JÜLICH
FORSCHUNGSZENTRUM

Bibliometric Analyses as a part of an information system in science management

26. April 2017 | Dr. Dirk Tunger

Agenda

JÜLICH
FORSCHUNGSZENTRUM

- Bibliometric analyses for monitoring research performance (for individuals, institutes and universities): Methods of Research Center Jülich

- Leiden Manifesto: What should be kept in mind and possible pitfalls for bibliometric analyses?

- Dynamic Topic Modelling: Identification of emerging fields in science. Discussion of a possible set of data sources as well as the combination of these data sources for the use as a trend recognition system in science. How does it work and where are bottlenecks?

Introduction

The focus is on research institutions as creators of a steadily growing, multidisciplinary scientific output. These compete with each other to rank among the leading institutions in their disciplines internationally and also to document their position through the perception of their publications.

(Price, 1963)

"Every enterprise and almost every organisation or corporation is confronted with the task to monitor and evaluate the performance [...] of its teams, or of the whole unit"

(Wagner-Doebler, 2003)

"Bibliometrics is the application of mathematical and statistical methods to publications and other media of communication […]"

(J. Gorraiz)

What is the value of single citations?

JÜLICH
FORSCHUNGSZENTRUM

- A Connection of 2 or more publications, indicating that knowledge has been transfered
- Indicator for acceptance of scientific results
- Paying attention to somebody´s work
- Possibility to show the knowledge about standard publications and the state of the art in a discipline

One citation is not much, but to achieve more citations than the average of other researchers.

Bibliometrics

- is creating indicators to meassure impact
- is deviding the science system into disciplines
- is comparing the impact of different units
- is looking for connections

STM-publications in WoS

JÜLICH FORSCHUNGSZENTRUM

Disziplin	Anzahl Publikationen	durchschnittliche Zitationsrate
Agricultural science	35.883	2,4
Biology and biotechnology	142.667	5,2
Chemistry	118.267	4,8
Energy	10.246	3,1
Engineering	92.090	2,4
Geosciences	59.291	3,0
Information and computer science	22.851	2,3
Materials science	53.064	3,4
Mathematics	33.495	1,6
Medicine	289.765	4,9
Multidisciplinary journals	8.938	15,5
Nanotechnology	10.973	6,7
Physics	116.663	4,3

STM-publications in WoS

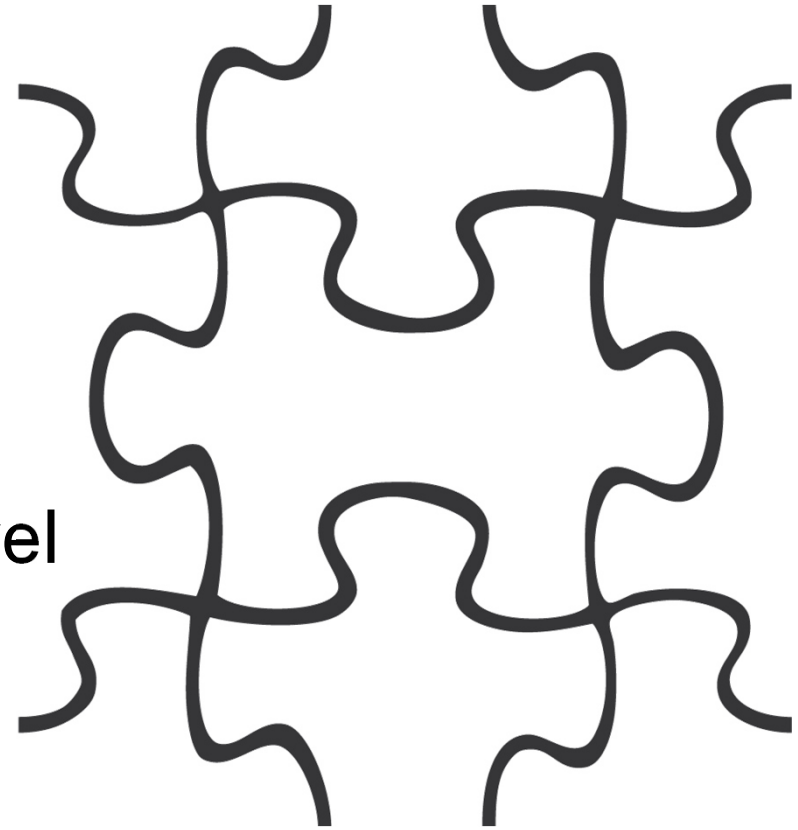
Discipline	Publications [%]		
	in journals	in WoS	WoS coverage
Molecular biology & biochemistry	96	97	92
Biological sciences (related to humans)	95	95	90
Chemistry	90	93	84
Clinical Medicine	93	90	84
Physics and astronomy	89	94	83
WoS average	*84*	*90*	*75*
Applied physics and chemistry	83	89	73
Biological sciences (animals and plants)	81	84	69
Psychology and psychiatry	75	88	66
Geosciences	77	81	62
Other social sciences (medicine and health)	75	80	60
Mathematics	71	74	53
Economics	59	80	47
Engineering	60	77	46
Social sciences	41	72	29
Humanities and arts	34	50	17

