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, pp. 46–54. ISBN: 1558605568 (cit. on p. 35).

[32]   David M. Blei, Andrew Y. Ng and Michael I. Jordan. 'Latent Dirichlet Allocation'. In: *J. Mach. Learn. Res.* 3 (Mar. 2003), pp. 993–1022. ISSN: 1532-4435. URL: http://dl.acm.org/citation.cfm?id=944919.944937 (cit. on pp. 36, 43, 129, 130, 137, 175).

[33]   Toine Bogers and Antal van den Bosch. 'Recommending Scientific Articles Using Citeulike'. In: *Proceedings of the 2008 ACM Conference on Recommender Systems*. RecSys '08. Lausanne, Switzerland: ACM, 2008, pp. 287–290. ISBN: 978-1-60558-093-7. DOI: 10.1145/1454008.1454053. URL: http://doi.acm.org/10.1145/1454008.1454053 (cit. on pp. 69, 84, 108, 208).

[34]   Kurt D. Bollacker, Steve Lawrence and C. Lee Giles. 'CiteSeer: An Autonomous Web Agent for Automatic Retrieval and Identification of Interesting Publications'. In: *Proceedings of the*

Second International Conference on Autonomous Agents. AGENTS '98. Minneapolis, Minnesota, USA: ACM, 1998, pp. 116–123. ISBN: 0-89791-983-1. DOI: 10.1145/280765.280786. URL: http://doi.acm.org/10.1145/280765.280786 (cit. on pp. 6, 69, 76, 80, 83, 84, 87, 106, 208).

[35] John S. Breese, David Heckerman and Carl Kadie. 'Empirical Analysis of Predictive Algorithms for Collaborative Filtering'. In: *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, pp. 43–52 (cit. on pp. 35, 41).

[36] Sergey Brin and Lawrence Page. 'The Anatomy of a Large-scale Hypertextual Web Search Engine'. In: *Proceedings of the Seventh International Conference on World Wide Web 7*. WWW7. Brisbane, Australia: Elsevier Science Publishers B. V., 1998, pp. 107–117 (cit. on p. 168).

[37] Simon Bruns et al. 'What Should I Read Next? A Personalized Visual Publication Recommender System'. In: *Human Interface and the Management of Information. Information and Knowledge in Context*. Ed. by Sakae Yamamoto. Cham: Springer International Publishing, 2015, pp. 89–100. ISBN: 978-3-319-20618-9 (cit. on pp. 69, 77, 82, 84, 106, 209).

[38] Robin Burke. 'Hybrid Recommender Systems: Survey and Experiments'. In: *User Modeling and User-Adapted Interaction* 12.4 (2002), pp. 331–370 (cit. on pp. 25, 45, 46, 48, 159).

[39] Robin Burke. 'Hybrid Web Recommender Systems'. In: *The Adaptive Web: Methods and Strategies of Web Personalization*. 2007, pp. 377–408 (cit. on pp. 46, 87).

[40] Pedro G. Campos, Fernando Díez and Iván Cantador. 'Time-Aware Recommender Systems: A Comprehensive Survey and Analysis of Existing Evaluation Protocols'. In: *User Modeling and User-Adapted Interaction* 24.1–2 (Feb. 2014), pp. 67–119. ISSN: 0924-1868. DOI: 10.1007/s11257-012-9136-x. URL: https://doi-org.libezproxy.open.ac.uk/10.1007/s11257-012-9136-x (cit. on p. 51).

[41] Iván Cantador. 'Exploiting the conceptual space in hybrid recommender systems: a semantic-based approach'. In: 2008 (cit. on pp. 26, 49).

[42] C. Caragea et al. 'Can't see the forest for the trees? A citation recommendation system'. In: Indianapolis, IN, 2013, pp. 111–114. ISBN: 9781450320764. DOI: 10.1145/2467696.2467743 (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 210).

[43] Òscar Celma and Perfecto Herrera. 'A New Approach to Evaluating Novel Recommendations'. In: *Proceedings of the 2008 ACM Conference on Recommender Systems*. RecSys '08. New York, NY, USA: Association for Computing Machinery, 2008, pp. 179–186. ISBN: 9781605580937. DOI: 10.1145/1454008.1454038. URL: https://doi.org/10.1145/1454008.1454038 (cit. on p. 58).

[44] T. Chakraborty et al. 'DiSCern: A diversified citation recommendation system for scientific queries'. In: vol. 2015-May. IEEE Computer Society, 2015, pp. 555–566. ISBN: 9781479979639. DOI: 10.1109/ICDE.2015.7113314 (cit. on pp. 69, 76, 80, 83, 84, 87, 106, 209).

[45] Tanmoy Chakraborty et al. 'FeRoSA: A Faceted Recommendation System for Scientific Articles'. In: *Proceedings, Part II, of the 20th Pacific-Asia Conference on Advances in Knowledge Discov-*

ery and Data Mining - Volume 9652. PAKDD 2016. Auckland, New Zealand: Springer-Verlag, 2016, pp. 528–541. ISBN: 978-3-319-31749-6. DOI: 10.1007/978-3-319-31750-2_42. URL: https://doi-org/10.1007/978-3-319-31750-2_42 (cit. on pp. 9, 69, 74–76, 80, 83, 84, 87, 102, 105, 114, 148, 157, 192, 210).

[46] K. Chandrasekaran et al. 'Concept-based document recommendations for CiteSeer authors'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5149 LNCS (2008), pp. 83–92. ISSN: 03029743. DOI: 10.1007/978-3-540-70987-9_11 (cit. on pp. 69, 84, 109, 210).

[47] Kannan Chandrasekaran et al. 'Concept-Based Document Recommendations for CiteSeer Authors'. In: *Adaptive Hypermedia and Adaptive Web-Based Systems*. Ed. by Wolfgang Nejdl et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 83–92. ISBN: 978-3-540-70987-9 (cit. on pp. 12, 69, 84, 109, 210).

[48] J. Chen and Z. Ban. 'Literature recommendation by researchers' publication analysis'. In: *2016 IEEE International Conference on Information and Automation (ICIA)*. 2016, pp. 1964–1969. DOI: 10.1109/ICInfA.2016.7832140 (cit. on pp. 69, 76, 80, 84, 87, 109, 208).

[49] T.T. Chen and M. Lee. 'Research paper recommender systems on big scholarly data'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11016 LNAI (2018), pp. 251–260. ISSN: 03029743. DOI: 10.1007/978-3-319-97289-3_20 (cit. on pp. 31, 69, 72, 76, 80, 84, 87, 109, 210).

[50] Xi Chen et al. 'Citation recommendation based on citation tendency'. In: *Scientometrics* 121 (2019), pp. 937 –956 (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 210).

[51] W. Choochaiwattana. 'Usage of tagging for research paper recommendation'. In: vol. 2. Chengdu, 2010, pp. V2439–V2442. ISBN: 9781424465408. DOI: 10.1109/ICACTE.2010.5579321 (cit. on pp. 69, 81, 84, 107).

[52] Mark Claypool et al. 'Combining Content-Based and Collaborative Filters in an Online Newspaper'. In: *In Proceedings of ACM SIGIR Workshop on Recommender Systems*. 1999 (cit. on p. 47).

[53] Arman Cohan et al. 'Structural Scaffolds for Citation Intent Classification in Scientific Publications'. In: *CoRR* abs/1904.01608 (2019). URL: http://arxiv.org/abs/1904.01608 (cit. on p. 102).

[54] Joseph Corneli et al. *Modelling serendipity in a computational context*. 2014. arXiv: 1411.0440 [cs.AI] (cit. on p. 57).

[55] J. Cortina. 'What Is Coefficient Alpha? An Examination of Theory and Applications'. In: 1993 (cit. on p. 139).

[56] Paolo Cremonesi, Yehuda Koren and Roberto Turrin. 'Performance of Recommender Algorithms on Top-n Recommendation Tasks'. In: *Proceedings of the Fourth ACM Conference on Recommender Systems*. RecSys '10. New York, NY, USA: Association for Computing Machinery, 2010, pp. 39–46. ISBN: 9781605589060. DOI: 10.1145/1864708.1864721. URL: https://doi-org.libezproxy.open.ac.uk/10.1145/1864708.1864721 (cit. on p. 32).

[57] Lee J. Cronbach. 'Coefficient alpha and the internal structure of tests'. In: *Psychometrika* 16.3 (1951), pp. 297–334. ISSN: 1860-0980.

DOI: `10.1007/BF02310555`. URL: `https://doi.org/10.1007/BF02310555` (cit. on p. 139).

[58] T. Cui, X. Tang and Q. Zeng. 'User network construction within online paper recommendation systems'. In: Beijing, 2010, pp. 361–366. ISBN: 9781424463572. DOI: `10.1109/SWS.2010.5607425` (cit. on pp. 69, 76, 80, 84, 85, 87, 107, 209).

[59] D. De Nart, F. Ferrara and C. Tasso. 'Personalized access to scientific publications: From recommendation to explanation'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7899 LNCS (2013), pp. 296–301. ISSN: 03029743. DOI: `10.1007/978-3-642-38844-6_26` (cit. on pp. 69, 84, 85, 108, 209).

[60] D. De Nart, F. Ferrara and C. Tasso. 'Personalized recommendation and explanation by using keyphrases automatically extracted from scientific literature'. In: Vilamoura, Algarve: SciTePress, 2013, pp. 96–103. ISBN: 9789898565754. URL: `https://www.scopus.com/inward/record.uri?eid=2-s2.0-84887723202&partnerID=40&md5=b23af27816c9825432794c76f2ead83e` (cit. on pp. 69, 84, 85, 108, 209).

[61] D. De Nart, F. Ferrara and C. Tasso. 'RES: A personalized filtering tool for CiteSeerX queries based on keyphrase extraction'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7899 LNCS (2013), pp. 341–343. ISSN: 03029743. DOI: `10.1007/978-3-642-38844-6_33` (cit. on pp. 69, 84, 85, 108, 209).

[62] Scott Deerwester et al. 'Indexing by latent semantic analysis'. In: *Journal of the American Society for Information Science* 41.6 (1990),

pp. 391–407. DOI: `10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9`. eprint: `https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291097-4571%28199009%2941%3A6%3C391%3A%3AAID-ASI1%3E3.0.CO%3B2-9`. URL: `https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-4571%28199009%2941%3A6%3C391%3A%3AAID-ASI1%3E3.0.CO%3B2-9` (cit. on p. 44).

[63]   M. Dhanda and V. Verma. 'Personalized recommendation approach for academic literature using high-utility itemset mining technique'. In: *Advances in Intelligent Systems and Computing* 519 (2018), pp. 247–254. ISSN: 21945357. DOI: `10.1007/978-981-10-3376-6_27` (cit. on pp. 69, 84, 109, 209).

