

ISSI

2021



18th INTERNATIONAL  
CONFERENCE ON  
SCIENTOMETRICS  
& INFORMETRICS

**ISSI2021**

Virtual event

12–15 July 2021

KU Leuven, Belgium

**PROCEEDINGS**





**2021**

**18th INTERNATIONAL  
CONFERENCE ON  
SCIENTOMETRICS  
& INFORMETRICS**

**ISSI2021**

**Virtual event**

**12–15 July 2021**

KU Leuven, Belgium

**PROCEEDINGS**

**18th INTERNATIONAL CONFERENCE  
ON SCIENTOMETRICS & INFORMETRICS**

**ISSI2021**

**12–15 July 2021**  
KU Leuven, Belgium

**PROCEEDINGS**

**Editors**

Wolfgang Glänzel, Sarah Heeffer, Pei-Shan Chi, Ronald Rousseau

- © Authors. No part of this book may be reproduced in any form without the written permission of the authors
- © International Society for Scientometrics and Informetrics (I.S.S.I.)

ISBN 9789080328228 – ISSN 2175-1935

July 2021 Printed in Belgium

## **Sponsors and Funding**

Diamond sponsor – Elsevier

Gold sponsor – Digital Science

Gold sponsor – RISIS

Silver sponsor – JDIS

Silver sponsor – MIT Press

Silver sponsor – Clarivate ISI

Bronze sponsor – Springer Nature

Bronze sponsor – FRONTIERS

Research Foundation – Flanders (FWO)

## **Organisers**

Koenraad Debackere, KU Leuven (Belgium), *Chair*

Wolfgang Glänzel, KU Leuven (Belgium)

Bart Thijs, KU Leuven (Belgium)

Tim Engels, UAntwerp (Belgium)

## **Programme Committee: Chairs**

Ronald Rousseau, KU Leuven (Belgium), *Chair*

Lin Zhang, Wuhan University (China)/KU Leuven (Belgium)

Henk F. Moed, Sapienza University of Rome (Italy)

## **Program Committee: Members**

Kevin Boyack, SciTech Strategies Inc. (USA), *Chair Workshops and tutorials*

Claudia Gonzalez Brambila, ITAM (Mexico)

Julie Callaert, KU Leuven (Belgium)

Pei-Shan Chi, KU Leuven (Belgium), *Poster Session*

Cinzia Daraio, Sapienza University of Rome (Italy)

Jacqueline Leta, UFRJ (Brazil), *Poster Session*

Raf Guns, UAntwerp (Belgium)

Stefanie Haustein & Juan Pablo Alperin, Ottawa (Canada)

Sarah Heeffter, KU Leuven (Belgium)

Omwoyo Bosire Onyancha, University of South Africa (South Africa)

Walter Ysebaert, VUB (Belgium)

Lin Zhang, Wuhan University (China)/KU Leuven (Belgium)

### **Doctoral Forum Committee**

Jens Peter Andersen, Aarhus University (Denmark), *Chair*

Guillaume Cabanac, University of Toulouse (France)

Vincent Larivière, Université de Montréal (Canada)

Andrea Scharnhorst, KNAW (Netherlands)

Cassidy Sugimoto, Indiana University (USA)

Alesia Zuccala, University of Copenhagen (Denmark)

### **Eugene Garfield Doctoral Dissertation Award 2021 Committee**

Birger Larsen, University of Aalborg (Denmark), *Chair*

Elías Sanz Casado, Carlos III University of Madrid (Spain)

Gemma Derrick, Lancaster University (UK)

Nees Jan van Eck, Leiden University (Netherlands)

### **Committee for the ISSI Student Travel Award**

Kevin Boyack, SciTech Strategies Inc. (USA), *Chair*

Aparna Basu, South Asian University (India)

Cinzia Daraio, Sapienza University of Rome (Italy)

## Scientific Committee

Abramo Giovanni	Caviggioli Federico	Gregori Martina
Adams Jonathan	Chang Ching-Chun	Guevara Miguel
Adie Euan	Chaurasia Neeraj	Gumpenberger Christian
Aguillo F. Isidro	Chavarro Diego	Gzoyan Edita
Ahlgren Per	Checchi Daniele	Halevi Gali
Ajiferuke Isola	Chen Chaomei	Hammarfelt Björn
Akbaritabar Aliakbar	Chen Dar-Zen	Haunschild Robin
Aman Valeria	Chen Yue	Hinze Sybille
Amez Lucy	Chincilla-Rodriguez Zaida	Holmberg Kim
Andersen Jens Peter	Ciarli Tommaso	Hörlesberger Marianne
Archambault Eric	Colavizza Giovanni	Hornbostel Stefan
Arvanitis Rigas	Confraria Hugo	Hu Xiaojun
Astrom Fredrik	Costas Rodrigo	Huang Mu-Hsuan
Ausloos Marcel	Côté Grégoire	Ingwersen Peter
Azagra-Caro Joaquín M.	Criado Regino	Ivancheva Ludmila
Badin Luiza	D'Angelo Ciriaco Andrea	Jana Siladitya
Baiget Tomas	Dastidar Prabir G.	Jeng Wei
Ball Rafael	De Bellis Nicola	Jimenez-Contreras Evaristo
Ballester-Gonzalez Omar	De Bordes Carole	Jonkers Koen
Bartol Tomaz	De Moya Felix	Jovanovic Milos
Bauer Guntram	Dehdari-rad Tahereh	Kajikawa Yuya
Benner Mats	Deinzer Gernot	Katz Sylvan
Bhattacharya Sujit	Díaz-Faes Adrián A.	Kelchtermans Stijn
Bianchi Federico	Ding Ying	Klavans Dick
Bollen Johan	Donner Paul	Konkiel Stacy
Bonaccorsi Andrea	Fatchur Rochim Adian	Kousha Kayvan
Borner Katy	Ferreira-Goncalves Márcia	Krauskopf Manuel
Bornmann Lutz	Franssen Thomas	Kulczycki Emanuel
Boshoff Nelius	Frietsch Rainer	Kumaravel J P S
Bouabid Hamid	Furner Jonathan	Lamers Wout
Bowman Timothy	Gallart Jordi Molas	Laudel Grit
Buehrer Susanne	Garcia Romero Antonio	Laurens Patricia
Cabezas-Clavijo Álvaro	Gauch Stephan	Lawson Cornelia
Calero-Medina Clara	Getz Daphne	Lemke Steffen
Campbell David	Ghiasi Gita	Leng Rhodri
Cañibano Carolina	Glänzel Wolfgang	Lepori Benedetto
Cao Cong	Gläser Jochen	Li Jiang
Carayol Nicolas	Gok Abdullah	Licea Judith
Casnici Niccolò	Gorraiz Juan	Liu Yuxian
Catalano Giuseppe	Gorry Philippe	Lo Szu-Chia

López-Illescas Carmen	Prozesky Heidi	Sterzi Valerio
Luwel Marc	Qiu Junping	Sun Yuan
Maisano Domenico	Raffo Julio D.	Tang Li
Markusova Valentina	Rafols Ismael	Tesch Jakob
Martín-Martín Alberto	Rao Ravichandra	Thelwall Mike
Maynard Diana	Reale Emanuela	Traag Vincent
Mayr Philipp	Rigby John	Tunger Dirk
Mègnigbêto Eustache	Rigolin Camila Carneiro Dias	Van den Besselaar Peter
Meho Lokman I.	Robinson Garcia Nicolas	Van der Meulen Barend
Melkers Julia	Roche Ivana	Van Leeuwen Thed
Meoli Michele	Romanello Matteo	Van Looy Bart
Milojevic Stasa	Ronda-Pupo Guillermo	Van Raan Anthony
Mongeon Philippe	Rons Nadine	Vancauwenbergh Sadia
Morillo Fernanda	Rotolo Daniele	Vargas-Quesada Benjamin
Moskaleva Olga	Rousseau Ronald	Velden Theresa
Mouton Maria	Rousseau Sandra	Vinkler Peter
Mugabushaka Alexis Michel	Ruggiero John	Visser Martijn
Mugnaini Rogerio	Ruocco Giancarlo	Wagner Caroline
Must Ülle	Sachwald Frederique	Waltman Ludo
Mutz Ruediger	Sahoo Bibhuti	Wang Lili
Nane Tina	Sainte-Marie Maxime	Wang Qi
Nederhof Anton	Salinas Daniel	Wang Xianwen
Nieminen Pentti	Sangam Shivappa Lingappa	Welch Eric
Noyons Ed	Sargsyan Shushanik	Winnink Jos
Ochsner Michael	Scherngell Thomas	Wolfram Dietmar
Olmeda-Gómez Carlos	Schiebel Edgar	Woolley Richard
Orozco Luis Antonio	Schmidt Marion	Wouters Paul
Ortega José Luis	Schmoch Ulrich	Wroblewski Angela
Pandiella Andres	Schneider Jesper	Wu Jinshan
Pellegrino Gabriele	Schubert Andrés	Wu Yishan
Perianes-Rodríguez Antonio	Schubert Torben	Yan Erjia
Peritz Cheila, Bluma	Seeber Marci	Yang Liying
Peset Fernanda	Shapira Philip	Ye Fred Y.
Peter Viola	Shen Zhesi	Yegros Alfredo
Peters Isabella	Shu Fei	Youtie Jan
Pinheiro Diogo	Sivertsen Gunnar	Zahedi Zohreh
Pislyakov Vladimir	Small Henry	Zhao Dangzhi
Pontille David	Smit Nynke-Jo	Zhao Yajuan
Porter Alan L.	Soós Sándor	Zhou Ping
Pouris Anastassios	Sorz Johannes	Zirulia Lorenzo
Prathap Gangan	Srivastava Divya	



## PREFACE

The 18th International Society of Scientometrics and Informetrics Conference is held on 12–15 July 2021 and organized by KU Leuven in close collaboration with the university of Antwerp under the auspices of ISSI the International Society for Scientometrics and Informetrics.

