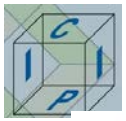
Appropriate and Inappropriate Uses of Journal Bibliometric Indicators (Why do we need more than one?)

Gianluca Setti
Department of Engineering, University of Ferrara
2013-2014 IEEE Vice President, Publication Services and Products
gianluca.setti@unife.it

Consiglio Scientifico GTTI
"Linee di indirizzo per una valutazione di qualità della produzione scientifica
Roma, 11 Gennaio 2016

IEEE Initiatives on Proper Use of Bibliometrics

1. Make clear that manipulation of any bibliometric indicator is **unethical**
2. Promote the **adoption of multiple bibliometric indicators** to evaluate the impact of scientific publications and of individual papers
3. **Educate the community** on the significance of all bibliometric indicator and their proper use
   a) panel discussion at the 2013 and 2014 IEEE Panel of Editors
   b) presentation on this subject and major IEEE conferences (so far ISCAS2013 ICIP2013, CDC2013, ISCAS2014, PES-GM 2014), NSF and to the Association of Heads of Electrical and Computer Engineering Departments (ECEDHA)



IEEE Initiatives on Proper Use of Bibliometrics

1. Make clear that manipulation of any bibliometric indicator is **unethical**
2. Promote the **adoption of multiple bibliometric indicators** to evaluate the impact of scientific publications and of individual papers
3. **Educate the community** on the significance of all bibliometric indicator and their proper use
   a) panel discussion at the 2013 and 2014 IEEE Panel of Editors
   b) presentation on this subject and major IEEE conferences (so far ISCAS2013 ICIP2013, CDC2013, ISCAS2014, PES-GM 2014), NSF and to the Association of Heads of Electrical and Computer Engineering Departments (ECEDHA)
4. **IEEE position statement on correct use of bibliometrics (approved by BoD in 09/2013)**

/www.ieee.org/publications_standards/publications/rights/bibliometrics_statement.html

IEEE Initiatives on Proper Use of Bibliometrics

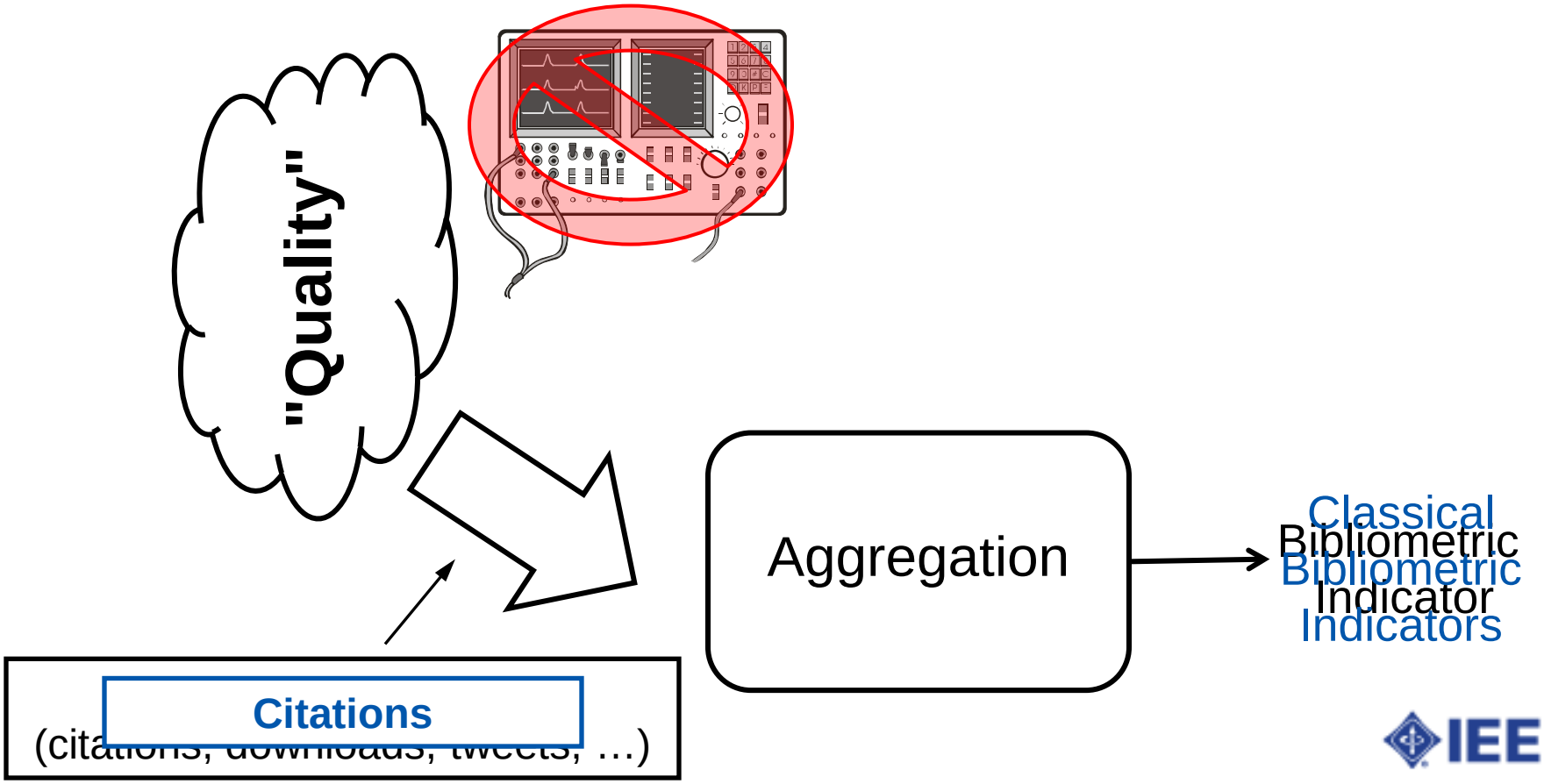
1. The use of **multiple complementary bibliometric indicators** is fundamentally important to offer an appropriate, comprehensive and balanced view of each journal in the space of scholarly publications.
2. **Any journal-based metric** is not designed to capture qualities of individual papers and **must therefore not be used as a proxy for single-article quality or to evaluate individual scientists**.
3. While bibliometrics may be employed as **a source of additional information** for quality assessment within a specific area of research, **the primary manner** for assessment of either the scientific quality of a research project or of an individual scientist **should be peer review**.
4. The IEEE **explicitly and firmly condemns** any practice aimed at influencing the number of citations to a specific journal with the sole purpose of artificially influencing the corresponding indices.

