ETH zürich

Quantifying scientific impact

Networks, measures, insights?

February 12-13, 2015

Villa Hatt, Zürich

Aims and Scope

Scientometrics, and quantitative studies of science in general, attract significant attention both from researchers and stakeholders. Importantly, they already impact the evaluation of researchers and the funding of proposals. Thanks to the availability of data from bibliographic databases, it has become particularly easy to construct and analyse collaboration and citation networks, develop novel measures of scientific impact, or even apply predictive analytics techniques to predict the future career of young scientists. Which quantitative measures convey the most information? What insights can we gain from these measures? Do, or will, such measures influence how we conduct science, how we collaborate or whom we cite? Which issues arise in the collection, preprocessing and analysis of bibliographic data? And should we quantify and/or predict scientific impact altogether?

In this workshop, we address novel research directions in this area, in particular network-based approaches for the quantitative evaluation of science. Participation is on an invitation-only basis and is restricted to 20 participants. It is the expressed goal of the organizers to bring together stakeholders dealing with the quantitative evaluation of science on a daily basis and scientists investigating novel ways to quantitatively study the structure and dynamics of scientific progress. As such, the list of participants includes prominent researchers in scientometrics from different disciplines, representatives of bibliometric data providers as well as key players in the provision of academic evaluation and ranking services.

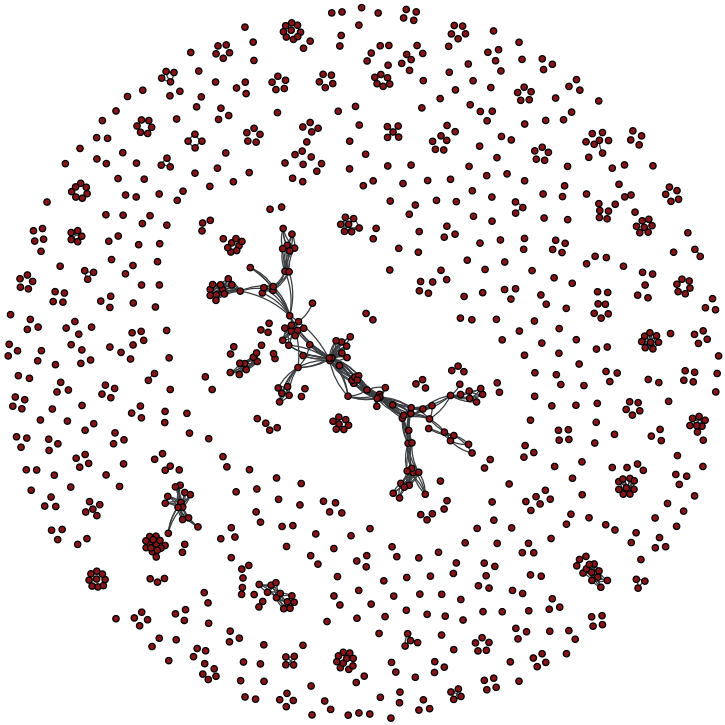
We are happy to welcome you at Villa Hatt, an exclusive seminar location offering beautiful views of Zurich, the lake and the mountains. We hope you will enjoy two inspiring and thought-provoking days in Zurich.

Frank Schweitzer
February 2015

cost
EUROPEAN COOPERATION
IN SCIENCE AND TECHNOLOGY

Participants

Judit Bar-Ilan
Sonja Berghoff
Peter van den Besselaar
Nees Jan van Eck
Martin Fenner
Robin Haunschild
Urs Hugentobler
Martin Juno
Evangelia Lipitakis
Matus Medo
Olesya Mryglod
Rüdiger Mutz
Vahan Nanumyan
Alexander Petersen
Filippo Radicchi
Martijn Roelandse
Martin Rosvall
Flaminio Squazzoni
Ingo Scholtes
Frank Schweitzer
Dirk Tunger