Due to the constraints imposed by the COVID-19 pandemic situation and the resulting risks for planning and organising a full-fledged conference with physical presence of attendees, the organisers decided to go virtual this time. This decision warranted a smooth organisation without unpredictable events and unnecessary modifications and adjustments during the preparation process. However, organising a virtual event of this size has posed a true challenge for the organisers. We were very pleased that despite the various different time-zones in the world, we could organise part of the conference, notably all keynotes, invited talks and plenary sessions as well as the workshops and tutorials and, last not least, all ceremonies as live events. The special tracks and the usual “parallel sessions” had, however, to be organised using pre-recorded on-line video streams. Making a virtue of necessity, we can provide all talks and the discussion for three months after the conference and thus make the conference a hopefully sustainable event.

The goal of ISSI2021 is to provide an international open forum for scholars and practitioners in the domain of informetrics, bibliometrics, scientometrics, webmetrics and altmetrics to discuss new research directions, advanced methods and theories, and to highlight the best research in this area. In order to achieve this goal, we asked researchers worldwide to submit original research manuscripts, particularly full papers, research-in-progress papers or posters, to propose and organise tutorials and workshops, with a special emphasis on the future of this area and on its interdisciplinary links with other fields. We succeeded in attracting a sufficient number of contributions to organise two special tracks, one on bibliometric approaches to measure and evaluate interdisciplinary research (IDR) organised by KU Leuven and one on the bibliometrics of social sciences and humanities (SSH) organised by our partner at University Antwerp. University Antwerp also took the opportunity to introduce their project of publishing on a new handbook on “Research Assessment in the Social Sciences”.

We would like to thank the three keynote speakers, who have accepted our invitation, namely Katy Börner from Indiana University in Bloomington (USA), Albert-László Barabási from Northeastern University in Boston (USA), and David Sweeney, Executive Chair of Research England (UK).

Despite the pandemic-induced constraints, we could accept about 230 contributions, roughly two thirds of which are full and research-in-progress papers and nearly one third is present in pre-recorded poster sessions.

The assignment of contributions to session topics was facilitated by the organisation as a virtual event since time constraints did not influence the schedule. We have chosen the following major themes.

- » Research evaluation & bibliometrics in support of science policy
- » Effects of research funding
- » Patent analysis
- » Individual-level bibliometrics
- » Collaboration and mobility
- » Domain studies and regional issues
- » Open Science, Open Access and editorial impact
- » Gender and equality studies
- » Webmetrics, altmetrics and media impact of research
- » Advanced methods in citation analysis
- » Data sources and data processing
- » Document and journal analysis
- » Network analysis

Furthermore, six workshops deal with the tracing of epistemic change, with creation and application of models, cited reference analysis, national and institutional research assessment, and collaborative archive and data research environment. This is completed by two tutorials. The organisers look forward to the future publication of the outcomes of these workshops.

All accepted presentations and posters are incorporated in the conference proceedings. The first part comprises invited and full papers as well as research-in-progress papers, while the second part is devoted to posters.

At this place, we would like to express our thanks to all participants and contributors for their understanding for the constraints due to the special situation and their active support. Our thanks are due to the ISSI board for its trust and support. We also thank the Leuven Congress Office – PC for their enormous effort and dedicated work and MEE-PLE for providing the online platform. In particular, we thank Liesbeth Michiels, Marie-Laure Bettens and Ann Moerenhout.

On behalf of the organisers and the programme committee of ISSI2021  
Wolfgang Glänzel

## INDEX OF PAPERS (FULL PAPERS AND RESEARCH IN PROGRESS)

Do hijacked journals attract dishonest authors? ..... 1 <i>Anna Abalkina</i>	1
Do the propensity and drivers of academics' engagement in research collaboration with industry vary over time? ..... 7 <i>Giovanni Abramo, Francesca Apponi and Ciriaco Andrea D'Angelo</i>	7
Gendered impact of COVID-19 pandemic on research production: a cross- country analysis based on bioRxiv ..... 19 <i>Giovanni Abramo, Ciriaco Andrea D'Angelo and Ida Mele</i>	19
Research Unit Size and Internal Co-publications ..... 31 <i>Hind Achachi and Hamid Bouabid</i>	31
Scholarly publishing at US federally funded research & development laboratories: influences on public-private science ..... 41 <i>Jeffrey M. Alexander, Cassidy R. Sugimoto and Vincent Larivière</i>	41
Governmental mobility programmes and academic performance in Spain ..... 47 <i>Patricia Alonso Álvarez, Jorge Mañana Rodríguez and Elías Sanz Casado</i>	47
A Visual Analytics Environment for the Assessment of Information Quality of Performance Models ..... 53 <i>Marco Angelini, Cinzia Daraio and Luca Urban</i>	53
Universities and patent trolls: an Empirical Study of University Patent Infringement Litigation in the United States ..... 59 <i>Grazia Sveva Ascione, Laura Ciucci, Claudio Detotto and Valerio Sterzi</i>	59
Measuring the impact of clinical data in terms of data citations by scientific publications ..... 71 <i>Yongmei Bai and Jian Du</i>	71
An empirical analysis of existence of Power Laws in social media mentions to scholarly articles ..... 81 <i>Sumit Kumar Banshal, Aparna Basu, Vivek Kumar Singh, Hiran H Lathabai, Solanki Gupta and Pranab K. Muburi</i>	81

Data sources and their effects on the measurement of open access. Comparing Dimensions with the Web of Science .....	93
<i>Isabel Basson, Marc-André Simard, Zoé Aubierge Ouangré, Cassidy R. Sugimoto and Vincent Larivière</i>	
Journal flipping to Open Access: The perception of Spanish journal managers .....	99
<i>Núria Bautista-Puig, Carmen López-Illescas and Henk F. Moed</i>	
The Place of Preprint Citations in the IMRaD Structure: a Study of PLOS Journals .....	105
<i>Marc Bertin and Iana Atanassova</i>	
Augmenting context-aware citation recommendations with citation and co-authorship history .....	115
<i>Anubrata Bhowmick, Ashish Singhal and Shenghui Wang</i>	
A systematic method for identifying references to academic research in grey literature .....	121
<i>Matthew S. Bickley, Kayvan Kousha and Michael Thelwall</i>	
Research funding and scientific performance: measuring the impact of a PhD scholarship program .....	133
<i>Adriana Bin, Sergio Salles-Filho, Ana Carolina Spatti, Jesús Pascual Mena-Chalco and Fernando Antonio Basille Colugnati</i>	
Gender differences in scientific careers: A large-scale bibliometric analysis .....	145
<i>Hanjo Boekhout, Inge van der Weijden and Ludo Waltman</i>	
How to do research on the societal impact of research? Studies from a semantic perspective .....	157
<i>Andrea Bonaccorsi, Filippo Chiarello, Nicola Melluso and Gualtiero Fantoni</i>	
Detecting interdisciplinarity in top-class research using topic modeling .....	169
<i>Andrea Bonaccorsi, Nicola Melluso and Francesco Alessandro Massucci</i>	
Stability through Constant Turnover: The Replacement Rate of the Scientific Workforce .....	181
<i>Clara Boothby, Staša Milojević, Vincent Larivière, John Walsh, Filippo Radicchi and Cassidy Sugimoto</i>	
Systematic characterisation and operationalisation of <i>cyber-physical convergence</i> in the context of Industry 4.0 .....	187
<i>Tausif Bordoloi, Philip Shapira and Paul Mativenga</i>	

