Master's Thesis

Semantic approaches to citation recommendation

Tarek Saier

Examiners: Prof. Dr. Georg Lausen
Prof. Dr. Christian Schindelhauer

Albert-Ludwigs-University Freiburg

Faculty of Engineering

Department of Computer Science

Chair of Databases and Information Systems

April 30th, 2019

Writing Period

15. 10. 2018 – 30. 04. 2019

First Examiner

Prof. Dr. Georg Lausen

Second Examiner

Prof. Dr. Christian Schindelhauer

Supervisor

Dr. Michael Färber

Master-Thesis

Semantic approaches to citation recommendation

Tarek Saier

Gutachter: Prof. Dr. Georg Lausen
Prof. Dr. Christian Schindelhauer

Albert-Ludwigs-Universität Freiburg

Technische Fakultät

Institut für Informatik

Lehrstuhl für Datenbanken und Informationssysteme

30. April 2019

Bearbeitungszeit

15. 10. 2018 – 30. 04. 2019

Erstgutachter

Prof. Dr. Georg Lausen

Zweitgutachter

Prof. Dr. Christian Schindelhauer

Betreuer

Dr. Michael Färber

Declaration

I hereby declare, that I am the sole author and composer of my thesis and that no other sources or learning aids, other than those listed, have been used. Furthermore, I declare that I have acknowledged the work of others by providing detailed references of said work. I hereby also declare, that my Thesis has not been prepared for another examination or assignment, either wholly or excerpts thereof.

Place, Date

Signature

Abstract

foo bar

Zusammenfassung

fu bar

Contents

1. Introduction **1**

1.1. Motivation . . . 1

1.2. Problem setting . . . 2

1.3. Method . . . 3

1.4. Contributions . . . 3

1.5. Document structure . . . 3

2. Related Work **5**

2.1. Semantic approaches to citation recommendation . . . 5

2.2. Local citation recommendation . . . 7

3. Background **11**

4. Data set **13**

4.1. Existing data sets . . . 13

4.2. Data set creation . . . 15

4.2.1. Used data sets . . . 15

4.2.2. Pipeline overview . . . 16

4.2.3. LaTeX Parsing . . . 17

4.2.4. Reference resolution . . . 19

4.2.5. Result format . . . 20

4.3. Evaluation of reference resolution . . . 21

4.4. Statistics and key figures 23
4.4.1. Creation process 23
4.4.2. Resulting data set 23

5. Approach **27**
5.1. Entity based recommendation 28
5.1.1. Fields of study in the MAG 28
5.1.2. Noun phrases 29
5.2. Claim based recommendation 29
5.2.1. Tools for extracting claims 30
5.2.2. A predicate-argument model 30

6. Evaluation **33**
6.1. Special considerations for citation recommendation 33
6.2. Offline evaluation 33
6.3. Online evaluation 34

7. Conclusion **35**

8. Future work **37**

Bibliography **37**

Appendices

A. Evaluation of reference resolution **A3**

List of Figures

1. Four types of numbers. 12

2. Schematic representation of the data set generation process. 17
3. Number of citing documents per cited document. 24
4. Number of citation contexts per reference. 24
5. Citation flow by field of study for 13.3 million reference items. 25

List of Tables

1. Overview of existing data sets. 14
2. Comparison of tools for parsing LaTeX. 18
3. Confidence intervals for a sample size of 300 with 297 positive results. . . . 21
4. Semantic constructs in citation contexts from a range of citation functions used in the field of citation context analysis. 31
5. Evaluation A4

List of Algorithms

1. Stochastic Gradient Descent: Neural Network 12

1. Introduction

1.1. Motivation

Citations are a central building block of scholarly discourse. They are the means by which scholars relate their research to existing work—be it in backing up claims, criticising, naming examples or engaging in any other form. Citing in a meaningful way requires an author to be aware of publications relevant to their work. Here, the ever increasing amount of new reseach publications per year poses a serious challenge. Even with academic search engines like Goolge Scholar and CiteSeerX at our disposal, identifying publications that are worthwhile to examine and appropriate to reference remains a time consuming task.

It is therefore not suprising that methods to aid researchers in these tasks have been and still are being actively researched. While diverse in nature, the common core of these efforts is the goal to utilize the automated processing of publications. This can be achieved by either extracting information from publications as they are [1, 2], or by introducing explicit semantic representations of their content or interrelations to facilitate automated processing [3–5]. Once processed, a typical method is to harvest human made citations, analyze them [6, 7] and use them for example to recommend papers [2] or aid in document exploration [8]. Although systems like this have existed for over 20 years [2, 9], there is not a lot of work looking into the use of explicit semantic representations for the recommendation of papers. This is why this thesis will investiage their application. More specifically, we will concentrate on the task of recommending papers for citation—as opposed to, for example, discovery. What this encompasses will be described in more detail in the following section.

1.2. Problem setting

In the boradest sense, recommending papers for citation means given an input text, suggest publications that can be referred to from within that text. In scale this can vary from specific recommendations for a section of a sentence (*local* or *context-based*), to general recommendations for a whole input document (*global*). The task can also include deciding whether or not the input contains parts that would justify inserting a citation in the first place [10]. In this thesis, we will focus on local citation recommendation with the assumption that the input always allows for/requires a citation to be put in.

Another distinction to be made is between personalized and general citation recommendation. Some approaches make use of user specific information such as an author's prior citations. Collaborative filtering approaches by nature include a user model and therfore fall into this category. While personalization can improve recommendation, it limits the approach to users that are willing to share personal information. We therefore limit ourselves to purely content based filtering approaches.

A last clarification has to be made concerning the term *explicit semantic representations*. This is to be understood as a differentiation from the mere use of unstructured text. A most prominent example for explicit semantic representations would be the structure of the Semantic Web [11]. In the context of citation recommendation as briefly outlined above this means representing citations in a semantically meaningful way as opposed to just relying on syntactical information like n-grams or bag-of-words representations.

The problem setting can be summarized as follows. To investigate is, the applicability of and requirements for the use of explicit semantic representations for content based, local citation recommendation. The following section will outline how this investigation is performed.

1.3. Method

In order to assess if and how explicit semantic representations can benefit citation recommendation we investigate the use of named entities as well as claim structures. For the evaluation of our models in a realistic setting we generate a large data set that allows for the extraction of precise citation marker positions. To ensure comparability with other approaches we also perform evaluations on existing data sets as far as possible.

Extend to mention offline and online eval

Extend moar

1.4. Contributions

The data set

Entity and claim based models

Insights into open problems with building claim models around citations (b/c of non-integral citation styles)

1.5. Document structure

foo bar

Copypasta of useful stuff below.

- Put a tilde (nbsp) in front of citations [12].
- **(TODO: Do this!)**
- **(EXTEND: Write more when new results are out!)**
- **(DRAFT: Hacky text!)**
- Chapter 1
- the colors of the Uni
  - UniBlue
  - UniRed
  - UniGrey
- a command for naming matrices $\mathbf{G}$, and naming vectors $\mathbf{a}$. This overwrites the default behavior of having an arrow over vectors, sticking to the naming conventions normal font for scalars, bold-lowercase for vectors, and bold-uppercase for matrices.
- named equations:

$$d(a, b) = d(b, a) \tag{1}$$

symmetry

- Use “these” for citing, not "these"
- If an equation is at the end of a sentence, add a full stop. If it's not the end, add a comma: $a = b + c \quad (1)$,
- `https://en.wikipedia.org`
- Do not overuse footnotes[1] if possible.

[1] `https://en.wikipedia.org`

2. Related Work

To the best of our knowledge there is, so far, almost no work investigating (1) the use of explicit senamtic representations for (2) the task of local citation recommendation. We will therefore present related work in two areas. First, semantic approaches to citation/paper recommendation in general (global as well as local). Second, local citation recommendation regardless of the specifics of the approach taken.

Note that SemCir [13] (see below) is the only case of a semantic approach to local citation recommendation we are aware of. The explicit semantic representations are, however, not generated from citation contexts (local) but from papers (global) that are textually (not necessarily semantically) similar to the citation contexts.

2.1. Semantic approaches to citation recommendation

At a point in time where publishing research papers online was an emerging trend, Middleton et al. [14] propose a system for paper recommendation making use of a topic ontology. Based on classifying papers into topics and recording which papers a researcher would access on the web, they employ content-based filtering, collaborative filtering and a feedback mechanism to suggest papers from new topics to users. Comparing the topic ontology to a flat list of topics in two user studies, they report 7–15% more user satisfaction for the ontology case.

In a similar vein, Zhang et al. [15] propose a hybrid recommender system for papers based on semantic concept similarity. They derive concepts from CiteULike[1] tags and use these to measure the semantic similarity of papers and users' interest. In their evaluation they compare different settings of the approach but do not compare to other work or alternative techniques.

Jiang et al. [16] use CiteULike tags as academic concepts to build a topic model applied to paper abstracts. In a content-based recommendation setting they let volunteers judge the relevance of recommendations for a test set of 30 papers. The evaluation includes a TFIDF baseline, latent Dirichlet allocation (LDA) and an approach combining LDA with their concept model. The reported MAP@5 and NDCG@5 values are best for the LDA+concept method.

In [17] Zarrinkalam et al. enrich their metadata on research papers using multiple Linked Open Data (LOD) sources to drive a hybrid recommender system. They compare a purely content-based method using only text similarity with a second method additionally utilizing collaborative filtering and a third method furthermore using the LOD enriched data. They report recall, co-cited probability and NDCG values for various cut-off values for which the LOD enriched method consistently achieves the best performance.

With SemCiR [13] Zarrinkalam et al. introduce a content-based, global citation recommendation approach that utilizes a semantic distance measure between papers. They furthermore introcude a method for extending the measure to determine the semantic distance between an input text and a paper, which is achieved by representing the input by textually similar papers. The distance measure suggested builds on six different relational features including shared authors, venue, and overlapping in- and outgoing citations. The approach is evaluated on a 12,500 paper subset of CiteSeerX [18] in a citation re-prediction setting, using as input a paper's title, abstract and contexts in other papers where it was referred to. An evaluation of different scenarios measuring recall, co-cited probability and NDCG leads the authors to conclude that recommendation results can be improved by

[1] See http://citeulike.org/.

using their semantic distance measure and including citation contexts in the measurement textual similarity.

2.2. Local citation recommendation

Probably one of the first investigations into local citation recommendation is the work of He et al. [19]. They propose a two-step system that first identifies recommendation candidates and then re-ranks them by concept similarity. While also discussing global citation recommendation in detail, for the local case they compare recommending for a single context and recommending for all contexts within a document simultaneously. In an evaluation on the CiteSeerX data set measured by recall, co-cited probability and NDCG they find that the single context task is harder, but also, that their approach to the all contexts task achieves results comparable to and even better than some global citation recommendation methods.

In a follow-up work Huang et al. [20] build upon above work by swapping out the computationally complex concept based re-ranking method with a translational model. In this model citation contexts are treated as the source language and cited papers as words in the target language. The resulting system, RefSeer, is evaluated on two smaller data sets (CiteULike and a CiteSeer subset) and one large one (all of CiteSeer). The authors report precision, recall, Bpref and MRR values for the two smaller data sets and conclude that their system can give correct recommendations in a realistic setting—such as when only the top 10 recommendations are shown.

Huang et al. improve RefSeer with a neural probabilistic model that learns distributed representations of words and documents in [21]. They evaluate their model for local citation recommendation on the whole of CiteSeer, splitting between train and test set at the year 2011 (9M contexts train, 1.5M contexts test). Measuring MAP, MRR and NDCG they show that their model outperforms 4 different state-of-the-art approaches. An analysis on the influence of papers' citation counts on recommendation performance shows that their approach especially exceeds other work in case of lesser cited papers (<100 citations).

In [22] Duma et al. test the effectiveness of using a variety of document internal and external text inputs to a TF-IDF model. Their data set is built from the ACL Anthology and contains 5446 citations. In a re-prediction setting the authors measure how reliably their models rank the correct paper at the top position. They conclude that a mixture of internal and external inputs outperforms either of the aforementioned used on their own.

The work presented in [23] by Duma et al. focusses on the rethorical function of sentences. The authors classify sentences using the Core Scientific Concepts (CoreSC) scheme and investigate how their distinction can be used to improve recommendation. Evaluating on one million papers from the PubMed Central Open Access Subset they measure NDCG values and find that for several classes of input sentences significant gains in recommendation quality can be made by focussing on certain rethoric passages of candidate documents when ranking text similarity.

The Neural Citation Network (NCN) proposed by Ebesu et al. in [24] is inspired by neural machine translation, learns citation context representations as well as auhtor representations and includes an attention mechanism. For their evaluation the authors use 4.5 million citation contexts from the RefSeer data set and report NDCG, MAP, MRR and recall values. They compare against a BM25 baseline, a citation translation model as well as two variations of their model that do not make use of author representations. While BM25 is outperformed by all of the other approaches to some degree the NCN's results lead the evaluation by a distinct margin.

Kobayashi et al. [25] describe a variation of local citation recommendation they call *context-based co-citation recommendation*. The input here is a citation context *plus* one publication referred to in that contexts. The goal then is to recommend other publications that also can be used as citations in that contexts. By classifying text sections into the discourse facets "objective", "method", and "result" the authors are able to train distributed vector representations per facet which are then used for the recommendation. They evaluate their approach on contexts from the ACL Anthology containing "enumerated co-citations" (e.g. *[27,42]*) and report NDCG values at a cut-off of 100. In comparison with two baseline methods their discourse facets are shown to be effective.

In [26] Jeong et al. introduce an appraoch to local citation recommendation using Graph Convolutional Networks (GCN) and Bidirectional Encoder Representations from Transformers (BERT). The GCN is used to capture information from the citation relationships between papers, while the pre-trained BERT is applied on the citation contexts themselves. The authors evaluate their approach on a subset of 6500 papers from the ACL Anthology and a self-created data set of close to 5000 papers. They report MAP, MRR and recall values at different cut-offs demostrating that their BERT+CGN approach outperforms several reduced versions of the aforementioned as well as a baseline model.