[64]   Mahak Dhanda and Vijay Verma. 'Recommender System for Academic Literature with Incremental Dataset'. In: *Procedia Computer Science* 89 (2016). Twelfth International Conference on Communication Networks, ICCN 2016, August 19– 21, 2016, Bangalore, India Twelfth International Conference on Data Mining and Warehousing, ICDMW 2016, August 19-21, 2016, Bangalore, India Twelfth International Conference on Image and Signal Processing, ICISP 2016, August 19-21, 2016, Bangalore, India, pp. 483 –491. ISSN: 1877-0509. DOI: `https://doi.org/10.1016/j.procs.2016.06.109`. URL: `http://www.sciencedirect.com/science/article/pii/S1877050916311747` (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 209).

[65]   T. G. Dietterich. 'Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms'. In: *Neural Computation* 10.7 (1998), pp. 1895–1923 (cit. on p. 176).

[66] S. Doerfel et al. 'Leveraging publication metadata and social data into folkrank for scientific publication recommendation'. In: Dublin, 2012, pp. 9–16. ISBN: 9781450316385. DOI: 10.1145/2365934.2365937 (cit. on pp. 6, 60, 69, 84, 108).

[67] Cailing Dong and Ulrich Schäfer. 'Ensemble-style Self-training on Citation Classification'. In: *Proceedings of 5th International Joint Conference on Natural Language Processing*. Chiang Mai, Thailand: Asian Federation of Natural Language Processing, Nov. 2011, pp. 623–631. URL: https://www.aclweb.org/anthology/I11-1070 (cit. on p. 102).

[68] R Dong, L Tokarchuk and A Ma. 'Digging friendship: paper recommendation in social network'. In: *Proceedings of Networking & Electronic Commerce Research Conference (NAEC 2009)*. 2009, pp. 21–28 (cit. on pp. 69, 84, 108).

[69] D. Duma et al. 'Applying core scientific concepts to context-based citation recommendation'. In: European Language Resources Association (ELRA), 2016, pp. 1737–1742. ISBN: 9782951740891. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84992450475&partnerID=40&md5=ee4a6e4b927e39715f63d149168dda3a (cit. on pp. 69, 72, 75, 76, 80, 83, 84, 87, 106, 210).

[70] Susan T. Dumais and Jakob Nielsen. 'Automating the Assignment of Submitted Manuscripts to Reviewers'. In: *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '92. Copenhagen, Denmark: Association for Computing Machinery, 1992, pp. 233–244. ISBN: 0897915232. DOI: 10.1145/133160.133205. URL: https://doi.org/10.1145/133160.133205 (cit. on p. 7).

[71]   Travis Ebesu and Yi Fang. 'Neural Citation Network for Context-Aware Citation Recommendation'. In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '17. Shinjuku, Tokyo, Japan: ACM, 2017, pp. 1093–1096. ISBN: 978-1-4503-5022-8. DOI: 10.1145/3077136.3080730. URL: http://doi.acm.org/10.1145/3077136.3080730 (cit. on pp. 9, 69, 76, 80, 83, 84, 87, 106, 208).

[72]   Michael D. Ekstrand et al. 'Automatically Building Research Reading Lists'. In: *4th ACM Conference on Recommender Systems*. 2010, pp. 159–166 (cit. on pp. 7, 12, 47, 69, 76, 77, 80, 82–84, 86, 87, 91, 105, 210).

[73]   Michael D. Ekstrand et al. 'User Perception of Differences in Recommender Algorithms'. In: *Proceedings of the 8th ACM Conference on Recommender Systems*. RecSys '14. California, USA: Association for Computing Machinery, 2014, pp. 161–168. ISBN: 9781450326681. DOI: 10.1145/2645710.2645737. URL: https://doi.org/10.1145/2645710.2645737 (cit. on p. 25).

[74]   Michael Färber and Tarek Saier. 'Semantic Modelling of Citation Contexts for Context-aware Citation Recommendation'. In: *Proceedings of the 42th European Conference on Information Retrieval (ECIR'20)*. Springer, 2020 (cit. on pp. 69, 76, 77, 80, 83, 84, 87, 106, 210).

[75]   F. Ferrara, N. Pudota and C. Tasso. 'A keyphrase-based paper recommender system'. In: *Communications in Computer and Information Science* 249 CCIS (2011), pp. 14–25. ISSN: 18650929. DOI: 10.1007/978-3-642-27302-5_2 (cit. on pp. 69, 77, 81, 84, 107, 209).

[76]   Joseph L. Fleiss and Jacob Cohen. 'The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability'. In: *Educational and Psychological Measurement* 33.3 (1973), pp. 613–619. DOI: 10.1177/001316447303300309. URL: https://doi.org/10.1177/0013164473033003 (cit. on p. 158).

[77]   John E. Freund. *Mathematical Statistics*. Englewood Cliffs, New Jersey, USA: Prentice-Hall International, Inc, 1992. ISBN: 0-13-565185-9 (cit. on pp. 135, 139).

[78]   Simon Funk. *Netflix Update: Try this at home*. 2006. URL: https://sifter.org/~simon/journal/20061211.html (cit. on pp. 44, 45).

[79]   M. Färber, A. Thiemann and A. Jatowt. 'CITEWERTs: A system combining cite-worthiness with citation recommendation'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10772 LNCS (2018), pp. 815–819. ISSN: 03029743. DOI: 10.1007/978-3-319-76941-7_82 (cit. on pp. 69, 76, 82–84, 106, 209).

[80]   Z. Gao. 'Examining influences of publication dates on citation recommendation systems'. In: Institute of Electrical and Electronics Engineers Inc., 2016, pp. 1400–1405. ISBN: 9781467376822. DOI: 10.1109/FSKD.2015.7382149 (cit. on pp. 69, 76, 77, 80, 83, 84, 87, 106, 210).

[81]   Mark Garzone and Robert E. Mercer. 'Towards an Automated Citation Classifier'. In: *Proceedings of the 13th Biennial Conference of the Canadian Society on Computational Studies of Intelligence: Advances in Artificial Intelligence*. AI '00. Berlin, Heidelberg: Springer-Verlag, 2000, pp. 337–346. ISBN: 3-540-67557-4. URL:

http://dl.acm.org/citation.cfm?id=647461.726271 (cit. on p. 103).

[82]   Mouzhi Ge, Carla Delgado-Battenfeld and Dietmar Jannach. 'Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity'. In: *Proceedings of the Fourth ACM Conference on Recommender Systems*. RecSys '10. New York, NY, USA: Association for Computing Machinery, 2010, pp. 257–260. ISBN: 9781605589060. DOI: 10.1145/1864708.1864761. URL: https://doi.org/10.1145/1864708.1864761 (cit. on pp. 25, 57).

[83]   Bela Gipp and Joran Beel. 'Identifying related documents for research paper recommender by CPA and COA'. In: *Proceedings of the World Congress on Engineering and Computer Science*. Vol. 1. 2009, pp. 20–22 (cit. on pp. 69, 76, 84, 105, 210).

[84]   Bela Gipp, Joran Beel and Christian Hentschel. 'Scienstein: A research paper recommender system'. In: *Proceedings of the international conference on emerging trends in computing (icetic'09)*. 2009, pp. 309–315 (cit. on pp. 6, 69, 76, 84, 105, 115, 128, 210).

[85]   Bela Gipp and Jöran Beel. 'Citation Proximity Analysis (CPA) : A New Approach for Identifying Related Work Based on Co-Citation Analysis'. In: *Proceedings of the 12th International Conference on Scientometrics and Informetrics, vol. 1*. Ed. by Birger Larsen. São Paulo: BIREME/PANO/WHO, 2009, pp. 571–575. URL: http://www.sciplore.org/wp-content/papercite-data/pdf/gipp09a.pdf (cit. on pp. 69, 74, 76, 84, 105, 114, 115, 118, 119, 138, 210).

[86] Stephanie Glen. *Cronbach's Alpha: Simple Definition, Use and Interpretation*. https://www.statisticshowto.com/cronbachs-alpha-spss/. Accessed: 2020 (cit. on pp. 139, 140).

[87] David Goldberg et al. 'Using Collaborative Filtering to Weave an Information Tapestry'. In: *Commun. ACM* 35.12 (Dec. 1992), pp. 61–70. ISSN: 0001-0782. DOI: 10.1145/138859.138867. URL: https://doi-org.libezproxy.open.ac.uk/10.1145/138859.138867 (cit. on p. 27).

[88] M. Gori and A. Pucci. 'Research Paper Recommender Systems: A Random-Walk Based Approach'. In: *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)*. 2006, pp. 778–781. DOI: 10.1109/WI.2006.149 (cit. on pp. 69, 76, 77, 80, 83, 84, 87, 105, 210).

[89] Marco Gori and Augusto Pucci. 'ItemRank: A Random-walk Based Scoring Algorithm for Recommender Engines'. In: *20th International Joint Conference on Artifical Intelligence*. IJCAI'07. Hyderabad, India: Morgan Kaufmann Publishers Inc., 2007, pp. 2766–2771 (cit. on p. 168).

[90] P. Guan and Y. Wang. 'Personalized scientific literature recommendation based on user's research interest'. In: *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. 2016, pp. 1273–1277. DOI: 10.1109/FSKD.2016.7603361 (cit. on pp. 69, 84, 109, 209).

[91] Z. Guan et al. 'Document recommendation in social tagging services'. In: 2010, pp. 391–400. ISBN: 9781605587998. DOI: 10.1145/1772690.1772731 (cit. on pp. 69, 84, 107, 209).

[92] L. Guo et al. 'Exploiting Fine-Grained Co-Authorship for Personalized Citation Recommendation'. In: *IEEE Access* 5 (2017),

pp. 12714–12725. ISSN: 21693536. DOI: 10.1109/ACCESS.2017.2721934 (cit. on pp. 69, 82, 84, 105, 207).

[93] Shanshan Guo, Wenyu Zhang and Shuai Zhang. 'A Pagerank-based Collaborative Filtering Recommendation Approach in Digital Libraries'. In: 2017 (cit. on pp. 69, 76, 80, 81, 84, 85, 87, 108, 210).

[94] Raja HABIB and Muhammad Tanvir Afzal. 'Paper recommendation using citation proximity in bibliographic coupling'. In: 2017 (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 210).

[95] Jiwoon Ha et al. 'Recommendation of Newly Published Research Papers Using Belief Propagation'. In: *2014 Conference on Research in Adaptive and Convergent Systems*. 2014, pp. 77–81 (cit. on pp. 8, 69, 72, 73, 76, 80, 81, 84, 87, 91, 98, 109, 145, 146, 156, 193, 210).

[96] K. Haruna and M.A. Ismail. 'Research paper recommender system evaluation using collaborative filtering'. In: vol. 1974. American Institute of Physics Inc., 2018. ISBN: 9780735416819. DOI: 10.1063/1.5041583 (cit. on pp. 69, 84, 106, 210).

[97] K. Haruna et al. 'A collaborative approach for research paper recommender system'. In: *PLoS ONE* 12.10 (2017). ISSN: 19326203. DOI: 10.1371/journal.pone.0184516 (cit. on pp. 31, 69, 76, 80, 83–85, 87, 105, 210).