Beyond the Web of Science: an overview of Brazilian papers indexed by regionally relevant databases .....	193
<i>André Brasil</i>	
An approach to evaluate the success of individual research articles based on the Anna Karenina Principle .....	205
<i>Zhen Cai and Wen Lou</i>	
GeoAcademy: algorithm and platform for the automatic detection and location of geographic coordinates in scientific articles .....	211
<i>Jesús Cascón-Katchadourian, Francisco Carranza-García, Carlos Rodríguez-Domínguez and Daniel Torres-Salinas</i>	
Star inventors: quantity and quality in the EOB model .....	217
<i>Federico Caviggioli and Boris Forthmann</i>	
Connecting Brain and Heart: Artificial Intelligence for Sustainable Development ..	229
<i>Diego Chavarro, Jaime Andrés Pérez-Taborda and Alba Ávila</i>	
A Fine-Grained Annotation Scheme for Research Contribution in Academic Literature .....	241
<i>Haihua Chen and Bhavya Nandana Kanuboddu</i>	
Measuring radical and incremental innovation of digital products Updates on text analysis of mobile apps .....	249
<i>Jin Chen, Xiaohu Li and Lele Kang</i>	
An Article-based Cross-disciplinary Study of Reference Literature for Indicator Improvement .....	259
<i>Pei-Shan Chi and Wolfgang Glänzel</i>	
Patterns of Knowledge Diffusion via Research Collaboration on a Global Level ....	269
<i>Zaida Chinchilla-Rodríguez, Jialin Liu and Yi Bu</i>	
Countries' research priorities in relation to the Sustainable Development Goals .....	281
<i>Hugo Confraria, Ed Noyons and Tommaso Ciarli</i>	
Dissension or consensus? Management and Business Research in Latin America and the Caribbean .....	293
<i>Julián D. Cortés</i>	
ISBNs as identifiers for books in research evaluation .....	299
<i>Eleonora Dagienė and Kai Li</i>	

An Ontology-Based Semantic Design for Good Evaluations of Research Practices .	305
<i>Cinzia Daraio and Alessio Vaccari</i>	
Open access policies and mandates and their practical implementation in Spanish public universities .....	317
<i>Daniela De Filippo and Jorge Mañana-Rodriguez</i>	
Excellence, Interdisciplinarity and Collaboration in Research Networks. Evidence from the Evaluation of the Special Research Programme (SFB) of the Austrian Science Fund (FWF) .....	329
<i>Michael Dinges, Barbara Heller-Schub, Robert Hawlik, Thomas Scherngell, Anna Wang, Bart Thijs and Wolfgang Glänzel</i>	
Co-link analysis as a monitoring tool: A webometric use case to map the web relationships of research projects .....	339
<i>Jonathan Dudek, David G. Pina and Rodrigo Costas</i>	
Early insights into the Arabic Citation Index .....	345
<i>Jamal El-Ouahi</i>	
Network Effects and Research Collaborations: Evidence from IMF Working Paper Co-authorship .....	357
<i>Dennis Essers, Francesco Grigoli and Evgenia Pugacheva</i>	
Clustering social sciences and humanities publications: Can word and document embeddings improve cluster quality? .....	369
<i>Joshua Eykens, Raf Guns and Tim Engels</i>	
Status and Challenges of Open Government Data in Nigeria: An Informetric Analysis of Websites of Government Ministries and Organizations .....	375
<i>Ifeanyi J. Ezema</i>	
User Engagement with Scholarly Twitter Mentions: A Large-scale and Cross-disciplinary Analysis .....	387
<i>Zhichao Fang</i>	
An Alternative Analysis Method for Measuring the Impact of Academic Papers Shared on Social Media and the Number of Citations Obtained based on a Support Vector Machine Algorithm .....	393
<i>Adian Fatchur Rochim, Faisal Rizky Rahadian and Dania Eridani</i>	

Doing as the Romans do? A comparative analysis of the thematic profiles of newcomers and already existing faculty at universities .....	401
<i>Márcia R. Ferreira and Rodrigo Costas</i>	
The Software Engineering Observatory Portal .....	407
<i>Mivian M. Ferreira, Bruno L. Sousa, Kecia A. M. Ferreira and Mariza A. S. Bigonha</i>	
Tracing epistemic effects of research governance: a mixed-method approach .....	413
<i>Thomas Franssen, Thed van Leeuwen, Ingeborg Meijer and Ismael Rafols</i>	
Using Citation Bias to Guide Better Sampling of Scientific Literature .....	419
<i>Yuanxi Fu, Jasmine Yuan and Jodi Schneider</i>	
Two-Dimensional Mapping of University Profiles in Research .....	425
<i>Joel Emanuel Fuchs and Thomas Heinze</i>	
Academic Leadership and University Performance: Do Russian Universities Improve after They Are Headed by Top Researchers? .....	435
<i>Daria Gerashchenko</i>	
Where international development and gender equality meet in science: a bibliometric analysis .....	447
<i>Gita Ghiasi, Matthew Harsh, Tanja Tajmel and Vincent Larivière</i>	
Improving the precision of subject assignment for disparity measurement in studies of interdisciplinary research .....	453
<i>Wolfgang Glänzel, Bart Thijs and Ying Huang</i>	
Measuring the excellence contribution at the journal level: An alternative to Garfield's impact factor .....	465
<i>Juan Gorraiz, Ursula Ulrych, Wolfgang Glänzel, Wenceslao Arroyo-Machado and Daniel Torres-Salinas</i>	
Revisiting the Obsolescence Process of Individual Scientific Publications: Operationalisation and a Preliminary Cross-discipline Exploration .....	477
<i>Zhenyu Gou, Fan Meng, Zaida Chinchilla-Rodríguez and Yi Bu</i>	
Peer Review versus Bibliometrics. Do the Two Methods Produce Identical Results in Measuring Academic Reputation? .....	489
<i>Katerina Guba, Mikhail Sokolov and Angelika Tsivinskaya</i>	

Use of Scientometrics in Evaluation of Grant Proposals .....	495
<i>Katerina Guba, Alexey Zheleznov, Elena Chechik and Angelika Tsvinskaya</i>	
Early detection of problems with scientific papers using Twitter data: A case study of three retracted COVID-19/SARS-CoV-2 papers .....	501
<i>Robin Haunschild and Lutz Bornmann</i>	
Examining International Research Collaboration during the COVID-19 Pandemic using arXiv Preprints .....	511
<i>Jianguan He, Erjia Yan and Chaoqun Ni</i>	
National and Organizational Patterns of Nobel Laureate Careers in Physiology/Medicine, Physics and Chemistry .....	517
<i>Thomas Heinze and Joel Emanuel Fuchs</i>	
Opening the black box of bibliometric mapping algorithms: Which algorithm should I choose? .....	527
<i>Matthias Held</i>	
A framework for measuring the knowledge diffusion impact of interdisciplinary research .....	533
<i>Ying Huang, Wolfgang Glänzel, Bart Thijs and Lin Zhang</i>	
Where to start reading? Introducing the Reference-Citation Plot .....	539
<i>Marcus John, Frank Fritsche and Christian Gülden</i>	
Reference Distributions of Reviews in the Web of Science – The Effects of Company Policy from 1992-2010 .....	545
<i>Miloš Jovanović, Philipp Baaden and Frank Fritsche</i>	
Detecting Global Publication Trends in Research Integrity and Research Ethics (RIRE) through Bibliometrics Analysis .....	551
<i>K. A. Khor, L. G. Yu and A. M. Soebartono</i>	
A Unique Approach to Characterizing Types of Prolific Authors .....	563
<i>Richard Klavans and Kevin W. Boyack</i>	
A model for Custom Bibliographic Databases Creation: Machine Learning Approach for Analogue Documents Inclusion .....	573
<i>Oskar Kosch and Marek Szarucki</i>	



Publish more or publish differently? New aspects of relationship between scientific mobility and performance of young researchers .....	585
<i>Maxim Kotsemir, Ekaterina Dyachenko and Alena Nefedova</i>	
Avoiding bias when inferring race using name-based approaches .....	597
<i>Diego Kozlowski, Dakota S. Murray, Alexis Bell, Will Hulsey, Vincent Larivière, Thema Monroe-White and Cassidy R. Sugimoto</i>	
Links from preprints to published papers in preprint metadata .....	609
<i>Bianca Kramer</i>	
Do impact-factor journals make blacklisted journals visible? A systematic citation study .....	615
<i>Emanuel Kulczycki, Marek Hołowiecki, Zehra Taşkın and Franciszek Krawczyk</i>	
Constructing models of academic funding system: relation between the amount of funding and the efficiency .....	621
<i>Lai Kwun Hang, Vincent Traag and Ludo Waltman</i>	
An appraisal of publication embedding techniques in the context of conventional bibliometric relatedness measures .....	633
<i>Wout S. Lamers, Nees Jan van Eck and Giovanni Colavizza</i>	
Exploring the Relationship between Qualities of Press Releases to Research Articles and the Articles' Impact .....	639
<i>Steffen Lemke, Julian Sakmann, Max Brede and Isabella Peters</i>	
Citing a retracted paper: the case of Wakefield's article that correlates vaccine and autism .....	645
<i>Jacqueline Leta, Kizi Araujo and Stephanie Treiber</i>	
Have the European Union programmes made a difference to biomedical research outputs? .....	651
<i>Grant Lewison</i>	
How are data paper abstracts constructed? Preliminary analysis of rhetorical moves in data paper abstracts from <i>Scientific Data</i> and <i>Data in Brief</i> .....	663
<i>Kai Li and Chenyue Jiao</i>	
Learning to represent for discipline .....	669
<i>Yuan Lin, Zhang Xie, Haifeng Liu, Kun Ding, Kan Xu and Shiyu Zhang</i>	