Outline

1. Overview on journal bibliometric indicators
2. Show that the "quality" of a **journal** as measured by journal bibliometric indicators is a multidimensional concept which **cannot be captured by any single indicator**
3. Show that the bibliometric indicators should not be misused by giving them **"more significance than they have":**
   a) the impact of an **individual paper cannot be measured** by the impact of the journal in which it has appeared
   b) there is **no strong correlation** between the Impact Factor of a journal and its **selectivity** (rejection rate)
   c) the Impact Factor of a journal **is not a good proxy** for the probability that **an individual paper will be highly cited**
4. Highlight that the misuse of journal bibliometric indicators has **undesired consequences** in the behavior of editors and individuals

Bibliometrics

- Definition: **Bibliometrics** is a set of methods to **quantitatively analyze** scientific and technological literature (it is part of Informetrics, which does the same for all information)



Journal Bibliometric Indicators, i.e. …numbers, numbers, numbers…

Many bibliometric indicators exist, each aiming to measure "journal quality"; they should:

1. ***Give a result which corresponds to the technical quality of the papers published in that journal***: *Nature*, *Science* or *Proceedings of the IEEE* and the "*Journal of Obscurity*" should have a very different value of the indicator
2. ***Be "fair" if applied to different areas***: different areas/communities may have different citation practices (e.g., long/short citation list)
3. ***Be immune to external manipulation***: it should be very difficult to artificially manipulate its value

Impact Factor and its criticisms - I

- Introduced by Eugene Garfield (1972) to help librarians understand how much a **journal** was being used (useful in renewal process)
- It is an **average** measure of usage across **an entire journal**
- It contains **no information** on the impact of an **individual paper**

- For a journal J_i in a year n

$$\mathrm{IF}(J_i, n) = \frac{\#\{\text{ citations to all items published in } J_i \text{ in } n-1 \text{ and } n-2 \}}{\#\{\text{articles and letter published in } J_i \text{ in } n-1 \text{ and } n-2 \}}$$

- **Pros**: simple, easy to compute, known and disseminated

Impact Factor and its criticisms - II

- **Cons/criticisms**:

1. Only 2 years of data to account for citations may not be enough in some areas to reach the citation peak ⇒ **IF varies very significantly among (sub)areas**

Ex: In SC Eng. E&E, $\boldsymbol{E}[IF_{2011}]$= 1.32; $\max[IF_{2011}]$=7

In SC Biology, $\boldsymbol{E}[IF_{2011}]$= 2.10; $\max[IF_{2011}]$=11.45

In SC Bioch and Molec. Bio $\boldsymbol{E}[IF_{2011}]$= 3.78; $\max[IF_{2011}]$=34.31

2. Citations are counted in the same way **independently of the source** (i.e. a citation obtained from *Science* is the same as the "*Journal of Obscurity*")

3. IF has an "non-consistent" definition: elements considered at the numerator are different than the denominator

4. IF is **liable to active manipulation**

Impact Factor: manipulation (1/3)

- How has IF been manipulated?

1. **Inconsistent definition**: citations to notes/"letters to the editor"/editorials count in the numerator but the same items are not counted in the denominator. They can be cited and, even more importantly, their citations count normally.


Neth Heart J (2012) 20:481–482
DOI 10.1007/s12471-012-0336-0
Netherlands Heart Journal

EDITOR'S COMMENT

The NHJ 2012 in retrospect: which articles are cited most?

Its bibliography contains 25 citations to the same journal, 24 of which count toward the 2012 IF

Impact Factor: manipulation (2/3)

2. **Coerce self-citations**: EiCs "force" authors to add citations to their journal (not necessarily to the authors) to increase IF

Coercive Citation in Academic Publishing

3 FEBRUARY 2012 VOL 335 SCIENCE www.sciencemag.org
Published by AAAS

Allen W. Wilhite*† and Eric A. Fong*

- EICs of 175/832 journals in the area of economics, sociology, psychology, and multiple business disciplines were found to "coerce" self-cites
- Coercing was more frequent with young authors than experienced ones
- Relation to area: if one journal coerces its authors other journals will most likely follow

Impact Factor: manipulation (3/3)

3. **Citation Cartel/Stacking**: EiCs or other members of editorial board of J_A and J_B:
   - publish in J_A a paper with (several) tens of citation to J_B
   - publish in another journal as authors to do the same

