Connecting the Dots: Research Discovery using Network Analysis Algorithms

Soma Prasad
IGNOU, somaprasad1994@gmail.com

Rupak Chakravarty
Panjab University, Chandigarh, India, rupak@pu.ac.in

Connecting the Dots: Research Discovery using Network Analysis Algorithms

1.   **Soma Prasad, MLIS, IGNOU. somaprasad1994@gmail.com**
2.   **Prof. Rupak Chakravarty, Department of Library &Information Science, Panjab University, Chandigarh, India. Email: rupak@pu.ac.in, ORCID: https://orcid.org/0000-0001-5046-1663**

Abstract

The purpose of the study is to examine the process of research discovery using network analysis algorithm. To assess the paper discovery module and other associated features, the paper titled "*Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication*" as selected as "seed paper" authored by American sociologist "Ronald C Kessler", Professor of Health Care Policy at Harvard Medical School. The findings of the experimental study revealed that unlike traditional citation databases which would yield results by keyword or topic search, Inciteful creates a graph of academic papers based on selected "seed papers" which provides greater insight from it very quickly while also providing information on similar papers, important papers as well as prolific authors, institutions and journals. Research discovery using network analysis algorithms is a novel concept being tested through the Inciteful web application.

The study demonstrates a novel tool way of exploring the literature through maps and filters. The method discussed goes beyond just keyword searching, thus avoiding the problem of missing papers if wrong keywords are entered by the researchers. The approach adopted by this study gives newer insights and allow researchers and student to find access to high quality content in a more intuitive and accurate way.

Keywords: Inciteful, Paper Discovery, Network Mapping, Interdisciplinary Research

Introduction

Research discovery is one of the most common steps of literature survey or research. Researchers use academic search engines and databases to search and retrieve relevant research papers from the huge corpus of scholarly content published by the global publishers. While the searching and discovery is conventional process, it can also be achieved by implementing network analysis with visual representations. In this paper, a new process of searching through academic literature through visual networks was undertaken using the Paper Discovery tool available in the web app Inciteful. Paper Discovery tool builds a network of papers from citations, uses network analysis algorithms to analyse the network, and provides the information to enhance the research accuracy while expediting the research process on a given topic. It helps in discovering the most similar papers, important papers as well as prolific authors and institutions.

Rationale:

The vast majority of academic search engines (ASEs) focus on "importance" (as measured by number of citations) and keyword matching to retrieve their results. ASE provide statistical about who the papers cite and who cites those papers. But often the value and information in the underlying structure citations provide is not taken into account. It would be an interesting exercise to build a citation network centered around the selected paper(s). This network can be investigated further to explore the most interesting data. Each network, and the data

within, is unique to the search based on the papers selected facilitating relevant results and expedites the literature survey.

Inciteful Data Sources

Inciteful follows a unique approach to provide the most "important" papers in the graph, along with the most similar ones. This similarity measurement is from a class of algorithms called "link prediction" algorithms. Link prediction algorithms are used most commonly in social networks to recommend researchers their potential peers. Inciteful applies the same algorithm for citations credited to the peer(s)/author(s) as suggested by the system. This approach tends to surface more recent literature and helps researchers focus on the state of the art more quickly than if just looking for the top cited papers.

Inciteful pulls data from other projects engaged in coordinating, parsing, and categorizing the academic literature. These projects include Microsoft Academic Graph, Semantic Scholar, UnPaywall, Crossref, and OpenCitations. The Microsoft Academic Graph is a heterogeneous graph containing scientific publication records, citation relationships between those publications, as well as authors, institutions, journals, conferences, and fields of study. Semantic Scholar is a project developed at the Allen Institute for Artificial Intelligence. Publicly released in November 2015, it is designed to be an AI-backed search engine for academic publications. The project uses a combination of machine learning, natural language processing, and machine vision to add a layer of semantic analysis to the traditional methods of citation analysis, and to extract relevant figures, entities, and venues from papers. In comparison to Google Scholar and PubMed, Semantic Scholar is designed to highlight the most important and influential papers, and to identify the connections between them. (Wikipedia) UnPaywall harvest Open Access content from over 50,000 publishers and repositories, and make it easy to find, track, and use. Crossref makes research outputs easy to find, cite, link, assess, and reuse. OpenCitations is an independent infrastructure organization for open scholarship dedicated to the publication of open bibliographic and citation data by the use of Semantic Web (Linked Data) technologies. It is also engaged in advocacy for open citations, particularly in its role as a key founding member of the Initiative for Open Citations (I4OC).