[98] K. Haruna et al. 'A citation-based recommender system for scholarly paper recommendation'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10960 LNCS (2018), pp. 514–525. ISSN: 03029743. DOI: 10.1007/978-3-319-95162-1_35 (cit. on pp. 69, 71, 76, 80, 83, 84, 87, 105, 210).

[99] Saeed-Ul Hassan, Anam Akram and Peter Haddawy. 'Identifying Important Citations Using Contextual Information from Full Text'. In: *Proceedings of the 17th ACM/IEEE Joint Conference on Digital Libraries*. JCDL '17. Toronto, Ontario, Canada: IEEE Press, 2017, pp. 41–48. ISBN: 978-1-5386-3861-3. URL: http://dl.acm.org/citation.cfm?id=3200334.3200340 (cit. on pp. 104, 158).

[100] Saeed-Ul Hassan et al. 'Deep context of citations using machine-learning models in scholarly full-text articles'. In: *Scientometrics* 117 (2018), pp. 1645–1662 (cit. on p. 104).

[101] Q. He et al. 'Citation recommendation without author supervision'. In: Hong Kong, 2011, pp. 755–764. ISBN: 9781450304931. DOI: 10.1145/1935826.1935926 (cit. on pp. 69, 77, 84, 105, 208).

[102] Qi He et al. 'Context-aware Citation Recommendation'. In: *Proceedings of the 19th International Conference on World Wide Web*. WWW '10. Raleigh, North Carolina, USA: ACM, 2010, pp. 421–430. ISBN: 978-1-60558-799-8. DOI: 10.1145/1772690.1772734. URL: http://doi.acm.org/10.1145/1772690.1772734 (cit. on pp. 7–9, 12, 69, 74, 76, 77, 80, 82–84, 87, 88, 90, 105, 120, 208).

[103] Jonathan L. Herlocker et al. 'An Algorithmic Framework for Performing Collaborative Filtering'. In: *SIGIR Forum* 51 (1999), pp. 227–234 (cit. on pp. 35, 40, 58).

[104] Jonathan L. Herlocker et al. 'Evaluating Collaborative Filtering Recommender Systems'. In: *ACM Trans. Inf. Syst.* 22.1 (Jan. 2004), pp. 5–53. ISSN: 1046-8188. DOI: 10.1145/963770.963772. URL: http://doi.acm.org/10.1145/963770.963772 (cit. on pp. 57, 62).

[105] Will Hill et al. 'Recommending and Evaluating Choices in a Virtual Community of Use'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '95. USA: ACM Press/Addison-Wesley Publishing Co., 1995, pp. 194–201. ISBN: 0201847051. DOI: 10.1145/223904.223929. URL: https://doi.org/10.1145/223904.223929 (cit. on p. 28).

[106] Jorge E Hirsch. 'An index to quantify an individual's scientific research output'. In: *National academy of Sciences* 102.46 (2005), pp. 16569–16572 (cit. on pp. 6, 7).

[107] Thomas Hofmann. 'Latent Semantic Models for Collaborative Filtering'. In: *ACM Trans. Inf. Syst.* 22.1 (Jan. 2004), pp. 89–115. ISSN: 1046-8188. DOI: 10.1145/963770.963774. URL: https://doi-org.libezproxy.open.ac.uk/10.1145/963770.963774 (cit. on p. 43).

[108] Thomas Hofmann and Jan Puzicha. 'Latent Class Models for Collaborative Filtering'. In: *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*. IJCAI '99. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999, pp. 688–693. ISBN: 1558606130 (cit. on p. 43).

[109] *Holdings of the ACM DL*. https://dl.acm.org/contents_guide.cfm?coll=portal&dl=GUIDE. Accessed: 28-01-2019. 2018 (cit. on p. 6).

[110] K. Hong, H. Jeon and C. Jeon. 'UserProfile-based personalized research paper recommendation system'. In: Gyeongju, 2012, pp. 134–138. ISBN: 9788994364179. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84874181840&partnerID=40&md5=b7dc44b599ff018da7d573fb5809cc13 (cit. on pp. 69, 81, 84, 108, 208).

[111] Liangjie Hong and Brian D. Davison. 'Empirical Study of Topic Modeling in Twitter'. In: *Proceedings of the First Workshop on Social Media Analytics*. SOMA '10. Washington D.C., District of Columbia: ACM, 2010, pp. 80–88. ISBN: 978-1-4503-0217-3. DOI: 10.1145/1964858.1964870. URL: http://doi.acm.org/10.1145/1964858.1964870 (cit. on p. 129).

[112] Maya Hristakeva et al. 'Building Recommender Systems for Scholarly Information'. In: *1st Workshop on Scholarly Web Mining*. 2017, pp. 25–32 (cit. on pp. 8, 61, 69, 84, 108, 208).

[113] Shen Huang et al. 'TSSP: A Reinforcement Algorithm to Find Related Papers'. In: *IEEE/WIC/ACM International Conference on Web Intelligence (WI'04)*. 2004, pp. 117–123. DOI: 10.1109/WI.2004.10038 (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 115, 210).

[114] W. Huang et al. 'RefSeer: A citation recommendation system'. In: Institute of Electrical and Electronics Engineers Inc., 2014, pp. 371–374. ISBN: 9781479955695. DOI: 10.1109/JCDL.2014.6970192 (cit. on pp. 69, 80, 84, 87, 106).

[115] Wenyi Huang et al. 'Recommending citations: translating papers into references'. In: *CIKM*. 2012 (cit. on p. 9).

[116] Wenyi Huang et al. 'A Neural Probabilistic Model for Context Based Citation Recommendation'. In: *29th AAAI Conference on Artificial Intelligence*. 2015, pp. 2404–2410 (cit. on pp. 7, 12, 69, 76, 80, 83, 84, 87, 106, 210).

[117] T. Huynh et al. 'Scientific publication recommendations based on collaborative citation networks'. In: Denver, CO, 2012, pp. 316–321. ISBN: 9781467313803. DOI: 10.1109/CTS.2012.6261069 (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 210).

[118] S.-Y. Hwang and S.-M. Chuang. 'Combining article content and Web usage for literature recommendation in digital libraries'. In: *Online Information Review* 28.4 (2004), pp. 260–272. ISSN: 14684527. DOI: 10.1108/14684520410553750 (cit. on pp. 33, 69, 84, 107, 208).

[119] S.-Y. Hwang, C.-P. Wei and Y.-F. Liao. 'Coauthorship networks and academic literature recommendation'. In: *Electronic Commerce Research and Applications* 9.4 (2010), pp. 323–334. DOI: 10.1016/j.elerap.2010.01.001 (cit. on pp. 69, 80, 84, 87, 107, 207).

[120] San-Yih Hwang et al. 'Combining Coauthorship Network and Content for Literature Recommendation'. In: *PACIS*. 2010, p. 40 (cit. on pp. 69, 82, 84, 209).

[121] San-Yih Hwang et al. 'Coauthorship network-based literature recommendation with topic model'. In: *Online Information Review* 41.3 (2017), pp. 318–336 (cit. on pp. 69, 84, 109, 207).

[122] Kris Jack et al. 'Mendeleys reply to the datatel challenge'. In: *Procedia Computer Science* 1.2 (2010), pp. 1–3 (cit. on p. 94).

[123] Dietmar Jannach and Gediminas Adomavicius. 'Recommendations with a Purpose'. In: *Proceedings of the 10th ACM Conference on Recommender Systems*. RecSys '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 7–10. ISBN: 9781450340359. DOI: 10.1145/2959100.2959186. URL: https://doi.org/10.1145/2959100.2959186 (cit. on pp. 7, 25, 59).

[124] Dietmar Jannach, Lukas Lerche and Markus Zanker. 'Recommending Based on Implicit Feedback'. In: *Social Information Access: Systems and Technologies*. Ed. by Peter Brusilovsky and Daqing He. Cham: Springer International Publishing, 2018,

pp. 510–569. ISBN: 978-3-319-90092-6. DOI: 10.1007/978-3-319-90092-6_14. URL: https://doi.org/10.1007/978-3-319-90092-6_14 (cit. on p. 33).

[125] Dietmar Jannach et al. *Recommender Systems: An Introduction*. 1st. USA: Cambridge University Press, 2010. ISBN: 0521493366 (cit. on pp. 25, 32, 35, 39, 40, 42–44, 60).

[126] J.G. Jardine. *Automatically Generating Reading Lists*. University of Cambridge, 2014. URL: https://books.google.co.uk/books?id=6TXsvgEACAAJ (cit. on p. 7).

[127] Chanwoo Jeong et al. *A Context-Aware Citation Recommendation Model with BERT and Graph Convolutional Networks*. 2019. arXiv: 1903.06464 [cs.CL] (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 120, 210).

[128] H. Jia and E. Saule. 'An analysis of citation recommender systems: Beyond the obvious'. In: Association for Computing Machinery, Inc, 2017, pp. 216–223. ISBN: 9781450349932. DOI: 10.1145/3110025.3110150 (cit. on pp. 69, 76, 80, 82, 84, 87, 105, 209).

[129] Haofeng Jia and Erik Saule. 'Graph Embedding for Citation Recommendation'. In: *ArXiv* abs/1812.03835 (2018) (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 210).

[130] Haofeng Jia and Erik Saule. 'Local Is Good: A Fast Citation Recommendation Approach'. In: *Advances in Information Retrieval*. Ed. by Gabriella Pasi et al. Cham: Springer International Publishing, 2018, pp. 758–764. ISBN: 978-3-319-76941-7 (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 210).

[131] Y. Jiang et al. 'Recommending academic papers via users' reading purposes'. In: Dublin, 2012, pp. 241–244. ISBN: 9781450312707.

DOI: `10.1145/2365952.2366004` (cit. on pp. 69, 76, 77, 80, 83, 84, 87, 105, 209).

[132]   Zhuoren Jiang, Xiaozhong Liu and Liangcai Gao. 'Chronological Citation Recommendation with Information-Need Shifting'. In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. CIKM '15. Melbourne, Australia: ACM, 2015, pp. 1291–1300. ISBN: 978-1-4503-3794-6. DOI: `10.1145/2806416.2806567`. URL: `http://doi.acm.org/10.1145/2806416.2806567` (cit. on pp. 69, 84, 106, 208).

[133]   Charles Jochim and Hinrich Schütze. 'Towards a Generic and Flexible Citation Classifier Based on a Faceted Classification Scheme'. In: *Proceedings of COLING 2012*. Mumbai, India: The COLING 2012 Organizing Committee, Dec. 2012, pp. 1343–1358. URL: `https://www.aclweb.org/anthology/C12-1082` (cit. on p. 102).

[134]   P. Jomsri, S. Sanguansintukul and W. Choochaiwattana. 'A framework for tag-based research paper recommender system: An IR approach'. In: Perth, 2010, pp. 103–108. ISBN: 9780769540191. DOI: `10.1109/WAINA.2010.35` (cit. on pp. 31, 69, 84, 107).

