Philipp Mayr
Ingo Frommholz
Guillaume Cabanac

March 26, 2018

7th BIR Workshop

Bibliometric-enhanced Information Retrieval

ECIR 2018, Grenoble, France

https://bit.ly/bir2018

#bir2018

The BIR Chairs

Philipp Mayr

@Philipp_Mayr

Team lead @GESIS_org. Semantic Search, Beyond Cranfield Paradigm

Cologne, Germany

philippmayr.github.io

Ingo Frommholz

@iFromm

Senior Lecturer at University of Bedfordshire. Mainly tweeting about Information Retrieval, Data Science and Digital Libraries. Retweet isn't agreement.

Luton, UK

frommholz.org

Guillaume Cabanac

@gcabanac

Associate Professor of Computer Science. Information Retrieval • Scientometrics — Currently studying #pirateLibraries #bibliogifts irit.fr/publis/IRIS/20…

University of Toulouse, France

irit.fr/~Guillaume.Cab…

The BIR Programme Committee

- Muhammad Kamran Abbasi, University of Bedfordshire, UK
- Karam Abdulahhad, GESIS – Leibniz Institute for the Social Sciences, Germany
- Iana Atanassova, CRIT, Université de Franche-Comté, France
- Joeran Beel, Trinity College Dublin, Ireland
- Patrice Bellot, Aix-Marseille Université - CNRS (LSIS), France
- Marc Bertin, Université Lyon 1, France
- Jose Borbinha, IST / INESC-ID, Portugal
- Cornelia Caragea, Kansas State University, USA
- Zeljko Carevic, GESIS - Leibniz Institute for the Social Sciences, Germany
- Muthu Kumar Chandrasekaran, National University of Singapore, Singapore
- Nicola Ferro, University of Padova, Italy
- Edward Fox, Virginia Polytechnic Institute and State University, USA
- Norbert Fuhr, University of Duisburg-Essen, Germany
- Behnam Ghavimi, GESIS – Leibniz Institute for the Social Sciences, Germany
- C. Lee Giles, The Pennsylvania State University, USA
- Bela Gipp, Universität Konstanz, Germany
- Gilles Hubert, University of Toulouse, France
- Peter Ingwersen, University of Copenhagen, Denmark
- Kokil Jaidka, University of Pennsylvania, USA
- Roman Kern, Know-Center GmbH, Germany
- Petr Knoth, The Open University, UK
- Marijn Koolen, Huygens Institute for the History of the Netherlands, Netherlands
- Rob Koopman, OCLC, The Netherlands
- Cyril Labbé, Grenoble University, France
- Vincent Larivière, EBSI-UdeM, Canada
- Haiming Liu, University of Bedfordshire, UK
- Stasa Milojevic, Indiana University Bloomington, USA
- Peter Mutschke, GESIS – Leibniz Institute for the Social Sciences, Germany
- Horacio Saggion, Universitat Pompeu Fabra, Spain
- Philipp Schaer, TH Cologne, Germany
- Andrea Scharnhorst, DANS-KNAW, The Netherlands
- Ralf Schenkel, University of Trier, Germany
- Vivek Singh, Banaras Hindu University, India
- Henry Small, SciTech Strategies, USA
- Cassidy Sugimoto, Indiana University Bloomington, USA
- Lynda Tamine, University of Toulouse, France
- Ludovic Tanguy, University of Toulouse, France
- Simone Teufel, Cambridge University, UK
- Ulrich Thiel, Fraunhofer IPA-PAMB, Germany
- Dietmar Wolfram, University of Wisconsin-Milwaukee, USA
- Haozhen Zhao, Navigant, USA

Activities

- Workshops

☐ 2014	BIR@ECIR	1st ed.	Amsterdam, NL
☐ 2015	BIR@ECIR	2nd ed.	Vienna, AT
☐ 2016	BIR@ECIR	3rd ed.	Padua, IT
☐ 2016	BIRNDL@JCDL	4th joint ed.	Newark, US
☐ 2017	BIR@ECIR	5th ed.	Aberdeen, UK
☐ 2017	BIRNDL@SIGIR	6th ed.	Tokyo, JP
☐ 2018	BIR@ECIR	7th ed.	Grenoble, FR
☐ 2018	BIRNDL@SIGIR	8th joint ed.	Ann Arbor, US

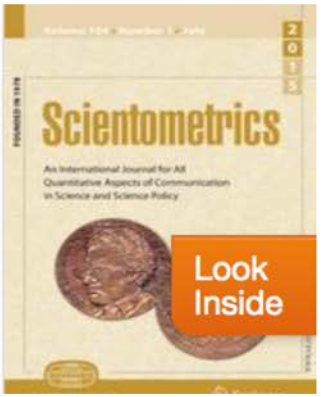
- Special Issues
  - ☐ 2018 *International Journal on Digital Libraries*
  - ☐ 2018 *Scientometrics*

Scientometrics

ISSN: 0138-9130 (Print) 1588-2861 (Online)

Description

Scientometrics is concerned with the quantitative features and characteristics of science and scientific research. Emphasis is placed on investigations in which the development and mechanism of science are studied by statistical mathematical methods.