Brazilian citation scheme outed

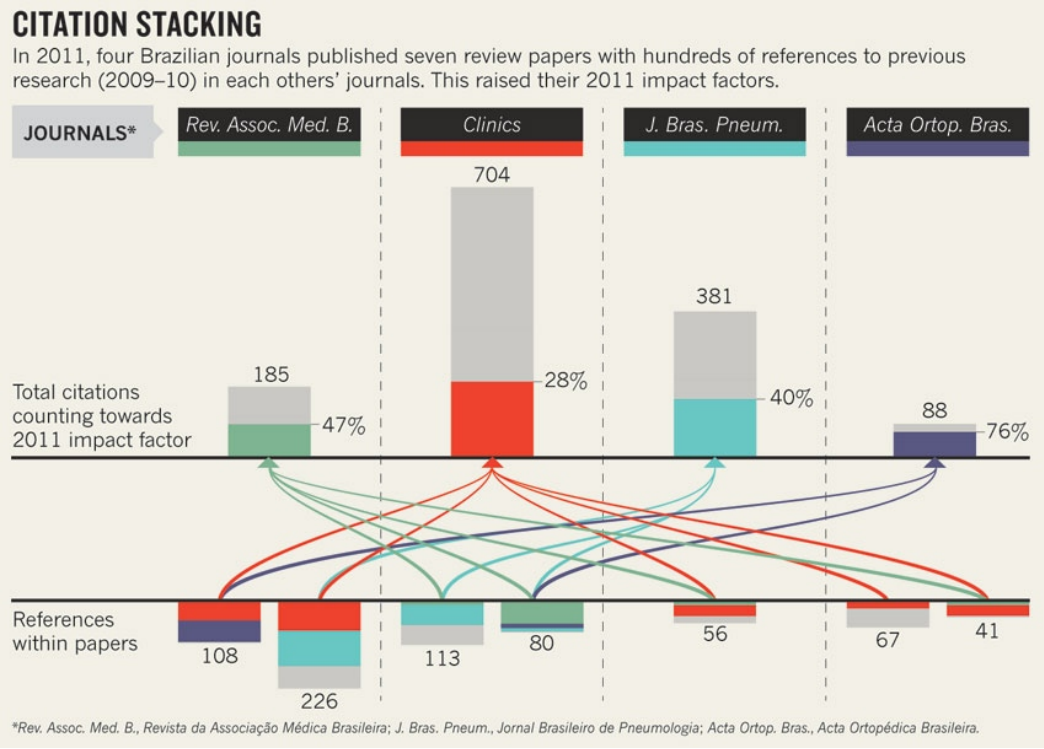
Thomson Reuters suspends journals from its rankings for 'citation stacking'.

Richard Van Noorden

nature

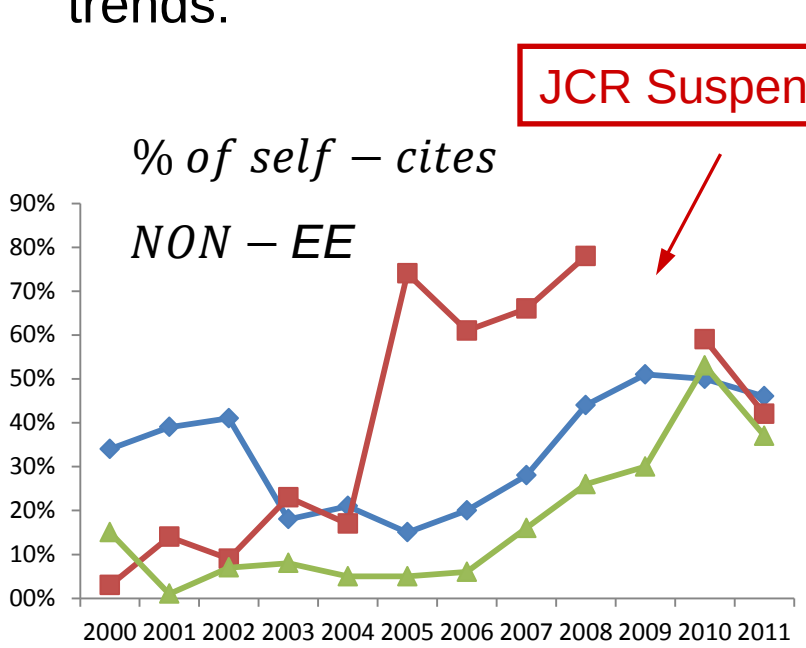
27 August 2013

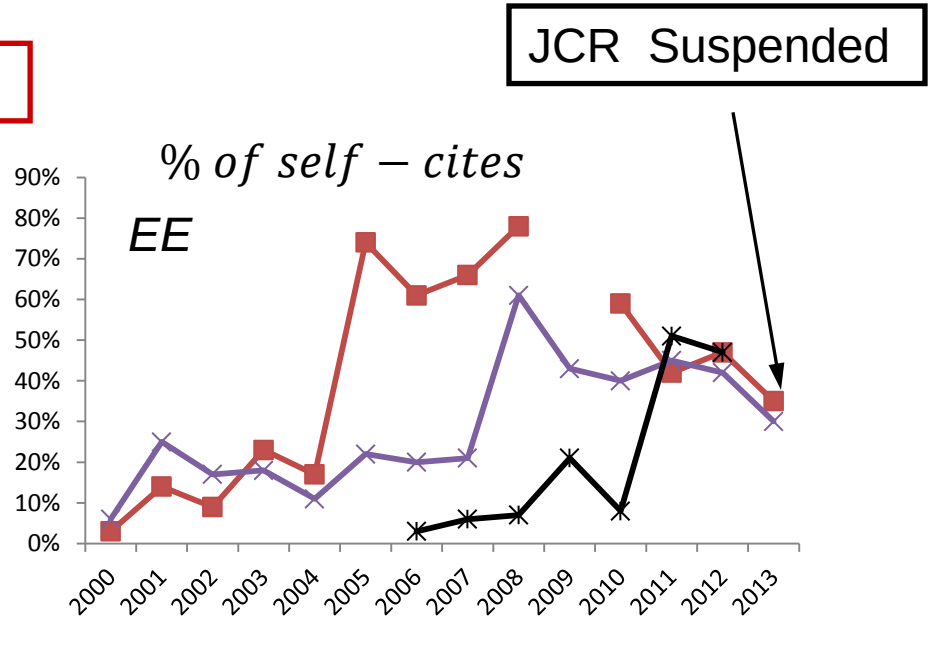
- Four Brazilian journals (Rev Assoc. Medic B, Clinics, J. Bras. Pneum, Acta Ortop Bras.) were found to establish a citation cartel
- Three Italian journals in the area of medicine (with the same EiC!)



