
STRIKE
DON'T
BE A
SCAB!

How to use bibliometric indices? (if you really must)

Denis Bouyssou
CNRS

Paris 2022

Outline

Outline

Academia

Globalization

- knowledge economy
- financial and economic crisis

Academia

Globalization

- knowledge economy
- financial and economic crisis

Globalization and academia

- budget cuts
- arrival of new players (China, India)
- increased mobility of staff & students
- industrialization of academia

Industrialization of academia

Symptoms

- evaluation & funding agencies
- students' debt crisis
- fraud & plagiarism
- proliferation of indices & rankings: "evaluation fever" (Y. Gingras)
  - bibliometric indices everywhere

Bibliometrics

Two extreme positions

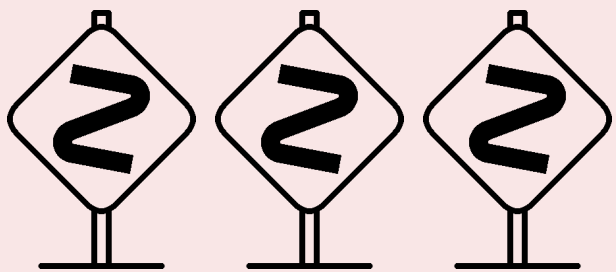
- bibliometrics is an absolute evil
- bibliometrics brings objectivity and fairness

Bibliometrics

Two extreme positions

- bibliometrics is an absolute evil
- bibliometrics brings objectivity and fairness

Both positions are plainly wrong!



Bibliometrics

Bibliometrics defined

- using mathematical and statistical techniques to study communication patterns

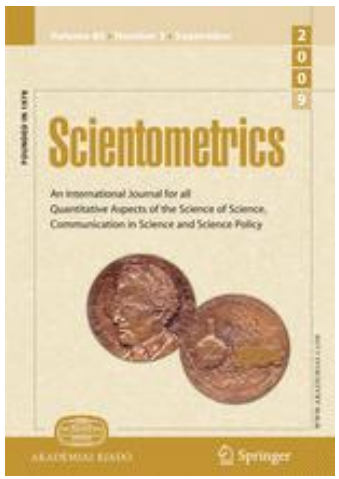
Bibliometrics

Bibliometrics defined

- using mathematical and statistical techniques to study communication patterns

The field of Bibliometrics

- active scientific field
  - journals: *Scientometrics*, *Journal of Informetrics*, *Journal of the Association for Information Science and Technology*
  - ISSI: International Society for Scientometrics and Informetrics