(Craig & Ferguson, 2009, S. 163)

JÜLICH
FORSCHUNGSZENTRUM

Our bibliometric toolbox:

- Publication analysis
- Cooperation analysis
- Benchmarking
- Topic analysis
- Individual researcher level
- Network analysis



Measuring of interdisciplinary impact – the J-Factor

JÜLICH FORSCHUNGSZENTRUM

$$J(I,R) = \sum_{S} \frac{cpp_I(S)}{cpp_R(S)} \bullet \frac{p_I(S)}{p_{I,ges}}$$

J(I,R): J-Factor of Institution I, in Relation to Benchmark R

S: Journal

$CPP_I(S)$: average Citationrate of the Publications by Institution I in Journal S

$CPP_B(S)$: average Citationrate of the Publications by Benchmark R in Journal S

$p_I(S)$: total number of publications by institution I in Journal S

$p_{I,ges}$: total number of publications by institution I

Ball, Mittermaier, Tunger, 2009)

Measuring of interdisciplinary impact – the J-Factor

JÜLICH FORSCHUNGSZENTRUM

			Benchmark			Institution			Calculation		
Doc. Type	PY	Journal	P	C	CPP	P	C	CPP	rel. CPP	Weighting	Inkre-ment
Article	2005	Journal of Applied Crystallography	130	1300	10	3	9	3	30%	14%	4%
Article	2007	Journal of Crystal Growth	75	450	6	5	15	3	50%	23%	11%
Review	2008	Nature Reviews Immunology	14	210	15	2	60	30	200%	9%	18%
Article	2008	The Journal of Immunology	150	1500	10	10	300	30	300%	45%	136%
Letter	2005	American Journal of Human Genetics	16	32	2	2	6	3	150%	9%	14%
		total	**395**			**22**				**100 %**	**184 %**

Measuring of interdisciplinary impact – the J-Factor

JÜLICH FORSCHUNGSZENTRUM

Pro

- Individual Benchmark for every research unit makes it easier to accept the result
- No problems with publications in interdisciplinary journals
- No problems with the definition of scientific fields

Contra

- Working with averages

Individual researchers level

- The age and the position of a scientist are responsible for his publication and citation data.
- Cross-disciplinary bibliometric comparisons are generally not directly possible due to different communication habits.
- At the level of individual researchers, bibliometric analyses must be handled with care due to low publication figures.

Leiden Manifesto 2015

1. Quantitative evaluation should support qualitative, expert assessment.
2. Keep data collection and analytical processes open, transparent and simple.
3. Allow those evaluated to verify data and analysis.
4. Be careful with variation by field in publication and citation practices.
5. Avoid misplaced concreteness and false precision.
6. Scrutinize indicators regularly and update them.

(The Leiden Manifesto for research metrics, 2015)

Dynamic Topic Modelling

Central Question

- What are the topics or buzzwords the publications are about?