Is the phenomenon widespread?

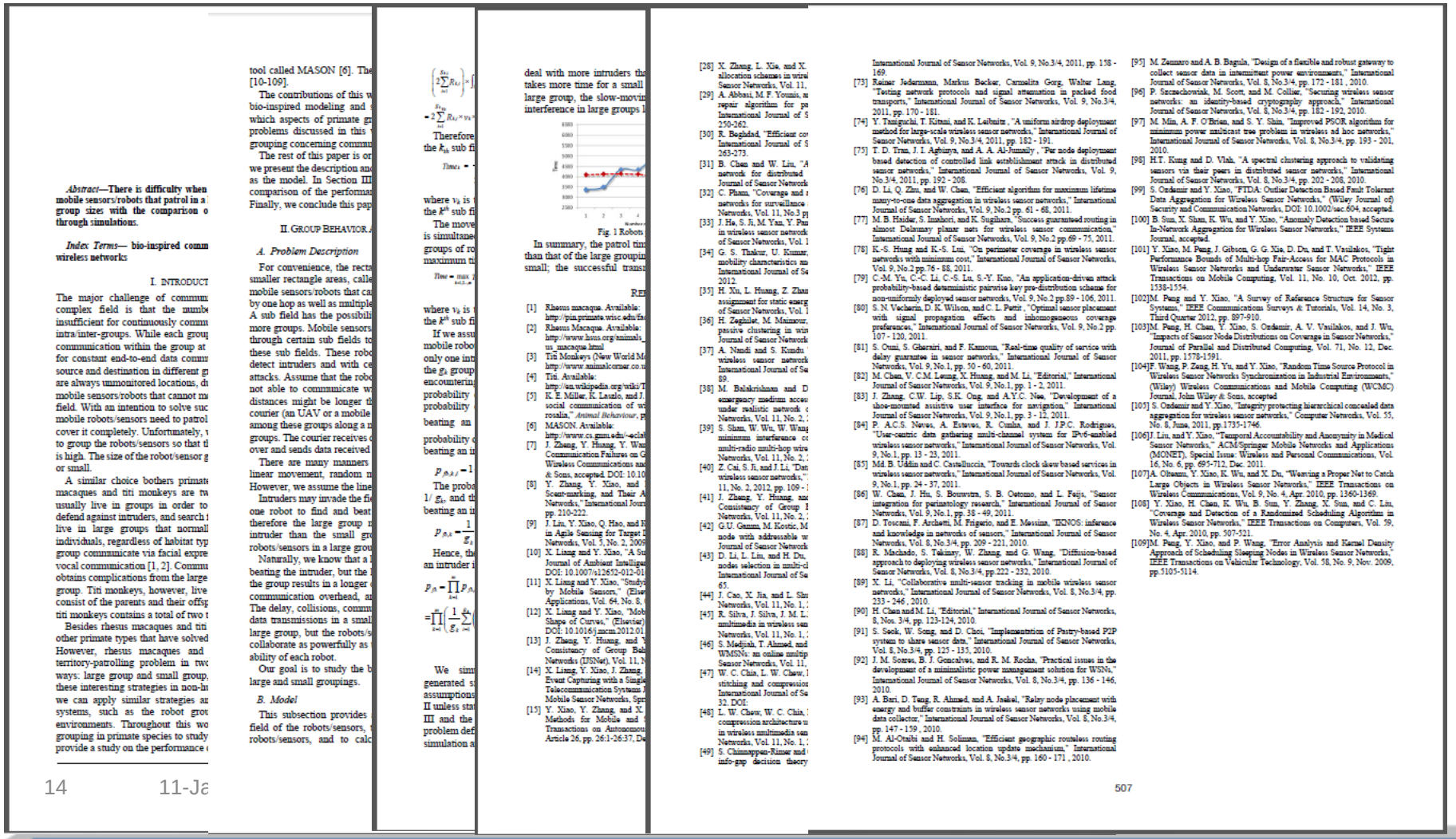
- No systematic study yet: one must use JCR data: For **citation cartels** the systematic analysis is very difficult, but one can rely on **self-citation** trends:





- Laser and Particles Beams (Phy Applied), Cortex (Neuroscience), Int. Journal of Hydrogen Energy (Energy and Fuels) show an increasing self-citation trend (and similar examples exist in many more areas)
- **Our Area**: Int J. Circuit Theory and Applications and Asian Journal of Control shows that we are not immune.

What is wrong with this conference paper?



What is wrong with this conference paper? - II

- The authors published 2 conference papers with 100+109 items in the reference list.
- There are 74+82 citations to the International Journal of Sensor Networks (IJSN)
- One of the 2 authors is the EiC of the IJSN
- IJSN was not included by Thomson in the 2013 Journal Citation Report since the above citations account for 82% of the total citations to IJSN.
- The addition of the citation was done **after** the review process was completed

What is wrong with this conference paper? - II

- The authors published 2 conference papers with 100+109 items in the reference list.
- There are 74+82 citations to the International Journal of Sensor Networks (IJSN)
- One of the 2 a
- IJSN was not nal Citation Report since t he total citations to IJ
- The addition c process was completed

Why this is happening?