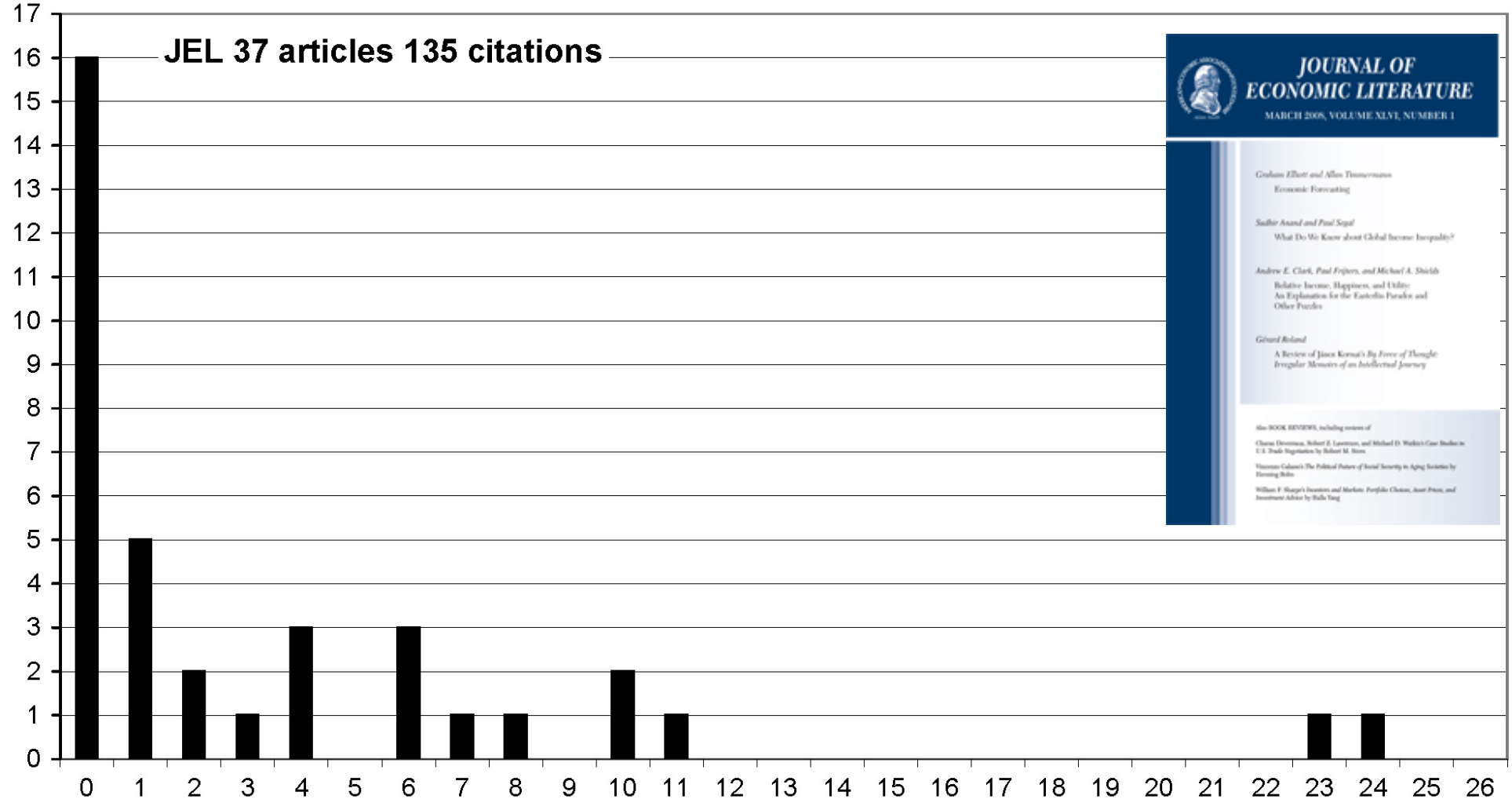


Bibliometrics

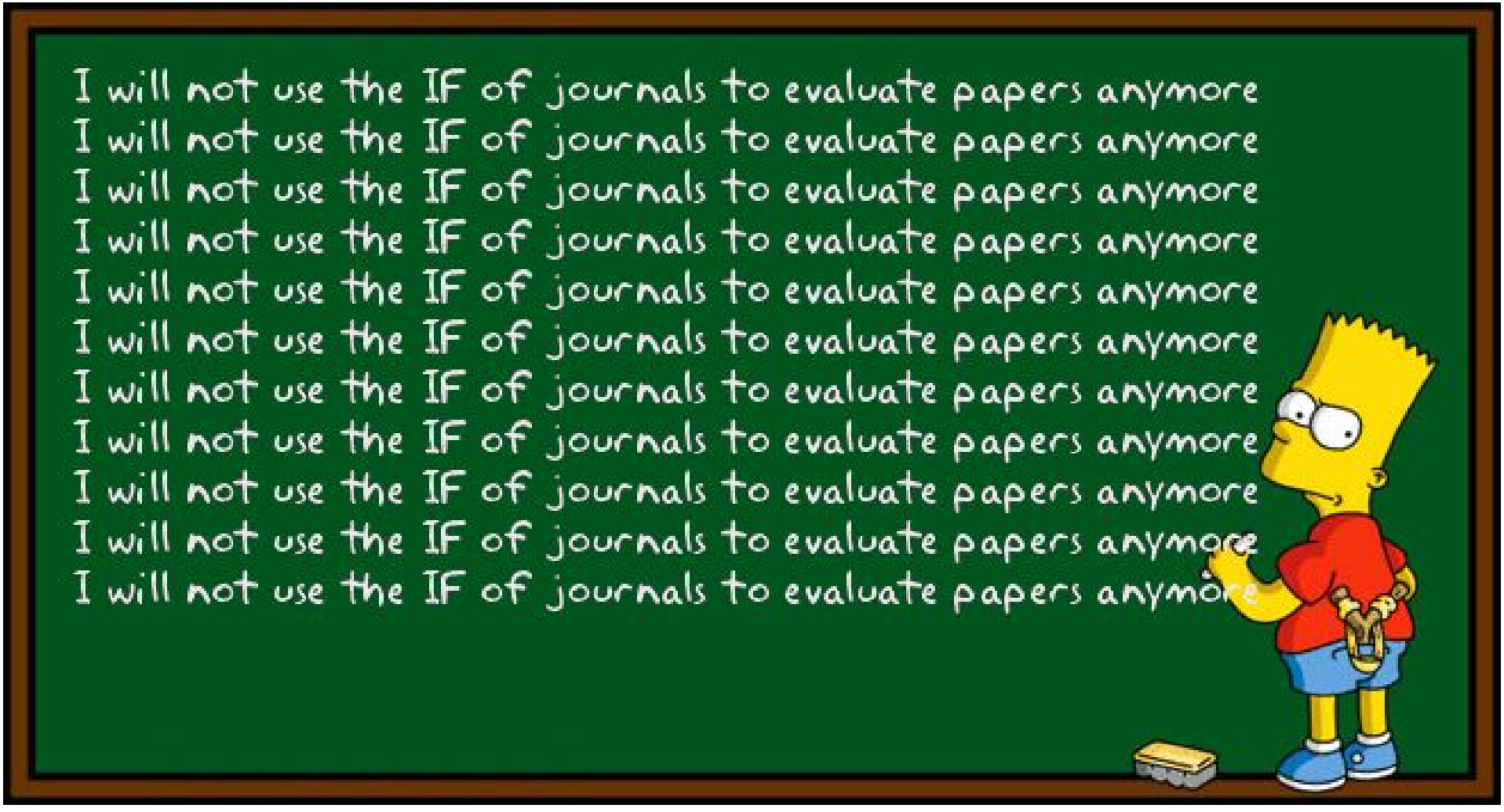
Some research questions

- bibliometric laws: Lotka, Bradford
- social network of {scientists, papers, fields}
- efficiency of research expenses
- optimal size of an academic institution
- factors influencing transfer of knowledge towards industry
- which journals should libraries subscribe to?
- impact of open access on diffusion on knowledge
- strong and weak research fields of a country
- emerging fields

Journal of Economic Literature 2008 IF (3.65 in 2008 / 5.410 in 2018)
(Using WoS, number of citations given by papers published in 2008 to papers published by JEL in 2006–2007 divided by the number of papers published by JEL in 2006–2007)



Bart knows!


I will not use the IF of journals to evaluate papers anymore
I will not use the IF of journals to evaluate papers anymore
I will not use the IF of journals to evaluate papers anymore
I will not use the IF of journals to evaluate papers anymore
I will not use the IF of journals to evaluate papers anymore
I will not use the IF of journals to evaluate papers anymore
I will not use the IF of journals to evaluate papers anymore
I will not use the IF of journals to evaluate papers anymore
I will not use the IF of journals to evaluate papers anymore
I will not use the IF of journals to evaluate papers anymore
I will not use the IF of journals to evaluate papers anymore

Evaluative bibliometrics and bibliometric indices

Evaluative bibliometrics

- publications in journals are the central research output
- citations to publications are important signs of recognition

"bibliometrically limited view of a complex reality" (van Raan, 2005)

Evaluative bibliometrics and bibliometric indices

Evaluative bibliometrics

- publications in journals are the central research output
- citations to publications are important signs of recognition

"bibliometrically limited view of a complex reality" (van Raan, 2005)

- count publications & citations
- summarize these counts by indices

Evaluative bibliometrics and bibliometric indices

Databases

- Web of Science (Clarivate aka Thomson Reuters aka ISI)
- Scopus (Elsevier)
- Google Scholar (Google or PoP)

Quality of data

Denis BOUYSSOU

- plain ASCII
- no LaTeX ligature
- no diacritical signs
- only one word
- no known scientific homonyms