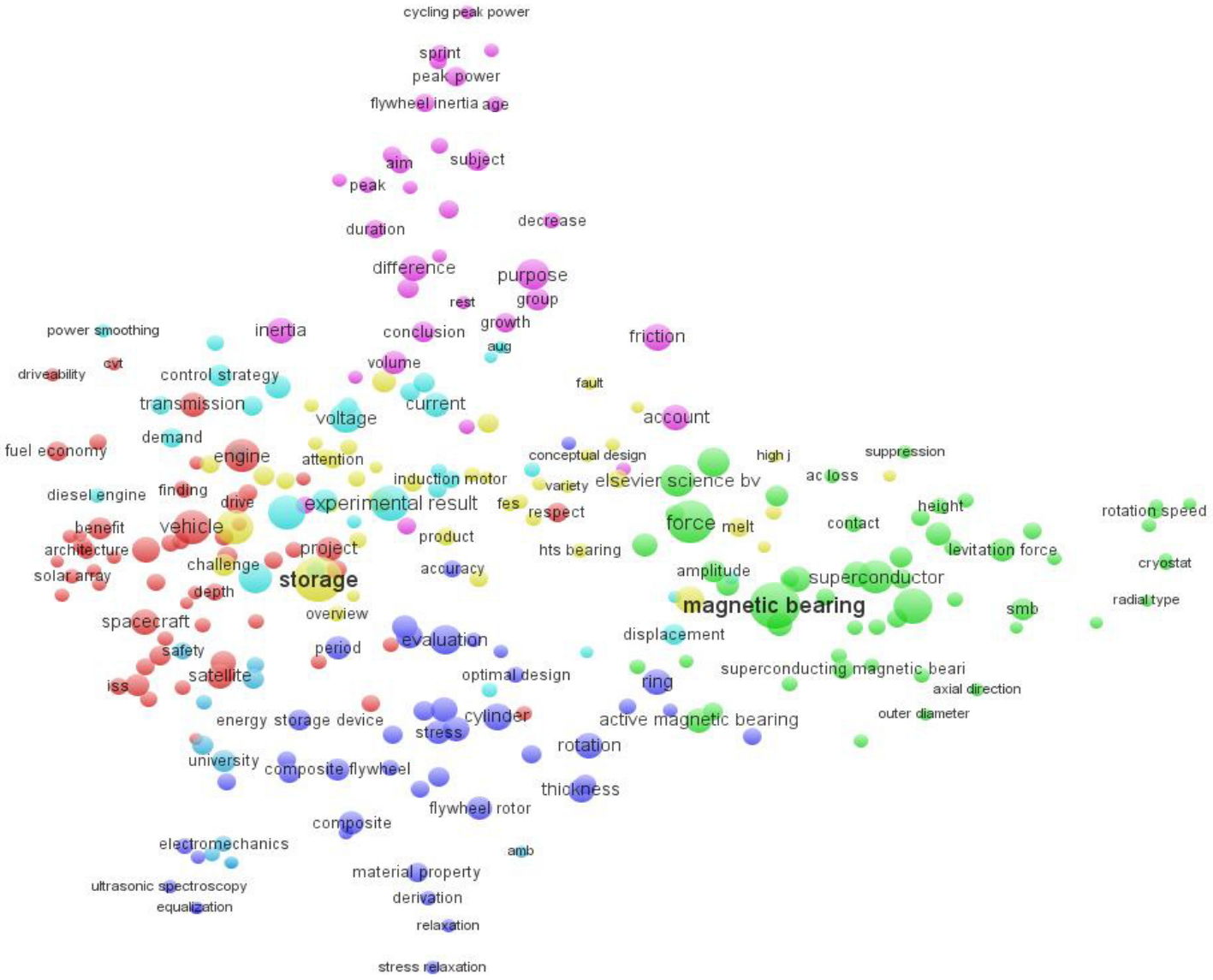
Dynamic Topic Modelling

The more central a term for the examined documents, the more central is the position of this term in the network. Centrality arises from the fact that a concept with a large number of the other terms occurs together in the same documents.