Program

Thursday, February 12

Time	
08:50 - 09:00	Frank Schweitzer **Opening Statement**
09:00 - 10:30	**Session: Bibliometrics** Speakers: Mutz, Haunschild, Tunger **Guiding questions:** • What is the value of single citations? How does the web-based increase in references affect this value? • Considering the algorithms/ranking/visualization methods you developed, what data would you wish for? • How can we better measure interdisciplinary impact?
10:30 - 11:00	Coffee Break
11:00 – 12:30	**Stakeholder Session: Ranking** Speakers: Berghoff, Juno, van Eck **Guiding questions:** • How do you determine the weights for ranking factors? How would you like to improve them? • To what extent are current ranking indicators biased to certain scientific cultures? • What are the risks and benefits of reputation surveys?
12:30 – 14:00	Lunch Break

Time	
14:00 – 15:30	**Session: Social Sciences** Speakers: van den Besselaar, Squazzoni, Bar-Ilan **Guiding questions:** • How does ranking influence the behavior of scientists and institutions? • Do services like Altmetric foster a tendency towards shallower research? • Do citiation alert mechanisms foster reciprocity, do they change citation cultures?
15:30 – 16:00	Coffee Break
16:00 – 17:30	**Session: Computer Science** Speakers: Radicchi, Rosvall, Scholtes **Guiding questions:** • What are the prospects of machine learning in rankings? Who decides about training data? • How can the time dimension be included in network-based ranking? • How can we improve name disambiguation methods?
19:00 – 21:00	**Joint Dinner in Restaurant Linde** (paid individually)

Program

Friday, February 13

Time	
08:50 - 09:00	Frank Schweitzer **Opening Statement**
09:00 - 10:30	**Stakeholder Session: Data** Speakers: Lipitakis, Fenner, Roelandse **Guiding questions:** • Given that you have the data, what kind of new quantitative indicators would you like to have? • How can access statistics (of publications, websites) be incorporated in ranking? How to prevent manipulations? • How do Open Access journals impact scientific quality? Is evaluation shifted from experts to the public?
10:30 - 11:00	Coffee Break
11:00 – 12:30	**Session: Statistical Analysis of Science Networks** Speakers: Medo, Mryglod, Petersen **Guiding questions:** • How can we model the feedback of bibliometrics (IF) on scientists› (career, journal) decisions? • Is fractional counting a solution to better capture the contribution of individuals? • Are the implicite assumptions of centrality measures justified in scientometrics?
12:30 – 14:00	Lunch Break

Time	
14:00 – 15:45	**Plenary discussion** Moderators: Hugentobler, Schweitzer **Guiding questions:** • How should bibliometric indicators (not) be used by funding agencies/ hiring committees (DORA project)? • Does bibliometric feedback lead to more specialized or more mainstream research? • Do alternative ranking schemes (U-Multirank) improve the situation or just shift the problems? • Can we use better quantitative approaches to ensure the quality of peer review/ institutional review?
15:45 – 16:00	Frank Schweitzer **Closing Statement**
16:00 – 16:30	**Coffee and Farewell**

Participant Statements

Prof. Judit Bar-Ilan
Department of Information Science, Bar-Ilan University, Israel
http://is.biu.ac.il/en/judit

Altmetrics , an emerging subfield of bibliometrics, looks for alternative and supplementary ways to assess impact and visibility of scholarly communication with an emphasis on social and societal impact. It relies on various Web based sources, among them the online reference manager Mendeley. It provides information on the number of users of the service that included the specific publication in their libraries (called "readers" of the publication). Mendeley has quite extensive coverage, thus allowing for comparisons between readership and citation counts. A number of studies found significant correlations of around .5 between these measures, indicating that readership counts capture an aspect of use of scholarly publications that are not completely reflected by citations. Thus readership counts are a good candidate to supplement traditional bibliometric measures. In the talk I will concentrate on articles published in JASIST - a leading information science journal; demonstrating the pros and cons of using Mendeley.

Dr. Sonja Berghoff
Centre for Higher Education (CHE), Gütersloh, Germany
http://www.che-ranking.de/cms/?getObject=613&getLang=en

CHE – Centre for Higher Education Development - is a private non-profit think tank for higher education. CHE established a ranking of German higher education institutions that was first published in 1998 and has been developed further since then in terms of scope and method. The CHE ranking is paradigmatic for field-based, multi-dimensional, interactive ranking that is following a grouping approach instead of league tables. CHE ranking has gained high international reputation. Since 2009 CHE is part of the consortium developing a multi-dimensional, user-driven approach to international ranking of higher

education institutions, called U-Multirank and prepared with funding from the European Union. First results were launched in May 2014. The dimensions it includes are teaching and learning, research, knowledge transfer, international orientation and regional engagement. Based on empirical data U-Multirank compares institutions with similar institutional profiles and allows users to develop personalised rankings by selecting performance measures/indicators in terms of their own preferences.