SCIENTOMETRICS: STATE-OF-THE-ART

Scientometrics,
Vol. 38, No. 1 (1997) 205–218

A. F. J. VAN RAAN

In this presentation we argue that the core research activities of scientometrics fall in four interrelated areas: science and technology indicators, information systems on science and technology, the interaction between science and technology, and cognitive as well as socio-organisational structures in science and technology.

The literature of bibliometrics, scientometrics, and informetrics

Scientometrics,
Vol. 52, No. 2 (2001) 291–314

WILLIAM W. HOOD, CONCEPCIÓN S. WILSON

Since Vassily V. Nalimov coined the term 'scientometrics' in the 1960s, this term has grown in popularity and is used to describe the study of science: growth, structure, interrelationships and productivity. Scientometrics is related to and has overlapping interests with bibliometrics and informetrics. The terms bibliometrics, scientometrics, and informetrics refer to component fields related to the study of the dynamics of disciplines as reflected in the production of their literature.

B + IR

(Academic) Search

Information Retrieval

The problem of directing a user to stored information, some of which may be unknown to him, is the problem of "information retrieval."

Calvin N. Mooers (1950, p. 572)

Exploratory Data Analysis

Scientometrics

In other words, scientometric research nowadays is at the crossroads among the social sciences, information science, and advanced computing with its efforts to capture patterns in 'big data.'

Leydesdorff et Milojević (2015, p. 323)