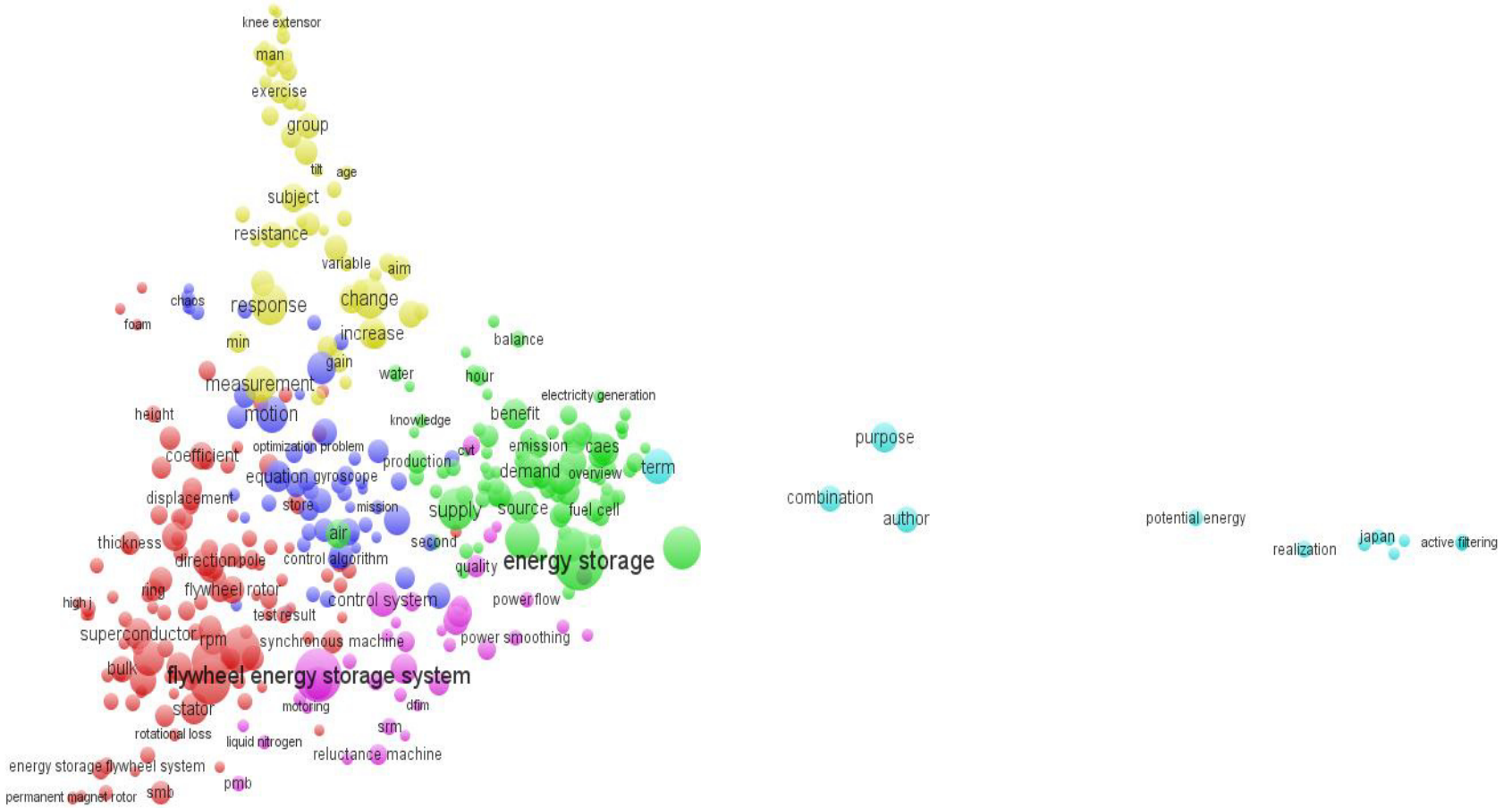
On the other hand, frequently occurring terms which are only mentioned together with a small number of other terms are further outwards in the representation. In addition, the terms which are most common together, are clustered together. Clusters are characterized by the same color of the circles and form thematic sub-areas of the analysis. The size of the circles is proportional to the occurrence of a single concept. A thesaurus does exclude non-meaningful terms and synonyms.

The diagram does not have any axes.