Prof. Peter van den Besselaar

Department of Organization Sciences, VU University Amsterdam, The Netherlands
http://www.vandenbesselaar.net

Quantifying scientific impact has many forms: rankings; citation and publication counts and related indicators; journal impact factors; and increasingly new (alt)metrics.
One hardly answered question is what these impact indicators actually measure, and therefore, what they can be used for. Or recursively, what quality dimensions are de facto important in decision-making and in selection processes, and how can these be adequately measured?
The second question is whether and how these metrics influence behavior – and that may be very different depending on e.g., career stage (individuals), and position in the higher education system (organizations). Do citation counts influence decisions of selection panels at all? And should they? Or is decision-making not influenced by scientific impact (metrics) at all, but dependent on social characteristics of knowledge production, and resource distribution?

Dr. Nees Jan van Eck

Head of ICT, Centre for Science and Technology Studies, Leiden University, The Netherlands
http://www.neesjanvaneck.nl

Nees Jan van Eck is a researcher in the field of bibliometrics affiliated with the Centre for Science and Technology Studies (CWTS) of Leiden University in the Netherlands. He is one of the developers (together with Ludo Waltman) of VOSviewer and CitNetExplorer, two well-known software tools for the analysis and visualization of bibliometric networks. Nees Jan also plays an important role in the production of the CWTS Leiden Ranking, a bibliometric ranking of major universities worldwide.
The Centre for Science and Technology Studies (CWTS) of Leiden University investigates the development of science, technology, and science-related innovation – often by using large-scale databases of scientific publications and patents. CWTS is a leading provider of science and technology indicators and of performance and benchmark studies of scientific groups and institutes, including bibliometric mapping and network analysis of science. CWTS has developed longstanding relationships with prestigious universities and is a recognized international leader in the field of bibliometrics, scientometrics, and informetrics. The CWTS information system offers full and unrestricted access to various in-house databases such as Web of Science, Scopus, and PATSTAT.

Dr. Martin Fenner

Technical Lead PLoS Article-level Metrics, Public Library of Science, United Kingdom
http://www.plos.org

The Public Library of Science (PLOS) is an open access publisher in the life sciences. Since 2009 we are collecting and displaying article-level metrics (ALM) for all our articles in the categories usage stats (views and downloads), citations, discussions (social media, wikipedia, etc.) and social bookmarks. Our guiding principles for this activity are the following:

- provide metrics for the invidual article instead of aggregated journal metrics
- provide a wide range of metrics not limited to citations
- make the data and the tools to generate the data openly available
- collaborate with other stakeholders (publishers, institutions, funders, etc.) on use cases, common tools, and best practices

Dr. Robin Haunschild

Information Retrieval Service, Max Planck Institute for Solid State Research, Stuttgart, Germany
http://www.ipc.kit.edu/theochem/456_715.php

After ten years of research in theoretical chemistry, I found my true calling in scientometric research and joined the Max Planck Institute for Solid State Research in September 2014. Since then, I am responsible for research evaluation and scientometric research. My current research interests include research evaluation, peer review, bibliometrics, altmetrics, and chemical/physical bibliometrics. My two main research areas (research evaluation and altmetrics) can be combined as the impulse to include altmetrics in research evaluation gets stronger. One of my contributions in research evaluation is my viewpoint regarding the Nature Index. A series of publications resulted in my interest in Mendeley reader counts. Regarding the interest of the scientific community in the peer review process, I analyzed the Mendeley bookmarking of F1000 evaluations. I also use my experience from scientometric research for research evaluation within the Max Planck Society.