[135]   Karen Sparck Jones. 'A statistical interpretation of term specificity and its application in retrieval'. In: *Journal of Documentation* 28.1 (1972), pp. 11–21 (cit. on pp. 38, 84, 138, 168).

[136]   David Jurgens et al. 'Measuring the Evolution of a Scientific Field through Citation Frames'. In: *Transactions of the Association for Computational Linguistics* 6 (2018), pp. 391–406. DOI: `10.1162/tacl_a_00028`. URL: `https://www.aclweb.org/anthology/Q18-1028` (cit. on pp. 102, 151, 153, 158, 161, 164, 185).

[137] N. Kapoor et al. 'TechLens - A researcher's desktop'. In: Minneapolis, MN, 2007, pp. 183–184. ISBN: 9781595937308. DOI: 10.1145/1297231.1297268 (cit. on p. 77).

[138] B. Kazemi and A. Abhari. 'A comparative study on content-based paper-to-paper recommendation approaches in scientific literature'. In: vol. 49. 2. The Society for Modeling and Simulation International, 2017, pp. 47–56. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85020621805&partnerID=40&md5=7f10a8dd6f447074175e203155133cd3 (cit. on pp. 69, 83–85, 106, 209).

[139] Anita Khadka. 'Citation-Context Dataset (C2D)'. In: (Aug. 2018). DOI: 10.21954/ou.rd.6865298. URL: https://figshare.com/articles/Citation-Context_Dataset_C2D_/6865298 (cit. on p. 136).

[140] Anita Khadka and Petr Knoth. 'Using Citation-context to Reduce Topic Drifting on Pure Citation-based Recommendation'. In: *12th ACM Conference on Recommender Systems*. 2018, pp. 362–366 (cit. on pp. 7, 69, 76, 80, 83–85, 87, 105, 114, 210).

[141] Jon M. Kleinberg. 'Authoritative Sources in a Hyperlinked Environment'. In: *J. ACM* 46.5 (Sept. 1999), pp. 604–632. ISSN: 0004-5411. DOI: 10.1145/324133.324140. URL: https://doi-org.libezproxy.open.ac.uk/10.1145/324133.324140 (cit. on p. 86).

[142] Bart P. Knijnenburg and Martijn C. Willemsen. 'Evaluating Recommender Systems with User Experiments'. In: *Recommender Systems Handbook*. Ed. by Francesco Ricci, Lior Rokach and Bracha Shapira. Boston, MA: Springer US, 2015, pp. 309–352. ISBN: 978-1-4899-7637-6. DOI: 10.1007/978-1-4899-7637-6_9. URL:

https://doi.org/10.1007/978-1-4899-7637-6_9 (cit. on p. 52).

[143]   Petr Knoth and Anita Khadka. 'Can we do better than co-citations? Bringing Citation Proximity Analysis from idea to practice in research articles recommendation'. In: *CEUR Workshop Proceedings*. Vol. 1888. 2017, pp. 14–25 (cit. on pp. 69, 76, 84, 85, 105, 210).

[144]   Yuta Kobayashi, Masashi Shimbo and Yuji Matsumoto. 'Citation Recommendation Using Distributed Representation of Discourse Facets in Scientific Articles'. In: *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. JCDL '18. Fort Worth, Texas, USA: ACM, 2018, pp. 243–251. ISBN: 978-1-4503-5178-2. DOI: 10.1145/3197026.3197059. URL: http://doi.acm.org/10.1145/3197026.3197059 (cit. on pp. 69, 76, 80, 83–85, 87, 105, 210).

[145]   A. Kodakateri Pudhiyaveetil et al. 'Conceptual recommender system for CiteSeerX'. In: New York, NY, 2009, pp. 241–244. ISBN: 9781605584355. DOI: 10.1145/1639714.1639758 (cit. on pp. 69, 81, 84, 108, 209).

[146]   Ron Kohavi et al. 'Controlled experiments on the web: survey and practical guide'. In: *Data Mining and Knowledge Discovery* 18.1 (2009), pp. 140–181. ISSN: 1573-756X. DOI: 10.1007/s10618-008-0114-1. URL: https://doi.org/10.1007/s10618-008-0114-1 (cit. on p. 52).

[147]   Joseph Konstan and John Riedl. 'Recommender systems: From algorithms to user experience'. In: *User Modeling and User-Adapted Interaction* 22 (Apr. 2012), pp. 101–123. DOI: 10.1007/s11257-011-9112-x (cit. on p. 115).

[148] Yehuda Koren, Robert Bell and Chris Volinsky. 'Matrix Factorization Techniques for Recommender Systems'. In: *Computer* 42.8 (Aug. 2009), pp. 30–37. ISSN: 0018-9162. DOI: 10.1109/MC.2009.263. URL: http://dx.doi.org/10.1109/MC.2009.263 (cit. on pp. 36, 43, 45, 86).

[149] William H. Kruskal and W. Allen Wallis. 'Use of Ranks in One-Criterion Variance Analysis'. In: *Journal of the American Statistical Association* 47.260 (1952), pp. 583–621. ISSN: 01621459. URL: http://www.jstor.org/stable/2280779 (cit. on p. 139).

[150] Onur Küçüktunç et al. 'Diversifying Citation Recommendations'. In: *ACM Trans. Intell. Syst. Technol.* 5.4 (Dec. 2014). ISSN: 2157–6904. DOI: 10.1145/2668106. URL: https://doi-org.libezproxy.open.ac.uk/10.1145/2668106 (cit. on pp. 12, 57, 60, 69, 76, 80, 84, 87, 108, 209).

[151] Onur Kucuktunc et al. 'TheAdvisor: A Webservice for Academic Recommendation'. In: *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*. JCDL '13. Indianapolis, Indiana, USA: ACM, 2013, pp. 433–434. ISBN: 978-1-4503-2077-1. DOI: 10.1145/2467696.2467752. URL: http://doi.acm.org/10.1145/2467696.2467752 (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 210).

[152] Onur Kúçúktunç et al. 'Recommendation on Academic Networks using Direction Aware Citation Analysis'. In: *CoRR* abs/1205.1143 (2012) (cit. on pp. 65, 69, 76, 80, 83, 84, 87, 105, 210).

[153] J. Richard Landis and Gary G. Koch. 'The Measurement of Observer Agreement for Categorical Data'. In: *Biometrics* 33.1

(1977), pp. 159–174. ISSN: 0006341X, 15410420. URL: http://www.jstor.org/stable/2529310 (cit. on p. 135).

[154] V. Le Anh et al. 'SciRecSys: A recommendation system for scientific publication by discovering keyword relationships'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8733 (2014), pp. 72–82. ISSN: 03029743. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84921341466&partnerID=40&md5=2f66b6deee101b3ed88bdfa4d9cb1d2e (cit. on pp. 69, 84, 106, 209).

[155] Joonseok Lee, Kisung Lee and Jennifer G. Kim. 'Personalized Academic Research Paper Recommendation System'. In: *CoRR* abs/1304.5457 (2013). arXiv: 1304.5457. URL: http://arxiv.org/abs/1304.5457 (cit. on pp. 37, 69, 81, 84, 109, 207).

[156] R. Lempel and S. Moran. 'The stochastic approach for link-structure analysis (SALSA) and the TKC effect1Abridged version1'. In: *Computer Networks* 33.1 (2000), pp. 387 –401. ISSN: 1389-1286. DOI: https://doi.org/10.1016/S1389-1286(00)00034-7. URL: http://www.sciencedirect.com/science/article/pii/S1389128600000347 (cit. on p. 86).

[157] Loet Leydesdorff and Staša Milojević. *Scientometrics*. 2012. arXiv: 1208.4566 [cs.DL] (cit. on p. 7).

[158] Xiang Li et al. 'Towards Fine-grained Citation Function Classification'. In: *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*. Hissar, Bulgaria: INCOMA Ltd. Shoumen, BULGARIA, Sept. 2013, pp. 402–407. URL: https://www.aclweb.org/anthology/R13-1052 (cit. on p. 103).

[159] S. Liang et al. 'A utility-based recommendation approach for academic literatures'. In: vol. 3. Lyon, 2011, pp. 229–232. ISBN: 9780769545134. DOI: 10.1109/WI-IAT.2011.110 (cit. on pp. 69, 76, 77, 80, 83, 84, 87, 105, 210).

[160] Yicong Liang, Qing Li and Tieyun Qian. 'Finding Relevant Papers Based on Citation Relations'. In: *Proceedings of the 12th International Conference on Web-age Information Management*. WAIM'11. Wuhan, China: Springer-Verlag, 2011, pp. 403–414. ISBN: 978-3-642-23534-4. URL: http://dl.acm.org/citation.cfm?id=2035562.2035609 (cit. on pp. 7, 12, 69, 72, 75–77, 80, 82–84, 86–88, 103, 105, 114, 157, 210).

[161] RA Likert. 'A Technique for Measurement of Attitudes'. In: *Archives of Psychology* 22 (Jan. 1932), pp. 1–55 (cit. on p. 137).

[162] Fengkun Liu and Hong Joo Lee. 'Use of Social Network Information to Enhance Collaborative Filtering Performance'. In: *Expert Syst. Appl.* 37.7 (July 2010), pp. 4772–4778. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2009.12.061. URL: http://dx.doi.org.libezproxy.open.ac.uk/10.1016/j.eswa.2009.12.061 (cit. on pp. 60, 88).

[163] H. Liu et al. 'CAR: Incorporating Filtered Citation Relations for Scientific Article Recommendation'. In: *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*. 2015, pp. 513–518 (cit. on pp. 69, 72, 76, 84, 108, 209).

[164] H. Liu et al. 'Context-Based Collaborative Filtering for Citation Recommendation'. In: *IEEE Access* 3 (2015), pp. 1695–1703. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2015.2481320 (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 210).

[165] X. Liu et al. 'Full-text based context-rich heterogeneous network mining approach for citation recommendation'. In: Institute of Electrical and Electronics Engineers Inc., 2014, pp. 361–370. ISBN: 9781479955695. DOI: `10.1109/JCDL.2014.6970191` (cit. on pp. 69, 76, 77, 80, 83, 84, 87, 106, 209).

[166] Xiaozhong Liu et al. 'Meta-Path-Based Ranking with Pseudo Relevance Feedback on Heterogeneous Graph for Citation Recommendation'. In: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. CIKM '14. Shanghai, China: ACM, 2014, pp. 121–130. ISBN: 978-1-4503-2598-1. DOI: `10.1145/2661829.2661965`. URL: `http://doi.acm.org/10.1145/2661829.2661965` (cit. on pp. 69, 76, 80, 83, 84, 87, 90, 106, 207).