- The IF was historically created **to give librarians tools for deciding renewals**, yet…
- It is currently **more and more used as the gold standard to evaluate the impact of an individual's research activity** (for hiring, tenure, promotion, salary increase…).
  - As an example, the Chinese government pays scientists for publication in high IF journals (see http://scholarlykitchen.sspnet.org/2011/04/07/paying-for-impact-does-the-chinese-model-make-sense/)

IF range	(0,1)	[1,3)	[3,5)	[5,10)	>10	Nature/Science
Increase in salary	$306	$458	$611	$764	$2139	$30562

Why this is happening?

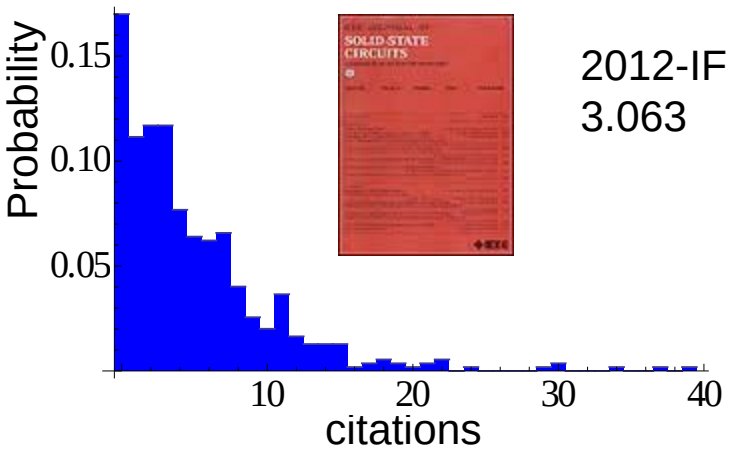
- The IF was historically created **to give librarians tools for deciding renewals**, yet…
- It is currently **more and more used as the gold standard to evaluate the impact of an individual's research activity** (for hiring, tenure, promotion, salary increase…)
- This use is commonly based on 2 main "assumptions". Assume that J_A has $IF_A \gg IF_B$ of J_B, then
  1. Any paper published in J_A has more impact (has received more citations) than any paper published in J_B
  2. The review process of J_A is more stringent than the one of J_B

Are these assumptions supported by data?

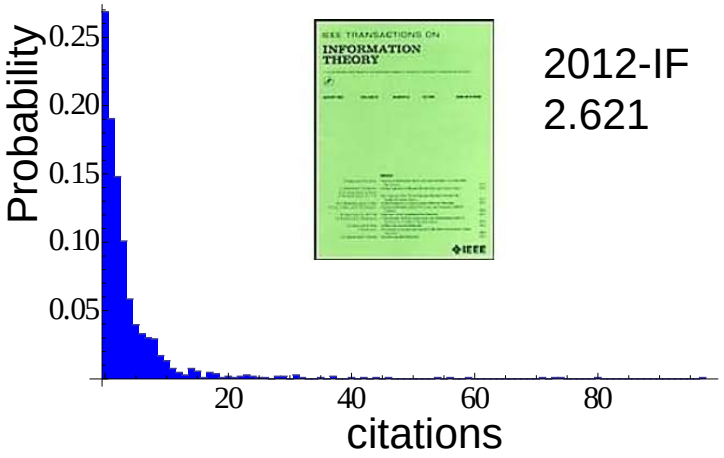
NO

Some data - I

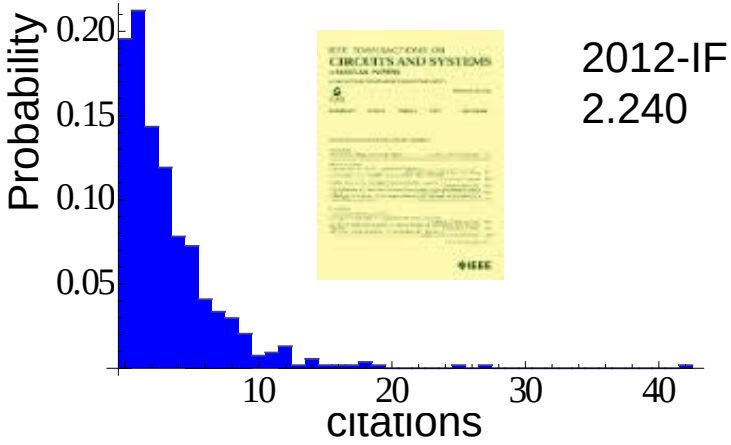
1. Evaluation of the impact of a **single** paper in a journal



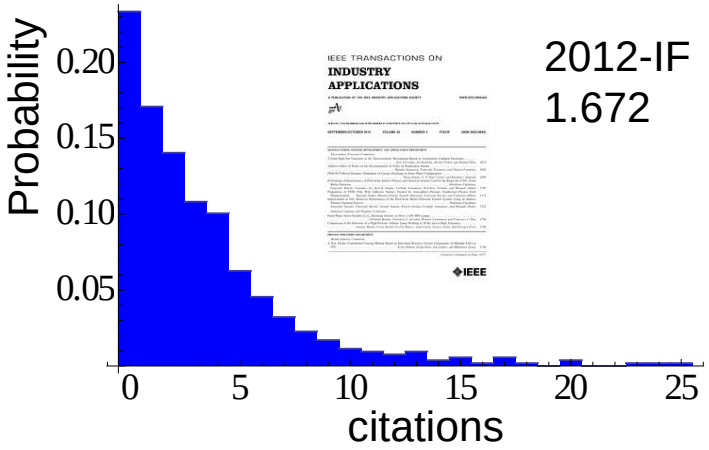
2012-IF 3.063



2012-IF 2.621



2012-IF 2.240



2012-IF 1.672

- JSSC, TIT, TCAS-I, and TIA distributions of citations for 2012 to papers of 2011 and 2010 show the same shape: **most papers are cited only a few times or never cited and only very few have high impact**