Dr. Urs Hugentobler

Head of Institutional Research, Controlling, ETH Zürich, Switzerland
https://www1.ethz.ch/fc/about/controlling/bereiche_co/ir

Urs Hugentobler is heading the Institutional Research Group in ETH Zurich's Controlling Department (which is part of the Executive Board domain for Finance and Controlling. Amongst others, the Controlling Department provides relevant information for the ETH Zurich management by supplying qualified data for decision-making processes.

... and what does it mean?

Quantitative measures on 'scientific performance' are more and more getting readily available, whereas the considerably methodological research in this field is providing a sound basis for discussing the validity of these measures. However, these measures provide information which in most cases is 'neutral' in itself. It is getting relevant only by setting the information in context of a frame of reference. Measures of scientific impact have to be evaluated in view of the questions raised by stakeholders, being a funding agency, politics or university management. At ETH Zurich, the Executive Board therefore uses a variety of measures and information - among them results of bibliometric analyses - as a basis for its dialogue with the Departments.

Martin Juno

Product Manager - Benchmarking Service, QS Quacquarelli Symonds Limited, London, United Kingdom
http://www.iu.qs.com

QS has been conducting research in a range of areas since 1990, beginning with a global survey of MBA employers. The QS World University Rankings®, the most established of the range of research projects that QS operates, have been in existence since 2004. The QS Intelligence Unit (QSIU) was formed in 2008 as a distinct and autonomous department. Some of its activities involve consultancy, benchmarking and rating services, and

the production of QS World University Rankings©, the QS Subject Rankings© and regional initiatives. QS covers 50 countries and evaluate over 3,000 of the world's leading higher education institutions. Furthermore, QS arguably conducts the most extensive annual academic and employer surveys in the world. In 2014, almost 64,000 academics and 29,000 employers contributed their opinion on the top institutions globally.

Martin Juno is the manager of the QS Benchmarking Service and is the leading analyst for the Latin American rankings.

Dr. Evangelia Lipitakis

Scientific & Scholarly Research, Thomson Reuters, London, United Kingdom
http://ip-science.thomsonreuters.com

Evangelia is a research analytics consultant for European, Middle Eastern and African (EMEA) higher educational institutions, research institutions and research assessment funding bodies (governmental/private) for the Scientific & Scholarly Research (SSR) division of Thomson Reuters. Evangelia holds a Ph.D. in Bibliometrics and a Master's degree in Management Science and specializes in the area of academic research performance evaluation and quantitative methodologies for measuring research performance. Before joining Thomson Reuters, Evangelia was conducting research in the area of Scientometrics and lecturing at the University of Kent, England.

Dr. Matus Medo
Physics Department, University of Fribourg, Switzerland
http://www.ddp.fmph.uniba.sk/~medo/physics/

I am research assistant at the University of Fribourg in the group of Yi-Cheng Zhang. I am currently most interested in studying the temporal patterns in the evolution of information and social data (bipartite user-item data produced by e-commerce platforms, citation networks, and others). To this end, I develop and validate network models, develop new temporal metrics, and study implications of the temporal patterns found in the real data on the algorithms that are typically used to analyze this kind of systems (PageRank, HITS, etc.). Information filtering (recommender and reputation systems) form the second direction of my research.

Dr. Olesya Mryglod
National Academy of Sciences of Ukraine, Lviv, Ukraine
http://www.icmp.lviv.ua/en/olesya_mryglod

While peer review and citations reflect opinion about a paper›s quality and scientific impact after reading, downloads rather reflect interest before reading. In other words, in addition to popularity and prestige, papers may be distinguished by their attractiveness. In such a classification, the overall number of citations measures popularity, the number of important citations is evidence of prestige, whereas the number of downloads reflects the level of attractiveness of a publication.

In our work the downloading statistics of publications in «Europhysics Letters» journal is analysed. We find that the journal is characterised by fast accumulation of downloads during the first couple of months after publication, followed by a slower rate thereafter. This behaviour can be modelled, so that the long-time download patterns for the journal can be predicted. We also find that individual papers can be classified in various ways according to their downloading statistics.

Dr. Rüdiger Mutz
Chair of Social Psychology and Research on Higher Education, ETH Zürich, Switzerland
http://www.psh.ethz.ch/people/mutz/

How to use bibliometric data to rank universities according to their research performance?