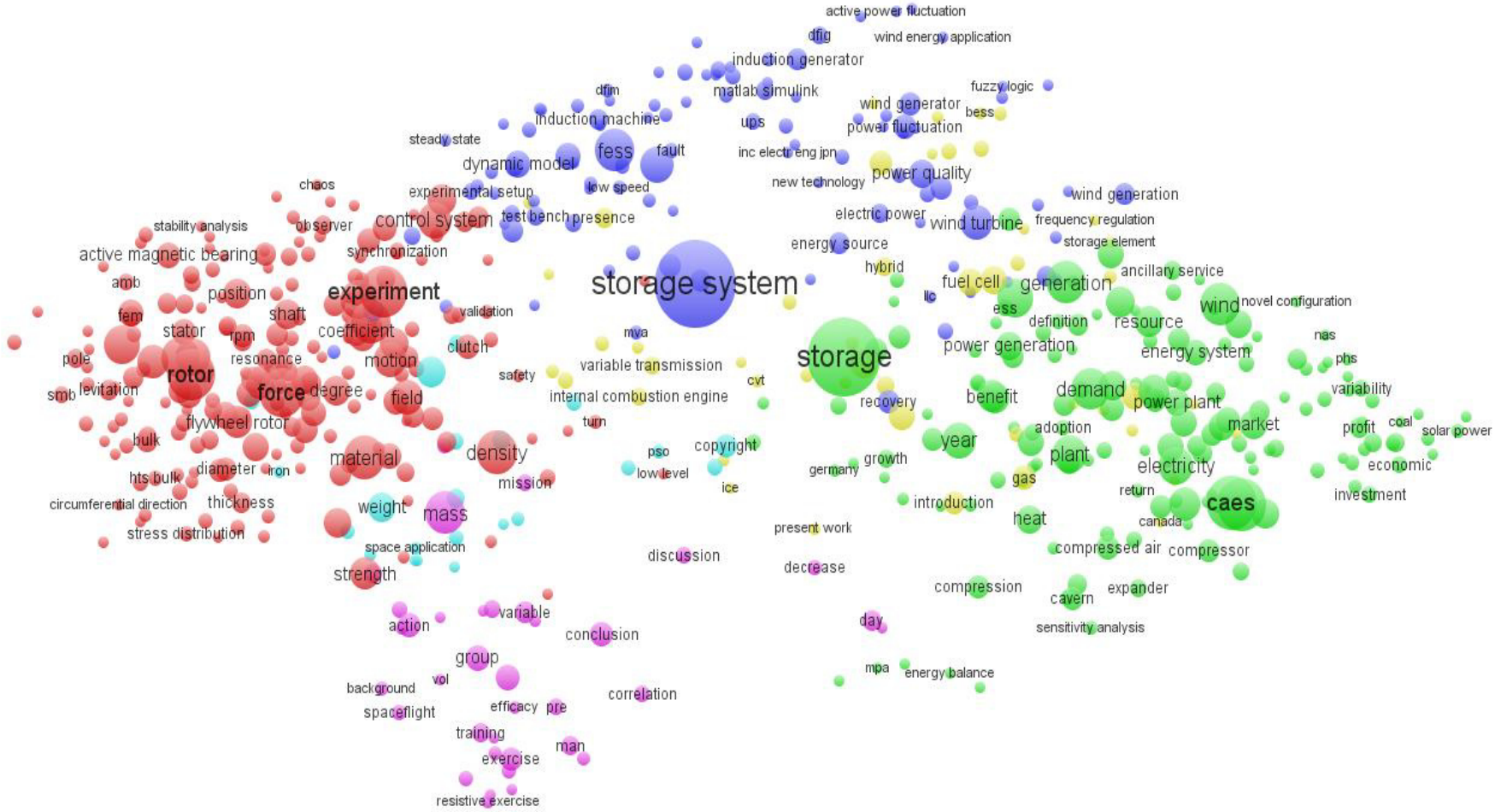
Mechanical energy storage 1999 - 2003



Mechanical energy storage 2004 - 2008



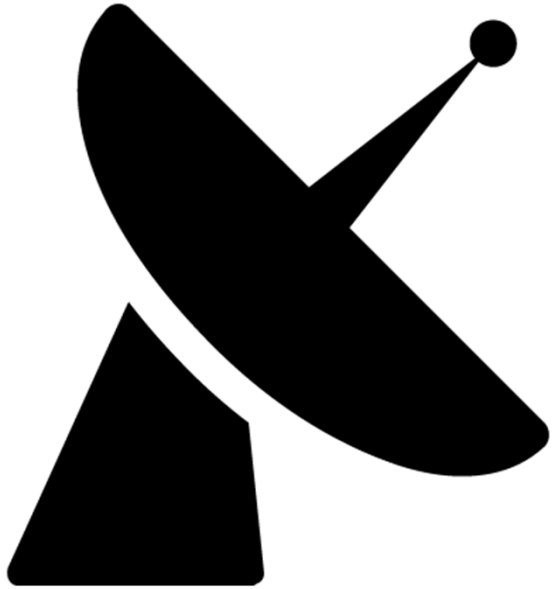
Mechanical energy storage 2009 - 2014



Trend recognition in science

How can quantitative and qualitative data be combined for trend recognition in science?

- New Journals
- News items in Nature / Science
- Topics from project databases
- Highly cited paper / hot paper



Dr. Dirk Tunger

Forschungszentrum Jülich
Bibliometrics Team
52425 Jülich

d.tunger@fz-juelich.de
www.bibliometrie.de
+49 2461 61-6198

Thank you!