3. Background

explain all the things.

global/local citation (for local especially explain harvesting citation contexts and comparing aggregrates to input)

citation marker

four types of numbers (citing/cited documents, reference items citation context)

reference resolution

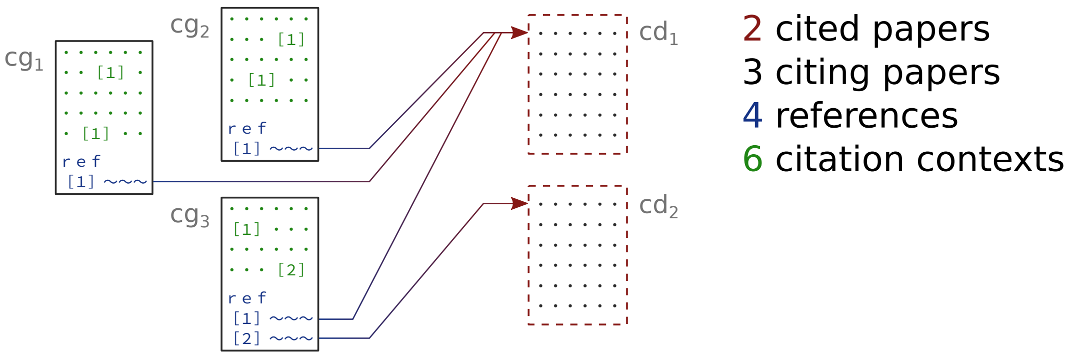

Figure 1.: Four types of numbers. A toy example with citation pairs $cg_1 \longrightarrow cd_1$, $cg_2 \longrightarrow cd_1$, $cg_3 \longrightarrow cd_1$, $cg_3 \longrightarrow cd_2$ resulting in 2 cited papers, 3 citing papers, 4 references and 6 citation contexts.

Algorithm 1 Stochastic Gradient Descent: Neural Network

Create a mini batch of m samples $\mathbf{x}_0 \ldots \mathbf{x}_{m-1}$
foreach sample $\mathbf{x}$ **do**
  $\mathrm{a}^{\mathrm{x},0} \longleftarrow \mathbf{x}$ ▹ Set input activation
  foreach Layer $l \in \{1 \ldots L-1\}$ **do** ▹ Forward pass
    $\mathrm{z}^{\mathrm{x},l} \longleftarrow \mathbf{W}^l \mathrm{a}^{\mathrm{x},l-1} + \mathbf{b}^l$
    $\mathrm{a}^{\mathrm{x},l} \longleftarrow \varphi(\mathrm{z}^{\mathrm{x},l})$
  end for
  $\delta^{\mathrm{x},L} \longleftarrow \nabla_{\mathrm{a}} C_{\mathrm{x}} \odot \varphi'(\mathrm{z}^{\mathrm{x},L})$ ▹ Compute error
  foreach Layer $l \in L-1, L-2 \ldots 2$ **do** ▹ Backpropagate error
    $\delta^{\mathrm{x},l} \longleftarrow ((\mathbf{W}^{l+1})^T \delta^{\mathrm{x},l+1}) \odot \varphi'(\mathrm{z}^{\mathrm{x},l})$
  end for
end for
foreach $l \in L, L-1 \ldots 2$ **do** ▹ Gradient descent
  $\mathbf{W}^l \longleftarrow \mathbf{W}^l - \frac{\eta}{m} \sum_{\mathrm{x}} \delta^{\mathrm{x},l} (\mathrm{a}^{\mathrm{x},l-1})^T$
  $\mathbf{b}^l \longleftarrow \mathbf{b}^l - \frac{\eta}{m} \sum_{\mathrm{x}} \delta^{\mathrm{x},l}$
end for

---

4. Data set

Recommender systems rely on data for their development, training and evaluation. It is therefore important to properly assess potential data sets in terms of their strengths and shortcomings—especially with regards to the task at hand. In citation recommendation, the goal is to identify papers relevant to a user input. Because of the large amount of available research, this means a recommender has to be able to find relevant publications in a large set of possible candidates in order to be considered fit for the task. As a comsequence, evaluation results are likely to be more meaningful when a large data set is used. Apart from the size, the quality of data is also crucial. For local citation recommendation this means that a clean citation context, precise position of citaiton markers and valid citation information are desirable. With these criteria in mind we assessed existing data sets and come to the conclusion that, for the relatively new task of local citation recommendation, the creation of a dedicated data set will bring considerable benefits.

The following sections describe the details of our assessment as well as the creation process and evaluation of our new data set.

4.1. Existing data sets

Table 1 gives an overview of relevant existing data sets. While various recommendation domains have established quasi standard data sets, this is not yet the case in citation

[1] `https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/`

Table 1.: Overview of existing data sets.
Listed are the number of papers, nature of citation contexts, covered disciplices of citing papers and the type of global reference identifiers.
(*extractable** is to indicate that extraction might be error-prone due to papers only being available as in PDF format.)

Data set	#Papers	Cit. context	Disciplines	Ref. IDs
CiteSeerX [18] / RefSeer [20]	5M	400 chars	(unrestricted)	internal
PubMed Central OAS[1]	2.3M	extractable	Biomed./Life Sci.	mixed
arXiv CS [27]	90K	1 sentence	CS	DBLP
Scholarly [28]	100K	extractable*	CS	-
ACL-AAN [29]	18K	extractable*	CS/Comp. Ling.	-
ACL-ARC [30]	11K	extractable*	CS/Comp. Ling.	-

recommendation. CiteSeerX is currently the most used in the field [2]. It is comparatively large, but many approaches only use subsets and generate them with varying filtering criteria. It includes pre-extracted citation contexts of 400 characters in length, whereby references are resolved to an internal set of identifiers. Unfortunately there are several quality issues with the data set. The main ones being inaccurate citaion information, noisy citation contexts and cut off words at the borders of citation contexts [31].

The PubMed Central Open Access Subset (PMC-OAS) is another large data set that has been used for citation based tasks [23, 32–34]. Contained publications are already processed and available in XML format. While the data set overall is comparatively clean, heterogeneous annotation of citations within the text and mixed usage of global reference identifiers (PubMed, MEDLINE, DOI, ...) make it difficult to retrieve high quality citation interlinkings of documents from the data set[2] [32].

Consistent global reference identifiers are given in the arXiv CS data set in the form of DBLP IDs. Linking to an existing repository of publication (meta) data has the advantage that

[2]To be more precise, the heterogeneity makes the usage of the data set *as is* problematic. Resolving references retrospectively would be an option but comparatively challenging in the case of PubMed because of the frequent usage of special notation in publication titles; see also: `http://www.sciplore.org/files/citrec/CITREC_Parser_Documentation.pdf`

information about cited papers in readily available. The choice of DBLP restricts resolved references to the field of computer science though. Citations to, for example, publications in maths or statistics can not be resolved to a DBLP ID. A strength of the data set is that it was generated from LaTeX source files, which makes it possible to get very clean data.

For the remainig the data sets—Scholarly, ACL-AAN and ACL-ARC—citing papers are only available in PDF format and references are not resolved. The two ACL sets have the additional drawback of being very small.

Above observations lead us to the conclusion that it would be worthwhile to tackle the creation of a data set that is large (in the order of CiteSeerX/RefSeer/PMC-OAS), clean (like the PMC-OAS and arXiv CS) and also offers consitent global reference IDs that don't restrict the data set to citations within the same discipline. The creation and evaluation of this data set is described in the following sections.

4.2. Data set creation

Scientific publications are usually distributed in formats targeted at human consumption (e.g. PDF) or, in cases like arXiv.org, also as source files *for* the aforementioned (e.g. LaTeX sources for generating PDFs). Citation-based tasks, such as local citation recommendation, in contrast, require automated processing of the publications' textual contents as well as the documents' interlinking through citations. The creation of a data set for such tasks therefore encompasses two main steps: extraction of plain text and resolution of references. In the following we will describe how we approached these two steps using arXiv publication sources and the Microsoft Academic Graph (MAG) [35].

4.2.1. Used data sets

The following two resources are the basis of the data set creation process.

arXiv.org hosts over 1.4 million submissions from August 1991 onward [36]. They are available not only as PDF, but (in most cases) also as LaTeX source files. The discipline most prominently represented is physics, followed by mathematics, with computer science seeing a continued increase in percentage of submissions ranking third (see Fig. 5). The availability of LaTeX sources makes arXiv submissions particularly well suited for extracting high quality plain text and accurate citation information. So much so, that it has been used to generate ground truths for the evaluation of PDF to text conversion tools [37]. Approaches to automatically extract citation interlinks from arXiv sources by parsing LaTeX files have existed for over 20 years [38]. Nevertheless we are not aware of any existing data sets for citation based tasks generated from the whole of arXiv.

Microsoft Academic Graph (MAG) is a very large[3], automatically generated data set on publications, related entities (authors, venues, etc.) and their interconnections through citation. While citation contexts are available to some degree, full text documents are not. The size of the MAG makes it a good target for matching reference items against it, especially given that arXiv spans several fields of study.

4.2.2. Pipeline overview

As depicted in Figure 2, we start out with arXiv sources to create the data set. From these we generate, per publication, a plain text file with the document's textual contents and a set of database entries reflecting the document's reference section. Association between reference items and citations in the text are preserved by placing citation markers in the text. In a second step, we then iterate through all reference items in the database and match them against paper metadata records in the MAG. The result of this process are MAG paper records associated with one or more reference items, that in turn are associated with citation contexts in the plain text files. In other words, we end up with cited documents

[3]At the time of writing the MAG contains data on over 200 million publications.

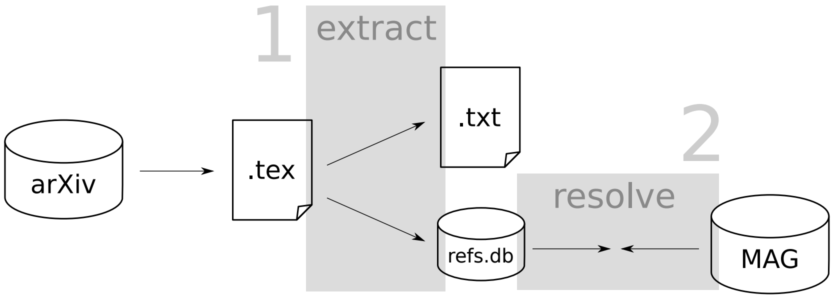

Figure 2.: Schematic representation of the data set generation process.

described by their MAG metadata and a distributed description of the document, consisting of citation contexts over many citing documents.

4.2.3. LaTeX Parsing

In the following we will describe the tools considered for parsing LaTeX, the challenges we faced in general and with regard to arXiv sources in particular, and our resulting approach.

Tools

We took several tools for the conversion from LaTeX to plain text or to intermediate formats into consideration and evaluated them. Table 2 gives an overview of our results. Half of the tools failed to produce any output for a large amount of arXiv submissions we used as test input and were therefore deemed not robust enough. *GrabCite* is able to parse 78.5% of arXiv CS submissions but integrates resolving references against DBLP into the parsing process and therefore would require significant modification to fit our system architecture. *LaTeXML* and *Tralics* are both robust and can be used as LaTeX conversion

[4] `https://github.com/tiarno/plastex`
[5] `https://github.com/alvinwan/texsoup`
[6] `https://github.com/pkubowicz/opendetex`
[7] `https://www.freebsd.org/cgi/man.cgi?query=detex`
[8] `https://github.com/brucemiller/LaTeXML`
[9] `https://www-sop.inria.fr/marelle/tralics/`

Table 2.: Comparison of tools for parsing LaTeX.

Tool	Output	Robust	Usable w/o modification
plastex[4]	DOM	no	yes
TexSoup[5]	document tree	no	yes
opendetex[6]/detex[7]	plain text	no	yes
GrabCite [27]	plain text + resolved ref.	yes	no
LaTeXML[8]	XML	yes	yes
Tralics[9]	XML	yes	yes

tools as is. On subsequent tests we note that *LaTeXML* needs on average 7.7 seconds (3.3 if formula environments are heuristically removed beforehand) to parse an arXiv submission while *Tralics* needs 0.09. Because the quality of their output seemed comparable we chose to use *Tralics*.

Challenges

Apart from the general difficulty of parsing LaTeX due to its feature richness and people's free-spirited use of it, we especially note difficulty in dealing with extra packages not included in submissions' sources[10]. While *Tralics* is supposed to for example deal with *natbib* citations[11], normalization of such citations lead to a decrease of citation markers not being able to be matched to their reference item from 30% to 5% in a sample of 565,613 citations we tested.

Resulting approach

Our LaTeX parsing solution consists of two steps. First, we flatten each arXiv submission's sources to a single LaTeX file using *latexpand*[12,13] and normalize `\cite` commands to prevent

[10] arXiv.org specifically suggest the omission of such (see `https://arxiv.org/help/submit_tex#wegotem`)

[11] See `https://www-sop.inria.fr/marelle/tralics/packages.html#natbib`

[12] See `https://ctan.org/pkg/latexpand`

[13] We also tested flatex (`https://ctan.org/pkg/flatex`) and flap (`https://github.com/fchauvel/flap`) but got the best results with `latexpand`.

parsing problems later on. In the second step, we then generate an XML representation of the LaTeX document using *Tralics*, replace formulas, figures, tables and non citation references with replacement tokens and extract the plain text. Furthermore, each reference item is assigned a unique identifier, its text is stored in a database and corresponding citation markers are placed in the plain text.

4.2.4. Reference resolution

Resolving references to globally consistent identifiers (e.g. detecting that the references (1), (2), and (3) in Listing 4.1 all reference the same document) is a challenging and still unsolved task [1]. Given it is, by itself, the most distinctive part of a publication, we base our reference resolution on the title of the cited work and use other pieces of information (e.g., authors' names) only in secondary steps. In the following we will describe the challenges we faced, matching arXiv.org submissions' reference items against MAG paper records and how we approached the task.

Listing 4.1: Examples of reference items.

```
(1) V. N. Senoguz and Q. Shafi, arXiv:hep-ph/0412102
(2) V.N. Senoguz and Q. Shafi, Phys. Rev. D 71 (2005) 043514.
(3) V. N. Şenoğuz and Q. Shafi, ‘‘Reheat temperature in supersymmetric
     hybrid inflation models,’’ Phys. Rev. D 71, 043514 (2005) [hep-ph
    /0412102].
(4) V.Sauli, JHEP 02, 001 (2003).
(5) Aaij, Roel, et al. "Search for the $B^{0}_{s} \to \eta^{\prime}\
    phi$ decay" Journal of High Energy Physics 2017.5 (2017): 158.
(6) According to the numerous discussions with my colleagues <removed>
     and <removed> an experimental verification of our theoretical
    predictions is feasible.
```

Challenges

Reference resolution can be challenging when reference items contain only minimal amounts of information, when formulas are used in titles or when they refer to non publications (e.g., Listing 4.1 (4)-(6)). A further problem we encountered was noise in the

MAG. Our mechanism matched 13,041 reference items like `K. Kondo, hep-th/0303251.` and `T. Heinzl, hep-th/9812190.` to a MAG paper with the title *“hep-th.”* with one of the author’s names being *“He”* (paper ID 2811252340).