Research on the Persona of Chinese Distinguished Young Scholars in Computer Science Area .....	681
<i>Yuan Lin, Xinhang Zhao, Jiaojiao Feng and Kun Ding</i>	
A retrospective analysis of China's science and technology evaluation policy since 1978 .....	693
<i>Xin Liu, Zhenyue Zhao and Jiang Li</i>	
The Effects of Rhetorical Structure on Citing Behavior in Research Articles .....	705
<i>Wen Lou, Zilong Su, Shaodan Zheng and Jiagen He</i>	
No Metrics for Postdocs: Precarious Labour in Science Policy .....	711
<i>Lai Ma</i>	
Studying researchers' institutional mobility: bibliometric evidence on academic inbreeding and internationalization .....	717
<i>Vít Macháček, Martin Srholec, Márcia R. Ferreira, Nicolas Robinson-Garcia and Rodrigo Costas</i>	
O Factor: Open-Open citation measure .....	729
<i>Devika P. Madalli</i>	
Article Processing Charges based publications: to which extent the price explains scientific impact? .....	741
<i>Abdelghani Maddi and David Sapinho</i>	
In-Cites research fronts and its relationship with citations per document. A case study of Carlos III University of Madrid .....	753
<i>Jorge Mañana-Rodríguez, Núria Bautista-Puig and Elías Sanz-Casado</i>	
A comparison of citation scores and Web of Science usage counts for Russian science papers, 2006-19 .....	759
<i>Valentina Markusova, Grant Lewison, Anna Zolotova, Ilya Libkind and Alexander Libkind</i>	
Large coverage fluctuations in Google Scholar: a case study .....	771
<i>Alberto Martín-Martín and Emilio Delgado López-Cózar</i>	
Brazilian scientific production on COVID-19: a bibliometric and altmetric analysis .....	783
<i>Isabella Maria Almeida Mateus and Cristian Berrío-Zapata</i>	

Co-Patents: What do trends reveal? An assessment of European innovation systems .....	795
<i>Max Meier, Marta Romagnoli, Julie Callaert, Paolo Landoni and Bart Van Looy</i>	
Locally global. Understanding the structural evolution of a bi-lingual national research landscape .....	807
<i>Matias Federico Milia, Ariadna Nebot Giralt and Rigas Arvanitis</i>	
RRI projects in the mirror of RRI indicators .....	813
<i>Ülle Must</i>	
Diversity and interdisciplinarity – Should variety, balance and disparity be combined as a product or better as a sum? A probabilitytheoretical approach .....	825
<i>Rüdiger Mutz</i>	
Datasets on DataCite – an Initial Bibliometric Investigation .....	837
<i>Anton Ninkov, Kathleen Gregory, Isabella Peters and Stefanie Hausteijn</i>	
A Generalized Power Law Model of Citations .....	843
<i>Rubeyan Nuermairaiti, Leonid V. Bogachev and Jochen Voss</i>	
The Challenges of Building Publication Collections in the Wild: The Case of German Art History .....	849
<i>Jenny Oltersdorf and Jochen Gläser</i>	
A Typology of Research Discovery Tools: Bibliometric Redundancy vs. Bibliometric Variety .....	859
<i>Andreas Pacher</i>	
World research and European news-media reports of cancer risk factors: implications for public health communication .....	871
<i>Elena Pallari and Grant Lewison</i>	
Are we there yet? Analyzing scientific research related to COVID-19 drug repurposing .....	883
<i>Namu Park, Hyeyoung Ryu, Ying Ding, Qi Yu, Yi Bu, Qi Wang, Jeremy J. Yang and Min Song</i>	
Scientific Awards: Women Laureates in STEM .....	895
<i>Angayar Kumari Pavanasam and Aparna Basu</i>	

Research performance and scholarly communication profile of competitive research funding: the case of Academy of Finland .....	901
<i>Janne Pölönen and Otto Auranen</i>	
Research performance indicators and management decision making: Using Collab-CNCI to understand institutional impact .....	913
<i>Ross W. K. Potter, Martin Szomszor and Jonathan Adams</i>	
Five decades of empirical research on the postdoc: a scoping review of the contribution of bibliometrics .....	921
<i>Heidi Prozesky</i>	
Why university affiliation discrepancies still exist in citation indexes and how to resolve them .....	933
<i>Philip J. Purnell</i>	
Media reporting of cancer research in China and the UK, 2014-19 .....	943
<i>Yuping Qu, Elizabeth Davies and Grant Lewison</i>	
Visualising plural mappings of science for Sustainable Development Goals (SDGs) .....	949
<i>Ismael Rafols, Ed Noyons, Hugo Confraria and Tommaso Ciarli</i>	
Protégé-advisor gender pairings in academic survival and productivity of German PhD graduates .....	955
<i>Andreas Rehs</i>	
Reporting transparency in analysis of variance assumption testing .....	967
<i>Maxime Sainte-Marie, Asger Dalsgaard Pedersen and Philippe Mongeon</i>	
Evaluating the impacts of international cooperation under institutional agreements involving funding agencies and research organizations .....	979
<i>Sergio Salles-Filho, Paulo Feitosa, Adriana Bin and Fernando Colugnati</i>	
The small world of editorships: A network on innovation studies .....	985
<i>Ana Teresa Santos and Sandro Mendonça</i>	
Journals' agendas versus actual publications: A first look at article dynamics in innovation journals .....	997
<i>Ana Teresa Santos and Sandro Mendonça</i>	
Quantum technology 2.0 – topics and contributing countries from 1980 to 2018 .....	1009
<i>Thomas Scheidsteeger, Robin Haunschild, Lutz Bornmann and Christoph Ettl</i>	

Does Wikipedia Cover the Relevant Literature on Major Innovations Timely? An Exploratory Case Study of CRISPR/Cas9 .....	1021
<i>Marion Schmidt, Wolfgang Kircheis, Arno Simons, Martin Potthast and Benno Stein</i>	
A Case Study of the Epistemic Function of Citations – Implications for Citation- based Science Mapping .....	1027
<i>Cara Seitz, Marion Schmidt, Nathalie Schwichtenberg and Theresa Velden</i>	
Geographic Differences in the Uptake of Open Access .....	1033
<i>Marc-André Simard, Gita Ghiasi, Philippe Mongeon and Vincent Larivière</i>	
From Citing Sentences to Causal Networks: The Causality Index .....	1039
<i>Henry Small</i>	
Bibliometric indicators to measure the internationalization of Brazilian research: a case study on common beans .....	1045
<i>Ana Carolina Spatti, Luiza Maria Capanema Bezerra, Adriana Bin, Carlos Eduardo Fredo and Sérgio Augusto Morais Carbonell</i>	
Differences between Web of Science, Scopus, and Dimensions in structure and citation networks affect German sectors' normalised citation impact .....	1051
<i>Stephan Stablschmidt and Dimity Stephen</i>	
PubMed Central Citation Context Dataset .....	1063
<i>Fengjun Sun, Yingqiu Li, Guojun Sheng and Xiaolin Yao</i>	
Visualizing Characteristic Interdisciplinary Research Collaborations Among Departments in a University .....	1075
<i>Tetsuya Takahashi, Marie Katsurai, Ikki Ohmukai and Hideaki Takeda</i>	
Editorial board member and prolific author status positively shorten publication time .....	1081
<i>Zebra Taşkın, Abdülkadir Taşkın, Güleda Doğan and Emanuel Kulczycki</i>	
Research managers as data experts? Task areas and competency profiles in IT-based research reporting in Germany .....	1091
<i>Christoph Thiedig, Stefan Schelske and Sabrina Petersohn</i>	
Comparing different implementations of similarity for disparity measures in studies on interdisciplinarity .....	1097
<i>Bart Thijs, Ying Huang and Wolfgang Glänzel</i>	