BibTex Import/Export:

Inciteful supports both BibTeX importing as well as exporting. The import functionality is used to "seed" a graph and the export functionality is used to save relevant papers. Inciteful recognizes BibTeX entries with a DOI associated with them. Imported references must have the appropriate DOIs. Inciteful works with Zotero and Mendeley as data file can be exported as *.bib file. Data from Inciteful can also a;so be exported in a BibTeX format containing DOI by clicking on the export button located at the bottom of every table. In order to add seed any of the following methods can be adopted:
- Import from BibTex
- Search for Papers
- Import from Zotero
- Import from Mendeley

The uploaded BibTex file must have Digital Object Identifiers (DOIs) matched for each item.

Paper Discovery using the Seed Paper:
Inciteful, the online academic paper finding tool allow researchers to:

i.insert an article of relevant domain as seed paper.

ii. the tool finds most relevant scholarly articles that cite that article along with those articles that it cites by pulling data from various data sources including Semantic Scholar (The source of all the abstracts rendered on the site), Unpaywall (The source for all the Open Access links), Microsoft Academic Graph (MAG is currently the basis for all of the underlying data used to construct the graphs), Open Citations (The original source of all of the citation data) and Crossref (original source for all the non-citation paper metadata).

iii. The web app generates a graph based on this logic. Customised graph can be generated using the query editor.

iv. Results can be further filtered by keywords and distance filters to look for relevant papers to add as seed paper or you can directly add seed papers by title or DOI.

v. filter can also be applied on the most recent few years giving insight into where prospective papers should be submitted.

vi. Two types of algorithms - "Importance" in the graph and "Similarity" to the seed paper to identify the most important papers in the graph, similarity, page raking etc.

vii. Inciteful by default produces the following lists of papers:-
  - Similar paper
  - Most important paper in the graph
  - Recent paper by the top 100 authors
  - The most important recent paper

viii. Export the organized results to the citation manager software like Zotero or Mendeley in BibTex.

Inciteful builds a network around the seed paper by finding all of the papers which these papers cites and which cite the seed paper.

Where the node labeled 0 represents the seed paper, the nodes labeled 1 represent the results of the first search, and the nodes labeled 2 represent the results of the second search. In this process we also capture all of the citations between the nodes in level two as well. From here we run a few different algorithms against this graph to find the most interesting papers.

Inciteful also supports paper discovery using multiple Seed Papers. Minimum 5 paper should be added as multiple seed paper in an iterative process for higher relevancy and accuracy. Thses papers can be added from the results retrieved as additional seed papers when searching for single seed paper or they can be imported in in a batch with a BibTex extension.

Distance Filters
For graphs that are centered on a single seed paper:
  - the seed paper has a distance of 0
  - papers which either cite the seed paper or are cited by the seed paper have a distance of 1
  - papers which either cite or are cited by the papers with a distance of 1, have a distance of 2
In visual form:

For papers which are centered around multiple seed papers:
  - the "fake paper" we create has a distance of 0
  - the seed papers have a distance of 1
  - papers which either cite or are cited by the seed papers have a distance of 2

Importance Metric: Citations vs PageRank

Conventionally, "Importance" centers around the number of citations a paper have. But papers tend to take a long time to accrue citations. Therefore the "Important" papers in a graph are always going to be the older papers who have had time to accrue the needed citations. Inciteful uses PageRank algorithm as the importance metric. PageRank doesn't just count the number of citations; it tries to score based on the importance of the papers making those citations. In order to calculate PageRank, inciteful cycle through the graph a few times, each time updating importance, in order to calculate it. It provides insight into the history of a topic than if just top cited papers are considered.

Similarity Algorithm:

Inciteful uses the class of algorithms called "Link Prediction" algorithms based on the principles of bibliographic coupling and co-citations. Inciteful implements the Adamic/Adar algorithm for its bibliographic coupling metric. For the co-citation metric it is uses the Salton Index.