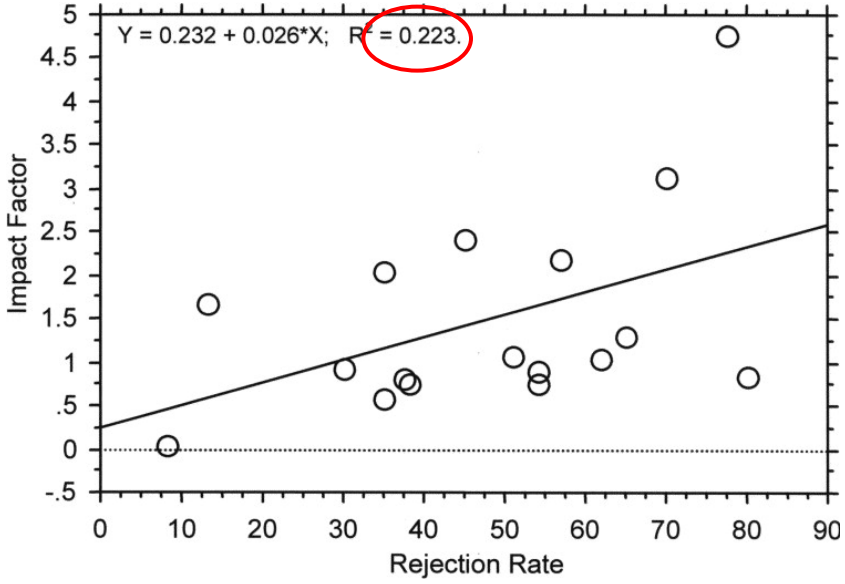
Some data - II

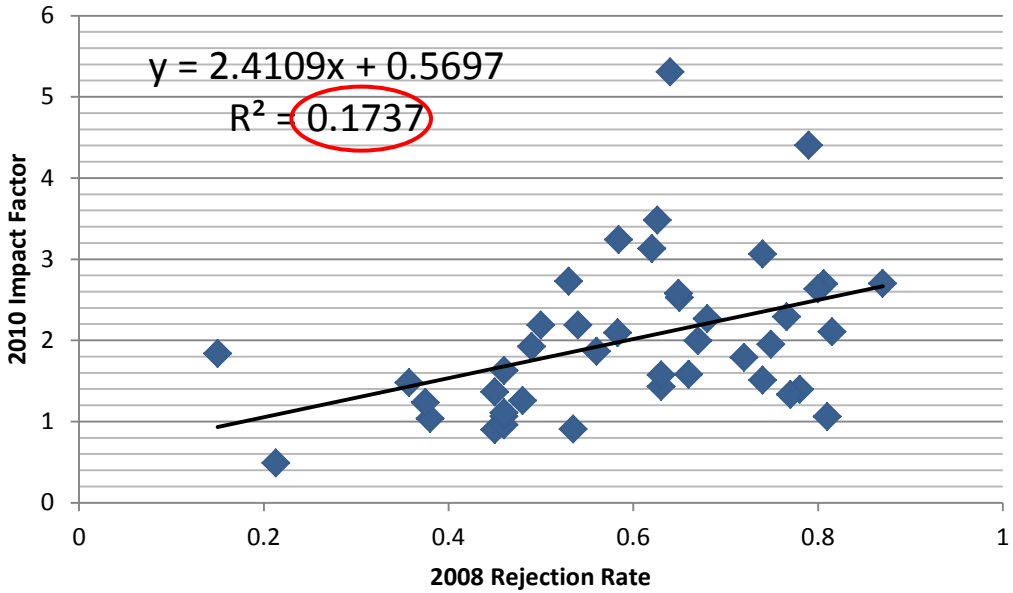
- **<u>Important</u>: regardless of IF, most papers in each journal are cited only a few times (if ever) and few papers are cited many times**
- Assuming that a randomly chosen paper in JSSC (IF=3.063) is better (has more citations) than one of TCAS-I (IF=2.240) **is wrong >36% of the time**
- Assuming that a randomly chosen paper in TIT (IF=2.612) is better than one of TIA (IF=1.672) **is wrong >43% of the time**

journal indicators are average quantities and give therefore **<u>no indication</u>** of the quality of any single published paper

Some data - III

- Indication of the **selectivity** of a journal: if the IF of a journal is large, is the review process "very strict"?
- This is **not supported by data** (at least if one assumes valid the equation "strict review process = high rejection rate"): the correlation coefficient **is on the order of 0.2**