Artificial Intelligence Research Planning Using Research Cluster Extreme Growth Forecasts: A Retrospective Calibration .....	1109
<i>Autumn Toney, Dewey Murdick, Kevin W. Boyack and Richard Klavans</i>	
The growth of COVID-19 scientific literature: A forecast analysis of different daily time series in specific settings .....	1121
<i>Daniel Torres-Salinas, Nicolas Robinson-Garcia, François van Schalkwyk, Gabriela F. Nane and Pedro Castillo-Valdivieso</i>	
The causal intricacies of studying gender in science .....	1131
<i>Vincent Traag and Ludo Waltman</i>	
The duality of disciplinary diversity for universities .....	1143
<i>Angelika Tsivinskaya</i>	
Writing style and success in research grant applications .....	1149
<i>Peter van den Besselaar and Charlie Mom</i>	
What leads to gender bias in review panels? .....	1159
<i>Peter van den Besselaar and Charlie Mom</i>	
Crossref as a source of open bibliographic metadata .....	1169
<i>Nees Jan van Eck and Ludo Waltman</i>	
The Uptake of a Label for Peer-reviewed Books: the Flemish GPRC label .....	1175
<i>Eline Vandewalle, Raf Guns and Tim Engels</i>	
Distinguishing Science Communication & Popularization from Research-Based Public Interventions .....	1181
<i>Florian Vanlee, Walter Ysebaert and Hans Jonker</i>	
Additional Context Helps! Leveraging Cited Paper Information To Improve Citation Classification .....	1187
<i>Kamal Kaushik Varanasi, Tirthankar Ghosal and Valia Kordoni</i>	
Thematic influence on academic impact according to types of collaboration: an analysis of the library and information science field from 2015 to 2019 .....	1193
<i>A. Velez-Estevez, P. Garcia-Sanchez, J.A. Moral-Munoz and M.J. Cobo</i>	
Assessing the potential effect of a unified system of journal classification based on citation impact indicators: the new Qualis .....	1205
<i>Eloisa Viggiani, Alexandre Prestes Uchoa and Jano Moreira Souza</i>	

Measuring Funding Programs Achievements in Fostering Cross-Disciplinary Research: The Potential Role of Integration and Collaboration Indicators in Capturing Distinct Support Mechanisms .....	1217
<i>Etienne Vignola-Gagné, Henrique Pinheiro, Maxime Rivest and David Campbell</i>	
A bibliometric strategy for identifying benchmark research units .....	1229
<i>Qi Wang and Tobias Jeppsson</i>	
International mobility and gender balance in academia. The case of Norway .....	1235
<i>Kaja Wendt and Hebe Gunnes</i>	
Hierarchical topic tree: A hybrid model comprising network analysis and density peak search .....	1241
<i>Mengjia Wu and Yi Zhang</i>	
Comparison of citation-based clustering and topic modeling: The case of cardiovascular disease research .....	1253
<i>Qianqian Xie, Ismael Rafols, Ludo Waltman and Alfredo Yegros</i>	
Interdisciplinarity vs. Unidisciplinarity: A Structural Comparison of Multi-generation Citations and References .....	1259
<i>Jiawei Xu, Chao Min, Win-bin Huang and Yi Bu</i>	
RStarRank: A Method for Identifying Rising Stars based on Academic Publication .....	1265
<i>Jing Xu, Chuan Tang and Lu Tang</i>	
Modeling Knowledge Diffusion in the Disciplinary Citation Network based on Differential Dynamics .....	1277
<i>Zenghui Yue, Haiyun Xu and Guoting Yuan</i>	
Network Analysis of Econometric Society Fellows .....	1289
<i>Tolga Yuret</i>	
Topic Identification Index of Disruptive Technology Based on Patent Characteristics .....	1295
<i>Jiawei Zhang, Kaibiao Wu and Yu Dong</i>	
Two New Field Normalization Indicators Considering the Reliability of Citation Time Window: Some Theoretical Considerations .....	1307
<i>Lihua Zhang and Xing Wang</i>	

Comparing Paper Level Classifications across Different Methods and Systems: An Investigation on <i>Nature</i> Publications .....	1319
<i>Lin Zhang, Beibei Sun, Fei Shu and Ying Huang</i>	
The prevalence and impact of different types of open access articles from China and USA .....	1325
<i>Lin Zhang, Yahui Wei, Ying Huang and Gunnar Sivertsen</i>	
Productivity patterns, collaboration and scientific careers of authors with retracted publications in clinical medicine .....	1337
<i>Qin Zhang and Hui-Zhen Fu</i>	
Enhanced Author Bibliographic Coupling Analysis Using Syntactic and Semantic Citation Information .....	1349
<i>Ruhao Zhang and Junpeng Yuan</i>	
Applying Bibliometrics Methods to Understanding Knowledge Domains on Wikipedia .....	1361
<i>Dangzhi Zhao and Andreas Strotmann</i>	
International Migration in Academia and Citation Performance: An Analysis of German-Affiliated Researchers by Gender and Discipline Using Scopus Publications 1996-2020 .....	1369
<i>Xinyi Zhao, Samin Aref, Emilio Zagheni and Guy Stecklov</i>	
The evolution of interdisciplinarity in five social sciences and humanities disciplines: relations to impact and disruptiveness .....	1381
<i>Hongyu Zhou, Raf Guns and Tim C. E. Engels</i>	
The Trend of Technological Convergence Based on Patent Analysis .....	1393
<i>Wenjing Zhu, Bohong Ma and Lele Kang</i>	
Impact of Model Settings on the Text-based Rao Diversity Index .....	1405
<i>Andrea Zielinski</i>	
Gendered feedback and negative adjective use in peer-reviewer reports .....	1417
<i>Alessandra Zimmerman, Helen Greaves, Richard Klavans, Jonathan Best and Gemma E. Derrick</i>	
A conceptual evaluation of measures of interdisciplinarity .....	1423
<i>Sander Zwanenburg, Maryam Nakhoda and Peter Whigham</i>	



## INDEX OF POSTERS

Does Publicity in the Science Press Drive Citations? .....	1433
<i>Manolis Antonoyiannakis</i>	
Scientometrics of complexity: an institutional approach .....	1435
<i>Ricardo Arencibia-Jorge, José Luis Jiménez Andrade, Ibis A. Lozano-Díaz, Javier García-García and Humberto Carrillo-Calvet</i>	
Thematic multidimensional profile of Mexican scientific output in Dimensions 2010-2019 .....	1437
<i>Ricardo Arencibia-Jorge, José Luis Jiménez-Andrade, Ibis A. Lozano-Díaz and Humberto Carrillo-Calvet</i>	
A new thematic dispersion index to assess multidisciplinary at different levels of aggregations .....	1439
<i>Ricardo Arencibia-Jorge, Rosa Lidia Vega-Almeida and Humberto Carrillo- Calvet</i>	
‘Larks’ and ‘Owls’: More on Eliciting Human Moods and Chronotypes from Psychometric Variables in Twitter Data .....	1441
<i>Aparna Basu</i>	
Robustness of citation networks retrieved from queries .....	1443
<i>Alexandre Benatti, Henrique Ferraz de Arruda, César Henrique Comin, Filipi Nascimento Silva and Luciano da Fontoura Costa</i>	
Extracting Sentences Concerning Research Results of Citations from Full Text of JASIST .....	1445
<i>Jiixin Bian, Si Shen and Dongbo Wang</i>	
Shifting the ARWU methodology to Q1 journals: what are the consequences for lower-tier universities? .....	1447
<i>Srđa Bjeladinović, Veljko Jeremić and Mladen Stamenković</i>	
Measuring Concentrations of Research Funding: A case of The Czech Science Foundation (GAČR) .....	1449
<i>Jiri Bures</i>	

Improving Coupling Metrics with Deep Neighborhood for Local Map of Science .....	1451
<i>Gaëlle Candell and David Naccache</i>	
Gender of Editors in Chief and Editorial Member Committees in journals of Women Studies .....	1453
<i>Lourdes Castelló-Cogollos, Andrea Sixto-Costoya, Adolfo Alonso-Arroyo, Juan Carlos Valderrama-Zurián and Rafael Aleixandre-Benavent</i>	
Differences between using institutional information from curriculum vitae and publications for tracking scientific mobility .....	1455
<i>Yu-Wei Chang</i>	
Mapping the Interdisciplinarity of the Arts, Humanities & Social Sciences by Paper-level Classifications of Science .....	1457
<i>Bikun Chen, Shiqin Wang, Yuxin Liu and Mengxia Cheng</i>	
WeChat Presence of Chinese Scholarly Journals: An analysis of CSSCI-indexed Journals .....	1459
<i>Ting Cong, Zhichao Fang and Rodrigo Costas</i>	
Understanding Scientific Success: A MacIntyrean Virtue Ethics Approach .....	1461
<i>Cinzia Daraio</i>	
Effect of Firm Size on Patent Maintenance Length .....	1463
<i>Huei-Ru Dong and Mu-Hsuan Huan</i>	
Metrics Literacies: On the State of the Art of Multimedia Scholarly Metrics Education .....	1465
<i>Isabelle Dorsch, Alyssa Jeffrey, Sanam Ebrahimzadeh, Lauren A. Maggio and Stefanie Hausteiner</i>	
How do peer reviewers (mis)interpret funding agency criteria? Results from a survey of Science Foundation Ireland grant reviewers .....	1467
<i>Thomas Feliciani, Junwen Luo and Kalpana Shankar</i>	
Impact of Secretariat Research Groups in Brazil .....	1469
<i>Rosimeri Ferraz Sabino, Fabio Gomes Rocha and Alejandro C. Frery</i>	
Altmetrics for evaluation of medical research in Germany .....	1471
<i>Nicholas Fraser, Paula Bräuer and Isabella Peters</i>	