Research Methodology:

Paper Discovery (PD) helps in finding the relevant research papers to be included in a literature review. The Incite search supports both title and the DOI of the paper. For identifying or finding pertinent and relevant research article either any chosen paper in any research domain can be fed to the Inciteful web application or by importing directly the BibTex file into the webapp. This chosen paper serves as the "seed paper" for the further analysis. It can be searched using the Research Paper title, Microsoft Academic ID, Digital Object Identifier (DOI), arXiv ID or PubMed ID.

For this study top cited paper written by Ronald C Kessler was selected as the seed paper. Kessler was selected from the online Webometrics ranking (h-index>100) of highly influential authors based on their h-indices as calculated by Google Scholar algorithm. The seed paper for this study was ''*Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication*''. As shown in the screen-shot taken from Google Scholar, this paper has been cited 20903 times.

Inciteful's Paper Discovery tool has an ability to craft a graph based on the specific topic. It can be used in a variety of ways depending on the need such as: getting familiar with a body of literature, finding literature, rounding out a literature review and others. This study covers a simple search process with the title of selected paper that covers a topic of interest. The selected paper can be searched by its Title of the topic or the DOI of the paper. The graph thus generated provided the information regarding the most relevant articles by using this network. The results retrieved provided information including the title, authors, journal it was published in, etc. "Mental Health" was used as the filters to narrow down the search results for more specificity. The terms was taken from the 1st paper appearing in the similar papers section.

Data Analysis and Results:

Information like title, author, journal in which it was published etc regarding seed paper is mention at the top of the page. It also include links to the open access version of the article as well as the abstract. In addition to those basic stats about how many papers are citing this paper, how many papers the seed paper cites, an open access link if applicable, as well as some information about the graph are provided.

Cited By	Citing	Published	Open	Papers in	Citations in	Graph

			Access	Graph	Graph	Depth
12,984	40	2005	Yes	13,024	49,806	1

In this instance, it is a depth 1 graph, with 13024 papers and 49806 citations. Important Papers

graph clustering similar research papers

Similar Papers

Similar papers cite the same papers as the selected paper. They tend to bias towards newer papers in the field.

SN	Similar Papers	similarity	published_year	num_cited_by
	Gender and mental health. Sarah Rosenfield, Dawne M. Mouzon	5.29295	2013	13
ii	A Handbook for the Study of Mental Health: Social Contexts, Theories, and Systems Teresa L. Scheid, Tony N. Brown	5.290415	2009	11
iii	Labeling and Stigma Bruce G. Link, Jo C. Phelan	5.28613	2013	8
iv	A Handbook for the Study of Mental Health: Psychological Approaches to Mental Illness Christopher Peterson	5.27886	2009	4
v	A Handbook for the Study of Mental Health: Well-Being across the Life Course John Mirowsky, Catherine E. Ross	5.270084	2009	1
vi	The measurement of mental disorder Jerome C. Wakefield, Mark F. Schmitz	5.270084	2009	1
vii	A Handbook for the Study of Mental Health: Work and Unemployment as Stressors Mary Clare Lennon, Laura Limonic	5.270084	2009	1
viii	A Handbook for the Study of Mental Health: An Overview of Sociological Perspectives on the Definitions, Causes, and Responses to Mental Health and Illness Allan V. Horwitz	5.261308	2009	0
ix	Mental Illness and the Criminal Justice System Virginia Aldigé Hiday, Padraic J. Burns	5.261308	2014	0
x	A Handbook for the Study of Mental Health: Integrating Service Delivery Systems for Persons with a Severe Mental Illness Gary S. Cuddeback, Joseph P. Morrissey	5.261308	2009	0

Most Important Papers in the Graph

These are the "Most Important" papers as ranked by PageRank. These tend to be the older papers in the field.