Resulting approach

Our reference resolution procedure can be broken down in two steps: title identification and matching. If possible, title identification is done by arXiv ID or DOI (where we retrieve the title from an arXiv metadata dump or via crossref.org[14]); otherwise we use Neural ParsCit [39]. The identified title is matched against the normalized titles of all publications in the MAG. Resulting candidates are considered, if at least one of the author’s names is present in the reference string. If multiple candidates remain, we judge by the citation count given in the MAG. The last step particularly helped mitigate rouge almost-duplicate entries in the MAG that often have few to no citations.

4.2.5. Result format

Listing 4.2 shows some example content from the data set. While the data set in and of itself consists of plain text files and a references database, citation contexts of successfully resolved references are straightforward to extract and use as input for a recommender. The bottom part of Listing 4.2 shows an example of a 3 sentence context with `cited doc MAG ID`, `MAG IDs of adjacent citations`, `citing doc arXiv ID` and `text` in a CSV format.

Listing 4.2: Excerpts from (top to bottom) a plain text file, corresponding data base entries in the references DB, entries in the MAG and extracted citation context CSV.

```
It has over 79 million images stored at the resolution of FORMULA .
   Each image is labeled with one of the 75,062 non-abstract nouns in
    English, as listed in the Wordnet{{cite:9ad20b7d-87d1-47f5-aeed
   -10a1cf89a2e2}}{{cite: 298db7f5-9ebb-4e98-9ecf-0bdda28a42cb}}
   lexical database.

[uuid]          [in_doc]   [mag_id]     [reference_string]
9ad20b7d-87d1   1412.3684  2081580037   George A. Miller (1995). WordNe
```

[14] See https://www.crossref.org/

```
-47f5-aeed-..                                t: A Lexical Database for Eng..
298db7f5-9ebb  1412.3684  2038721957  Christiane Fellbaum (1998), ""W
-4e98-9ecf-..                                ordNet: An Electronic Lexical..

[paperid]    [originaltitle]                           [publisher]  ...
2038721957   WordNet : an electronic lexical database  MIT Press    ...
2081580037   WordNet: a lexical database for English   ACM          ...

2038721957|2081580037|1412.3684|It has over 79 million images stored
    at the resolution of FORMULA . Each image is labeled with one of
    the 75,062 non-abstract nouns in English, as listed in the Wordnet
     CIT MAINCIT lexical database. It has been noted that many of the
    labels are not reliable CIT .
```

4.3. Evaluation of reference resolution

To evaluate the quality of our reference resolution results, we take a random sample of 300 matched reference items and manually check if the correct record in the MAG was identified by our method. Given the 300 items, we obtained 3 errors, giving us an accuracy estimate of 96% at the worst, as shown in Table 3.

Table 3.: Confidence intervals for a sample size of 300 with 297 positive results as given by Wilson score interval and Jeffreys interval [40].

Confidence level	Method	Lower limit	Upper limit
0.99	Wilson	0.9613	0.9975
	Jeffreys	0.9666	0.9983
0.95	Wilson	0.9710	0.9966
	Jeffreys	0.9736	0.9972

The three incorrectly identified references were as follows (MAG IDs in square brackets):

1. *"Eddy, J.A.: 1983, The maunder minimum - a reappraisal. Solar Phys. 89, 195. ADS."*
    - matched: [2024069573]
      "The Maunder Minimum" (John A. Eddy; 1976)

    - correct: [2080336740]

      "The Maunder Minimum: A reappraisal" (John A. Eddy; 1983)

2. *"J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani. 1-norm support vector machines. In Advances in Neural Information Processing Systems (NIPS), volume 16, pages 49–56, 2004."*
    - matched: [2249237221]

      "Support Vector Machines" (Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani; 2013)
    - correct: [2130698119]

      "1-norm Support Vector Machines" (Ji Zhu, Saharon Rosset, Robert Tibshirani, Trevor J. Hastie; 2003)
3. *"D. T. Limmer and D. Chandler. The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water. The Journal of Chemical Physics, 135(13):134503, 2011."*
    - matched: [2599889364]

      "The Putative Liquid-Liquid Transition is a Liquid-Solid Transition in Atomistic Models of Water" (David Chandler, David Limmer; 2013)
    - correct: [1977410206]

      "The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water. II" (David T. Limmer, David Chandler; 2011)

In all three cases the misidentified document's title is contained in the correct document's title, and there is a large or complete author overlap between correct and actual match. This shows that authors sometimes title follow-up work very similarly, which leads to hard to distinguish cases. Another observation that can be made, is that longer titles are more distincitive. As certain publication titles might be sub-strings of of other publication's title, a matching mechanic should always try to first match a long title before trying shorter candidates. The full details of the evaluation can be found in Appendix A.

4.4. Statistics and key figures

4.4.1. Creation process

We used an arXiv source dump containing all submissions up until the end of 2017 (1,340,770 documents). 100,240 of these were only available in PDF format, leaving 1,240,530 sources. Our pipeline output 1,151,707 (92.8%) plain text files, 1,018,976 (82.1%) of which contained citation markers. The number of reference items identified is 35,053,329, for which 56,077,906 citation markers were placed within the plain text files. This first part of the process took 59 hours to run, unparallalized on a 8 core Intel Core i7-7700 3.60GHz machine with 60 GB of memory.

Of the 35,053,329 reference items, we were able to match 14,046,239 (40.07%). For 33.14% of the reference items we could neither find an arXiv ID or DOI, nor was Neural ParsCit able to identify a title. For the remaining 26.79% a title was identified but could not be matched with the MAG. Of the matched 14 million items' titles, 50.67% were identified via Neural ParsCit, 29.67% by DOI and 19.66% by arXiv ID. Of the identified DOIs 26.8% were found as is while 73.2% were heuristically determined[15]. The matching process took 103 hours, run in 10 parallel processes on a 64 core Intel Xeon Gold 6130 2.10GHz machine with 500 GB of memory.

4.4.2. Resulting data set

The resulting data set consists of *2,343,585 cited papers, 926.644 citing papers, 13,303,373 references and 24,558,560 citation contexts.* Note that references with no citation markers (due to parsing errors) are not counted here.

Figure 3 shows the number of citing documents for all cited documents. There is one document with close to 10,000 citations, another three with more than 5,000 and another

[15] This was possible because the DOIs of articles in journals of the American Physical Society follow predictable patterns.

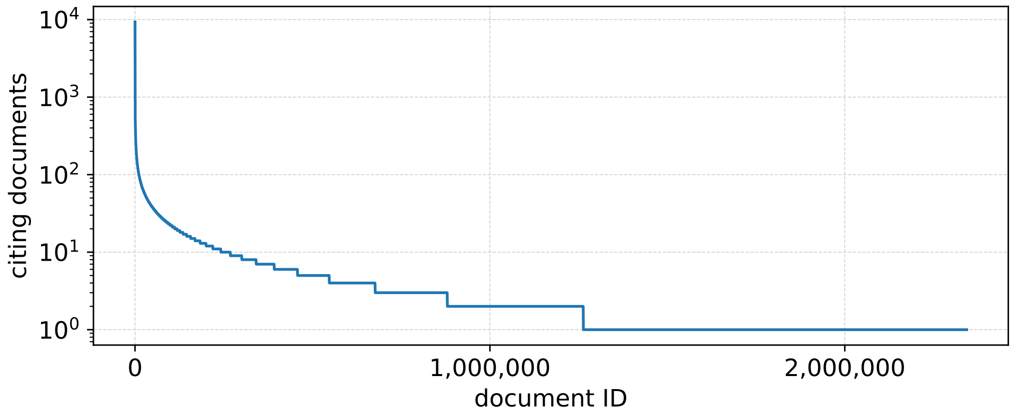

Figure 3.: Number of citing documents per cited document.

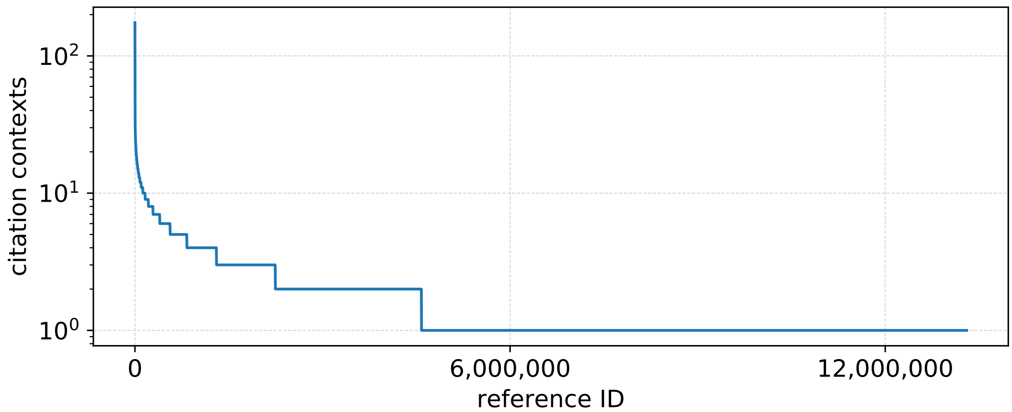

Figure 4.: Number of citation contexts per reference.

ten with more than 3,000. 1,262,861 (53.89%) of the documents have at least two citations, 547,036 (23.34%) have at least five. The mean number of citations is 5.68 (SD 26.82). Figure 4 shows the number of citation context per reference. 8,722,795 (65.57%) references have only one citation context, the maximum is 278, the mean 1.85 (SD 2.02). This means a cited document is described by on average $1.85 \times 5.68 \approx 10.5$ citation contexts.

Figure 5 depicts the flow of citations by field of study for all 13.3 million matched reference items. Fields of study with very small numbers of references are combined to *other* for legibility reasons. For the citing document's side, these are economics, electrical engineering

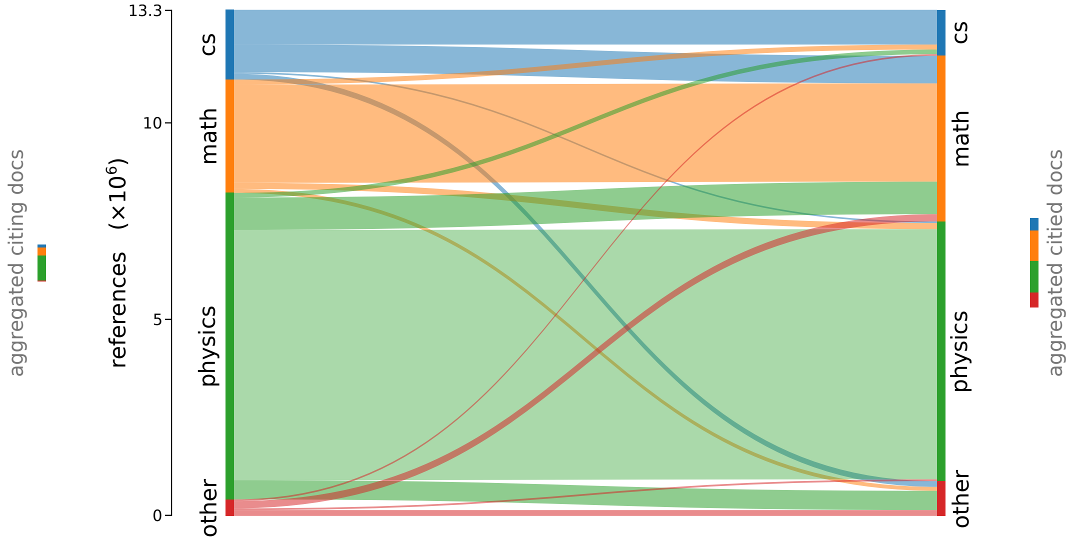

Figure 5.: Citation flow by field of study for 13.3 million reference items. For reference, the number of citing and cited documents per field of study are plotted on the sides.

and systems science, quantitative biology, quantitative finance and statistics. Combined on the cited document's side are chemistry, biology, engineering, materials science, economics, geology, psychology, medicine, business, geography, sociology, political science, philosophy, environmental science and art. To no surprise, publications in each field are cited the most from within the field itself. Notable is, however, that the incoming citations in mathematics are the most varied (physics and computer science combined make up 38% of the citations).

By generating our data set from LaTeX sources we were able to ensure clean text output as well as accurate citation information and exact citation marker positions. In terms of size it is closer to CiteSeerX and PMC-OAS than the smaller data sets available. The fact that the data set spans multiple disciplines also allows for comparisons in citing behaviour between these disciplices. Because references are already resolved to MAG IDs, the data set is readily usable for recommendation tasks and allows for the use of rich metadata on both the citing and cited document side. Lastly, the embedding of citation markers in the full plain text of papers instead of pre-extraction enables users of the data set to choose and compare citation context lenghts and variations at will.

5. Approach

In order to investigate the use of explicit semantic representations for the task of local citation recommendation we first need to decide which kinds of semantic constructs we want to model. As a starting point for this we looked into the field of citation context analysis [41]. A common task in this area is the classification of citation contexts by their polarity (positive/neutral/negative) and function (often based on the four dimensions identified by Moravcsik et al. [12], conceptual/operational, evolutionary/juxtapositional, organic/perfunctory, conformative/negational). Such approaches are primarily concerned with the *intent* of the author rather than the *content* of what is being cited. We can therefore not expect to derive types of semantic constructs directly from citation functions. Starting from an established typology of citation functions will, however, ensure that we consider a wide range of different citations rather than cherry picking those that fit our preconceptions.

Table 4 lists categories of citation functions along with the kinds of semantic constructs that can be found in such citation contexts. The list of citation functions is taken from [45] (and therein built upon [44]). This study was selected because it gives an overview of previous attempts to classify citations, presents their new typology with extensive explanation as well as example contexts, and does not mix polarity into its function categories. Examining contexts from each of the eight functions we identify two types of semantic constructs: named entities (NE) and claims (or statements). The rationale behind these two is as follows. Named entities can identify reference publications for a certain data set/tool/concept (see *Attribution*, *Statement of use* and *Application* in Table 4) as well as a method/field of study common to a selection of publications (see *Exemplification* in Table 4). Claims can identify publications that can be cited to back or support the very claim contained in a citation context. Note that the example contexts listed to have no construct ("-" in the *Construct* column) may contain named entities and claims as well (e.g. "DBLP" or "Lamers et al. base their definition of the author's name"), but these are (in the case of NEs) not representative of the cited work or (in the case of claims) just statements *about* a publication rather than statements being backed by the cited work. A third semantic construct that can be considered, but would require considering a larger citation context, is argumentative structures. To keep the scope of this thesis at a reasonable level we will, however, limit our investigation to named entities and claims.