The three-dimensional activity index .....	1473
<i>Joel Emanuel Fuchs</i>	
Reasons for Retraction Affect the Impact on Citations .....	1475
<i>John Gibson and Keruma Gibson</i>	
Altmetrics System for the Evaluation of Chinese Scholars .....	1477
<i>Guo Ying and Xiao Xiantao</i>	
Publishing contribution of editorial board members to bibliometric indices of Library and Information Science journals .....	1479
<i>Nikolay A. Mazov and Vadim N. Gureyev</i>	
Automatic Recognition and Classification of Future Work Sentences in Academic Articles in a Specific Domain .....	1481
<i>Wenke Hao, Yuchen Qian, Zhicheng Li, Yuzhuo Wang and Chengzhi Zhang</i>	
International Cooperation in Times of Crises – Differences in Cooperation patterns between COVID-19 and non-COVID-19 Trials .....	1483
<i>Gerrit Hirschfeld and Christian Thiele</i>	
Changes in the Funding Composition of Czech Science .....	1485
<i>Radim Hladik</i>	
Using an Entymetric Approach to Trace Novel Chemical Compounds for Research Impact .....	1487
<i>David E. Hubbard</i>	
A Bibliometric Mapping of Research on Quality Assurance in African Higher Education: a systematic literature review (from 1999 to 2019) .....	1489
<i>Luis M. João and Patrício Langa</i>	
Measuring Altmetric Events: The need for Longer Observation Period and Article Level Computations .....	1491
<i>Mousumi Karmakar, Sumit Kumar Banshal and Vivek Kumar Singh</i>	
Tutmetrics and usability: The footprint of science mapping tools in tutorial space .....	1493
<i>Veslava Osinska and Radoslaw Klimas</i>	
Comparing citations, usage, and interdisciplinarity in cover and noncover papers .....	1495
<i>Ling Kong and Dongbo Wang</i>	

Knowledge Positioning of Patent Assignees .....	1497
<i>Chung-Huei Kuan and Dar-Zen Chen</i>	
The lockdown effect: Research in the age of a pandemic .....	1499
<i>Barbara S. Lancho Barrantes and Nick Sheppard</i>	
Recognition the Citation Author from JASIST Full Text .....	1501
<i>Zongyi Li, Dongbo Wang and Si Shen</i>	
Recognizing Sentences Concerning Research Questions from the Full Text of <i>JASIST</i> .....	1503
<i>Chang Liu, Dongbo Wang and Si Shen</i>	
Research Methods Sentence Identification in Academic Full Text Based on Deep Learning .....	1505
<i>Jiangfeng Liu, Dongbo Wang and Si Shen</i>	
Does Technology of Climate Change Mitigation Transfer Follow UNFCCC? ....	1507
<i>Paoling Liu and Chun-Chieh Wang</i>	
Countries multidimensional scientometric performance profiles compared with a self-organized neural network .....	1509
<i>Ibis A. Lozano-Díaz, Ricardo Arencibia-Jorge and Humberto Carrillo-Calvet</i>	
Analysis of Topic Evolvments of COVID-19 Scientific Literatures based on Probabilistic Modelling .....	1511
<i>Ma Lili, Zhao Yanqiang, Yue Mingliang and Ma Tingcan</i>	
Are Non-Journal Citations More Important for Books? Study of Types of Citing Documents in Scopus .....	1513
<i>Ashraf Maleki</i>	
Identifying research themes in Finland using topic modeling and word embeddings .....	1515
<i>Katja Mankinen and Yrjö Leino</i>	
International collaboration of post-Soviet countries: what has changed in post-Soviet period? .....	1517
<i>Nataliya Matveeva, Andrey Lovakov and Ivan Sterligov</i>	
Building Multi-level Aspects of Peer Reviews for Academic Articles .....	1519
<i>Minghui Meng, Yuzhuo Wang and Chengzhi Zhang</i>	

Multilingualism of Ukrainian Humanities: how it is seen globally .....	1521
<i>Serhii Nazarovets and Olesya Mryglod</i>	
Rigor and Transparency Index for Systematic Literature Reviews: a first stage approach .....	1523
<i>Gerson Pech and Catarina Delgado</i>	
Users of National Open Access Journals: The case of Journal.fi platform .....	1525
<i>Janne Pölönen, Sami Syrjämäki, Antti-Jussi Nygård and Björn Hammarfelt</i>	
Exploring the Distribution Regularities of Referees' Comments in IMRaD Structure of Academic Articles .....	1527
<i>Chenglei Qin and Chengzhi Zhang</i>	
References to Literature in Patent Documents: A Case Study of Electromobility Research .....	1529
<i>Zhao Qu</i>	
The scientific productivity of German PhD graduates: A machine learning–based author name disambiguation and record linkage approach .....	1531
<i>Andreas Rehs</i>	
Do COVID-related articles pass through the fast lane? The case of the International Journal of Infectious Diseases .....	1533
<i>Ronald Rousseau</i>	
Editorial recommendations in Nature and Science .....	1535
<i>Jasmin Sadat and João Martins</i>	
Watching over innovation studies: Profiling the gatekeepers .....	1537
<i>Ana Teresa Santos and Sandro Mendonça</i>	
Keyword Distance Ratio: Evaluating Keyword Assignment with Word Embeddings .....	1539
<i>Brandon Sepulvado</i>	
Ranking Universities via Clustering .....	1541
<i>George Stoupas, Antonis Sidiropoulos, Dimitrios Katsaros and Yannis Manolopoulos</i>	
RCC: NSFC's new review mechanism: its missions and challenges .....	1543
<i>Shiyang Tang, Weiwei Li, Quan Liu and Yuxian Liu</i>	

Broad shouldered – publication formats served by doctoral students .....	1545
<i>Jakob Tesch</i>	
Words Connecting Scientific Communities The case of ‘argumentation’ .....	1547
<i>Natalija Todorovic and Benedetto Lepori</i>	
Linking the Web of Science and Scopus to a Publication Repository .....	1549
<i>Dirk Tunger, Roland Erwin Suri, Andreas la Roi and David Johann</i>	
Measuring «Science Teaching» in Higher Education – An Empirical Study Of Teaching Content .....	1551
<i>Carolin Vollenberg, Julian Koch, Christopher Still and André Coners</i>	
Do UTAUT variables influence the adoption and use of open access scholarly publishing in Kenya? A case study of selected universities .....	1553
<i>Mercy W. Waitbaka and Omwoyo Bosire Onyancha</i>	
Patterns of Innovation Team of Pharmaceutical in Countries .....	1555
<i>Chun-Chieh Wang, Dar-Zen Chen and Mu-Hsuan Huang</i>	
Using Full-text Content of Academic Articles to Classify Research Methods in Library and Information Science .....	1557
<i>Ruping Wang, Liang Tian and Chengzhi Zhang</i>	
Algorithm Entities Usage in Chinese Academic Articles from The Domain of Information Science .....	1559
<i>Yuzhuo Wang, Heng Zhang and Chengzhi Zhang</i>	
Breakthrough Foresight Based on Citation Diffusion Features .....	1561
<i>Haiyun Xu, Jos Winnink, Xin Zhang, Zenghui Yue, Jing Li and Chen Liang</i>	
Analysis of selected journals in Excellence Action Plan for China STM Journals Based on Bibliometrics .....	1563
<i>Yang Rui and Wang Baoji</i>	
Does knowledge of Library and information science field transform to technology? .....	1565
<i>Xiao Yang, Lingzi Feng and Junpeng Yuan</i>	
Artificial Intelligence in Pharmaceuticals: Bibliometric and Collaboration Network Analysis of Patents .....	1567
<i>Yang Xin</i>	

Who Shares Scholarly Output on Facebook? .....	1569
<i>Houqiang Yu, Wei Zhang, Yue Wang and Tingting Xiao</i>	
Where, how and why are scientific products mentioned in policy documents? .....	1571
<i>Houqiang Yu, Bieizat Murat, Jiatong Li and Longfei Li</i>	
Is <i>Dimensions</i> a reliable data source of funding and funded publications? .....	1573
<i>Lin Zhang, Yinxin Zheng, Wenjing Zhao and Ying Huang</i>	
The effect of international academic mobility on international cooperation: from the perspective of overseas institutions .....	1575
<i>Zhang Ying, Liu Xiaomin and Sun Yuan</i>	
Research on international patent technology layout in high-speed railway field ....	1577
<i>Rongying Zhao, Ruru Chang, Xiaoyu Wang and Zhaoyang Zhang</i>	
A Bipartite Author-algorithm Network for Exploring Versatile Scholars .....	1579
<i>Yi Zhao, Yuzhuo Wang and Chengzhi Zhang</i>	
Citation Time Entity Recognition Based on Deep Learning .....	1581
<i>Yufei Zhu, Si Shen and Dongbo Wang</i>	