The Field

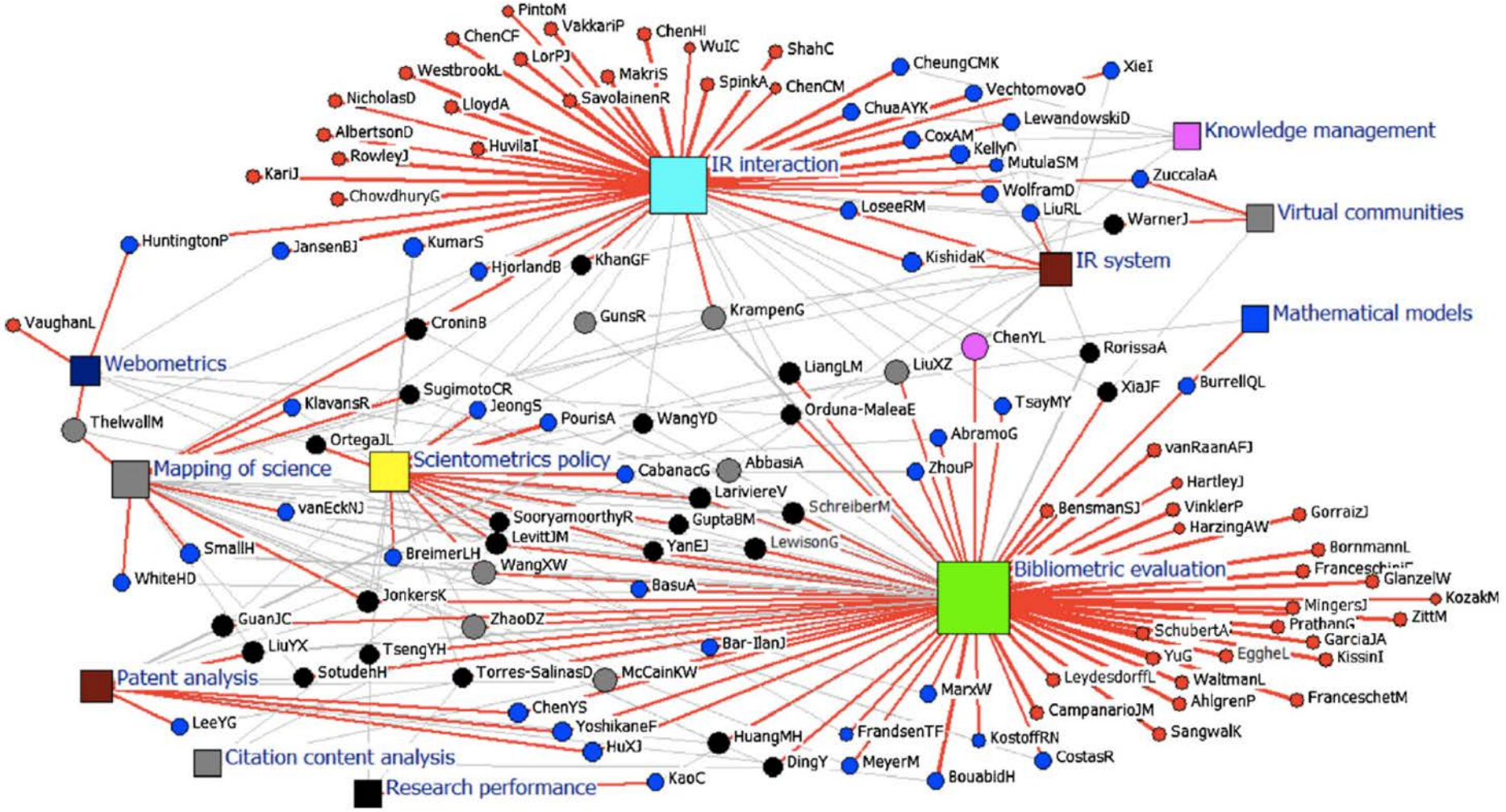


Fig. 2. Structure of the IS field in 2006–2015 as seen from the first-author AKCA.

Yang, S., Han, R., Wolfram, D., & Zhao, Y. (2016). Visualizing the intellectual structure of information science (2006–2015): Introducing author keyword coupling analysis. *Journal of Informetrics*, 10(1), 132–150. doi:10.1016/j.joi.2015.12.003

Agenda: Morning

Time	Speaker	Title
9:30	Introduction by organizers	
9:40	Cyril Labbé (Keynote)	Trends in Gaming Indicators: On Failed Attempts at Deception and their Computerised Detection

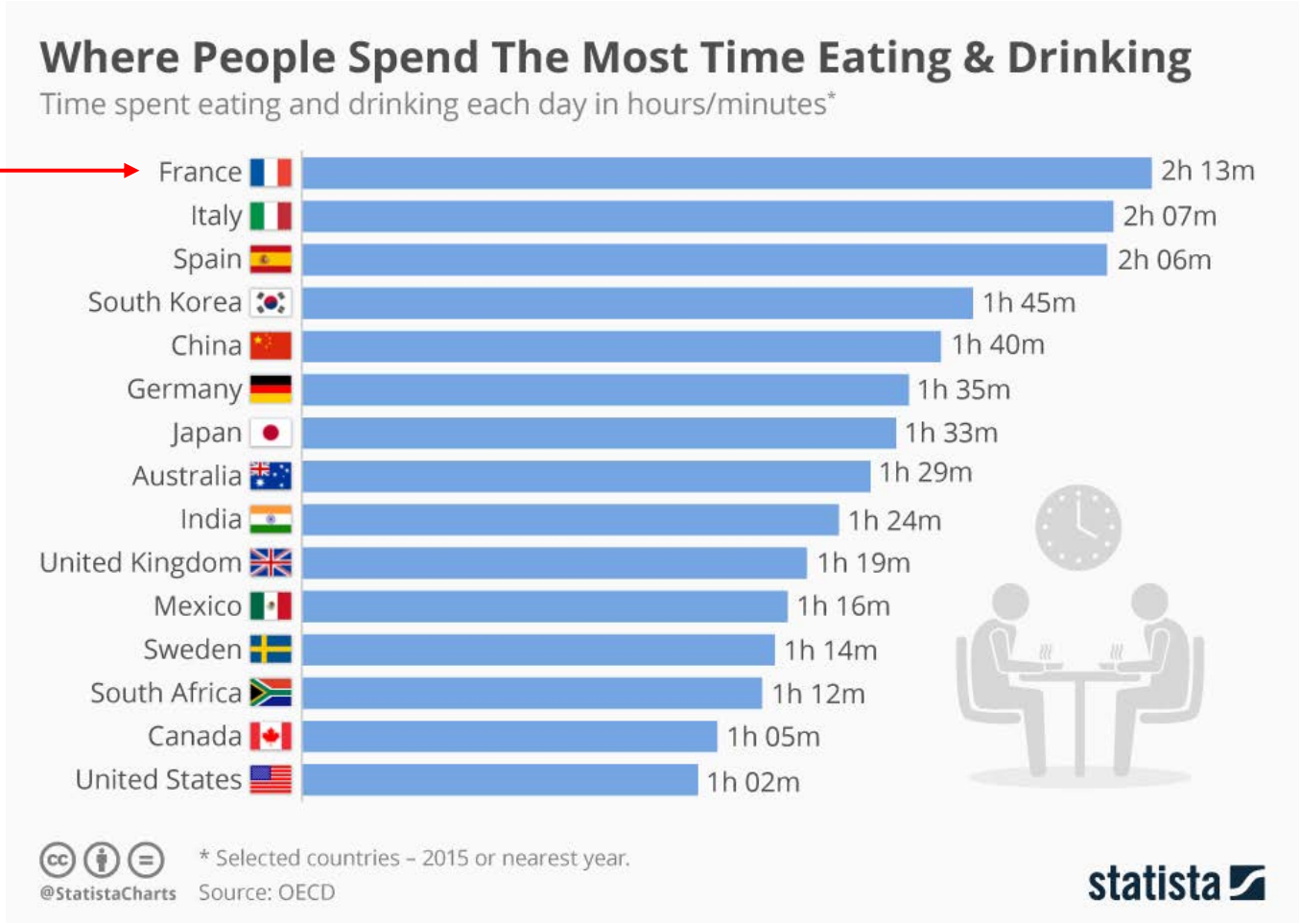
Counting articles and citations, analyzing citations and co-authors graphs have become ways to assess researchers and institutions performance. Fairly enough, these measures are becoming targets for institutions and individual researchers thus triggering new behaviors. As a matter of fact, scientometrics and informetrics systems of all kinds have to separate the grain form the chaff. Among others, fields like **Information Retrieval, network analysis and natural language processing may offers answers to deal with this kind of problems**. Through several emblematic case studies (fake researcher, generated papers, paper mills), we show **evidences of attempts to game indicators together with automatic ways to detect them** (automatic detection of generated papers, errors detection).