Publication and citation data is often used to make a statement about the productivity and the citation impact of institutions. The talk is about some problems and solutions to the use of bibliometric data for analyzing research performance of universities in order to finally rank them: statistical instead of numerical viewpoint, confounding of rankings, visualization. Two empirical applications should give an impression about possible solutions. First, we have statistically reformulated the Leiden Ranking 2011/2012 by means of a multilevel regression model. Second, the web application "Excellence mapping" (Bornmann, Stefaner, de Moya Anegón, and Mutz, 2014) will be shortly outlined, where bibliometric data, graphical design and statistical concepts are combined. The web application visualizes institutional performance within specific subject areas as ranking lists and on custom tile-based maps. Scopus data were used which have been collected for the SCImago Institutions Ranking.

Vahan Nanumyan
Chair of Systems Design, ETH Zürich, Switzerland
https://www.sg.ethz.ch/people/vnanumyan/

The PhD programme I am undertaking has two key aspects. One is extending the methodology of multi-layer and multiplex networks. The other aspect is the application of these methods to problems of sociotechnical systems, which can be described as comprising information and social components, or layers. Citation and collaboration networks in science are a prominent example for such systems.

An ongoing project is aiming at delivering a minimalistic model of coupled formation of citation and collaboration networks. The hypothesis is that such model can describe the stylized facts about these networks better than conventional models of similar complexity, but developed only for one of the two components. Preliminary results hint that the citation formation is dependent on the collaboration dynamics, the strength of this dependence varying for different disciplines. The study is supposed to extend our understanding of biases in citations and add yet another argument against purely citation based measures in scientometrics.

Prof. Alexander Petersen
IMT Lucca, Italy
http://physics.bu.edu/~amp17/webpage_files/publications.html

Research does not produce itself. Instead, there are idiosyncratic individuals involved, characterized by diverse backgrounds, interests, behaviors, strategies, and goals. As such, science is an extremely complex socio-economic innovation system. I use data-driven computational methods to analyze and model the science of science, where the unit of analysis can vary across multiple scales, from publications, to individuals (careers), to teams, and large institutions such as countries. Against this multilevel backdrop, questions motivated from the theories of complex systems, management

& organization science, labor economics, and research policy are often the starting point. Are there quantifiable patterns of scientific success? Are they useful in the career evaluation process? Are there ways to improve the sustainability of science careers while at the same time maintaining a high level of competitive selection? How do metrics for individual achievement depend on collaboration factors? How might paradigm shifts in science affect science careers?

Prof. Filippo Radicchi
Center for Complex Networks and Systems Research, Indiana University, Bloomington, USA
http://filrad.homelinux.org

Bibliographic data can be naturally modelled as directed graphs. This representation allows us to measure the influence of papers, scientists and journals with alternative metrics imported from network science. We provide an application of these techniques to the weighted network of citations between authors of papers published in the Physical Review collection. We compare the results obtained with our algorithm with those obtained by local measures such as the citation count, and provide a statistical analysis of the assignment of major career awards in the area of physics. We further propose a recipe to use the longitudinal nature of bibliographic data to generate dynamical impact metrics that describe career trajectories of individual scholars.

Dr. Martijn Roelandse
Senior Editor, Biomed, Springer, Heidelberg
http://www.springer.com/gp/about-springer/media/press-releases/corporate/springer-and-peerage-of-science-team-up-/39456

Springer Science+Business Media is one of the largest scholarly publishers with the largest open access portfolio worldwide. In 2013, we published more than 8,400 new book titles and some 2,200 English-language journals, including 450 open access journals. A newly developed manuscript transfer service will decrease peer review pressure and increase manuscript quality, for both OA and non-OA manuscripts.

We have embraced article-level metrics in various parts of the company. Not on our article pages, we also use top cited, top downloaded and top shared article metrics in our marketing message, either for a particular journal or across a discipline. These access metrics can also help us to identify upcoming topics and new leads. Ultimately, all types of scholarly output should be treated equally, being articles, chapter, data, software, etc. as these metrics are pivotal to assess research output and benchmark gainst other (in whichever way).