# An appraisal of publication embedding techniques in the context of conventional bibliometric relatedness measures

Wout S. Lamers<sup>1,\*</sup>, Nees Jan van Eck<sup>1</sup> and Giovanni Colavizza<sup>2</sup>

<sup>1</sup> {w.s.lamers, ecknjpvan}@cwts.leidenuniv.nl

Leiden University, Centre for Science and Technology Studies (CWTS), Leiden (The Netherlands)

<sup>2</sup> g.colavizza@uva.nl

University of Amsterdam, Faculty of Humanities, Department Mediastudies, Amsterdam (The Netherlands)

## Abstract

Modern natural language processing techniques have given rise to embedding techniques that can represent documents based on their content or context, and several papers have operationalized these to perform bibliometric tasks. The relationship between these embeddings and conventional citation based or title and abstract based mappings remains unclear. Contrary to citation-based or term-based relatedness, embedding-based relatedness is not immediately interpretable. We consider four embedding-derived publication relatedness measures, based on: 1) word2vec embeddings of citation labels, sentence embeddings using 2) BERT and 3) SciBERT, and 4) title and abstract embeddings using SPECTER, and compare them with conventional bibliometric publication relatedness measures derived from citation relations and title and abstract noun phrases. We show that there is stronger overlap between these embedding-derived relatedness measures and citation-based relatedness than with title and abstract noun phrase-based relatedness, and that embedding-derived relatedness measures outperform conventional techniques when used to cluster publications cited with the same citation intent.

## Introduction

The notion of relatedness of scientific publications is at the heart of many open problems in bibliometrics. Publication clustering and mapping techniques rely on publication-to-publication relatedness measures, conventionally based on citation relations, textual similarities of titles and abstracts, or occasionally combinations of these. It should come as no surprise that different relatedness measures produce different results (Gläser et al., 2017), which makes the informed choice of appropriate relatedness measures an imperative aspect of many bibliometric research tasks.

Meanwhile, advances in natural language processing (NLP) and information retrieval have introduced a host of new text-based methods for establishing relatedness of not only documents to one another, but also documents to terms in the vocabulary of these documents. The introduction of word2vec (Mikolov, Chen, et al., 2013; Mikolov, Sutskever, et al., 2013) marked a substantial advance, allowing vector representations of words to be learned that not only allow for easy computation of semantic similarities, but also preserve—to some extent—the semantic and syntactic relationships between terms. Doc2vec (Le & Mikolov, 2014) adapted this method, allowing for the generation of similar vector representations for larger bodies of text. Further advances in the field since then such as the emergence of large pre-trained language models like BERT (Devlin et al., 2018) have only expanded the suite of options for establishing relatedness between documents.

The case for using NLP techniques in science mapping and clustering is compelling. At their core lies the distributional hypothesis, the assertion that terms with similar distributions have similar meaning, i.e., similar terms are used in similar context. What is citation, if not the deliberate act of putting previous work in the context of a new author's design? Hence, papers that are cited in similar context, or that introduce similar language, can be assumed to be related to each other, and embedding techniques can capture this relatedness. Another important benefit of these embedding methods is that publications can be embedded along with terms, allowing

researchers to learn relationships between embedded literature and concepts in the vocabulary, directly from the data.

Several recent publications have leveraged these new NLP techniques to explore new ways of representing publications, visualize research landscapes, and facilitate document search and retrieval. Approaches range from representing documents using the text in their abstracts (e.g. Hain et al., 2020) to embeddings derived from citation context (e.g. Berger et al., 2017; He & Chen, 2018), and a number of studies have explored combining text-based approaches with citation-based approaches (e.g. Cohan et al., 2020; Ganguly & Pudi, 2017). However, while many of these papers find that their representations of scientific literature capture useful and relevant aspects of publication context, topic, purpose, or overall relatedness, and often outperform traditional techniques in specific tests selected by the authors, embedding-derived relatedness measures remain difficult to interpret, and their resulting vector representations of papers are essentially black boxes. The nature of the resulting relatedness measures is difficult to compare to traditional bibliometric techniques, and this ambiguity makes the use of these new embedding techniques in science mapping and publication clustering less attractive.

Our research aims to compare a set of common NLP-based publication representations and their resulting relatedness measures with traditional citation-based and abstract-based publication relatedness measures. We follow the methodology introduced by Waltman et al. (2020), allowing us to compare the accuracy of clustering results obtained using various publication relatedness measures based on a selected baseline relatedness measure. We compare four methods for generating publication embeddings with two established bibliometric relatedness measures, one based on publication titles and abstracts, and one based on citations, in two data sets. Our goal is to establish whether embedding-derived publication relatedness measures more strongly resemble established text-based relatedness measures or citation-based relatedness measures.

## Methods

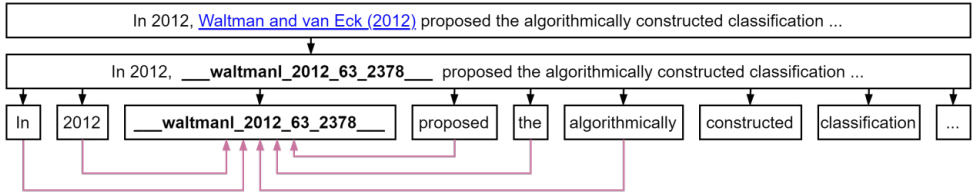
Our first data set is selected from the Elsevier ScienceDirect corpus previously analyzed in detail by Boyack et al. (2018). We select scientometrics-related publications using the meso-level of the CWTS hierarchical publication classification system (Waltman & van Eck, 2012). This produced a total of 7825 publications, that are cited in 51003 full-text sentences (not limited to the scientometrics field). 3169 of these publications feature in at least five citing sentences, for a total of 44464 citing sentences. If a sentence features multiple citations, only the first citation is considered, to avoid associating the exact same context with multiple works.

The second dataset is the SciCite citation intent dataset introduced by Cohan et al. (2019). This consists of 6427 publications and 10817 sentences citing these publications, as well as annotated citation intent labels noting whether a citing sentence cites a paper as *background*, for its *methods*, or to compare *results*. Titles and abstracts of these publications, as well as publications cited by or citing them, were retrieved using the Semantic Scholar API.

### *Word2vec*

Our first embedding-derived relatedness measure is a simple word2vec embedding of citation labels. In each citing sentence, the label referencing the cited publication is replaced with a token unique to the cited work, and a word2vec embedding is trained on the resulting corpus using the *gensim* python library (Rehurek & Sojka, 2010). For a schematic representation, see figure 1. Relatedness of publications is calculated as the cosine similarities of the unique publication tokens. The word2vec architecture has a few adjustable hyperparameters. We varied

the embedding size, window size, epochs, and negative sample size, and tried both the CBOW and skip-gram methods. Skip-gram with a window size encompassing the entire sentence, keeping other parameters set to their default values, produced the most stable publication-to-publication similarities over multiple model runs.



**Figure 1: example of word2vec embedding of citation labels, window size three.**

### *Sentence-BERT*

Entire sentences can be embedded using Sentence-BERT (Reimers & Gurevych, 2019). Averaging the embeddings of sentences that cite the same paper then offers us another way to generate representations for cited publications based on their citation context. In this procedure, we again associated sentences only with the first cited publication, and omitted citation labels entirely from the sentence. Since the Sentence-BERT architecture can use a variety of language models, we repeated this process twice, once using the original BERT base model (Devlin et al., 2018), and once using the SciBERT model (Beltagy et al., 2019) which was trained specifically on scientific text.

### *SPECTER*

Our final embedding-based relatedness measure uses SPECTER (Cohan et al., 2020) to generate publication representations based on titles and abstracts alone. SPECTER’s core feature is that it generates document-level representations using only titles and abstracts, in a manner informed by pre-training on a citation graph. This allows it to outperform other architectures that rely solely on title and abstract information in a variety of performance benchmarks such as citation prediction and topic classification (Cohan et al., 2020).

### *Conventional relatedness measures and comparison*

Waltman et al. (2020) introduce a methodology for comparing publication relatedness measures based on the accuracy of clustering solutions, compared to a baseline relatedness measure. Following their work, we use the BM25 text similarity measure based on title and abstract noun phrases, and the combined direct citation, bibliographic coupling and co-citation (DC-BC-CC) relatedness measure, to represent baseline text-based and citation-based relatedness measures, respectively. All relatedness measures are also simplified and normalized, first by keeping only the 20 most strongly related connections for each document, and subsequently normalizing their weights by dividing the relatedness of each outgoing edge by the total weight of outgoing edges for each node.