Agenda: Morning

Time	Speaker	Title
9:30	Introduction by organizers	
9:40	Cyril Labbé (Keynote)	Trends in Gaming Indicators: On Failed Attempts at Deception and their Computerised Detection
10:30	Maria Janina Sarol, Linxi Liu and Jodi Schneider	Testing a Citation and Text-Based Framework for Retrieving Publications for Literature Reviews
11:00	Coffee break 	
11:30	Anaïs Ollagnier, Sébastien Fournier and Patrice Bellot	BIBLME RecSys: Harnessing Bibliometric Measures for Scholarly Paper Recommender System
12.00	André Rattinger, Jean-Marie Le Goff and Christian Guetl	Local Word Embeddings for Query Expansion based on Co-Authorship and Citations
12:30	Mario Cataldi, Luigi Di Caro and Claudio Schifanella	All for One or One for All? Analyzing Collaboration Patterns in Research Environments (Demo) Demo-Session 1

Agenda: Lunch

Time	Speaker	Title
13:00-14:30	Lunch break	



Agenda: Afternoon

Time	Speaker	Title
14.30	Ralf Schenkel (Keynote)	Integrating and exploiting public metadata sources in a bibliographic information system

Bibliographic information systems need to rely on metadata provided by various sources in various forms and with various quality. The talk gives some insights how **the dblp bibliography** as an example for such a system is maintained and improved. It shows how metadata can be automatically harvested from publisher websites and **how the harvesting process can be steered**. It also discusses some **open sources of bibliographic metadata** and how they can be used to enrich existing bibliographic data, but also how varying their quality is.

Agenda: Afternoon

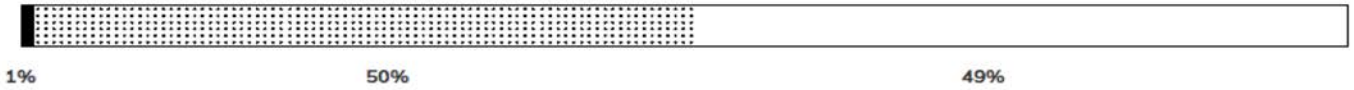
Time	Speaker	Title
14.30	Ralf Schenkel (Keynote)	Integrating and exploiting public metadata sources in a bibliographic information system
15:30	Marc Bertin and Iana Atanassova	InTeReC : an In-text Reference corpus for applying Natural language processing to Bibliometrics
16:00	Coffee break	
16:30	Ameni Kacem Sahraoui and Philipp Mayr	Users are not influenced by high impact and core journals while searching
17:00	Marie-Noëlle Bessagnet	A particular research data: a set of tweets (Demo) Demo-Session 2
17:30	Discussions	
18:00	End of the workshop	
18:30	Reception	

Hot Topics

Initiative for Open Citations

https://i4oc.org

How many citations are open today?