The following sections will describe our investigation of entity based and claim based models for local citation recommendation.

5.1. Entity based recommendation

The intuition behind an entity based approach is, that there exists a reference publication for a named entity. Examples would be a data set ("CiteSeerX [18]"), a tool ("Neural ParsCit [39]") or a concept ("Semantic Web [11]"). In a more loose sense this can also include publications being referred to as examples ("approaches to local citation recommendation [19–26]"). For the identification of such named entities we take two approaches. A more strict one based on the fields of study given in the MAG, and a more loose one based on noun phrases.

5.1.1. Fields of study in the MAG

Along with papers, authors, venues etc. the MAG data schema also includes fields of study (FoS) that are associated with papers and interlinked in a child-parent manner. At the time of writing there are 229,716 FoS at 6 levels of granularity. They range from the most coarse level 0 (example entries being *mathematics*, *sociology*, *computer science*) to more and more fine grained entities (*theoretical computer science*(1)⟶*encryption*(2)⟶*on-the-fly encryption*(3)⟶*filesystem-level encryption*(4)⟶*FileVault*(5)). The levels of granularity don't seem to follow a globally consistent pattern though. *WordNet*, for example, being a particular piece of data in the same way *FileVault* is a particular piece of software, is not at level 5 but level 2 (*computer science*(0)⟶*artificial intelligence*(1)⟶*WordNet*(2)). Another noteworthy aspect is that a FoS can have multiple parents (*FileVault* for example also has the parent *multiple encryption*).

annotated whole data set

experimented a lot

did a small manual evaluation and realized that seemingly normal terms can be specialized vocab in a particular field (e.g. motivation in psychology)

experimented more

tried directly in front of the citation marker

use jaccard similarity btw.

ended up with bow+superminifosboost model that does consistently improve eval score but not much

5.1.2. Noun phrases

5.2. Claim based recommendation

how citations are embedded in sentences (integral/non-integral [43, 44, 46–48])

actually look into automatic classification [49, 50]

5.2.1. Tools for extracting claims

tools tools

also: Survey on open information extraction [42]

context specific claim detection [51]

if only papers where semantically annotated as proposed in [3]

5.2.2. A predicate-argument model

predpatt [52, 53]

unfeasibility of use of PredPatt output as is

loosened predicate:parameter model

predicates could be grouped/clustered to represent functions as in [54]

alternative view: model gives a selective citation context derived from claim structure (cf. concept of reference scope as sub part of citation context sentence [50, 55]

Table 4.: Semantic constructs in citation contexts from a range of citation functions used in the field of citation context analysis.

Function	Construct	Examples (semantic construct *highlighted*)
Attribution	claim	“Berners-Lee et al. [11] argue that *structured collections of information and sets of inference rules are prerequesites for the semantic web to function*.”
	NE	“A variation of this task is ‘*context-based co-citation recommendation*’ [25].”
	-	“In [22] Duma et al. test the effectiveness of using a variety of document internal and external text inputs to a TF-IDF model.”
Exemplification	NE	“We looked into approaches to *local citation recommendation* such as [19–26] for our investigation.”
Further reference	-	“See [42] for a comprehensive overview.”
Statement of use	NE	“We use *CiteSeerX* [18] for our evaluation.”
Application	NE	“Using this mechanism we perform ‘*context-based co-citation recommendation*’ [25].”
Evaluation	-	“The use of DBLP in [27] restricts their data set to the field of computer science.”
Establishing links between sources	claim	“A common motivation brought forward for research on citation recommendation is that *finding proper citations is a time consuming task* [10, 19, 24, 25].”
	-	“Lamers et al. [43] base their definition on the author’s name whereas Thompson [44] focusses on the grammatical role of the citation marker.”
Comparison of own work with sources	claim	“Like [27] we find that, albeit written in a structured language, *parsing LaTeX sources is a non trivial task*.”

6. Evaluation

evaluate evaluate

implemetation pain and bad evaluation scores [56]

6.1. Special considerations for citation recommendation

train/test splitting (per cited doc, temporal, ...), re-recommendation, number of contexts describing a recommendation item, ...

a cited doc's role (how it is cited) can develop over time [57, 58]

relevance of time [59]

candidates are only citations within current paper [22]

6.2. Offline evaluation

pre-filtering experiments (knn [34], lsi, lda, fos, ...)

different evaluation settings (all, CSonly, comparison to MAG, ACL (data from [60])...)

FoS alone, restrictively combined w/ BOW, only directly preceeding, ...

PP model alone, combined, ...

-> not *generally* applicable/beneficial but for certain citation types ...

also mention [25] here b/c they specifically target cases where more than one citation is applicable (could be interpreted as either *multiple (simultaneously)* for one context or *several options that are all valid by themselves but in the end a single one is to be chosen* for one contexts)

6.3. Online evaluation

online online

7. Conclusion

conclude conclude.

8. Future work

As a first step identify types of citations more systematically.

For different types, different models.

Proper claim model. (that could also include assessing credibility [61])

Argumentative structures. (Argumentation mining [62–64])

also data set is already extended with 2018 data

for data where no precise citation marker can be guaraneed (and so far NPmarker was not tested) look into heuristically identifying marker position

Bibliography

[1] Z. Nasar, S. W. Jaffry, and M. K. Malik, "Information extraction from scientific articles: a survey," *Scientometrics*, vol. 117, pp. 1931–1990, Dec 2018.

[2] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems: a literature survey," *International Journal on Digital Libraries*, vol. 17, pp. 305–338, Nov 2016.

[3] S. Buckingham Shum, E. Motta, and J. Domingue, "Scholonto: an ontology-based digital library server for research documents and discourse," *International Journal on Digital Libraries*, vol. 3, pp. 237–248, Oct 2000. r (ch 1-3).

[4] J. Schneider, T. Groza, and A. Passant, "A review of argumentation for the social semantic web," *Semant. web*, vol. 4, pp. 159–218, Apr. 2013.

[5] M. Y. Jaradeh, S. Auer, M. Prinz, V. Kovtun, G. Kismihók, and M. Stocker, "Open research knowledge graph: Towards machine actionability in scholarly communication," *CoRR*, vol. abs/1901.10816, 2019.

[6] A. Abu-Jbara, J. Ezra, and D. Radev, "Purpose and polarity of citation: Towards nlp-based bibliometrics," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 596–606, Association for Computational Linguistics, 2013.

[7] S. Teufel, A. Siddharthan, and D. Tidhar, "Automatic classification of citation function," in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, (Stroudsburg, PA, USA), pp. 103–110, Association for Computational Linguistics, 2006.

[8] M. Berger, K. McDonough, and L. M. Seversky, "Cite2vec: Citation-driven document exploration via word embeddings," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, pp. 1–1, 01 2016.

[9] K. D. Bollacker, S. Lawrence, and C. L. 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, (New York, NY, USA), pp. 116–123, ACM, 1998.

[10] Q. He, D. Kifer, J. Pei, P. Mitra, and C. L. Giles, "Citation recommendation without author supervision," in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, (New York, NY, USA), pp. 755–764, ACM, 2011.

[11] T. Berners-Lee, J. Hendler, O. Lassila, *et al.*, "The semantic web," *Scientific american*, vol. 284, no. 5, pp. 28–37, 2001.

[12] M. J. Moravcsik and P. Murugesan, "Some results on the function and quality of citations," *Social Studies of Science*, vol. 5, no. 1, pp. 86–92, 1975.

[13] F. Zarrinkalam and M. Kahani, "Semcir: A citation recommendation system based on a novel semantic distance measure," *Program: Electronic Library and Information Systems*, vol. 47, pp. 92–112, 2013.

[14] S. E. Middleton, D. D. Roure, and N. Shadbolt, "Capturing knowledge of user preferences: ontologies in recommender systems," in *K-CAP*, 2001.

[15] M. Zhang, W. Wang, and X. Li, "A paper recommender for scientific literatures based on semantic concept similarity," in *Digital Libraries: Universal and Ubiquitous Access*

to Information (G. Buchanan, M. Masoodian, and S. J. Cunningham, eds.), (Berlin, Heidelberg), pp. 359–362, Springer Berlin Heidelberg, 2008.

[16] Y. Jiang, A. Jia, Y. Feng, and D. Zhao, "Recommending academic papers via users' reading purposes," *RecSys'12 - Proceedings of the 6th ACM Conference on Recommender Systems*, 09 2012.

[17] F. Zarrinkalam and M. Kahani, "A multi-criteria hybrid citation recommendation system based on linked data," in *2012 2nd International eConference on Computer and Knowledge Engineering (ICCKE)*, pp. 283–288, IEEE, 2012.

[18] C. Caragea, J. Wu, A. Ciobanu, K. Williams, J. Fernández-Ramírez, H.-H. Chen, Z. Wu, and L. Giles, "Citeseerx: A scholarly big dataset," in *Advances in Information Retrieval* (M. de Rijke, T. Kenter, A. P. de Vries, C. Zhai, F. de Jong, K. Radinsky, and K. Hofmann, eds.), (Cham), pp. 311–322, Springer International Publishing, 2014.

[19] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles, "Context-aware citation recommendation," in *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, (New York, NY, USA), pp. 421–430, ACM, 2010.

[20] W. Huang, , P. Mitra, and C. L. Giles, "Refseer: A citation recommendation system," in *IEEE/ACM Joint Conference on Digital Libraries*, pp. 371–374, Sep. 2014.

[21] W. Huang, Z. Wu, C. Liang, P. Mitra, and C. L. Giles, "A neural probabilistic model for context based citation recommendation," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pp. 2404–2410, AAAI Press, 2015.

[22] D. Duma and E. Klein, "Citation resolution: A method for evaluating context-based citation recommendation systems," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 2, pp. 358–363, 2014.

[23] D. Duma, E. Klein, M. Liakata, J. Ravenscroft, and A. Clare, "Rhetorical classification of anchor text for citation recommendation," *D-Lib Magazine*, vol. 22, 2016.

[24] T. Ebesu and Y. 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, (New York, NY, USA), pp. 1093–1096, ACM, 2017.

[25] Y. Kobayashi, M. Shimbo, and Y. 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, (New York, NY, USA), pp. 243–251, ACM, 2018. r (ch 1-2).

[26] C. Jeong, J. Sion, H. Shin, E. Park, and S. Choi, "A context-aware citation recommendation model with bert and graph convolutional networks," 02 2019.

[27] M. Färber, A. Thiemann, and A. Jatowt, "A High-Quality Gold Standard for Citation-based Tasks," in *Proceedings of the 11th International Conference on Language Resources and Evaluation*, LREC 2018, 2018. r.

[28] K. Sugiyama and M.-Y. Kan, "Exploiting potential citation papers in scholarly paper recommendation," in *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '13, (New York, NY, USA), pp. 153–162, ACM, 2013.

[29] D. R. Radev, P. Muthukrishnan, V. Qazvinian, and A. Abu-Jbara, "The acl anthology network corpus," *Language Resources and Evaluation*, vol. 47, pp. 919–944, Dec 2013.

[30] S. Bird, R. Dale, B. J. Dorr, B. R. Gibson, M. T. Joseph, M. Kan, D. Lee, B. Powley, D. R. Radev, and Y. F. Tan, "The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics," in *Proceedings of the International Conference on Language Resources and Evaluation*, LREC 2008, 2008.

[31] D. Roy, K. Ray, and M. Mitra, "From a Scholarly Big Dataset to a Test Collection for Bibliographic Citation Recommendation," SBD'16, 2016.

[32] B. Gipp, N. Meuschke, and M. Lipinski, "Citrec : An evaluation framework for citation-based similarity measures based on trec genomics and pubmed central," in *iConference 2015 Proceedings*, iSchools, 2015.

[33] L. Galke, F. Mai, I. Vagliano, and A. Scherp, "Multi-modal adversarial autoencoders for recommendations of citations and subject labels," in *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, UMAP '18, (New York, NY, USA), pp. 197–205, ACM, 2018.

[34] C. Bhagavatula, S. Feldman, R. Power, and W. Ammar, "Content-based citation recommendation," in *NAACL-HLT*, 2018.

[35] A. Sinha, Z. Shen, Y. Song, H. Ma, D. Eide, B.-J. P. Hsu, and K. Wang, "An overview of microsoft academic service (mas) and applications," in *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, (New York, NY, USA), pp. 243–246, ACM, 2015. r.

[36] P. Ginsparg, "First steps towards electronic research communication," *Computers in Physics*, vol. 8, pp. 390–396, July 1994.

[37] H. Bast and C. Korzen, "A benchmark and evaluation for text extraction from pdf," in *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pp. 1–10, June 2017.

[38] H. Nanba, "Towards multi-paper summarization using reference information," Master's thesis, Japan Advanced Institute of Science and Technology, 2 1998. (in Japanese).

[39] A. Prasad, M. Kaur, and M.-Y. Kan, "Neural parscit: A deep learning based reference string parser," *International Journal on Digital Libraries*, vol. 19, pp. 323–337, 2018.

[40] L. D. Brown, T. T. Cai, and A. DasGupta, "Interval estimation for a binomial proportion," *Statistical Science*, vol. 16, no. 2, pp. 101–133, 2001.

[41] M. Hernández-Alvarez and J. M. Gomez, “Survey about citation context analysis: Tasks, techniques, and resources,” *Natural Language Engineering*, vol. 22, no. 3, p. 327–349, 2016. "r".

[42] C. Niklaus, M. Cetto, A. Freitas, and S. Handschuh, “A survey on open information extraction,” in *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 3866–3878, Association for Computational Linguistics, 2018.

[43] W. Lamers, N. J. v. Eck, L. Waltman, and H. Hoos, “Patterns in citation context: the case of the field of scientometrics,” in *STI 2018 Conference proceedings*, pp. 1114–1122, Centre for Science and Technology Studies (CWTS), 2018.

[44] P. Thompson, *A pedagogically-motivated corpus-based examination of PhD theses: Macrostructure, citation practices and uses of modal verbs.* PhD thesis, University of Reading, 2001.