SN	Most Important Papers	num_cited_by	published_year	page_rank
	Lifetime and 12-Month Prevalence of DSM-III-R Psychiatric Disorders in the United States: Results From the National Comorbidity Survey Ronald C. Kessler, ... Kenneth S. Kendler	773	1994	0.047128
	Psychiatric Disorders in America: The Epidemiologic Catchment Area Study Darrel A. Regier, Lee N. Robins	119	1990	0.029821
	Psychoactive drug use in a general population sample, Sweden: correlates with perceived health, psychiatric diagnoses, and mortality in an automated record-linkage study. C Allgulander	17	1989	0.026019
	Limitations of Diagnostic Criteria and Assessment Instruments for Mental Disorders: Implications for Research and Policy Darrel A. Regier, ... Grayson Norquist	67	1998	0.022336
	Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication Ronald C. Kessler, ... Ellen E. Walters	1748	2005	0.021725
	Effect of Intensive Compared With Moderate Lipid-Lowering Therapy on Progression of Coronary Atherosclerosis A Randomized Controlled Trial Steven E. Nissen, ... Anthony N. DeMaria	35	2004	0.012875
	The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). Ronald C. Kessler, T. Bedirhan Üstün	402	2004	0.01279
	Introduction To Variance Estimation Kirk M. Wolter	62	1985	0.009552
	Revised Prevalence Estimates of Mental Disorders in the United States: Using a Clinical Significance Criterion to Reconcile 2 Surveys' Estimates William E. Narrow, ... Darrel A. Regier	90	2002	0.009192
	The National Comorbidity Survey Replication (NCS-R): background and aims.	192	2004	0.008989

Ronald C. Kessler, Kathleen R. Merikangas			

The effect of similar papers is the inverse of important papers. Since papers which are published after the seed paper have a better chance of citing the same papers, these papers tend to be more recent papers in the field. Under the Similar Papers section there is an Other Data section. This section is used to highlight other interesting information about the graph. Such as most important authors, institutions, and journals. It helps to get an idea of where the work in this space is being done, by whom, and in what journals. By adding seed paper and/or applying filters is an iterative process for achieving a graph with better precision. keyword filters help to find papers relevant to the query to make the graph more tailored to search. To ensure good results, it is best to add at least five papers which cover the specific topic. This new graph should be more focused on the topic than the previous one. In this study, after using keyword filter, 7 more papers are added to the graph. "Mental Health" was used as the filters to narrow down the search results for more specificity. The terms was taken from the 1st paper appearing in the similar papers section. After creating new graph will find a similar papers table again. And it is much more targeted to the previous one. These papers cite the same papers as a selected paper. They tend to bias towards newer papers in the field.

Similar paper	num_cited_by	similarity	published_year
Volunteer Work and Hedonic, Eudemonic, and Social Well-BeingJoonmo Son, John Wilson	1	0.643998	2012
Cultural definitions of emotional problems: Impact on problem experience, care seeking, sources of care, and satisfaction with care among Latino community members living in the United StatesFrederick O. Rasmussen	0	0.643998	2011
A Handbook for the Study of Mental Health: Social Contexts, Theories, and SystemsTeresa L. Scheid, Tony N. Brown	460	0.487144	2009
Gender and mental health.Sarah Rosenfield, Dawne M. Mouzon	257	0.487144	2013
The descriptive epidemiology of commonly occurring mental disorders in the United States.Ronald C. Kessler, Philip S. Wang	141	0.487144	2008
Psychiatric epidemiology: it's not just about counting anymore.Thomas R. Insel, Wayne S. Fenton	59	0.487144	2005
Prevalence and treatment of mental disorders in Lebanon: a national epidemiological surveyElie G. Karam, ... Ronald C. Kessler	79	0.487144	2006
Errors in Assessing DSM-IV Substance Use DisordersBridget F. Grant, ... George E. Woody	34	0.487144	2007
Te Rau Hinengaro: The New Zealand Mental Health SurveyMark Oakley-Browne, ... T. B. Üstün	35	0.487144	2008
A Handbook for the Study of Mental Health: Psychological Approaches to Mental IllnessChristopher Peterson	29	0.487144	2009

Table: Similar article after keyword filtering and multiple seed paper

Important papers

These are the "Most Important" papers as ranked by PageRank. These tend to be the older papers in the field.