[167] Xuan-You Liu and Been-Chain Chien. 'Applying Citation Network Analysis on Recommendation of Research Paper Collection'. In: *Proceedings of the 4th Multidisciplinary International Social Networks Conference*. MISNC '17. Bangkok, Thailand: ACM, 2017, 30:1–30:6. ISBN: 978-1-4503-4881-2. DOI: `10.1145/3092090.3092138`. URL: `http://doi.acm.org/10.1145/3092090.3092138` (cit. on pp. 69, 71, 76, 80, 83, 84, 87, 106, 208).

[168] Avishay Livne et al. 'CiteSight: Supporting Contextual Citation Recommendation Using Differential Search'. In: ACM – Association for Computing Machinery, 2014. URL: `https://www.microsoft.com/en-us/research/publication/citesight-supporting-contextual-citation-recommendation-using-differential-search/` (cit. on pp. 69, 76, 77, 80, 82–84, 87, 105, 207).

[169] S. Loh et al. 'Identifying similar users by their scientific publications to reduce cold start in recommender systems'. In: Lisbon, 2009, pp. 593–600. ISBN: 9789898111814. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-67650527543&partnerID=40&md5=0f0f678e339fff3eed1f05e8d27e9797 (cit. on pp. 69, 84, 85, 109, 208).

[170] Patrice Lopez. 'GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications'. In: *Research and Advanced Technology for Digital Libraries: 13th European Conference, ECDL 2009, Corfu, Greece, September 27 - October 2, 2009. Proceedings*. Ed. by Maristella Agosti et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 473–474. ISBN: 978-3-642-04346-8. DOI: 10.1007/978-3-642-04346-8_62. URL: http://dx.doi.org/10.1007/978-3-642-04346-8_62 (cit. on pp. 132, 136).

[171] Pasquale Lops, Marco de Gemmis and Giovanni Semeraro. 'Content-based Recommender Systems: State of the Art and Trends'. In: *Recommender Systems Handbook*. Ed. by Francesco Ricci et al. Boston, MA: Springer US, 2011, pp. 73–105. ISBN: 978-0-387-85820-3. DOI: 10.1007/978-0-387-85820-3_3. URL: https://doi.org/10.1007/978-0-387-85820-3_3 (cit. on pp. 30, 37, 39).

[172] Qiuxia Lu et al. 'Serendipitous Personalized Ranking for Top-N Recommendation'. In: *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 01*. WI-IAT '12. IEEE Computer Society, 2012, pp. 258–265. ISBN: 9780769548807 (cit. on pp. 58, 114).

[173] K. Ma, T. Lu and A. Abraham. 'Hybrid parallel approach for personalized literature recommendation system'. In: Institute of Electrical and Electronics Engineers Inc., 2014, pp. 31–36. ISBN: 9781479959402. DOI: 10.1109/CASoN.2014.6920428 (cit. on pp. 69, 84, 85, 108, 208).

[174] Shutian Ma, Chengzhi Zhang and Xiaozhong Liu. 'A review of citation recommendation: from textual content to enriched context'. In: *Scientometrics* 122.3 (2020), pp. 1445–1472. DOI: 10.1007/s11192-019-03336-. URL: https://ideas.repec.org/a/spr/scient/v122y2020i3d10.1007_s11192-019-03336-0.html (cit. on pp. 70, 72).

[175] X. Ma and R. Wang. 'Personalized Scientific Paper Recommendation Based on Heterogeneous Graph Representation'. In: *IEEE Access* 7 (2019), pp. 79887–79894. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2923293 (cit. on pp. 69, 84, 109, 208).

[176] M.B. Magara, S. Ojo and T. Zuva. 'Toward altmetric-driven research-paper recommender system framework'. In: vol. 2018-January. Institute of Electrical and Electronics Engineers Inc., 2018, pp. 63–68. ISBN: 9781538642832. DOI: 10.1109/SITIS.2017.21 (cit. on pp. 69, 84, 109).

[177] M.B. Magara, S.O. Ojo and T. Zuva. 'Towards a Serendipitous Research Paper Recommender System Using Bisociative Information Networks (BisoNets)'. In: Institute of Electrical and Electronics Engineers Inc., 2018. ISBN: 9781538630600. DOI: 10.1109/ICABCD.2018.8465475 (cit. on pp. 69, 84, 85, 106, 208).

[178] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze. *Introduction to Information Retrieval*. New York, NY,

USA: Cambridge University Press, 2008. ISBN: 0521865719, 9780521865715 (cit. on pp. 37, 54–56).

[179]   Benjamin M. Marlin and Richard S. Zemel. 'Collaborative Prediction and Ranking with Non-Random Missing Data'. In: *Proceedings of the Thirs ACM Conference on Recommender Systems*. RecSys 2009. 2009, pp. 5–12. ISBN: 9781605584355 (cit. on p. 33).

[180]   Nikolaos F. Matsatsinis, Kleanthi Lakiotaki and Pavlos Delias. *A System based on Multiple Criteria Analysis for Scientific Paper Recommendation*. 2007 (cit. on pp. 69, 82, 84, 106, 209).

[181]   P. Mayr. 'How do practitioners, PhD students and postdocs in the social sciences assess topic-specific recommendations?' In: vol. 1610. CEUR-WS, 2016, pp. 84–92. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84977591081&partnerID=40&md5=69282a7c58b11db2fe4c1eb88fff9084 (cit. on pp. 69, 84, 106, 208).

[182]   S.M. McNee, N. Kapoor and J.A. Konstan. 'Don't look stupid: Avoiding pitfalls when recommending research papers'. In: Banff, AB, 2006, pp. 171–180. ISBN: 1595932496; 9781595932495. DOI: 10.1145/1180875.1180903 (cit. on pp. 62, 69, 76, 80, 83, 84, 86, 105, 209).

[183]   S.M. McNee et al. 'On the recommending of citations for research papers'. In: New Orleans, LA, 2002, pp. 116–125. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0038336876&partnerID=40&md5=8826c02491485fa014b8af3e2081be94 (cit. on pp. 6, 12, 31, 69, 73, 76, 77, 80, 82–84, 86, 87, 91, 105, 114, 156, 192, 210).

[184]   F. Meng et al. 'A unified graph model for personalized query-oriented reference paper recommendation'. In: San Francisco,

CA, 2013, pp. 1509–1512. ISBN: 9781450322638. DOI: 10.1145/2505515.2507831 (cit. on pp. 69, 77, 84, 106, 209).

[185] Stephen Merity, Tara Murphy and James R. Curran. 'Accurate Argumentative Zoning with Maximum Entropy Models'. In: *Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries*. NLPIR4DL '09. Suntec, Singapore: Association for Computational Linguistics, 2009, pp. 19–26. ISBN: 978-1-932432-58-9. URL: http://dl.acm.org/citation.cfm?id=1699750.1699754 (cit. on p. 102).

[186] Robert K. Merton. 'The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property'. In: *Isis* 79.4 (1988), pp. 606–623. ISSN: 00211753, 15456994. URL: http://www.jstor.org/stable/234750 (cit. on p. 90).

[187] Adam Meyers. 'Contrasting and Corroborating Citations in Journal Articles'. In: *RANLP*. 2013 (cit. on p. 103).

[188] S.E. Middleton, N.R. Shadbolt and D.C. De Roure. 'Capturing interest through inference and visualization: Ontological user profiling in recommender systems'. In: Association for Computing Machinery, Inc, 2003, pp. 62–69. ISBN: 1581135831; 9781581135831. DOI: 10.1145/945645.945657 (cit. on pp. 33, 69, 81, 84, 107).

[189] Stuart E. Middleton, David C. De Roure and Nigel R. Shadbolt. 'Capturing Knowledge of User Preferences: Ontologies in Recommender Systems'. In: *1st International Conference on Knowledge Capture*. 2001, pp. 100–107 (cit. on pp. 7, 12, 33, 69, 81, 84, 107).

[190] Tomas Mikolov et al. 'Distributed Representations of Words and Phrases and Their Compositionality'. In: *Proceedings of*

the 26th International Conference on Neural Information Processing Systems - Volume 2. NIPS'13. Lake Tahoe, Nevada: Curran Associates Inc., 2013, pp. 3111–3119. URL: http://dl.acm.org/citation.cfm?id=2999792.2999959 (cit. on p. 130).

[191] H.A. Mohamed Hassan. 'Personalized research paper recommendation using deep learning'. In: Association for Computing Machinery, Inc, 2017, pp. 327–330. ISBN: 9781450346351. DOI: 10.1145/3079628.3079708 (cit. on pp. 69, 84, 108, 208).

[192] D. Mu et al. 'Query-Focused Personalized Citation Recommendation with Mutually Reinforced Ranking'. In: *IEEE Access* 6 (2017), pp. 3107–3119. ISSN: 21693536. DOI: 10.1109/ACCESS.2017.2787179 (cit. on pp. 69, 76, 80, 83, 84, 87, 90, 106, 207).

[193] Hidetsugu Nanba, Noriko Kando and Manabu Okumura. 'Classification of research papers using citation links and citation types: Towards automatic review article generation.' In: *Advances in Classification Research Online* 11.1 (2000), pp. 117–134. ISSN: 2324-9773. URL: https://journals.lib.washington.edu/index.php/acro/article/view/12774 (cit. on p. 103).

[194] Hidetsugu Nanba and Manabu Okumura. 'Towards Multipaper Summarization Reference Information'. In: *Proceedings of the 16th International Joint Conference on Artificial Intelligence - Volume 2*. IJCAI'99. Stockholm, Sweden: Morgan Kaufmann Publishers Inc., 1999, pp. 926–931. URL: http://dl.acm.org/citation.cfm?id=1624312.1624351 (cit. on p. 103).

[195] C. Nascimento et al. 'A source independent framework for research paper recommendation'. In: Ottawa, ON, 2011, pp. 297–306. ISBN: 9781450307444. DOI: 10.1145/1998076.1998132 (cit. on pp. 69, 77, 82, 84, 106, 114, 192, 208).

[196] K.V. Neethukrishnan and K.P. Swaraj. 'Ontology based research paper recommendation using personal ontology similarity method'. In: Institute of Electrical and Electronics Engineers Inc., 2017. ISBN: 9781509032389. DOI: 10.1109/ICECCT.2017.8117833 (cit. on pp. 69, 84, 106, 209).

[197] C. Nishioka and A. Scherp. 'Profiling vs. time vs. content: What does matter for top-k publication recommendation based on Twitter profiles?' In: vol. 2016-September. Institute of Electrical and Electronics Engineers Inc., 2016, pp. 171–180. ISBN: 9781450342292. DOI: 10.1145/2910896.2910898 (cit. on pp. 60, 69, 81, 84, 109, 208).

[198] M. Ohta, T. Hachiki and A. Takasu. 'Related paper recommendation to support online-browsing of research papers'. In: *Fourth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2011)*. 2011, pp. 130–136. DOI: 10.1109/ICADIWT.2011.6041413 (cit. on pp. 69, 84, 107, 208).