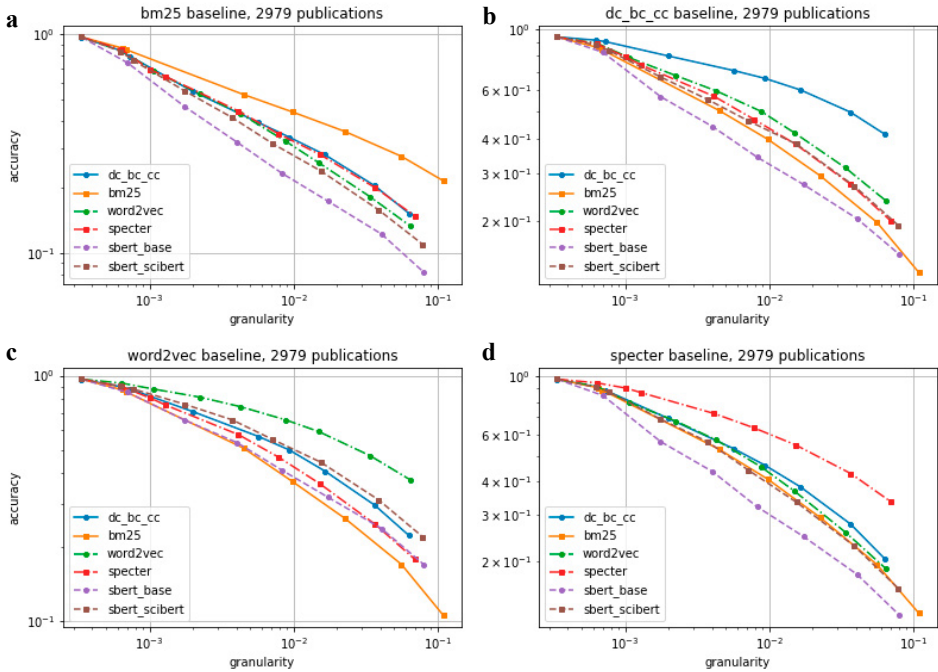
Waltman et al. (2020) then define the accuracy of a clustering solution based on relatedness measure X, evaluated against measure C, as

$$A^{X|C} = \frac{1}{N} \sum_{ij} (c_i^X = c_j^X) r_{ij}^C$$

where  $(c_i^X = c_j^X)$  equals 1 if the clustering  $c$  of document  $i$  is the same as that of document  $j$ , and 0 otherwise, with  $r_{ij}^C$  the relatedness between documents  $i$  and  $j$  per measure  $C$ . The normalization of the relatedness measures constrains this accuracy score between 0 and 1. Accuracy is computed for a variety of clustering solutions obtained using the Leiden algorithm (Traag et al., 2019), using a range of clustering resolution parameters, and the obtained accuracy is then plotted against the granularity of the resulting clustering solution in a granularity-accuracy plot, allowing us to easily compare the clustering accuracy of relatedness measures over a range of resolutions.

## Results

Figure 2a displays the accuracy of the clustering solutions resulting from the various relatedness measures, using the BM25 title and abstract based relatedness as a baseline, over different clustering granularities. This analysis is limited to only the 2979 publications that are connected by relatedness in each of the six networks. Figure 2b uses the DC-BC-CC citation relatedness baseline, while 2c and 2d use word2vec and SPECTER-derived relatedness instead.



**Figure 2: granularity-accuracy plots, scientometrics data set.**

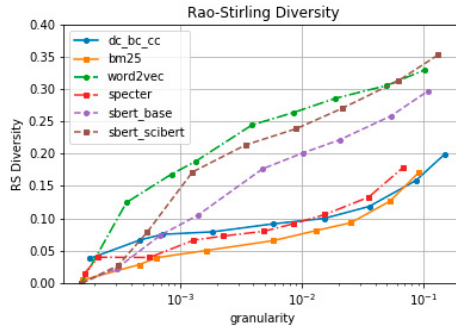
Figure 2c shows us that, in our scientometrics data set, the word2vec embedding derived from citing context resembles the citation-based relatedness measures more strongly than it does the title and abstract text-based relatedness measure. This is perhaps surprising, as it is fundamentally text-based, though this text originates from sentences that cite, which would explain the similarities to citation relatedness patterns. More surprising, figure 2d shows that the SPECTER-embeddings, despite being based on the same title and abstract text as the BM25 relatedness, behave more like the DC-BC-CC citation relatedness. Indeed, figure 2a shows that SPECTER and DC-BC-CC are approximately equally accurate when compared directly against the BM25 relatedness. Finally, figure 2b shows that word2vec, SPECTER and the SciBERT-based sentence embedding more closely resemble citation relatedness than BM25, which was

previously found to be the best-performing title and abstract relatedness measure, at least in our sample in which publications with 5 or more in-text citations are considered. When also including publications with fewer than 5 in-text citations (not pictured), SPECTER and BM25 are the most similar to DC-BC-CC, followed by word2vec and SciBERT sentence embedding.

The SciCite dataset, moreover, contains not only citation context for each publication, but is hand-labeled with citation intent. These labels (*background*, *methods*, and *results*) make for a poor relatedness measure but can be assessed at the cluster level. We can express the diversity of these labels over the clusters using Rao-Stirling diversity (see e.g. Leydesdorff et al., 2019)

$$\Delta = \sum_{i,j} d_{ij}(p_i p_j)$$

in which  $d_{ij}$  is a distance measure between clusters  $i$  and  $j$ , and  $p_i$  is the proportion of nodes in cluster  $i$ . To compute a distance measure, we take the numbers of citing sentences labeled *background*, *methods* and *results* in each cluster as a three-length vector and take the cosine distance of these vectors between each cluster pair. This means that diversity is maximized if publications cited with the same intent are clustered together. We vary the clustering resolution and plot the resulting diversity of SciCite intent labels against cluster granularity in Figure 3.



**Figure 3: diversity-granularity plot of citation intent labels, SciCite dataset.**

Figure 3 shows us that, with some margin, the citation context derived embeddings outperform the conventional relatedness measures, and that SPECTER improves only marginally over BM25 when it comes to clustering citation intent. Even BERT base, while lagging in figure 2, improves markedly on the conventional relatedness measures.

### Discussion

Our research shows that embedding representations of publications have the potential to resemble conventional citation-based relatedness structures more accurately than the best-available conventional title and abstract-based relatedness measure. This holds true even for a simple word2vec embedding, provided enough citation context is available. Context-based embeddings also show a stronger correlation to citation intent labels in the SciCite data set than conventional citation-based or abstract-based relatedness measures.

It must be noted that the embedding-based relatedness measures used in this research might be further explored and optimized for the task of establishing publication relatedness. Especially BERT-based representations are typically fine-tuned for downstream tasks. Nevertheless, our research demonstrates the base feasibility of using embedding representations of publications

and provides important insight into how they relate to traditional publication relatedness measures.

## References

- Beltagy, I., Lo, K. & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*, 3615–3620.
- Berger, M., McDonough, K. & Seversky, L. M. (2017). cite2vec: Citation-Driven Document Exploration via Word Embeddings. *IEEE Transactions on Visualization and Computer Graphics*, 23(1), 691–700.
- Boyack, K. W., van Eck, N. J., Colavizza, G. & Waltman, L. (2018). Characterizing in-text citations in scientific articles: A large-scale analysis. *Journal of Informetrics*, 12(1), 59–73.
- Cohan, A., Ammar, W., Van, M. & Cady, Z. F. (2019). Structural Scaffolds for Citation Intent Classification in Scientific Publications. *Proceedings of NAACL-HLT*, 3586–3596.
- Cohan, A., Feldman, S., Beltagy, I., Downey, D. & Weld, D. (2020). SPECTER: Document-level Representation Learning using Citation-informed Transformers. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2270–2282.
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186.
- Ganguly, S. & Pudi, V. (2017). Paper2vec: Combining graph and text information for scientific paper representation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10193 LNCS, 383–395.
- Gläser, J., Glänzel, W. & Scharnhorst, A. (2017). Same data—different results? Towards a comparative approach to the identification of thematic structures in science. *Scientometrics*, 111(2), 981–998.
- Hain, D., Jurowetzki, R., Buchmann, T. & Wolf, P. (2020). *Text-based Technological Signatures and Similarities: How to create them and what to do with them*. <http://arxiv.org/abs/2003.12303>
- He, J. & Chen, C. (2018). Temporal Representations of Citations for Understanding the Changing Roles of Scientific Publications. *Frontiers in Research Metrics and Analytics*, 3, 27.
- Le, Q. V. & Mikolov, T. (2014). Distributed Representations of Sentences and Documents. *31st International Conference on Machine Learning, ICML 2014*, 4, 2931–2939.
- Leydesdorff, L., Wagner, C. S. & Bornmann, L. (2019). Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal of Informetrics*, 13(1), 255–269.
- Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013, January 16). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems*, 26.
- Rehurek, R. & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50.
- Reimers, N. & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*, 3982–3992.
- Traag, V. A., Waltman, L. & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1), 1–12.
- Waltman, L., Boyack, K. W., Colavizza, G. & Van Eck, N. J. (2020). A principled methodology for comparing relatedness measures for clustering publications. *Quantitative Science Studies*, 1(2), 691–702.
- Waltman, L. & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, 63(12), 2378–2392.