Important papers	page_rank	published_year	num_cited_by
Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey ReplicationRonald C. Kessler, ... Ellen E. Walters	0.061045	2005	12985
Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey ReplicationRonald C. Kessler, ... Ellen E. Walters	0.061045	2005	12985
National Institute of Mental Health diagnostic interview schedule: Its history, characteristics, and validity.Lee N. Robins, ... Kathryn S. Ratcliff	0.028117	1981	304
The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI).Ronald C. Kessler, T. Bedirhan Üstün	0.016181	2004	3446
Lifetime and 12-Month Prevalence of DSM-III-R Psychiatric Disorders in the United States: Results From the National Comorbidity SurveyRonald C. Kessler, ... Kenneth S. Kendler	0.012681	1994	1377
Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys.Koen Demyttenaere, ... Somnath Chatterji	0.012502	2004	2453
Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey ReplicationRonald C. Kessler, ... Ellen E. Walters	0.008893	2005	2272
A twenty-two item screening score of psychiatric symptoms indicating impairment.Thomas S. Langner	0.007494	1962	71
Twelve-Month Use of Mental Health Services in the United States Results From the National Comorbidity Survey ReplicationPhilip S. Wang, ... Ronald C. Kessler	0.006425	2005	2107
Utilization of Health and Mental Health Services: Three Epidemiologic Catchment Area SitesSam Shapiro, ... Darrel A. Regier	0.006295	1984	85

Table: Most important article after keyword filtering and multiple seed paper

The top papers tend to focus on the "DSM IV disorder", which makes sense because we started with that paper. Finally, the other data section shows that who are publishing the paper, their affiliations, as well as the publication venues. It also helps to identify who the field experts are? and potential journal where we can submit the manuscript to. These Other data section like top author, institute and journal are then compare with the previous result in terms of num_paper and its page rank to observe the differences and analyses the result.

Comparing top 10 Author's num_paper and PageRank before and after keyword filtering and adding multiple seed paper:

Author's num_paper indicated to the number of related papers published by that author and page rank is the PageRank of the author based on selected seed paper. From the table it can be identified that num_page of particular author is continuously increasing after the process and PageRank varies depending on the seed paper. There are also two author named Michael R. Hughes and Suzann Eshleman are not listed in the top author section after the process. It may be because their contribution does not remains related to required paper after filter application and adding multiple papers based on the requirement.

AUTHOR	Num_paper		Total Page rank	
	BEFORE	AFTER	before	after
Ronald C. Kessler	108	520	0.049127	0.063287
Darrel A. Regier	4	22	0.025719	0.011739
Hans-Ulrich Wittchen	44	109	0.007423	0.011313
Kenneth S. Kendler	33	68	0.015913	0.007082
Katherine A. McGonagle	1	3	0.001964	0.00068
Shanyang Zhao	1	6	0.001964	0.001677
Michael R. Hughes	1	-	0.001964	-
Suzann Eshleman	1	-	0.001964	-
Christopher B. Nelson	1	4	0.001964	0.000948
Olga Demler	5	10	0.002377	0.004686

Table: Companions of Author's num_page and PageRank

Comparing top 10 Institute's num_page and PageRank before and after filter and adding multiple seed paper:

Institute num_page indicated to the number of papers published in that institute and page rank is the PageRank of that institute based on selected seed paper. From the table it can be identified that num_page of particular institute is continuously increasing after the process and PageRank varies depending on the seed paper.

Affiliation	Num_paper	Total Page rank

	BEFORE	AFTER	before	after
Harvard University	530	1182	0.112771	0.17827
University of Michigan	285	673	0.080886	0.058864
Max Planck Society	60	97	0.064048	0.038775
National Institutes of Health	250	422	0.059554	0.054681
Karolinska Institutet	104	146	0.030231	0.006706
University of California, San Diego	127	260	0.024389	0.014403
University of Pittsburgh	165	275	0.022178	0.019122
University of Washington	189	432	0.020399	0.021775
Columbia University	265	623	0.019654	0.045356
Washington University in St. Louis	99	151	0.01813	0.05188

Table: Comparisons of Institute's num_page and PageRank

Comparing top 10 Journal's num_page and PageRank before and after filter and adding multiple seed paper:

Journal's num_paper indicated to the number of related papers published in that journal and page rank is the PageRank of the journal based on selected seed paper. From the table it can be identified that num_page of particular journal is continuously increasing after the process and PageRank varies depending on the seed paper.