[199] A. Ollagnier, S. Fournier and P. Bellot. 'BIBLME RecSys: Harnessing bibliometric measures for a scholarly paper recommender system'. In: vol. 2080. CEUR-WS, 2018, pp. 34–45. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85045380851&partnerID=40&md5=c334a7a3410e39fd8fd133bcf2fa6d43 (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 210).

[200] Lawrence Page et al. *The PageRank Citation Ranking: Bringing Order to the Web.* Technical Report 1999-66. Previous number = SIDL-WP-1999-0120. Stanford InfoLab, 1999. URL: http://ilpubs.stanford.edu:8090/422/ (cit. on p. 86).

[201] C. Pan and W. Li. 'Research paper recommendation with topic analysis'. In: vol. 4. Qinhuangdao, Hebei, 2010, pp. V4264–

V4268. ISBN: 9781424471638. DOI: 10.1109/ICCDA.2010.5541170
(cit. on pp. 69, 77, 84, 85, 108).

[202] L. Pan et al. 'Academic paper recommendation based on hetero-
geneous graph'. In: *Lecture Notes in Computer Science (including
subseries Lecture Notes in Artificial Intelligence and Lecture Notes
in Bioinformatics)* 9427 (2015), pp. 381–392. ISSN: 03029743. DOI:
10.1007/978-3-319-25816-4_31 (cit. on pp. 6, 69, 76, 80, 83,
84, 87, 105, 210).

[203] I.C. Paraschiv et al. 'A paper recommendation system with
readerbench: the graphical visualization of semantically related
papers and concepts'. In: *Lecture Notes in Educational Technology*
9789812878663 (2016), pp. 445–451. ISSN: 21964963. DOI: 10.
1007/978-981-287-868-7_53 (cit. on pp. 69, 77, 84, 106, 209).

[204] D. Parra-Santander and P. Brusilovsky. 'Improving collaborat-
ive filtering in social tagging systems for the recommendation
of scientific articles'. In: vol. 1. Toronto, ON, 2010, pp. 136–142.
ISBN: 9780769541914. DOI: 10.1109/WI-IAT.2010.261 (cit. on
pp. 69, 81, 84, 107).

[205] Bhavik Pathak et al. 'Empirical Analysis of the Impact of
Recommender Systems on Sales.' In: *Journal of Management
Information Systems* 27.2 (2010), pp. 159–188. ISSN: 07421222.
URL: http://libezproxy.open.ac.uk/login?url=https:
//search.ebscohost.com/login.aspx?direct=true&db=bth&
AN=55139359&site=ehost-live&scope=site (cit. on p. 28).

[206] Michael J. Pazzani. 'A Framework for Collaborative, Content-
Based and Demographic Filtering'. In: *ARTIFICIAL INTELLI-
GENCE REVIEW* 13 (1999), pp. 393–408 (cit. on pp. 37, 61).

[207] Michael J. Pazzani and Daniel Billsus. 'Content-Based Recommendation Systems'. In: *The Adaptive Web: Methods and Strategies of Web Personalization*. Ed. by Peter Brusilovsky, Alfred Kobsa and Wolfgang Nejdl. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 325–341. ISBN: 978-3-540-72079-9. DOI: `10.1007/978-3-540-72079-9_10`. URL: `https://doi.org/10.1007/978-3-540-72079-9_10` (cit. on p. 37).

[208] Michael Pazzani and Daniel Billsus. 'Learning and Revising User Profiles: The Identification OfInteresting Web Sites'. In: *Mach. Learn.* 27.3 (June 1997), pp. 313–331. ISSN: 0885-6125. DOI: `10.1023/A:1007369909943`. URL: `https://doi.org/10.1023/A:1007369909943` (cit. on pp. 37, 39).

[209] Jeffrey Pennington, Richard Socher and Christopher D. Manning. 'GloVe: Global Vectors for Word Representation'. In: *Empirical Methods in Natural Language Processing (EMNLP)*. 2014, pp. 1532–1543. URL: `http://www.aclweb.org/anthology/D14-1162` (cit. on p. 130).

[210] M.S. Pera and Y.-K. Ng. 'A personalized recommendation system on scholarly publications'. In: Glasgow, 2011, pp. 2133–2136. ISBN: 9781450307178. DOI: `10.1145/2063576.2063908` (cit. on pp. 33, 69, 81, 84, 88, 107).

[211] M.S. Pera and Y.-K. Ng. 'Exploiting the wisdom of social connections to make personalized recommendations on scholarly articles'. In: *Journal of Intelligent Information Systems* 42.3 (2014), pp. 371–391. ISSN: 09259902. DOI: `10.1007/s10844-013-0298-8` (cit. on pp. 60, 69, 81, 84, 88, 107, 208).

[212] Son Bao Pham and Achim Hoffmann. 'A New Approach for Scientific Citation Classification Using Cue Phrases'. In: *AI*

2003: Advances in Artificial Intelligence. 2003, pp. 759–771. DOI: 10. 1007/978-3-540-24581-0_65. URL: https://app.dimensions. ai/details/publication/pub.1026864792 (cit. on p. 104).

[213] Sheila Pontis et al. 'Keeping up to date: An academic researcher's information journey'. In: *JASIST* 68 (2017), pp. 22–35 (cit. on p. 7).

[214] H.-E. Popa et al. 'DL-agentrecom - A multi-agent based recommendation system for scientific documents'. In: Timisoara, 2008, pp. 320–324. ISBN: 9780769535234. DOI: 10.1109/SYNASC. 2008.79 (cit. on pp. 33, 69, 81, 84, 101, 107).

[215] A.S. Raamkumar, S. Foo and N. Pang. 'Comparison of techniques for measuring research coverage of scientific papers: A case study'. In: Institute of Electrical and Electronics Engineers Inc., 2016, pp. 132–137. ISBN: 9781467391511. DOI: 10.1109/ ICDIM.2015.7381881 (cit. on pp. 69, 84, 106, 209).

[216] Sandeep K. Raghuwanshi and R. K. Pateriya. 'Recommendation Systems: Techniques, Challenges, Application, and Evaluation'. In: *Soft Computing for Problem Solving*. Ed. by Jagdish Chand Bansal et al. Singapore: Springer Singapore, 2019, pp. 151–164. ISBN: 978-981-13-1595-4 (cit. on p. 59).

[217] X. Ren et al. 'ClusCite: Effective citation recommendation by information network-based clustering'. In: New York, NY: Association for Computing Machinery, 2014, pp. 821–830. ISBN: 9781450329569. DOI: 10.1145/2623330.2623630 (cit. on pp. 69, 76, 80, 83, 84, 87, 90, 105, 208).

[218] Steffen Rendle. 'Factorization Machines'. In: *Proceedings of the 2010 IEEE International Conference on Data Mining*. ICDM '10.

IEEE Computer Society, 2010, pp. 995–1000. ISBN: 978-0-7695-4256-0. DOI: 10.1109/ICDM.2010.127 (cit. on pp. 47, 48, 170).

[219]   Paul Resnick and Hal R Varian. 'Recommender systems'. In: *Communications of the ACM* 40.3 (1997), pp. 56–58 (cit. on p. 40).

[220]   Paul Resnick et al. 'GroupLens: an open architecture for collaborative filtering of netnews'. In: *CSCW '94*. 1994 (cit. on pp. 27, 28, 35, 40, 41).

[221]   Francesco Ricci. *Recommender Systems Research Challenges*. Lugano, Italy, 2017. URL: https://www.inf.unibz.it/~ricci/Slides/Recommender-Systems-Ricci-IIR2017.pdf (cit. on p. 59).

[222]   Francesco Ricci et al. 'Case-based travel recommendations.' In: 2006 (cit. on p. 25).

[223]   Francesco Ricci et al. *Recommender Systems Handbook*. 1st. Berlin, Heidelberg: Springer-Verlag, 2010. ISBN: 0387858199, 9780387858197 (cit. on pp. 25, 31–33, 39, 40, 46).

[224]   Anna Ritchie, Simone Teufel and Stephen Robertson. 'How to Find Better Index Terms Through Citations'. In: *Proceedings of the Workshop on How Can Computational Linguistics Improve Information Retrieval?* CLIIR '06. Sydney, Australia: Association for Computational Linguistics, 2006, pp. 25–32. URL: http://dl.acm.org/citation.cfm?id=1629808.1629813 (cit. on pp. 6, 88, 120, 127).

[225]   Alan Said and Alejandro Bellogín. 'Rival: A Toolkit to Foster Reproducibility in Recommender System Evaluation'. In: *8th ACM Conference on Recommender Systems*. 2014, pp. 371–372 (cit. on p. 171).

[226] Badrul M. Sarwar et al. 'Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System'. In: *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*. CSCW '98. Seattle, Washington, USA: Association for Computing Machinery, 1998, pp. 345–354. ISBN: 1581130090. DOI: 10.1145/289444.289509. URL: https://doi.org/10.1145/289444.289509 (cit. on p. 58).

[227] Badrul M. Sarwar et al. 'Application of Dimensionality Reduction in Recommender System – A Case Study'. In: *IN ACM WEBKDD WORKSHOP*. 2000 (cit. on p. 44).

[228] Badrul Sarwar et al. 'Item-based Collaborative Filtering Recommendation Algorithms'. In: *10th International Conference on World Wide Web*. 2001, pp. 285–295 (cit. on pp. 35, 41, 42, 51, 159).

[229] J. Ben Schafer et al. 'Collaborative Filtering Recommender Systems'. In: *The Adaptive Web: Methods and Strategies of Web Personalization*. Ed. by Peter Brusilovsky, Alfred Kobsa and Wolfgang Nejdl. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 291–324. ISBN: 978-3-540-72079-9. DOI: 10.1007/978-3-540-72079-9_9. URL: https://doi.org/10.1007/978-3-540-72079-9_9 (cit. on pp. 33, 40).

[230] Malte Schwarzer et al. 'Evaluating Link-Based Recommendations for Wikipedia'. In: *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*. JCDL '16. Newark, New Jersey, USA: Association for Computing Machinery, 2016, pp. 191–200. ISBN: 9781450342292. DOI: 10.1145/2910896.2910908. URL: https://doi.org/10.1145/2910896.2910908 (cit. on p. 74).

[231]   Giovanni Semeraro et al. 'Discovering User Profiles from Semantically Indexed Scientific Papers'. In: *From Web to Social Web: Discovering and Deploying User and Content Profiles*. Ed. by Bettina Berendt, Andreas Hotho and Dunja Mladenic. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 61–81. ISBN: 978-3-540-74951-6 (cit. on pp. 69, 84, 108, 208).

[232]   A. Sesagiri Raamkumar, S. Foo and N. Pang. 'Rec4LRW-scientific paper recommender system for literature review and writing'. In: *Frontiers in Artificial Intelligence and Applications* 275 (2015), pp. 106–119. ISSN: 09226389. DOI: `10.3233/978-1-61499-503-6-106` (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 208).