As of January 2018, the fraction of publications with open references has grown from 1% to more than 50% out of 38 million articles with references deposited with Crossref.

NEWS · 16 JANUARY 2018

nature
International journal of science

Science search engine links papers to grants and patents

The Dimensions database promises a financial perspective on scholarly literature.

Richard Van Noorden

Dimensions

https://dimensions.ai

RESEARCH

Fortunato *et al.*, *Science* **359**, 1007 (2018) 2 March 2018

REVIEW SUMMARY

SCIENCE COMMUNITY

Science of science

Santo Fortunato,* Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, Alessandro Vespignani, Ludo Waltman, Dashun Wang, Albert-László Barabási*

Merci !

#bir2018

Keynote #1

Dr. Cyril LABBÉ, Université Grenoble Alpes, France

Trends in Gaming Indicators:
On Failed Attempts at Deception and their Computerised Detection

Counting articles and citations, analyzing citations and co-authors graphs have become ways to assess researchers and institutions performance. Fairly enough, these measures are becoming targets for institutions and individual researchers thus triggering new behaviors. As a matter of fact, scientometrics and informetrics systems of all kinds have to separate the grain form the chaff. Among others, fields like **Information Retrieval, network analysis and natural language processing may offers answers to deal with this kind of problems**. Through several emblematic case studies (fake researcher, generated papers, paper mills), we show **evidences of attempts to game indicators together with automatic ways to detect them** (automatic detection of generated papers, errors detection).

Keynote #1 – Selected Highlights from LABBÉ's Research

Journal of Quantitative Linguistics
2001, Vol. 8, No. 3, pp. 213–231

0929-6174/01/0803-213$16.00

Inter-Textual Distance and Authorship Attribution Corneille and Molière*

Cyril Labbé[1] and Dominique Labbé[2]
[1]France Télécom, Meylan, France, and [2]CERAT-IEP, Grenoble, France

ABSTRACT

The calculation proposed in this paper measures neighbourhood between several texts. It leads to a normalized metric and a distance scale which can be used for authorship attribution. An experiment is presented on one of the famous cases in French literature: Corneille and Molière. The calculation clearly makes the difference between the two works but it also demonstrates that Corneille contributed to many of Molière's masterpieces.

SCIENCE VS. FRAUDULENCE

IKE ANTKARE, ONE OF THE GREAT STARS IN THE SCIENTIFIC FIRMAMENT

ISSI NEWSLETTER VOL. 6. NR. 2.

2010

CYRIL LABBÉ
University of Grenoble, LIG Laboratory. France
E-mail: Cyril.Labbe[at]imag.fr

Keynote #1 – Selected Highlights from LABBÉ's Research

Scientometrics (2013) 94:379–396
DOI 10.1007/s11192-012-0781-y

Duplicate and fake publications in the scientific literature: how many SCIgen papers in computer science?

Cyril Labbé · Dominique Labbé

NATURE | NEWS

nature

Publishers withdraw more than 120 gibberish papers

Conference proceedings removed from subscription databases after scientist reveals that they were computer-generated.

Richard Van Noorden

24 February 2014 | Updated: 25 February 2014

Scientometrics (2017) 110:1471–1493
DOI 10.1007/s11192-016-2209-6

Striking similarities between publications from China describing single gene knockdown experiments in human cancer cell lines

Jennifer A. Byrne[1,2] · Cyril Labbé[3]

Nature's 10

Jennifer Byrne
Error sleuth

Ten people who mattered this year.

Her tenacious work is now making waves. Journals have retracted nine papers as a result of Byrne's work — seven this year. And in October, she and French computer scientist Cyril Labbé released an online program called Seek & Blastn to help automatically detect similar problems. "When I'm on my deathbed, I'll look back and be really proud of this work," she says.

Keynote #2

Prof. Dr. Ralf SCHENKEL, University of Trier, Germany

Integrating and exploiting public metadata sources in a bibliographic information system

Bibliographic information systems need to rely on metadata provided by various sources in various forms and with various quality. The talk gives some insights how **the dblp bibliography** as an example for such a system is maintained and improved. It shows how metadata can be automatically harvested from publisher websites and **how the harvesting process can be steered**. It also discusses some **open sources of bibliographic metadata** and how they can be used to enrich existing bibliographic data, but also how varying their quality is.