Prof. Martin Rosvall
Umea University, Sweden
http://www.tp.umu.se/~rosvall/

Machine learning for robust rankings
Journal rankings derived from the citation network between a selection of journals unavoidably depend on this selection. However, little is known about how robust rankings are to the selection of included journals. In my presentation, I will compare the robustness of three journal rankings based on network flows induced on citation networks. They model pathways of researchers navigating scholarly literature, stepping between journals and remembering their previous steps to different degree: zero-step

memory as impact factor, one-step memory as Eigenfactor, and two-step memory, which requires citation data from twice as long a time period. Using machine-learning techniques, we conclude that two-step memory gives slightly more robust ranking, because it combines the advantages of the lower-order models: perturbations remain local and citation weights depend on journal importance.

I will also show our latest visual tools for better search and discovery of scientific articles available on http://infobaleen.com.

Prof. Flaminio Squazzoni

Research Group on Experimental and Computational Sociology, University of Brescia, Italy
http://www.eco.unibs.it/gecs/squazzoni.html

Competition, serious "gamification" and scientist misbehaviour: Can quantitative indicators and rankings be neutral and context-free measures of the quality of science?

In this talk I shall discuss the impact of quantitative indicators on scientist behaviour and institutional policies. First, recent findings on peer review and scientists' publication strategies revealed possible misuses of science indicators and bias in allocation of resources. Secondly, recent reports from important agencies in Italy and Norway indicated potential trade-offs between measuring the quality of scientific outputs and needs for preserving context-dependent aspects. Finally, rankings have serious pros and cons that should be considered. On the one hand, they can fruitfully inform scientist decisions and help scientific institutions in monitoring and incentivising the quality of their output. On the other hand, rankings are social artefacts that meet an elementary social need of humans and may trigger emotional, irrational behaviour, even nurturing animal spirits that, in some cases, can be detrimental for the long-time evolution of the scientific community.

Dr. Ingo Scholtes
Chair of Systems Design, ETH Zürich, Switzerland
http://www.ingoscholtes.net

The overabundance of information is a common feature of knowledge spaces emerging in Online Social Networks, web-based communication and collaboration tools as well as scholarly publications. The design of measures and algorithms targeting an automated extraction of relevant knowledge from such systems requires a solid understanding not only of the structure and dynamics of the underlying knowledge spaces, but also of the social processes that create and shape them.

I will introduce a methodology aimed at improving our understanding of the social processes involved in the collective filtering of information. Our approach is based on the analysis of evolving multiplex network structures, that combine the social layer of collaborations with a semantic layer of interlinked information. Interpreting it as the outcome of a collective information filtering process, we measure the success of a particular piece of information in the semantic layer and study the social position of the actor that has contributed that particular piece of information. We then apply statistical inference and machine learning techniques to uncover complex statistical dependencies which allow to quantify whether the social positions of actors are indicative for the success of information they contribute. We use a large-scale data set of 100,000 scholarly publications and study how centrality in the coauthorship network differs between authors who have highly cited papers and those who do not. We find that collaboration structures are indicative for the success of publications in terms of citations.

Prof. Frank Schweitzer

Chair of Systems Design, ETH Zürich, Switzerland
https://www.sg.ethz.ch/team/people/fschweitzer/

«Science is done by people» - this is the opening sentence of Werner Heisenberg›s autobiography. Question is to what extent we realise this in our quantitative assessment of scientific outcome. Our research focuses on the role of social networks in science and how they can be used to predict the success of scientific papers. From a broader perspective, we are interested in colloaboration networks in general where the nodes of the network, the agents, can be scientists, firms, software developers etc. and links represent their time bounded interactions. Monitoring the structure and dynamics of such collaboration networks over time, what can we learn about the success of individual agents or the performance of the network as a system? Are there early indicators of failure or success? And how can we design interactions such that the performance of collaboration networks can be improved?

To answer such questions, it takes more than «big data». We develop refined methods, by building on machine learning, network analysis, and agent-based modeling, and we underpin our data-driven modeling approach using concepts of social organizations, cooperation, collective decision processes.