[233]   A. Sesagiri Raamkumar, S. Foo and N. Pang. 'Evaluating a threefold intervention framework for assisting researchers in literature review and manuscript preparatory tasks'. In: *Journal of Documentation* 73.3 (2017), pp. 555–580. ISSN: 00220418. DOI: `10.1108/JD-06-2016-0072` (cit. on pp. 69, 76, 80, 83, 84, 87, 106, 209).

[234]   A. Sesagiri Raamkumar, S. Foo and N. Pang. 'User evaluation of a task for shortlisting papers from researcher's reading list for citing in manuscripts'. In: *Aslib Journal of Information Management* 69.6 (2017), pp. 740–760. ISSN: 20503806. DOI: `10.1108/AJIM-01-2017-0020` (cit. on pp. 69, 76, 80, 83, 84, 87, 106, 208).

[235]   A. Sesagiri Raamkumar, S. Foo and N. Pang. 'Using author-specified keywords in building an initial reading list of research papers in scientific paper retrieval and recommender systems'. In: *Information Processing and Management* 53.3 (2017), pp. 577–

594. ISSN: 03064573. DOI: 10.1016/j.ipm.2016.12.006 (cit. on pp. 69, 84, 106, 209).

[236] A. Sesagiri Raamkumar, S. Foo and N. Pang. 'Can I have more of these please?: Assisting researchers in finding similar research papers from a seed basket of papers'. In: *Electronic Library* 36.3 (2018), pp. 568–587. ISSN: 02640473. DOI: 10.1108/EL-04-2017-0077 (cit. on pp. 7, 69, 83, 84, 105, 209).

[237] Cyrus Shahabi and Yi-Shin Chen. 'An Adaptive Recommendation System without Explicit Acquisition of User Relevance Feedback'. In: *Distributed and Parallel Databases* 14.2 (2003), pp. 173–192 (cit. on pp. 34, 81).

[238] Guy Shani and Asela Gunawardana. 'Evaluating Recommendation Systems'. In: *Recommender Systems Handbook*. Ed. by Francesco Ricci et al. Boston, MA: Springer US, 2011, pp. 257–297. ISBN: 978-0-387-85820-3. DOI: 10.1007/978-0-387-85820-3_8. URL: https://doi.org/10.1007/978-0-387-85820-3_8 (cit. on pp. 49, 57, 59, 62).

[239] Upendra Shardanand and Pattie Maes. 'Social Information Filtering: Algorithms for Automating "Word of Mouth"'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '95. USA: ACM Press/Addison-Wesley Publishing Co., 1995, pp. 210–217. ISBN: 0201847051. DOI: 10.1145/223904.223931. URL: https://doi.org/10.1145/223904.223931 (cit. on pp. 28, 29, 35, 41).

[240] R. Sharma, D. Gopalani and Y. Meena. 'Concept-Based Approach for Research Paper Recommendation'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10597 LNCS

(2017), pp. 687–692. ISSN: 03029743. DOI: `10.1007/978-3-319-69900-4_87` (cit. on pp. 69, 84, 90, 101, 106).

[241] Masashi Shimbo, Ito Takahiko and Yuji Matsumoto. 'Evaluation of kernel-based link analysis measures on research paper recommendation'. In: *7th ACM International Conference on Digital Libraries*. 2007, pp. 354–355 (cit. on p. 12).

[242] S. Siebert, S. Dinesh and S. Feyer. 'Extending a research-paper recommendation system with scientometric measures'. In: vol. 1823. CEUR-WS, 2017, pp. 112–121 (cit. on pp. 69, 81, 84, 109, 209).

[243] Henry Small. 'Co-citation in the scientific literature: A new measure of the relationship between two documents'. In: *Journal of the American Society for Information Science* 24.4 (1973), pp. 265–269. DOI: `10.1002/asi.4630240406` (cit. on pp. 133, 137).

[244] J. Son and S.B. Kim. 'Academic paper recommender system using multilevel simultaneous citation networks'. In: *Decision Support Systems* 105 (2018), pp. 24–33. ISSN: 01679236. DOI: `10.1016/j.dss.2017.10.011` (cit. on pp. 69, 71, 76, 80, 83, 84, 87, 105, 210).

[245] T. Sripadh and G. Ramesh. 'Personalized research paper recommender system'. In: *Lecture Notes in Computational Vision and Biomechanics* 28 (2018), pp. 437–446. ISSN: 22129391. DOI: `10.1007/978-3-319-71767-8_36` (cit. on pp. 69, 84, 107, 146).

[246] T. Sripadh and G. Ramesh. 'Personalized research paper recommender system'. In: *Lecture Notes in Computational Vision and Biomechanics* 28 (2018), pp. 437–446. ISSN: 22129391. DOI: `10.1007/978-3-319-71767-8_36` (cit. on pp. 81, 84).

[247] Laura Steinert, Irene-Angelica Chounta and H. Ulrich Hoppe. 'Where to Begin? Using Network Analytics for the Recommend-

ation of Scientific Papers'. In: *Collaboration and Technology*. Ed. by Nelson Baloian et al. Cham: Springer International Publishing, 2015, pp. 124–139. ISBN: 978-3-319-22747-4 (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 210).

[248] Trevor Strohman, W. Bruce Croft and David Jensen. 'Recommending Citations for Academic Papers'. In: *30th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2007, pp. 705–706 (cit. on pp. 7, 12, 69, 76, 77, 80, 82–84, 87, 90, 105, 210).

[249] Xuan Su et al. 'Neural Multi-Task Learning for Citation Function and Provenance'. In: *CoRR* abs/1811.07351 (2018). URL: http://arxiv.org/abs/1811.07351 (cit. on p. 103).

[250] K. Sugiyama and M.-Y. Kan. 'Serendipitous recommendation for scholarly papers considering relations among researchers'. In: Ottawa, ON, 2011, pp. 307–310. ISBN: 9781450307444. DOI: 10.1145/1998076.1998133 (cit. on pp. 69, 76, 80, 81, 84, 87, 109, 210).

[251] K. Sugiyama and M.-Y. Kan. 'A comprehensive evaluation of scholarly paper recommendation using potential citation papers'. In: *International Journal on Digital Libraries* 16.2 (2015), pp. 91–109. ISSN: 14325012. DOI: 10.1007/s00799-014-0122-2 (cit. on pp. 69, 70, 72, 76, 77, 79–81, 84, 87, 90, 109, 200, 208).

[252] Kazunari Sugiyama and Min-Yen Kan. 'Scholarly Paper Recommendation via User's Recent Research Interests'. In: *10th Joint Conference on Digital Libraries*. 2010, pp. 29–38 (cit. on pp. 8, 12, 34, 58, 69, 72, 76, 77, 80–82, 84, 85, 87, 109, 115, 147, 210).

[253] Kazunari Sugiyama and Min-Yen Kan. 'Exploiting Potential Citation Papers in Scholarly Paper Recommendation'. In: *13th*

ACM/IEEE-CS Joint Conference on Digital Libraries. 2013, pp. 153–162 (cit. on pp. 12, 69, 72, 74, 76, 77, 79–82, 84, 87, 88, 109, 115, 147, 148, 157, 208).

[254] Kazunari Sugiyama and Min-Yen Kan. '"Towards Higher Relevance and Serendipity in Scholarly Paper Recommendation" by Kazunari Sugiyama and Min-Yen Kan with Martin Vesely As Coordinator'. In: *SIGWEB Newsl.* Winter (Feb. 2015), 4:1–4:16. ISSN: 1931-1745. DOI: 10.1145/2719943.2719947. URL: http://doi.acm.org/10.1145/2719943.2719947 (cit. on pp. 6, 9, 69, 72, 76, 79–81, 84, 85, 87, 109, 210).

[255] J. Sun et al. 'A Novel Approach for Personalized Article Recommendation in Online Scientific Communities'. In: *2013 46th Hawaii International Conference on System Sciences.* 2013, pp. 1543–1552. DOI: 10.1109/HICSS.2013.48 (cit. on pp. 69, 84, 109, 208).

[256] J. Sun et al. 'A hybrid approach for article recommendation in research social networks'. In: *Journal of Information Science* 44.5 (2018), pp. 696–711. ISSN: 01655515. DOI: 10.1177/0165551517728449 (cit. on pp. 69, 84, 108, 208).

[257] Jianshan Sun et al. 'Leveraging Content and Connections for Scientific Article Recommendation in Social Computing Contexts'. In: *Comput. J.* 57 (2014), pp. 1331–1342 (cit. on pp. 69, 81, 84, 107, 208).

[258] X. Tang and Q. Zeng. 'Keyword clustering for user interest profiling refinement within paper recommender systems'. In: *Journal of Systems and Software* 85.1 (2012), pp. 87–101. ISSN: 01641212. DOI: 10.1016/j.jss.2011.07.029 (cit. on pp. 69, 84, 85, 107).

[259] W. Tanner, E. Akbas and M. Hasan. 'Paper Recommendation Based on Citation Relation'. In: *2019 IEEE International Conference on Big Data (Big Data)*. 2019, pp. 3053–3059. DOI: `10.1109/BigData47090.2019.9006200` (cit. on pp. 69, 73, 76, 80, 83, 84, 87, 105, 192, 210).

[260] A. Tejeda-Lorente et al. 'REFORE: A recommender system for researchers based on bibliometrics'. In: *Applied Soft Computing Journal* 30 (2015), pp. 778–791. ISSN: 15684946. DOI: `10.1016/j.asoc.2015.02.024` (cit. on pp. 69, 84, 109, 208).

[261] A. Tejeda-Lorente et al. 'Using bibliometrics and fuzzy linguistic modeling to deal with cold start in recommender systems for digital libraries'. In: *Advances in Intelligent Systems and Computing* 643 (2018), pp. 393–404. ISSN: 21945357. DOI: `10.1007/978-3-319-66827-7_36` (cit. on pp. 69, 84, 109, 209).

[262] Loren Terveen et al. 'PHOAKS: A System for Sharing Recommendations'. In: *Commun. ACM* 40.3 (Mar. 1997), pp. 59–62. ISSN: 0001-0782. DOI: `10.1145/245108.245122`. URL: `https://doi.org/10.1145/245108.245122` (cit. on p. 28).

[263] Simone Teufel. 'Argumentative zoning information extraction from scientific text'. In: 1999 (cit. on p. 102).

[264] Simone Teufel, Advaith Siddharthan and Dan Tidhar. 'Automatic Classification of Citation Function'. In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. EMNLP '06. Sydney, Australia: Association for Computational Linguistics, 2006, pp. 103–110. ISBN: 1-932432-73-6. URL: `http://dl.acm.org/citation.cfm?id=1610075.1610091` (cit. on p. 103).

[265] G. Tian and L. Jing. 'Recommending scientific articles using bi-relational graph-based iterative RWR'. In: Hong Kong, 2013, pp. 399–402. ISBN: 9781450324090. DOI: 10.1145/2507157.2507212 (cit. on pp. 69, 84, 108, 208).