[45] B. Petrić, “Rhetorical functions of citations in high- and low-rated master’s theses,” *Journal of English for Academic Purposes*, vol. 6, no. 3, pp. 238 – 253, 2007.

[46] J. Swales, *Genre analysis: English in academic and research settings.* Cambridge University Press, 1990.

[47] K. Hyland, “Academic attribution: citation and the construction of disciplinary knowledge,” *Applied Linguistics*, vol. 20, no. 3, pp. 341–367, 1999.

[48] A. Okamura, “Citation forms in scientific texts: Similarities and differences in l1 and l2 professional writing,” *Nordic Journal of English Studies*, vol. 7, no. 3, pp. 61–81, 2008.

[49] M. Whidby, D. Zajic, and B. Dorr, “Citation handling for improved summarization of scientific documents,” tech. rep., 2011.

[50] A. Abu-Jbara and D. Radev, “Reference scope identification in citing sentences,” in *Proceedings of the 2012 Conference of the North American Chapter of the Association for*

Computational Linguistics: Human Language Technologies, NAACL HLT '12, (Stroudsburg, PA, USA), pp. 80–90, Association for Computational Linguistics, 2012.

[51] R. Levy, Y. Bilu, D. Hershcovich, E. Aharoni, and N. Slonim, "Context dependent claim detection," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, (Dublin, Ireland), pp. 1489–1500, Dublin City University and Association for Computational Linguistics, August 2014. r.

[52] A. S. White, D. Reisinger, K. Sakaguchi, T. Vieira, S. Zhang, R. Rudinger, K. Rawlins, and B. Van Durme, "Universal decompositional semantics on universal dependencies," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1713–1723, Association for Computational Linguistics, 2016.

[53] S. Zhang, R. Rudinger, and B. V. Durme, "An evaluation of predpatt and open ie via stage 1 semantic role labeling," in *IWCS 2017 — 12th International Conference on Computational Semantics — Short papers*, 2017.

[54] K. Gábor, D. Buscaldi, A.-K. Schumann, B. QasemiZadeh, H. Zargayouna, and T. Charnois, "Semeval-2018 task 7: Semantic relation extraction and classification in scientific papers," in *Proceedings of The 12th International Workshop on Semantic Evaluation*, pp. 679–688, Association for Computational Linguistics, 2018.

[55] R. Jha, A.-A. Jbara, V. Qazvinian, and D. R. Radev, "Nlp-driven citation analysis for scientometrics," *Natural Language Engineering*, vol. 23, no. 1, p. 93–130, 2017.

[56] J. Beel and S. Dinesh, "Real-world recommender systems for academia: The pain and gain in building, operating, and researching them [long version]," *CoRR*, vol. abs/1704.00156, 2017.

[57] J. Swales, "Citation analysis and discourse analysis," *Applied Linguistics*, vol. 7, no. 1, pp. 39–56, 1986.

[58] J. He and C. Chen, “Temporal representations of citations for understanding the changing roles of scientific publications,” in *Front. Res. Metr. Anal.*, 2018.

[59] J. Beel, “It’s time to consider "time" when evaluating recommender-system algorithms [proposal],” *CoRR*, vol. abs/1708.08447, 2017.

[60] M. Färber, A. Thiemann, and A. Jatowt, “To Cite, or Not to Cite? Detecting Citation Contexts in Text,” in *Proceedings of the 40th European Conference on Information Retrieval*, ECIR 2018, 2018.

[61] K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum, “Credibility assessment of textual claims on the web,” in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM ’16, (New York, NY, USA), pp. 2173–2178, ACM, 2016.

[62] C. Stab and I. Gurevych, “Parsing argumentation structures in persuasive essays,” *CoRR*, vol. abs/1604.07370, 2016.

[63] M. Lippi and P. Torroni, “Argumentation mining: State of the art and emerging trends,” *ACM Trans. Internet Technol.*, vol. 16, pp. 10:1–10:25, Mar. 2016.

[64] I. Habernal and I. Gurevych, “Argumentation mining in user-generated web discourse,” *Comput. Linguist.*, vol. 43, pp. 125–179, Apr. 2017.

Appendices

A. Evaluation of reference resolution

The sample of 300 matched reference items was acquired from the reference data base as shown in Listing A.1. The table `bibitem` holds most of the information on reference items. The table `bibitemmagidmap` contains per row the UUID of a reference item and the MAG ID it was matched to.

Listing A.1: SQL query used to acquire the sample

```
select b.bibitem_string, m.mag_id
    from bibitem as b
        join
         (select *
            from bibitemmagidmap
                order by random()
                limit 300
         ) as m
        on b.uuid = m.uuid;
```

Table 5 shows the full evaluation. Wrongly matched items are highlited in red. Note that there are seven cases where the reference item refers to more than one publication. In such cases our method only captures the first one and is cosequetually evaluated on the first one. Note further that in one case a reference item named a PhD thesis while the match was for the very same thesis published two years later. This was deemed a correct match. Lastly, there was one case where a book was cited with the date indicating its second edition while the matched record in the MAG has a date indiciating the books thrid edition. This also was deemed a correct match.

Table 5.: Evaluation

Reference item	MAG ID	
V. N. Senoguz and Q. Shafi, "Reheat temperature in supersymmetric hybrid inflation models," Phys. Rev. D 71, 043514 (2005) [hep-ph/0412102].	2075392245	✓
Keiding, N. and Nielsen, J.E. (1975) Branching processes with varying and random geometric offspring distributions. J. Appl. Prob. 12, 135–141.	2332540167	✓
H. Izeki and S. Nayatani, Combinatorial harmonic maps and discrete-group actions on Hadamard spaces, Geom. Dedicata 114 (2005), 147–188.	2017711173	✓
T. Adamo, M. Bullimore, L. Mason and D. Skinner, "Scattering Amplitudes and Wilson Loops in Twistor Space," J. Phys. A 44 (2011) 454008 [arXiv:1104.2890 [hep-th]].	2002091616	✓
Eren Mehmet Kıral and Matthew Young, The fifth moment of modular FORMULA -functions, arXiv preprint arXiv:1701.07507 (2017).	2582886839	✓
T. Baumgarte and S. Shapiro, Numerical Relativity: Solving Einstein's Equations on the Computer. Cambridge University Press, 2010. http://books.google.co.uk/books?id=dxU1OEinvRUC.	2566410267	✓
R.E. Renfordt, D. Schall, R. Bock, R. Brockmann, J.W. Harris, A. Sandoval, R. Stock, H. Ströbele, D. Bangert, W. Rauch, G. Odyniec, H.G. Pugh, and L.S. Schroeder, Phys. Rev. Lett. 53 (1984) 763.	2001418221	✓
T. A. Porter, I. V. Moskalenko, A. W. Strong, E. Orlando and L. Bouchet, arXiv:0804.1774 [astro-ph].	2129746122	✓
Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. arXiv preprint arXiv:1609.05807, 2016.	2522104760	✓
R.M. Fernandes, L. H. VanBebber, S. Bhattacharya, P. Chandra, V. Keppens, D. Mandrus, M.A. McGuire, B.C. Sales, A.S. Sefat, and J. Schmalian, Phys. Rev. Lett. 105, 157003 (2010)	2143202785	✓
Fomin, S., Wei, P., Chugunov, V., 1995. Contact melting by a non-isothermal heating surface of arbitrary shape. Int. J. Heat Mass Transfer 38 (17), 3275–3284.	2030656707	✓
R.F. Lebed, arXiv:1507.05867v1 [hep-ph].	1844403609	✓
R. Billinton, R. Karki, Y. Gao, D. Huang, P. Hu, and W. Wangdee, "Adequacy Assessment Considerations in Wind Integrated Power Systems," IEEE Trans. Power Syst., vol. 27, no. 4, pp. 2297–2305, 2012.	2024825567	✓
C. Morningstar and M. Peardon, Phys. Rev. D 56, 4043 (1997).	2109255696	✓
I. B. S. Passi, M. Singh and M. K. Yadav, Automorphisms of abelian group extensions, J. Algebra 324 (2010), 820–830.	2051680489	✓
S. Dimopoulos, P. W. Graham, J. M. Hogan, M. A. Kasevich, and S. Rajendran, Phys. Rev. D 78, 122002 (2008).	2749889157	✓
P.Jaworski, Value at Risk in the presence of the power laws, Acta Physica Polonica B 36 (2005) 2575-2587.	2566822874	✓
L. L. Chau, M. L. Ge and Y. S. Wu, Phys. Rev. D 25, 1086 (1982); L. L. Chau and Wu Yong-Shi, Phys. Rev. D 26, 3581 (1982); L. L. Chau, M. L. Ge, A. Sinha and Y. S. Wu, Phys. Lett. B 121, 391 (1983).	2075757391	✓
E. C. Blomberg, M. a. Tanatar, R. M. Fernandes, I. I. Mazin, B. Shen, H.-H. Wen, M. D. Johannes, J. Schmalian, and R. Prozorov, Nat. Commun. 4, 1914 (2013).	2081425933	✓
A. Arenas, A. Díaz-Guilera, C. J. Pérez-Vicente, Synchronization reveals topological scales in complex networks, Phys. Rev. Lett. 96 (11) (2006) 114102.	2100240966	✓
M. Alizadeh, A. Greenberg, D. A. Maltz, J. Padhye, P. Patel, B. Prabhakar, S. Sengupta, and M. Sridharan. Data Center TCP (DCTCP). ACM SIGCOMM, 2010.	2164740236	✓
M. Nishiyama, T. Okabe, Y. Sato, and I. Sato. Sensation-based photo cropping. In ACM Multimedia, pages 669–672, 2009.	2013339738	✓
L. Landau and E. Lifchitz, Classical theory of fields, Butterworth-Heinemann, Oxford, 1994, p. 87.	119088996	✓
Peregrine, D., 2003. Water-wave impact on walls. Annu. Rev. Fluid Mech. 35, 23–43.	2118259172	✓
M. S. Khalil, S. Gladchenko, M. J. A. Stoutimore, F. C. Wellstood, A. L. Burin, and K. D. Osborn, Phys. Rev. B 90, 100201 (2014).	1998393159	✓
F. Gabbiani, E. Gabrielli, A. Masiero and L. Silvestrini, Nucl. Phys. B 477, 321 (1996) [arXiv:hep-ph/9604387].	2133327165	✓
Abramowitz, M. and Stegun, I. A., Handbook of Mathematical Functions, (Dover, New York, 1965).	2120062331	✓
N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16(1):321–357, 2002.	2148143831	✓
B. H. Lee, W. Lee, R. MacKenzie, M. B. Paranjape, U. A. Yajnik and D. h. Yeom, Phys. Rev. D 88, 085031 (2013).	2074315988	✓
D. Nishiguchi, K. H. Nagai, H. Chaté, and M. Sano, Long-range nematic order and anomalous fluctuations in suspensions of swimming filamentous bacteria. arXiv preprint arXiv:1604.04247 (2016).	2336881770	✓
Huelga, S. F. et al., Improvement of frequency standards with quantum entanglement. Phys. Rev. Lett. 79, 3865 (1997).	2015876000	✓
M. Henneaux and C. Teitelboim, Asymptotically anti-de Sitter spaces, Comm. Math. Phys. 98, 391 (1985).	2134856726	✓
D. Boucher, W. Geiselmann, and F. Ulmer, Skew-cyclic codes, Applicable Algebra in Engineering, Communication and computing, (18) (4), (2007), 379-389.	2151359594	✓
R. C. Brower, H. Nastase, H. J. Schnitzer and C.-I. Tan, arXiv:0801.3891 [hep-th].	1995185450	✓