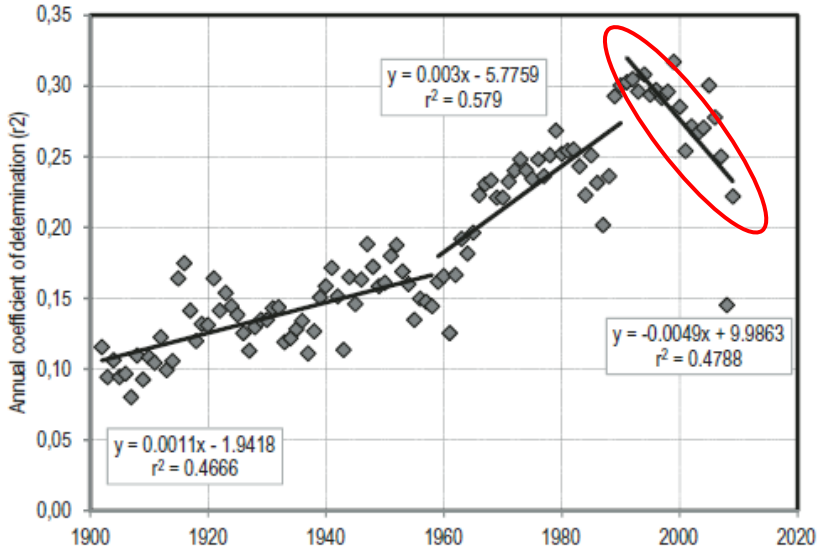


A. Kurmin, T. Krimis, "Exploring the Relationship Between Impact Factor and Manuscript Rejection Rates in Radiologic Journals, Acad Radiol 2006; 13:77–83

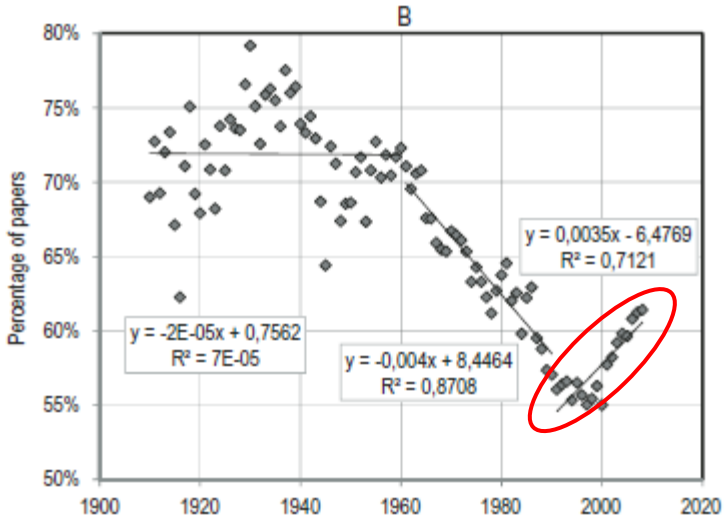
43 IEEE titles, Rejection Rate obtained by internal reports

Some data IV

- **Assumption**: the IF of a journal is large, papers published there are highly cited, if I publish there my paper has an higher probability to be highly cited
- This is **not supported by data** (neither in terms of correlation nor of probability) [G. A. Lozano *et al*., "The Weakening Relationship Between the Impact Factor and Papers' Citations in the Digital Age", J. American Society for Information Science and Technology, 63(11):2140–2145, 2012]



"Correlation coefficient" between IF in year of publication and citation rate in the following 2 years



Percentage of papers which are in the top 5% of the distribution citation in a given year which were NOT published in a journal in the top 5% of the IF ranking

Why this is happening?

- While the IF was historically created to help **librarians**, it is **misused** to evaluate **individual's research activity** (for hiring, tenure, promotion…)

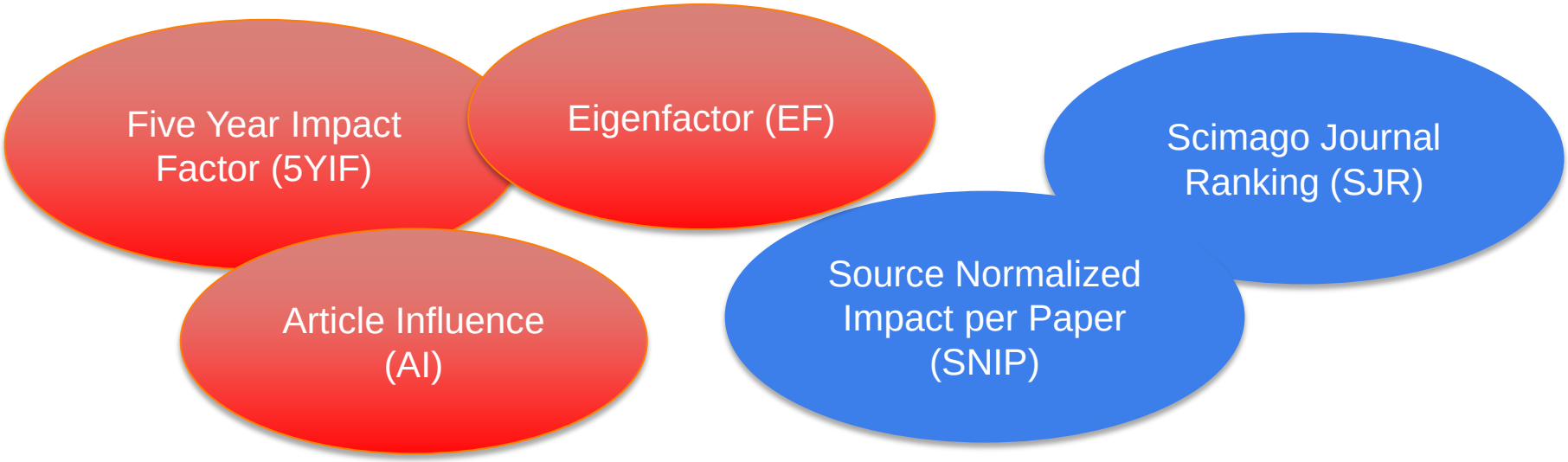
The unintended use of the IF **made it the target and not the measure** and created incentive for its manipulation