Name of Journal	Num_paper		Total Page rank	
	BEFORE	AFTER	before	after
Archives of General Psychiatry	40	159	0.12346	0.185866
International Journal of Methods in Psychiatric Research	21	99	0.040308	0.038988
American Journal of Public Health	29	138	0.031076	0.012645
JAMA	13	60	0.024258	0.04546
Psychological Medicine	138	274	0.020918	0.021272
The New England Journal of Medicine	7	27	0.011168	0.00947
Biological Psychiatry	88	112	0.011045	0.007074
American Journal of Psychiatry	42	159	0.010747	0.027541
Journal of Affective Disorders	265	495	0.00977	0.012028

Psychiatry Research-neuroimaging	168	260	0.008508	0.005776

Table: Comparisons of Journal's num_page and PageRank

Discussion:

Huge differences have been discovered before and after the results. num_cited of the paper after adding multiple seed paper and filtering shows tremendous growth in number. This implies the relevancy of paper based on requirement. The num_paper of Author, institute and Journal is continuously increasing after keyword filtering and adding multiple seed papers whereas page rank in most cases goes down but, in some cases, slightly increased. It means top authors, institutes and journals list suggested after keyword filtering and adding multiple seeds have published maximum number of related seed paper and are more précised and relevant to the requirement as compared to without filtering and single title. PageRank is a complementary method to citation analysis which allows to identify publications referenced by highly cited articles. It depends on significance of the article. There is no direct relation of the PageRank of an article with keyword filter and multiple seed paper. Inciteful offers an innovative way of getting familiar with a body of most recent literature centered around a topic that is directly or indirectly related to current work or a paper in progress. It provides useful information regarding similar papers have the greatest number of overlapping citations with the prospective paper and which may not have been cited. Inciteful help researchers to visualise academic papers as academic graphs. It generates one big graph with the papers serving as the nodes and citations serving as the edges connecting the vertices. This way of conceptualizing a body of literature corresponds to applying graph theory a field of Mathematics, to academic literature addresses complicated search strategies involving multiple Boolean operators.

Conclusion

Inciteful assists to locate the scholarly publication related to the particular topics. To explore and categories the academic literature this free online tool uses the citation. After using the paper of interest, the inciteful surfaces the most relevant academic literature based on open metadata and citations. Paper discovery need one paper as seed paper, but it work best if at least five more papers can import in a batch with a BibTex File. Similar paper can also be used as additional seed papers for more relevant results. Tools produces the results like similar papers, most important papers in the graph based on PageRank, Recent papers by top authors and the most important recent papers. It can be furthered filtered to look for relevant papers to add as seed papers. Num_page and PageRank of author, institute and journal varies after keyword filter and adding multiple paper to seed paper. The novel approach of identifying and retrieving research papers starting with seed paper(s), graph cluster mapping and by adopting iterations was found to be practically useful for the researchers for research discovery.

References:

i. Belford, R. (2014). Evaluating Library Discovery tools through a music lens. *Library resources & technical services , 58* (1), 49-72.

ii. Laura Icela Gonzalez Perez, M. S.-P. (2016). Discovery tools for open access respositiories: a literature mapping. *Proceedings of the Fourth International Conference on Technological Ecosystems for enhancing Multiculturality* , 299-305p.

iii. Shneiderman, B. (2002). Inventing Discovery tools: combining information visualization with data mining. *Information visualisation , 1* (1), 5-12.

iv. inciteful. (n.d.). Inciteful.Xyz. Retrieved September 16, 2021, from https://inciteful.xyz/

Appendix

Technical Terms	Description
distance	The distance away (as an undirected graph) this paper lies from the seed paper.
page_rank	The PageRank of this paper.
adamic_adar	The Adamic/Adar score of the paper. Adamic/Adar is a link prediction algorithm which, when used in this context, can help to detect similarity between papers.
resource_allocation	Similar to Adamic/Adar, the Resource Allocation score of the paper is a link prediction algorithm which, when used in this context, can help to detect similarity between papers.
cocite	The co-citation score between two papers. We use the Salton Index to calculate the score.
conference_series_name	The name of the conference in which the paper was presented.
conference_instance_name	The instance of the conference in which the paper was presented. This is more specific than the conference_series_name
sequence	The sequence of the author in the paper.
paper_id	The unique id of the paper.
title	The text which was searched.
bm25(title_search)	Returns a real value indicating how well the current row matches the full-text query. The better the match, the numerically smaller the value returned.
term	The porter-stemmed terms which appear in the titles.
doc_count	The number of papers in which these terms appear.