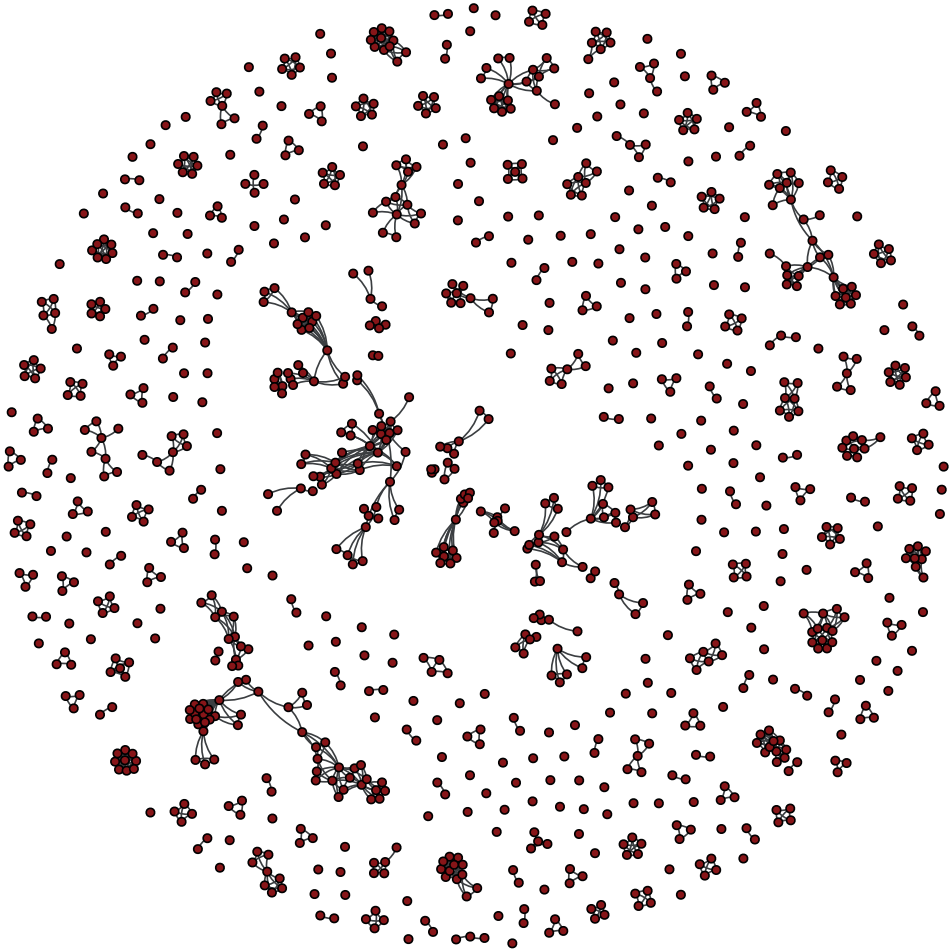
Dr. Dirk Tunger

Head of the Bibliometry Team, Forschungszentrum Jülich, Germany
http://www.bibliometrie.de

Dirk Tunger studied information sciences at Hamburg University and received his PhD in 2007 from the University of Regensburg in information science. His PhD thesis focused on the use of bibliometrics for trend recognition. This is a topic, he is still interested in. Another major interest is the use of bibliometrics for science evaluation.

Dirk Tunger has worked as an information scientist at the Central Library of Forschungszentrum Jülich since 2003. Therefore, his perspective on bibliometric methods is the information science perspective. Databases as the source of bibliometric data are one of the fundamentals for bibliometric analyses. That makes it interesting for Dirk Tunger to look at the databases, their content and their errors to answer questions about the need of data correction. As Scopus was introduced by Elsevier, he was one of the first authors, who compared Scopus and Web of Science from a content and a bibliometric point of view.
should be considered. On the one hand, they can fruitfully inform scientist decisions and help scientific institutions in monitoring and incentivising the quality of their output. On the other hand, rankings are social artefacts that meet an elementary social need of humans and may trigger emotional, irrational behaviour, even nurturing animal spirits that, in some cases, can be detrimental for the long-time evolution of the scientific community.

Notes



Chair of Systems Design
ETH Zürich
D-MTEC, WEV
Weinbergstrasse 56/58
8092 Zürich
Switzerland

www.sg.ethz.ch

Cover image: ETH Division for Events and Location Development