S. Sarkar, Big bang nucleosynthesis and physics beyond the standard model, Rept. Prog. Phys. 59 (1996) 1493–1610, [hep-ph/9602260].	2013442457	✓
M. Iskin and J. K. Freericks, Phys. Rev. A 80, 053623 (2009).	2076979008	✓
M. L. Skoge & T. W. Baumgarte, Phys. Rev. D 66 107501 (2002).	2016175541	✓
A. Lozano, A. M. Tulino, and S. Verdú, “Optimum power allocation for parallel Gaussian channels with arbitrary input distributions,” IEEE Trans. Inf. Theory, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.	2097695636	✓
F. Paci, A. Gruppuso, F. Finelli, A. De Rosa, N. Mandolesi, and P. Natoli, MNRAS 434, 3071 (Oct. 2013), arXiv:1301.5195	1976439668	✓
Samuel Brody and Noemie Elhadad. 2010. An unsupervised aspect-sentiment model for online reviews. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 804–812. Association for Computational Linguistics.	2113786470	✓
R. Islam, C. Senko, W. C. Campbell, S. Korenblit, J. Smith, A. Lee, E. E. Edwards, C.-C. J. Wang, J. K. Freericks, and C. Monroe, Emergence and frustration of magnetism with variable-range interactions in a quantum simulator, Science 340, 583 (2013).	1606923443	✓
I. Ermolli, K. Matthes, T. Dudok de Wit, N.A. Krivova, K. Tourpali, M. Weber, Y.C. Unruh, L. Gray, U. Langematz, P. Pilewskie, E. Rozanov, W. Schmutz, A. Shapiro, S.K. Solanki, T.N. Woods, Recent variability of the solar spectral irradiance and its impact on climate modelling. Atmos. Chem. Phys. 13, 3945–3977 (2013). doi:10.5194/acp-13-3945-2013	2158950390	✓
Tauchen G. and Zhou, H. (2011), “Realized Jumps on Financial Markets and Predicting Credit Spreads,” Journal of Econometrics, 160, 102–118	2113380547	✓
S. Yoon, W. Ye, J. Heidemann, B. Littlefield, and C. Shahabi. Swats: Wireless sensor networks for steamflood and waterflood pipeline monitoring. Network, IEEE, 25(1):50–56, January 2011.	2054094122	✓
J.Y. Vaishnav and C.W. Clark, Phys. Rev. Lett. 100, 153002 (2008).	2752272568	✓
L. J. Hall, T. Moroi and H. Murayama, Phys. Lett. B 424, 305 (1998). [arXiv:hep-ph/9712515].	2078593377	✓
Movassaghi, Samaneh and Abolhasan, Mehran and Smith, David, Smart spectrum allocation for interference mitigation in Wireless Body Area Networks, IEEE International Conference on Communications (ICC), pages 5688-5693, 2014	2083602816	✓
M. Apollonio et al. [CHOOZ Collaboration], Phys. Lett. B 466, 415 (1999) [arXiv:hep-ex/9907037].	1571701324	✓
G. ’t Hooft, “Magnetic Monopoles in Unified Gauge Theories,” Nucl.Phys. B79 (1974) 276–284.	2027710569	✓
D. C. Cabra, M. D. Grynberg, P. C. W. Holdsworth, A. Honecker, P. Pujol, J. Richter, D. Schmalfß, and J. Schulenburg, Phys. Rev. B 71, 144420 (2005).	2032654705	✓
V.A. Dolgushev, Erratum to: ""A Proof of Tsygan’s Formality Conjecture for an Arbitrary Smooth Manifold"", arXiv:math/0703113.	132267651	✓
Eddy, J.A.: 1983, The maunder minimum - a reappraisal. Solar Phys. 89, 195. ADS.	2024069573	×
Bouliotis, G. and Billingham, L. (2011) Crossing survival curves: alternatives to the log-rank test. Trials, 12, 1.	2136058878	✓
I. Peschel and V. Eisler, Reduced density matrices and entanglement entropy in free lattice models, J. Phys. A 42 504003 (2009).	2141130841	✓
C. Gundlach, Critical phenomena in gravitational collapse, submitted to Adv. Theor. Math. Phys., preprint gr-qc/9712084.	2103342599	✓
A. Blumen, G. Zumofen, and J. Klafter, Target annihilation by random walkers, Phys. Rev. B 30, 5379 (1984).	2060897997	✓
M. Alidoust, K. Halterman, and J. Linder, Phys. Rev. B 89, 054508 (2014).	2012436162	✓
G. Ciavola, L. Celona, S. Gammino, M. Presti, L. Ando, S. Passarello, X.Zh. Zhang, F. Consoli, F. Chines, C. Percolla, V. Calzona and M. Winkler, A version of the Trasco Intense Proton Source optimized for accelerator driven system purposes. Rev. Sci. Instrum. 75 (2004) 1453–1456,.	2075428341	✓
D.Litim Phys. Rev. Lett. 92 (2004) 201301, hep-th/0312114	2000661521	✓
P.K. Kovtun, D.T. Son, and A.O. Starinets, Phys. Rev. Lett. 94, 111601 (2005).	2097909025	✓
Jones, J. A. et al. Magnetic Field Sensing Beyond the Standard Quantum Limit Using 10-Spin NOON States. Science 324, 1166–1168 (2009).	2022149792	✓
G.E.Brown and M.Rho, Phys.Rev.Lett. 66, (1991) 2720;	1971702030	✓
K. Banaszek and K. Wódkiewicz, “Operational theory of homodyne detection,” Phys. Rev. A 55, 3117 (1997).	1985850877	✓
Barbara Di Eugenio, Pamela W. Jordan, and Liina Pylkkänen. 1998. The COCONUT project: Dialogue annotation manual. Technical Report 98-1, University of Pittsburgh, Intelligent Systems Program. [www.isp.pitt.edu/intgen/coconut.html].	116082719	✓
Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). Statistics of extremes: Theory and Applications. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd.	1598342322	✓
Guillaumin, M., Mensink, T., Verbeek, J.J., Schmid, C.: Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. In: IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 - October 4, 2009. (2009) 309–316	2536305071	✓
T. Faulkner et al., “Gravitation from entanglement in holographic CFTs,” (2013), arXiv:1312.7856v1.	2102970467	✓
René Thom, Quelques propriétés globales des variétés différentiables, Comment. Math. Helv. 28 (1954), 17–86.	1989427081	✓
A. Schikorra. Regularity of n/2-harmonic maps into spheres. PhD-Thesis, arXiv:1003.0646v1, 2010.	2029831933	✓
M. J. Simpson, J. A. Sharp, and R. E. Baker. Survival probability for a diffusive process on a growing domain. Phys. Rev. E 91, 042701 (2015).	1988447698	✓

D. Bergamini, N. Descoubes, C. Joubert & R. Mateescu (2005): Bisimulator: A Modular Tool for On-the-Fly Equivalence Checking. In: Proc. of TACAS'05, Lecture Notes in Computer Science 3440, Springer-Verlag, pp. 581–585.	1503429725	✓
J. A. Baldwin, O. Plamenevskaya, Khovanov homology, open books, and tight contact structures. math.GT/0808.2336	2054234174	✓
Eisenberger, P., et al., 1972, Phys. Rev. B 6, 3671.	1980522055	✓
O. Viehmann, C. Eltschka, and J. Siewert, Appl. Phys. B 106, 533 (2012).	2076107066	✓
A. Rényi. Representations for real numbers and their ergodic properties. Acta Math. Acad. Sci. Hungar 8 (1957), 477–493.	2089164015	✓
D. Chicharro and R. G. Andrzejak, Phys. Rev. E 80, 026217 (2009).	2078979753	✓
Kenward, M. G., Jones, B., 1992. Alternative approaches to the analysis of binary and categorical repeated measurements. Journal of Biopharmaceutical Statistics 2 (2), 137–170.	1967449535	✓
J. Nešetřil and V. Rödl. The partite construction and Ramsey set systems. Discrete Mathematics, 75(1-3):327–334, 1989.	2062861878	✓
E. Seiler, Gauge Theories as a Problem of Constructive Quantum Field Theory and Statistical Mechanics, Lecture Notes in Physics Vol. 159 (Springer, Berlin, 1982).	1608098855	✓
M. Gaye, Y. Chitour, and P. Mason. Properties of barabanov norms and extremal trajectories associated with continuous-time linear switched systems. In Proceedings of the 52nd IEEE Conference on Decision and Control, pages 716–721, Florence, Italie, 2013.	2090302562	✓
Kováčik R. and Ederer C., Phys. Rev. B 80 (2009) 140411; Kim M. et al., Phys. Rev. B 81 (2010) 100409.	1758648405	✓
Morandi, G., Ferrario, C., Lo Vecchio, G., Marmo, G. and Rubano, C. (1990). The inverse problem in the calculus of variations and the geometry of the tangent bundle, Phys. Rep. 188, 147.	2026992500	✓
G. Da Prato and J. Zabczyk, Ergodicity for infinite dimensional systems, London Mathematical Society Lecture Note Series, 229, Cambridge University Press, 1996.	1530927473	✓
R. Trotta, Bayes in the sky: Bayesian inference and model selection in cosmology, Contemp. Phys. 49 (2008) 71–104, [arXiv:0803.4089].	2021748112	✓
Christof, J., M. Gebhardt, A. E.-M. Clemen, J. Jaud, , and M. Rief, 2006. Myosin-V is a mechanical ratchet. Proc. Natl. Acad. Sci. U.S.A. 103:8680–8685.	2095633207	✓
G.M. Molchan: Burgers equation with self-similar Gaussian initial data: tail probabilities. J. of Stat. Phys. 88 (1997) 1139–1150.	2000957076	✓
Arata, I., Y. Ohno, F. Matsukura, and H. Ohno, 2001, "Temperature dependence of electroluminescence and I-V characteristics of ferromagnetic/non-magnetic semiconductor pn junctions," Physica E 10, 288–291.	1987870048	✓
I. Zlatev, L. Wang and P.J. Steinhardt, Phys. Rev. Lett. 82, 896 (1999); Phys. Rev. D 59, 123504 (1999).	2032901690	✓
A. Valentini, in: Bohmian Mechanics and Quantum Theory: an Appraisal, eds. J. T. Cushing et al. (Kluwer, Dordrecht, 1996).	207861407	✓
R. Killip, S. Kwon, S. Shao, and M. Visan. On the mass-critical generalized KdV equation. Discrete Contin. Dyn. Syst., 32(1):191–221, 2012.	1975802612	✓
Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. arXiv preprint arXiv:1505.05770, 2015.	299440670	✓
D. Rossi and G. Rossini, "On sizing CCN content stores by exploiting topological information," in Proc. IEEE NOMEN, 2012.	1985355206	✓
F. Bezrukov, A. Magnin, M. Shaposhnikov and S. Sibiryakov, JHEP 1101 (2011) 016 [arXiv:1008.5157].	2064410211	✓
M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed quantum states: linear contractions approach. preprint archiv quant-ph/0206008.	1668368460	✓
S. White, Phys. Rev. B 48, 10345 (1993).	2016407890	✓
A. Brandt et al., [UA8 Collaboration], Evidence for a Super-Hard Pomeron Structure, submitted to Phys. Lett. 1992.	1983143801	✓
Fan, T.-H. & Berger, J. O. (2000). Robust Bayesian displays for standard inferences concerning a normal mean. Computational Statistics & Data Analysis 33 381–399.	2030654698	✓
S. Ryu, J. E. Moore, and A. W. W. Ludwig, Phys. Rev. B 85, 045104 (2012), arXiv:1010.0936.	2083123179	✓
P. Balaz, V. K. Dugaev, and J. Barnaś Phys. Rev. B 85, 024416 (2012)	2322343165	✓
T. Giamarchi and A. Tsvelik, Phys. Rev. B 59, 11398 (1999).	2069018114	✓
M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. Combin. Probab. Comput., 7(3):295–305, (1998).	2129918926	✓
D. W. Sivers, Phys. Rev. D 41, 83 (1990); Phys. Rev. D 43, 261 (1991).	2174029682	✓
J. Boronat and J. Casulleras, Phys. Rev. B 49, 8920 (1994).	1982967539	✓
J. N. Bandyopadhyay and A. Lakshminarayan, Phys. Rev. E 69, 016201 (2004).	1977839735	✓
Priest ER, Forbes TG (2002) The magnetic nature of solar flares. 10:313–377, DOI 10.1007/s001590100013	2014209718	✓
W. Woerndl, C. Schueller, and R. Wojtech. A hybrid recommender system for context-aware recommendations of mobile applications. In Proceedings of ICDEW '07, pages 871–878, Washington, DC, USA, 2007. IEEE Computer Society.	2112166834	✓
[auto:STB|2013/06/05|13:45:01] Worsley, K. J.K. J., Liao, C. H.C. H., Aston, J.J., Petre, V.V., Duncan, G. H.G. H., Morales, F.F. Evans, A. C.A. C. (2002). A general statistical analysis for fMRI data. NeuroImage 15 1–15. imsref	1975938737	✓
L.H. Ford and N.F. Svaiter, Phys. Rev. D 58, 065007 (1998), quant-ph/9804056.	1963985219	✓
H. Häffner, S. Gulde, M. Riebe, G. Lancaster, C. Becher, J. Eschner, F. Schmidt-Kaler, R. Blatt, Precision measurement and compensation of optical Stark shifts for an ion-trap quantum processor, Phys. Rev. Lett. 90 (2003) 143602.	2129198554	✓

I. Jeon, K. Lee, J.-H. Park, and Y. Suh, Stringy Unification of Type IIA and IIB Supergravities under N=2 D=10 Supersymmetric Double Field Theory, Phys.Lett. B723 (2013) 245–250, [arXiv:1210.5078].	1967998606	✓
Z.A. Anastassi and T.E. Simos: A Trigonometrically-Fitted Runge-Kutta Method for the Numerical Solution of Orbital Problems, New Astronomy, 10, 301-309 (2005)	2024993485	✓
Fletcher, A. 2010, in Astronomical Society of the Pacific Conference Series, Vol. 438, The Dynamic Interstellar Medium: A Celebration of the Canadian Galactic Plane Survey, ed. R. Kothes, T. L. Landecker, & A. G. Willis, 197	1671679100	✓
V. A. Belinsky, I. M. Khalatnikov, and E. M. Lifshitz. Oscillatory approach to a singular point in the relativistic cosmology. Adv. Phys., 19:525–573, 1970.	2048737175	✓
Allen, D. A. et al., 1993. IRIS – an Infrared Imager and Spectrometer for the Anglo-Australian Telescope. Proceedings of the Astronomical Society of Australia 10, 298.	91162570	✓
D. Cooper, Automorphisms of free groups have finitely generated fixed point sets. J. Algebra, 111 (1987), no. 2 453–456	2076181847	✓
J.-L. Lehners and P. J. Steinhardt, “Intuitive understanding of non-gaussianity in ekpyrotic and cyclic models,” Phys.Rev. D78 (2008) 023506, arXiv:0804.1293 [hep-th].	2592352904	✓
Z.-B. Wu, Global transposable characteristics in the complete DNA sequence of the yeast. Physica A 389 (2010) 5698.	1548592618	✓
E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. In Computer Vision–ECCV 2006, pages 490–503. Springer, 2006.	2171896402	✓
Benjamin C. Pierce. Types and programming languages: The next generation. LICS’03, 2003.	1951034176	✓
G. Tardos and G. Tóth. Multiple coverings of the plane with triangles. Discrete & Computational Geometry, 38(2):443–450, 2007.	2043718124	✓
T. Rivière, Analysis aspects of Willmore surfaces, Invent. Math., Vol. 174, (2008), 1–45.	1606077524	✓
H. Mabuchi, Phys. Rev. A 85, 015806 (2012).	1620088716	✓
I. Schienbein, J. Y. Yu, K. Kovarik, C. Keppel, J. G. Morfin, F. Olness and J. F. Owens, Phys. Rev. D 80, 094004 (2009)	2050053974	✓
D.Q. Goldin, S.A. Smolka, P.C. Attie, E.L. Sonderegger, Turing machines, transition systems and interaction, manuscript, 2003.	2048671682	✓
P. Collet and J.P. Eckmann, Iterated Maps on the Interval as Dynamical Systems, (Birkhäuser, Basel, 1980).	1573241742	✓
M. F. Maghrebi, R. L. Jaffe, and M. Kardar, Phys. Rev. Lett. 108, 230403 (2012).	2025571896	✓
A. Minami and A. Onuki, Phys. Rev. B 70, 184114 (2004); Acta Mater. 55, 2375 (2007).	1514590689	✓
Kfir Blum, Anson Hook, and Kohta Murase, “High energy neutrino telescopes as a probe of the neutrino mass mechanism,” (2014), arXiv:1408.3799 [hep-ph] .	1628049864	✓
O. E. Buryak, Phys. Rev. D 53 (1996) 1763 [gr-qc/9502032].	2069968655	✓
Kolb, E. W.; Turner, M. S. The Early Universe, AddisonWesley Publishing Company: California, USA, 1990.	2595419339	✓
G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn, Phys. Rev. B 39, 4828 (1989).	2043072234	✓
JC Baygents and DA Saville. The circulation produced in a drop by an electric field: a high field strength electrokinetic model. In AIP Conference Proceedings, volume 197, pages 7–17. AIP, 1990.	1612371284	✓
E. Altman and R. Vosk, Annual Review of Condensed Matter Physics 6, 383 (2015).	2166587989	✓
J.-M. Souriau, Structure des systèmes dynamiques (Dunod, 1970).	108534386	✓
F. Horn. Explicit Muller games are PTIME. In Proc. 28th Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS’08), LIPIcs 2, p. 235–243. Leibniz-Zentrum für Informatik, 2008.	2240543079	✓
A. G. Izergin and V. E. Korepin, “The Inverse Scattering Method Approach To The Quantum Shabat-Mikhailov Model,” Commun. Math. Phys. 79 (1981) 303.	1977168974	✓
Balzano, V. A. 1983 Star-burst Galactic Nuclei. ApJ 268, 602–627.	2059760395	✓
M. Ishak, A. Upadhye, D. N. Spergel, Phys. Rev. D 74, 043513 (2006) [arXiv:astro-ph/0507184].	2044422425	✓
M. Lynker and R. Schimmrigk, Landau–Ginzburg theories as orbifolds, Phys. Lett. B249 (1990) 237	2030907161	✓
S. T. Petcov, T. Shindou and Y. Takanishi, Nucl. Phys. B 738, 219 (2006) [arXiv:hep-ph/0508243].	2007551251	✓
J. E. Lye, L. Fallani, C. Fort, V. Guarrera, M. Modugno, D. S. Wiersma, and M. Inguscio, Phys. Rev. A 75, 061603(R) (2007).	2015993116	✓
J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996).	1981368803	✓
J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Physical Review Letters 23, 880 (1969), URL http://doi.org/10.1103/PhysRevLett.23.880.	2028815089	✓
M. Horodecki and P. Horodecki, Phys. Rev. A 59, 4206 (1999).	2000407553	✓
B. Mazur, Rational isogenies of prime degree. Invent. Math. 44 (1978), 129–162.	2011844852	✓
Arnold, B.C, Balakrishnan, N., Nagaraja H.N., (1992), A First course in order statistics, Wiley and sons.	2318245334	✓
K. Bamba and S. D. Odintsov, JCAP 0804, 024 (2008) [arXiv:0801.0954 [astro-ph]].	2065428968	✓
Mujherjee N. P. and Bhattacharya, P., Fuzzy Groups Some Group-Theoretic Analogs, Information Science39,247-268 (1986).	1999360086	✓
Karen Suzanne Oberhauser, M. J. S. The monarch butterfly: biology and conservation. Cornell university press, 2004.	570971783	✓
A. Belloni, V. Chernozhukov, and C. Hansen. Inference for high-dimensional sparse econometric models. Advances in Economics and Econometrics: The 2010 World Congress of the Econometric Society, 3:245–295, 2013.	1720842782	✓
M. I. Aroyo, A. Kirov, C. Capillas, J. M. Perez-Mato, and H. Wondratschek, Acta Cryst. A 62, 115 (2006b).	1996820002	✓