[266] Ignacio Tobías. 'Matrix factorization models for cross-domain recommendation: Addressing the cold start in collaborative filtering'. In: 2017 (cit. on p. 26).

[267] Roberto Torres et al. 'Enhancing Digital Libraries with Tech-Lens+'. In: *4th ACM/IEEE-CS Joint Conference on Digital Libraries*. 2004, pp. 228–236 (cit. on pp. 7, 12, 47, 69, 72, 76, 80, 82, 84, 87, 91, 108, 114, 146, 147, 210).

[268] A. Tsolakidis et al. 'Research publication recommendation system based on a hybrid approach'. In: Association for Computing Machinery, 2016. ISBN: 9781450347891. DOI: 10.1145/3003733.3003805 (cit. on pp. 69, 84, 109).

[269] Marco Valenzuela, Vu A. Ha and Oren Etzioni. 'Identifying Meaningful Citations'. In: *AAAI Workshop: Scholarly Big Data*. 2015 (cit. on pp. 97, 104, 158).

[270] Saúl Vargas and Pablo Castells. 'Improving Sales Diversity by Recommending Users to Items'. In: *Proceedings of the 8th ACM Conference on Recommender Systems*. RecSys '14. California, USA: Association for Computing Machinery, 2014, pp. 145–152. ISBN: 9781450326681. DOI: 10.1145/2645710.2645744. URL: https://doi-org.libezproxy.open.ac.uk/10.1145/2645710.2645744 (cit. on pp. 57, 62).

[271] A. Vellino and D. Zeber. 'A hybrid, multi-dimensional recommender for journal articles in a scientific digital library'. In: Silicon Valley, CA, 2007, pp. 111–114. ISBN: 0769530281;

9780769530284. DOI: 10.1109/WIIATW.2007.4427552 (cit. on pp. 69, 84, 100, 108, 209).

[272] Chong Wang and David M Blei. 'Collaborative topic modeling for recommending scientific articles'. In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2011, pp. 448–456 (cit. on pp. 8, 69, 72, 84, 88, 91, 98, 108, 137, 146, 169, 183, 193, 208).

[273] Fan Wang, Ning Shi and Ben Chen. 'A comprehensive survey of the reviewer assignment problem'. In: *International Journal of Information Technology & Decision Making* 9.04 (2010), pp. 645–668 (cit. on p. 6).

[274] Q. Wang et al. 'Academic paper recommendation based on community detection in citation-collaboration networks'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9932 LNCS (2016), pp. 124–136. DOI: 10.1007/978-3-319-45817-5_10 (cit. on pp. 69, 76, 80, 83, 84, 87, 106, 209).

[275] Y. Wang et al. 'Personalized paper recommendation based on user historical behavior'. In: *Communications in Computer and Information Science* 333 CCIS (2012), pp. 1–12. ISSN: 18650929. DOI: 10.1007/978-3-642-34456-5_1 (cit. on pp. 69, 81, 84, 107, 208).

[276] Z. Wang et al. 'A personalization-oriented academic literature recommendation method'. In: *Data Science Journal* 14 (2015). ISSN: 16831470. DOI: 10.5334/dsj-2015-017 (cit. on pp. 69, 81, 84, 107, 209).

[277] S. Watanabe et al. 'A paper recommendation mechanism for the research support system Papits'. In: vol. 2005. Tokyo, 2005,

pp. 71–80. ISBN: 076952401X; 9780769524016. DOI: `10.1109/DEEC.2005.3` (cit. on pp. 69, 84, 108, 209).

[278] S.-S. Weng and H.-L. Chang. 'Using ontology network analysis for research document recommendation'. In: *Expert Systems with Applications* 34.3 (2008), pp. 1857–1869. ISSN: 09574174. DOI: `10.1016/j.eswa.2007.02.023` (cit. on pp. 69, 84, 107, 210).

[279] J. D. West, I. Wesley-Smith and C. T. Bergstrom. 'A Recommendation System Based on Hierarchical Clustering of an Article-Level Citation Network'. In: *IEEE Transactions on Big Data* 2.2 (2016), pp. 113–123. ISSN: 2332-7790. DOI: `10.1109/TBDATA.2016.2541167` (cit. on pp. 69, 71, 76, 78, 80, 83, 84, 87, 105, 210).

[280] Ryen W. White, Ian Ruthven and Joemon M. Jose. 'A Study of Factors Affecting the Utility of Implicit Relevance Feedback'. In: *28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2005), Salvador, Brazil,* Aug. 2005, pp. 35–42. URL: `https://www.microsoft.com/en-us/research/publication/study-factors-affecting-utility-implicit-relevance-feedback/` (cit. on p. 33).

[281] F. Xia et al. 'Scientific Article Recommendation: Exploiting Common Author Relations and Historical Preferences'. In: *IEEE Transactions on Big Data* 2.2 (2016), pp. 101–112. ISSN: 2332-7790. DOI: `10.1109/TBDATA.2016.2555318` (cit. on pp. 69, 84, 108, 209).

[282] Z. Xiao et al. 'Increasing serendipity of recommender system with ranking topic model'. In: *Applied Mathematics and Information Sciences* 8.4 (2014), pp. 2041–2053. ISSN: 19350090. DOI: `10.12785/amis/080463` (cit. on pp. 69, 77, 84, 106, 208).

[283] H. Xue et al. 'Personalized paper recommendation in online social scholar system'. In: Institute of Electrical and Electronics Engineers Inc., 2014, pp. 612–619. ISBN: 9781479958771. DOI: 10.1109/ASONAM.2014.6921649 (cit. on pp. 69, 76, 80, 84, 85, 87, 108, 208).

[284] Libin Yang et al. 'A LSTM Based Model for Personalized Context-Aware Citation Recommendation'. In: *IEEE Access* 6 (2018), pp. 59618–59627 (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 208).

[285] M. Yang, Y.-M. Li and Z.M. Zhang. 'Scientific articles recommendation with topic regression and relational matrix factorization'. In: *Journal of Zhejiang University: Science C* 15.11 (2014), pp. 984–998. ISSN: 18691951. DOI: 10.1631/jzus.C1300374 (cit. on pp. 69, 84, 109, 208).

[286] W.-S. Yang and Y.-R. Lin. 'A task-focused literature recommender system for digital libraries'. In: *Online Information Review* 37.4 (2013), pp. 581–601. ISSN: 14684527. DOI: 10.1108/OIR-10-2011-0172 (cit. on pp. 69, 76, 80, 81, 84, 87, 109, 208).

[287] P. Yin, M. Zhang and X. Li. 'Recommending scientific literatures in a collaborative tagging environment'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Afrtificial Intelligence and Lecture Notes in Bioinformatics)* 4822 LNCS (2007), pp. 478–481. ISSN: 03029743. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-38149029112&partnerID=40&md5=85878767c03acc02f3b4d2febbf0ae67 (cit. on pp. 69, 76, 80, 84, 87, 107, 209).

[288] F. Zarrinkalam and M. Kahani. 'A multi-criteria hybrid citation recommendation system based on linked data'. In: 2012,

pp. 283–288. ISBN: 9781467344760. DOI: 10.1109/ICCKE.2012. 6395393 (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 210).

[289]   F. Zarrinkalam and M. Kahani. 'SemCiR: A citation recommendation system based on a novel semantic distance measure'. In: *Program* 47.1 (2013), pp. 92–112. ISSN: 00330337. DOI: 10.1108/00330331311296320 (cit. on pp. 69, 76, 80, 82–84, 87, 106, 208).

[290]   M. Zhang, W. Wang and X. Li. 'A paper recommender for scientific literatures based on semantic concept similarity'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5362 LNCS (2008), pp. 359–362. ISSN: 03029743. DOI: 10.1007/ 978-3-540-89533-6-44 (cit. on pp. 69, 81, 84, 107).

[291]   Ye Zhang et al. 'A Novel Personalized Citation Recommendation Approach Based on GAN'. In: *International Symposium on Methodologies for Intelligent Systems*. Springer. 2018, pp. 268–278 (cit. on pp. 69, 71, 76, 80, 82–85, 87, 106, 208).

[292]   Z. Zhang and L. Li. 'A research paper recommender system based on spreading activation model'. In: Hangzhou, 2010, pp. 928–931. ISBN: 9781424480968. DOI: 10.1109/ICISE.2010. 5689417 (cit. on pp. 69, 84, 108).

[293]   Z.P. Zhang, L.N. Li and H.Y. Yu. 'A hybrid document recommender algorithm based on random walk'. In: *Applied Mechanics and Materials* 336-338 (2013), pp. 2270–2276. ISSN: 16609336. DOI: 10.4028/www.scientific.net/AMM.336-338.2270 (cit. on p. 88).

[294]   W. Zhao et al. 'Research Paper Recommendation Based on the Knowledge Gap'. In: *2015 IEEE International Conference*

on Data Mining Workshop (ICDMW). 2015, pp. 373–380. DOI: `10.1109/ICDMW.2015.40` (cit. on pp. 69, 84, 107, 208).

[295] Weidong Zhao, Ran Wu and Haitao Liu. '"Paper recommendation based on the knowledge gap between a researcher's background knowledge and research target"'. In: *Information Processing & Management* 52.5 (2016), pp. 976 –988. ISSN: 0306-4573. DOI: `https://doi.org/10.1016/j.ipm.2016.04.004`. URL: `http://www.sciencedirect.com/science/article/pii/S030645731630070X` (cit. on pp. 69, 84, 107, 208).

[296] Qianru Zheng, Chi-Kong Chan and Horace H. S. Ip. 'An Unexpectedness-Augmented Utility Model for Making Serendipitous Recommendation'. In: *Advances in Data Mining: Applications and Theoretical Aspects*. Ed. by Petra Perner. Cham: Springer International Publishing, 2015, pp. 216–230. ISBN: 978-3-319-20910-4 (cit. on p. 57).

[297] D. Zhou et al. 'Learning multiple graphs for document recommendations'. In: 2008, pp. 141–150. ISBN: 9781605580852. DOI: `10.1145/1367497.1367517` (cit. on pp. 69, 76, 80, 83, 84, 87, 105, 209).

[298] Quan Zhou, Xiuzhen Chen and Changsong Chen. 'Authoritative scholarly paper recommendation based on paper communities'. In: *2014 IEEE 17th International Conference on Computational Science and Engineering (CSE)*. IEEE. 2014, pp. 1536–1540 (cit. on pp. 69, 76, 80, 83, 84, 87, 106, 210).

[299] Cai-Nicolas Ziegler et al. 'Improving Recommendation Lists through Topic Diversification'. In: *Proceedings of the 14th International Conference on World Wide Web*. WWW '05. New York, NY, USA: Association for Computing Machinery, 2005, pp. 22–

32. ISBN: 1595930469. DOI: 10.1145/1060745.1060754. URL: https://doi-org.libezproxy.open.ac.uk/10.1145/1060745. 1060754 (cit. on pp. 57, 58).