According to the 2013 Nature article of Richard Van Noorden the EiCs of the 4 journals involved in a citation cartel created it because

"In Brazil, an agency in the education ministry, called CAPES, evaluates graduate programmes in part by the impact factors of the journals in which students publish research"

Other measures to solve IF issues for Journal evaluation

Several "successful" new indicators: 5 in either WoS or Scopus



- ***Increase the citation window*** : 3 or 5 years
- ***Introduce subject field normalization***: explicit (SNIP) or implicit (EF, AI, SJR)
- ***Exclude all (or most) self-cites***: eliminate the inflation issue (EF, AI, SJR)
- ***Only count "equivalent scientific" documents both at numerator and denominator***: eliminate another cause of inflation (EF, AI, SJR, SNIP)

Popularity vs Prestige

- An important distinction is between indicators measuring *popularity* or *prestige*

1. *Popularity indicators*: are based on an algebraic formula and count citations directly **independently of their source** (IF, 5YIF, SNIP)
2. *Prestige indicators*: are based on an recursive formula and weight the influence of citations **depending on their source** (EF, AI, SJR)

They evaluate **different aspects of Journal Impact**

At the very minimum, one needs to use **both** popularity (ex. IF, 5YIF) and prestige (ex. AI, SJR) indicators

Addressing the issues: the rest of the landscape

In approving the statement IEEE joins several other research agencies and professional organizations in the area of Physics, Medical Sciences, Biology, ….

European Physical Society
www.eps.org

VETENSKAPSRÅDET
THE SWEDISH RESEARCH COUNCIL

Australian Government
National Health and Medical Research Council
N H M R C

INSTITUT DE FRANCE
Académie des sciences

Council of Canadian Academies
Conseil des académies canadiennes

San Francisco
DORA
Declaration on Research Assessment

Some Don'ts (1/3)

1. Journal Bibliometrics indicators have been designed to **evaluate journal impact** but **cannot be employed** as a single measure of the **quality of single papers** or to evaluate the **quality of a scientists.**
   - This is particularly problematic for the IF but applies **to all journal indicators**

     Examples:

     a. Do not rank faculty candidates using the IF of the journal they publish in

2. The application of **aggregation or filter operations** to Journal or Individual Bibliometric indicators makes their use to rank scientists even a **worse abuse**
   - *Examples*:
     a. Do not use the *sum of publication IFs* or use the *average of publication IFs* to rank candidates
     b. Do not apply a threshold to IF to make a particular publication count for raises (say first quarter in a specific subject category of JCR)

Some Do's

1. Journal Bibliometric indicators exist, each aiming to measure the journal **scientific impact** and **they measure it in different ways**
   - One **cannot use a single indicator** (neither IF, nor any other) to measure journal impact. At the very least, one needs to use
     a. One popularity indicator (e.g. the IF, or the 5YIF)
     b. One prestige indicator (e.g. the AI)
   - Use of multiple indicators provides a much more accurate evaluation of a journal's impact and can also **make evident existing anomalies**
2. Individual Bibliometric indicators are statistical quantities and if the faculties/candidates have a sufficiently large publication output, citation analysis **can be used (with caution) as an additional source of information for evaluation**
   - *Examples*:
     a. Different career progression dynamics may (will) exist
     b. Benchmarking is fundamental especially for multidisciplinary research
     c. Read the contribution and apply your own judgment!!

For more info

DOI: 10.1109/ACCESS.2013.2261115

IEEE Access
practical innovations : open solutions

Received April 5, 2013, accepted April 27, 2013, published May 10, 2013.

Digital Object Identifier 10.1109/ACCESS.2013.2261115

Bibliometric Indicators: Why Do We Need More Than One?

GIANLUCA SETTI, (Fellow, IEEE)
Department of Engineering (ENDIF), University of Ferrara, Ferrara 44100, Italy, and also with the Advanced Research Center on Electronic Systems, University of Bologna, Bologna 40125, Italy (gianluca.setti@unife.it)

ABSTRACT This paper provides an overview of the main features of several bibliometric indicators which were proposed in the last few decades. Their pros and cons are highlighted and compared with the features of the well-known impact factor (IF) to show how alternative metrics are specifically designed to address the flaws that the IF was shown to have, especially in the last few years. We also report the results of recent studies in the bibliometric literature showing how the scientific impact of journals as evaluated by bibliometrics is a very complicated matter and it is completely unrealistic to try to capture it by any single indicator, such as the IF or any other. As such, we conclude that the adoption of more metrics, with complementary features, to assess journal quality would be very beneficial as it would both offer a more comprehensive and balanced view of each journal in the space of scholarly publications, as well as eliminate the pressure on individuals and their incentive to do metric manipulation which is an unintended result of the current (mis)use of the IF as the gold standard for publication quality.

Some Information to EiCs, PC and PRAC members about current bibliometric measures

Gianluca Setti, Stephen Yurkovich, Jacek Zurada

July 9, 2012

EigenFactor and Article Influence in a nutshell

As it will be clarified in the following, Impact Factor (IF), EigenFactor (EF), and Article Influence (AI) extract information from the collection of citations of a given set of journals, each in a different way, with the aim of measuring the quality and influence of each publication. More specifically:

- AI and IF are a measure of quality *per-article*, whereas EF measures the quality of the *overall journal*. As such, the latter tends to be larger for journals publishing many articles per year.
- EF and AI *weight citations in a different way depending on the reputation of the source*; on the contrary the IF considers a citation coming from the Proceedings of the IEEE and one coming from a low quality journal at the same level.

Email: gianluca.setti@unife.it