V. Del Duca and C. R. Schmidt, Dijet Production At Large Rapidity Intervals, 4919944510 [9311290].	2003427747	✓
G. F. Giudice and A. Strumia, Nucl. Phys. B 858 (2012) 63 [arXiv:1108.6077 [hep-ph]].	2015208992	✓
C. D. Herrera, J. Kannala, P. Sturm, and J. Heikkila. A learned joint depth and intensity prior using markov random fields. In 3DTV-Conference, 2013 International Conference on, pages 17–24. IEEE, 2013.	1994295411	✓
J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani. 1-norm support vector machines. In Advances in Neural Information Processing Systems (NIPS), volume 16, pages 49–56, 2004.	2249237221	×
M. Beynon, B. Curry, and P. Morgan, ""The Dempster-Shafer theory of evidence:an alternative approach to multicriteria decision modelling"", Omega, vol. 28, no. 1, pp. 37–50, 2000.	2121042048	✓
S. Flach, M. V. Ivanchenko, and O. I. Kanakov, Phys. Rev. Lett. 95, 064102/1-4 (2005).	2004622729	✓
P. Carrasco, A. M. Cegarra, and A. R. Garzon. Nerves and classifying spaces for bicategories, 2010.	2108966476	✓
R. Sandhu, S. Dambreville, A. Tannenbaum, Particle Filtering for Registration of 2D and 3D Point Sets with Stochastic Dynamics. Pro. of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8.	2146847221	✓
S. Ji, Y. Xue and L. Carin, “Bayesian compressive sensing,” IEEE Trans. Signal Process., vol. 56, no. 6, pp. 2346–2356, 2008.	2071284784	✓
Jun Li. A degeneration formula of GW-invariants. J. Differential Geom., 60(2):199–293, 2002.	1484479264	✓
J. Milnor, Dynamics in One Complex Variable, Vieweg, Göttingen, 2000.	1603977374	✓
Marcheselli, M., Baccini, A. and Barabesi, L. (2008). Parameter estimation for the discrete stable family. Communications in Statistics - Theory and Methods 37 815–830.	2035009392	✓
A. Chantasri, J. Dressel, and A. N. Jordan, Action principle for continuous quantum measurement, Phys. Rev. A 88, 042110 (2013).	2025224750	✓
A. Strominger, Black hole entropy from near horizon microstates, JHEP 9802 (1998) 009, [hep-th/9712251].	2058588165	✓
C. I. Lazaroiu, “On the structure of open-closed topological field theory in two dimensions,” Nucl. Phys. B 603, 497 (2001) arXiv:hep-th/0010269.	2033747261	✓
Liao, L. Z., Qi, H., & Qi, L. (2004). Neurodynamical optimization. Journal of Global Optimization, 28(2), 175-195.	2340554656	✓
J. L. Lions, Quelques méthodes de résolution des problèmes aux limites non linéaires, Dunod, Paris, 1969.	1519031678	✓
M. Scully and W. E. Lamb, Jr. Quantum theory of an optical maser. Phys. Rev. Lett., 16(19):853–855, 1966.	2015518403	✓
G. Lanckriet, M. Deng, N. Cristianini, M. Jordan, W. Noble, Kernel-based data fusion and its application to protein function prediction in yeast, in: Proceedings of the Pacific Symposium on Biocomputing, Vol. 9, 2004, pp. 300–311.	2013502943	✓
Aharonson, O., Hayes, A. G., Lunine, J. I., Lorenz, R. D., Allison, M. D., Elachi, C., Dec. 2009. An asymmetric distribution of lakes on Titan as a possible consequence of orbital forcing. Nature Geoscience 2, 851–854.	2083408367	✓
K. Ito and S. S. Ravindran. A reduced-order method for simulation and control of fluid flows. Journal of Computational Physics, 143(2):403–425, 1998.	2045627558	✓
P. You, Y. Sun, J. Pang, S. Low, and M. Chen, “Battery swapping assignment for electric vehicles: A bipartite matching approach,” SIGMETRICS Performance Evaluation Review, vol. 45, no. 2, pp. 85–87, 2017.	2762188191	✓
Panjer H. (1981). Recursive evaluation of a family of compound distributions. ASTIN Bulletin 12, 22-26.	2156812602	✓
M. Ibrahim, S. Muralidharan, Z. Deng, A. Vahdat, and G. Mori. A hierarchical deep temporal model for group activity recognition. In Computer Vision and Pattern Recognition, 2016.	2259801182	✓
B. G. Saar, C. W. Freudiger, J. Reichman, C. M. Stanley, G. R. Holtom, and X. S. Xie, “Video-rate molecular imaging in vivo with stimulated raman scattering,” Science 330, 1368–1370 (2010).	2096138335	✓
J. L. Feng, C. F. Kolda and N. Polonsky, Nucl. Phys. B 546, 3 (1999) [arXiv:hep-ph/9810500]; J. Bagger, J. L. Feng and N. Polonsky, Nucl. Phys. B 563, 3 (1999) [arXiv:hep-ph/9905292]; J. A. Bagger, J. L. Feng, N. Polonsky and R. J. Zhang, Phys. Lett. B 473, 264 (2000) [arXiv:hep-ph/9911255]; H. Baer, C. Balazs, M. Brhlik, P. Mercadante, X. Tata and Y. Wang, Phys. Rev. D 64, 015002 (2001) [arXiv:hep-ph/0102156].	1997759872	✓
R. Fei, V. Tran, and L. Yang, Phys. Rev. B 91, 195319 (2015).	1590844150	✓
G. Thalhammer et al., Phys. Rev. Lett. 100, 210402 (2008)	1601643666	✓
D. S. Petrov, G. V. Shlyapnikov, and J. T. M. Walraven, Phys. Rev. Lett. 87, 050404 (2001).	2072707005	✓
Z. Bern, L. J. Dixon, D. C. Dunbar and D. A. Kosower, “Fusing gauge theory tree amplitudes into loop amplitudes,” Nucl. Phys. B 435, 59 (1995) [hep-ph/9409265].	2021404114	✓
D. Telnov and S.-I. Chu, Phys. Rev. A 79, 041401(R) (2009).	2046622440	✓
S. Gupta, Phys. Rev. D 64 (2001) 034507 [hep-lat/0010011].	2616732687	✓
T. Alazard, J.M. Delort, Global solutions and asymptotic behavior for two dimensional gravity water waves, Preprint, 2013.	1585883756	✓
M. Gastpar and M. Vetterli, “On the capacity of wireless networks: the relay case,” in Proc. IEEE Infocom, June 2002.	2097463269	✓
G. Vidal, “Efficient classical simulation of slightly entangled quantum computations,” Phys. Rev. Lett. 91 (2003).	2036604884	✓
I. Frank and J. Friedman. A statistical view of some chemometrics regression tools (with discussion). Technometrics, 35:109–148, 1993.	2079775628	✓
S. Rajagopalan and V. Vazirani. Primal-dual rnc approximation algorithms for set cover and covering integer programs. SIAM Journal of Computing, 28(2):525–540, 1998.	1988837529	✓
E. N. Parker: Cosmical Magnetic Fields: Their Origin and Their Activity, (Clarendon, Oxford 1979)	1661725509	✓

Li, J., Xin, Z.: Some uniform estimates and blowup behavior of global strong solutions to the Stokes approximation equations for two-dimensional compressible flows. J. Differ. Eqs. 221(2), 275-308 (2006).	2012319434	✓
L. Visinelli, Observational Constraints on Monomial Warm Inflation, JCAP 07 (2016) 054.	2403695403	✓
T. Kimura and V. Pestun, arXiv:1608.04651	2513775828	✓
S. Benvegnù, L. Dąbrowski: Relativistic point interaction, Lett. Math. Phys. 30 (1994), 159–167.	2053324072	✓
M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions (Dover, New York, 1970).	2120062331	✓
A. Korobeinikov, P. K. Maini, A Lyapunov function and global properties for SIR and SEIR epidemiological models with nonlinear incidence, Math. Biosci. Eng. 1 (1) (2004) 57–60.	2160057076	✓
S. J. Weidenschilling and F. Marzari. Gravitational scattering as a possible origin for giant planets at small stellar distances. , 384:619–621, December 1996.	2068108425	✓
I. Carusotto, D. Gerace, H. E. Tureci, S. De Liberato, C. Ciuti, and A. Imamoglu, Fermionized Photons in an Array of Driven Dissipative Nonlinear Cavities, Phys. Rev. Lett. 103, 033601 (2009).	2100065221	✓
I.Ya. Aref'eva, P.B. Medvedev, A.P. Zubarev, "New representation for string field solves the consistency problem for open superstring field theory,"" Nuclear Physics B, Volume 341, Issue 2.	2098143658	✓
H. Baer and X. Tata, Weak scale supersymmetry: From superfields to scattering events, . Cambridge, UK: Univ. Pr. (2006) 537 p.	1575963702	✓
Ron Kimmel and Nahum Kiryati. Finding shortest paths on surfaces by fast global approximation and precise local refinement. International Journal of Pattern Recognition and Artificial Intelligence, 10(6):643–656, 1996.	2019758632	✓
W.B. Kilgore, One-Loop Single-Real-Emission Contributions to FORMULA at Next-to-Next-to-Next-to-Leading Order, Phys. Rev. D89 (2014) 073008 [arXiv:1312.1296].	2057405276	✓
Kielpinski D Phys. Rev. A. 73 063407 (2006)	2039261148	✓
Gavrilov,L.A. FORMULA Gavrilova N.S (1991) , The Biology of life span: a quantitative approach, N.Y.:Harwood Academic Publisher.	1979363640	✓
M. Scadron, Phys. Rev. D 26, 239 (1982).	2141883609	✓
Y. Aoki, Z. Fodor, S. Katz, and K. Szabo, Phys.Lett. B643, 46 (2006), arXiv:hep-lat/0609068 [hep-lat] .	2020173052	✓
M. Glück, E. Reya, M. Stratmann, and W. Vogelsang, Phys. Rev. D 53 4775 (1996).	1550005211	✓
J.-Y. Courtois, G. Grynberg, B. Lounis and P. Verkerk, Phys. Rev. Lett. 72, 3017 (1994).	2074496196	✓
U. Leonhardt, ""Measuring the quantum state of light"", Cambridge University press, Cambridge, 1997.	1996720084	✓
M. T. Glossop and K. Ingersent, Phys. Rev. Lett. 95, 067202 (2005); Phys. Rev. B 75, 104410 (2007).	2013030742	✓
A. Giveon, A. Konechny, A. Pakman, and A. Sever, Type 0 strings in a 2-d black hole, JHEP 10 (2003) 025, [hep-th/0309056].	2070010107	✓
Hubbard, P. M., 1996. "Approximating polyhedra with spheres for time-critical collision detection". ACM Transac, 15(3), pp. 179–210.	2053212688	✓
T. Xu, L. Ma and G. Sternberg, "Practical interference alignment and cancellation for MIMO underlay cognitive radio networks with multiple secondary users," IEEE GLOBECOM, pp. 1031–1036, Dec. 2013.	2073404794	✓
T. S. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," IEEE Trans. Inform. Theory, vol. 35, no. 1, pp. 2–14, January 1989.	2071851353	✓
P. Fendley, F. Lesage, and H. Saleur, Journal of Statistical Physics 85, 211 (1996).	2007477100	✓
Akimasa Miyake and Miki Wadati. Geometric strategy for the optimal quantum search. Phys. Rev. A, 64:042317, Sep 2001.	2145249001	✓
Griffiths, G.A., Estimation of landform life expectancy, Geology, 21, 403-406, 1993.	1983389789	✓
J.-P. Kahane, Random coverings and multiplicative processes, In Fractal geometry and stochastics, II (Greifswald/Koserow, 1998), Progr. Probab. 46, 125–146, Birkhäuser, 2000.	2125142036	✓
D. C. Cox, "Antenna diversity performance in mitigating the effects of portable radiotelephone orientation and multipath propagation," IEEE Trans. Commun., vol. 31, pp. 620–628, May 1983.	2120721514	✓
Doron Zeilberger, The algebra of linear partial difference operators and its applications, SIAM J. Math. Anal. 11 (1980), no. 6, 919–932.	1991175354	✓
D. T. Limmer and D. Chandler. The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water. The Journal of Chemical Physics, 135(13):134503, 2011.	2599889364	×
M. G. Santos et al., "Cosmology with a SKA HI intensity mapping survey," PoS AASKA14 (2015) 019 [arXiv:1501.03989 [astro-ph.CO]].	1596200123	✓
L. Castillejo, R.H. Dalitz and F.J. Dyson, Low's Scattering Equation for the Charged and Neutral Scalar Theories, Phys. Rev. 101 (1956) 453-458.	2028108845	✓
S. Jennewein, M. Besbes, N. J. Schilder, S. D. Jenkins, C. Sauvan, J. Ruostekoski, J.-J. Greffet, Y. R. P. Sortais, and A. Browaeys, Phys. Rev. Lett. 116, 233601 (2016).	2345644887	✓
M. Chernicoff, J. A. Garcia, A. Guijosa and J. F. Pedraza, "Holographic Lessons for Quark Dynamics," J. Phys. G 39, 054002 (2012) [arXiv:1111.0872 [hep-th]].	2106350856	✓
C. E. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. The Annals of Statistics, 2(6):1152–1174, 1974.	1967687583	✓
R.D. Astumian : Symmetry relations for trajectories of a brownian motor, Phys. Rev. E 76, 020102 (2007).	2066644069	✓
Forest, S., Micromorphic approach for gradient elasticity, viscoplasticity, and damage. Journal of Engineering Mechanics, 135, 117-131 (2009).	2157979983	✓

L. Sironi and A. Spitkovsky, Astrophys. J. 726, 75 (2011) [arXiv:1009.0024 [astro-ph.HE]].	2032048508	✓
T. L. Barklow, arXiv:hep-ph/0312268.	2142208316	✓
M. Sharir and J. Solymosi, Distinct distances from three points, to appear in Combinatorics, Probability and Computing. Also in arXiv:1308.0814.	2100718844	✓
Ferrari, A. C., et al. Raman spectrum of graphene and graphene layers. Phys. Rev. Lett. 97, 187401 (2006).	2136334331	✓
M. Grützmann, T. Strobl, General Yang-Mills type gauge theories for p-form gauge fields: From physics-based ideas to a mathematical framework OR From Bianchi identities to twisted Courant algebroids, arXiv:1407.6759.	2107310943	✓
K. Maeda, M. Natsuume and T. Okamura, Vortex lattice for a holographic superconductor, Phys. Rev. D 81 (2010) 026002.	2111816020	✓
Y. Lai and T. Tél, Transient Chaos - Complex Dynamics on Finite-Time Scales (Springer, 2011).	654916661	✓
T. Gehrmann, Nucl. Phys. B 534, 21 (1998) [arXiv:hep-ph/9710508].	2073871748	✓
J. W. Fisher and S. Montgomery, Semiprime skew group rings, J. Algebra, 52, (1978), no. 1, 241-247	1977258644	✓
CMS Collaboration, "Search for new physics in events with same-sign dileptons and b-tagged jets in pp collisions at FORMULA TeV", JHEP 08 (2012) 110, doi:tt10.1007/JHEP08(2012)110, arXiv:1205.3933.	2790162885	✓
Oded Goldreich and Madhu Sudan. Locally testable codes and pcps of almost-linear length. J. ACM, 53(4):558–655, 2006.	2022381972	✓
J.-P. Serre, Algebraic groups and class fields, vol. 117 of Graduate Texts in Mathematics, Springer-Verlag, New York, 1988. Translated from the French.	1572771201	✓
H. Cheung and E. K. Riedel, "Energy spectrum and persistent current in one-dimensional rings," Physical Review B, vol. 40, no. 14, pp. 9498–9501, Nov. 1989.	2006367954	✓
G.-B. Huang, Q.-Y. Zhu, K. Mao, C.-K. Siew, P. Saratchandran, and N. Sundararajan, "Can threshold networks be trained directly?" IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 53, no. 3, pp. 187–191, 2006.	2161055889	✓
Hauser, Oliver P, Rand, David G, Peysakhovich, Alexander, and Nowak, Martin A. Cooperating with the future. Nature, 511(7508):220–223, 2014.	2072455842	✓
P. Arnoux, C. Mauduit, I. Shiokawa, and J. Tamura. Complexity of sequences defined by billiard in the cube. Bull. Soc. Math. France, 122(1):1–12, 1994.	2089093075	✓
C. Klix, F. Ebert, F. Weysser, M. Fuchs, G. Maret, and P. Keim, Phys. Rev. Lett. 109, 178301 (2012).	2163069494	✓
M. Srednicki, "Entropy and area," Phys. Rev. Lett. 71, 666 (1993) [hep-th/9303048].	2053387157	✓
Hughes, I.G. Velocity selection in a Doppler-broadened ensemble of atoms interacting with a monochromatic laser beam, J. Mod. Opt. 2017.	2617563275	✓
A. B. Aceves, J. V. Moloney , and A. C. Newell, Phys. Rev. A. 39 (1989) 1809.	2005701305	✓
H. Yabuki, "Feynman path integrals in the young double-slit experiment," International Journal of Theoretical Physics 25, 159–174 (1986).	2028663815	✓
L. F. Santos and M. Rigol, Phys. Rev. E 81, 036206 (2010).	2000628151	✓
W. H. Kleiner, L. M. Roth, and S. H. Autler, Phys. Rev. 133, A1226 (1964).	2047723053	✓
H.G. Bock, M. Diehl, E.A. Kostina, and J.P. Schlöder. Constrained optimal feedback control of systems governed by large differential algebraic equations. In L. Biegler, O. Ghattas, M. Heikenschloss, D. Keyes, and B. Bloemen Waanders, editors, Real-Time PDE-Constrained Optimization, pages 3–22. SIAM, 2007.	2477791619	✓
D. Larson et. al., Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-Derived Parameters, Astrophys. J. Suppl. 192 (2011) 16, [arXiv:1001.4635].	2105687315	✓
G. Leon, Y. Leyva and J. Socorro, Quintom phase-space: beyond the exponential potential, Phys. Lett. B732 (2014) 285–297, [1208.0061].	2015773494	✓
Jennings E., Baugh C. M., Li B., Zhao G.-B., Koyama K., 2012, ArXiv:1205.2698 [astro-ph.CO]	1836372023	✓
P. Hu and D. Ramanan. Bottom-up and top-down reasoning with hierarchical rectified gaussians. In Proc. IEEE Conf. Comp. Vis. Patt. Recogn., pages 5600–5609, 2016.	2346846221	✓
Whitall MWG, Gehring GA. Path integral approach to methyl group rotation. J Phys C 1987;20:1619-1639.	2084791981	✓
S.S. Gubser and I.R. Klebanov, "Absorption by Branes and Schwinger Terms in the World Volume Theory", Phys. Lett. B413 (1997) 41, hep-th/9708005.	2064448358	✓
R. Ferraro and F. Fiorini, Phys. Lett. B702, 75 (2011); R. Ferraro and F. Fiorini, arXiv:1106.6349 [gr-qc].	2052693432	✓
Gert Almkvist and George E. Andrews, A Hardy-Ramanujan formula for restricted partitions, Journal of Number Theory 38 (1991), no. 2, 135 – 144, Dedicated to the Memory of S. Ramanujan.	2016968046	✓
D. H. Lyth and D. Wands, Phys. Lett. B 524, 5 (2002) [arXiv:hep-ph/0110002].	1979147028	✓
Rapoport A, Chammah A M (1966) The game of chicken, American Behavioral Scientist 10:10-14	2050880379	✓
M. Proebster, M. Kaschub, T. Werthmann and S. Valentin, "Context-aware resource allocation for cellular wireless networks," EURASIP Journal on Wireless Communications and Networking, DOI: 10.1186/1687-1499-2012-216, Jul 2012.	2123032963	✓
Audinot, M., Pinchinat, S., Kordy, B.: Is my attack tree correct? In: ESORICS. LNCS, Springer (2017), (to appear)	2744888375	✓
A. Kuhn, M. Hennrich, and G. Rempe, Phys. Rev. Lett. 89, 067901 (2002).	2001597879	✓
I. F. Blake, Codes over certain rings, Inform. Control 20(1972), 396-404.	2016367126	✓

E. Braaten and H. W. Hammer, Phys. Rev. Lett. 91, 102002 (2003) [arXiv:nucl-th/0303038].	2157831165	✓
J. Dai, Y. Li, K. He, and J. Sun. R-FCN: Object detection via region-based fully convolutional networks. In NIPS, 2016.	2407521645	✓
Theano Development Team. 2016. Theano: A Python framework for fast computation of mathematical expressions. arXiv e-prints, abs/1605.02688.	2384495648	✓
C.W.J. Granger. Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society, pages 424–438, 1969.	2178225550	✓
M. Eckstein and P. Werner, Phys. Rev. B 82, 115115 (2010).	2075658311	✓
H. Halpern, V. Kaftal and G. Weiss, The Relative Dixmier Property in Discrete Crossed Products, J. Functional Anal. 68(1986),	1993094219	✓
H. Li and F. D. M. Haldane, Phys. Rev. Lett. 101, 010504 (2008).	2082257352	✓
Bouchaud J.-P., Potters M., Meyer M. Apparent multifractality in financial time series Eur. Phys. J. B 13 (2000) 595-599	2110077303	✓
H Nakada, Phys Rev C 68, 014316 (2003)	2005421586	✓
P. D. B. Collins, An Introduction to Regge Theory and High-Energy Physics, . Cambridge 1977, 445p.	2054027276	✓
G. Schütz, S. Sandow, Non–Abelian symmetries of stochastic processes: Derivation of correlation functions for random–vertex models and disordered–interacting–particle systems, Phys. Rev. E 49, 2726 (1994) DOI:10.1103/PhysRevE.49.2726	2012762389	✓
Waxman, E. 2010, arXiv:1010.5007	2134348647	✓
W. Wang, S. G. Hanson, Y. Miyamoto, and M. Takeda, Phys. Rev. Lett. 94, 103902 (2005).	2066919669	✓
W. Schoutens. Lévy Processes in Finance: Pricing Financial Derivatives. Wiley series in probability and statistics, Wiley, Chichester, 2003.	2002530172	✓
K. K. Kwong, Some sharp Hodge Laplacian and Steklov eigenvalue estimates for differential forms, Calc. Var. Partial Differential Equations 55 (2016), no. 2, Art. 38, 14. MR 3478292	2309623648	✓
T. Iwashyna, J. Christie, J. Kahn, and D. Asch. Uncharted paths: hospital networks in critical care. CHEST, 135(3):827–833, 2009.	2101874017	✓
A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: A unifying theory and numerical methods for characterization," IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications 47, 655–674 (2000).	2111079473	✓
]csiszar78a Csiszár, I. and J. Körner, 1978, Broadcast Channels with Confidential Messages, IEEE Trans. Inf. Theory IT-24, 339.	2144007657	✓
F. Beaudoin, J. M. Gambetta, and A. Blais, Phys. Rev. A 84, 043832 (2011).	2074456944	✓
R. Ablamowicz, B. Fauser, On the transposition anti-involution in real Clifford algebras III: The automorphism group of the transposition scalar product on spinor spaces, Linear and Multilinear algebra, 60(6), 2012, 621-644.	2080980639	✓
H. Kanao, S. Tanaka, S. Oka, M. Hirata, S. Yoshida, K. Chayama, Narrow-band imaging magnification predicts the histology and invasion depth of colorectal tumors., Gastrointest Endosc 69 (3 Pt 2) (2009) 631–636.	2038420951	✓
S.Q. Murphy et al., Phys. Rev. Lett. 72, 728 (1994).	2088746433	✓
Y. Kubo, C. Grezes, A. Dewes, T. Umeda, J. Isoya, H. Sumiya, N. Morishita, H. Abe, S. Onoda, T. Ohshima, V. Jacques, A. Dréau, J. -F. Roch, I. Diniz, A. Auffeves, D. Vion, D. Esteve, and P. Bertet. Phys. Rev. Lett. 107, 220501 (2011).	2064756338	✓
Uriel Frisch, Turbulence: The Legacy of A.N. Kolmogorov. (Cambridge University Press, Cambridge, 1995).	1611318213	✓
Joe, H. (1997), Multivariate Models and Multivariate Dependence Concepts, Boca Raton, FL: CRC Press.	2507039649	✓
A. Hanke, F. Schlesener, E. Eisenriegler, and S. Dietrich, Phys. Rev. Lett. 81, 1885 (1998b).	2054116307	✓
O. Hohm, D. Lüst and B. Zwiebach, Fortsch. Phys. 61 (2013) 926 [arXiv:1309.2977 [hep-th]].	2139390505	✓
F. Karsch and E. Laermann, Phys. Rev. D 50, 6954 (1994) .	2132089715	✓
Lev Kapitanski. Global and unique weak solutions of nonlinear wave equations. Math. Res. Lett., 1(2):211–223, 1994.	2060831697	✓
S.-W. Wei snd Y.-X. Liu, Critical phenomena and thermodynamic geometry of charged Gauss-Bonnet AdS black holes, Phys. Rev. D 87 044014 (2013). arXiv:1209.1707 [gr-qc].	2095004813	✓
S. A. Haine and J. J. Hope, Phys. Rev. A 72, 033601 (2005).	2092020630	✓
T. Cochran, A. Gerges, and K. Orr, Dehn surgery equivalence relations on 3-manifolds, Math. Proc. Camb. Phil. Soc. 131 (2001) 97-127.	2163237620	✓
K. Chou and X.-J. Wang, The Lp-Minkowski problem and the Minkowski problem in centroaffine geometry, Adv. Math. 205 (2006), 33—83.	2082611711	✓
C. S. O'Hern, D. A. Egolf, and H. S. Greenside, Phys. Rev. E 53, 3374 (1996).	2002846857	✓
Vincent, O.; Szenicer, A.; Stroock, A. D. Capillarity-driven flows at the continuum limit. Soft Matter 2016, 12, 6656–6661.	2191910464	✓