Meltem Öztürk-Escoffier, Zhāng Wěi, Włodzimierz Łukaszewski, Kim Seo-yoon

Quality of data

Denis BOUYSSOU

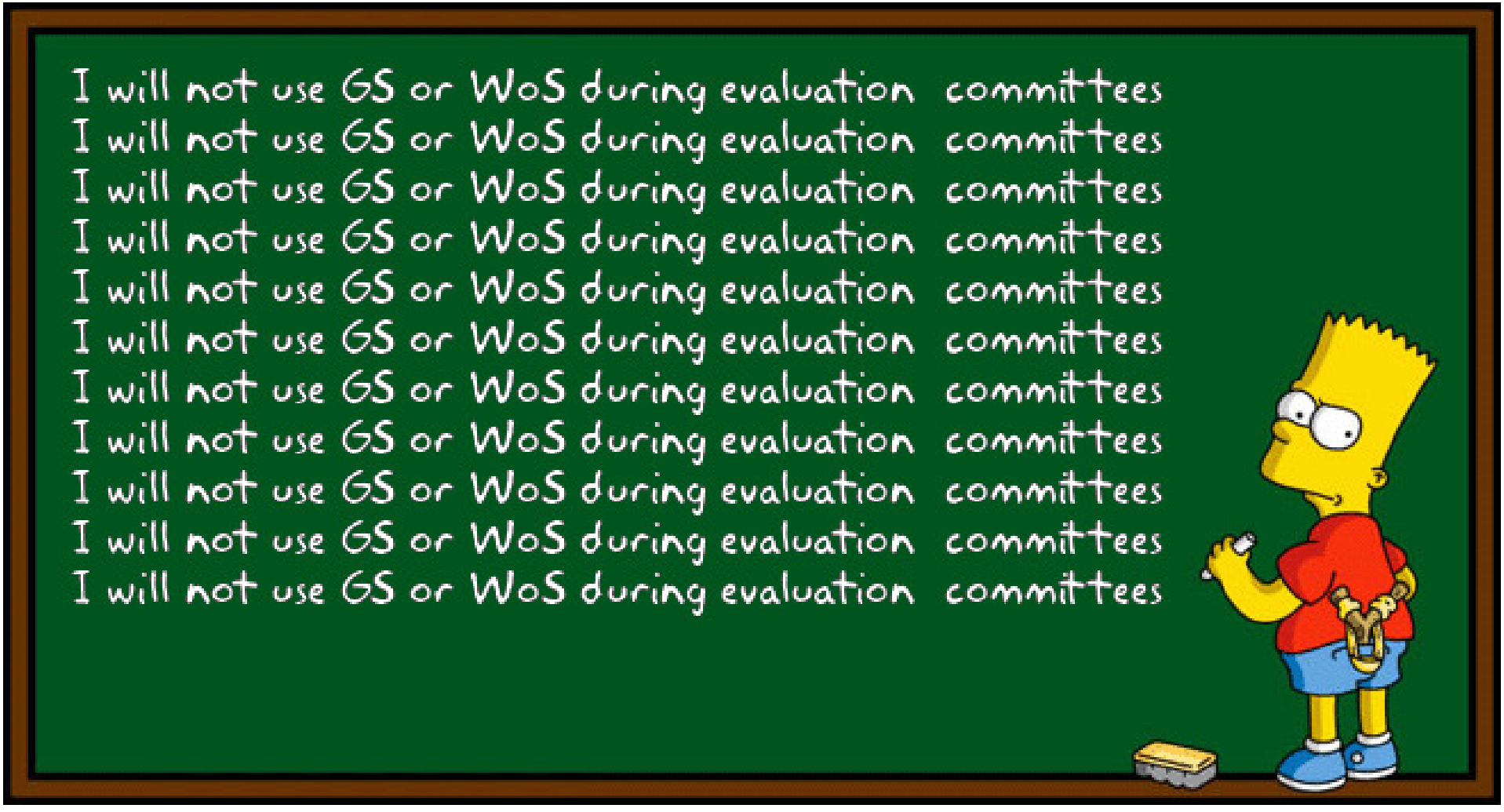
- plain ASCII
- no LaTeX ligature
- no diacritical signs
- only one word
- no known scientific homonyms

Meltem Öztürk-Escoffier, Zhāng Wěi, Włodzimierz Łukaszewski, Kim Seo-yoon

Denis BOUYSSOU (checked: 5 September 2022)

GS	280 papers, 8870 citations, h-index 41
Scopus	83 papers, 1667 citations, h-index 22
WoS	77 papers, 875 citations, h-index 19

Bart knows!

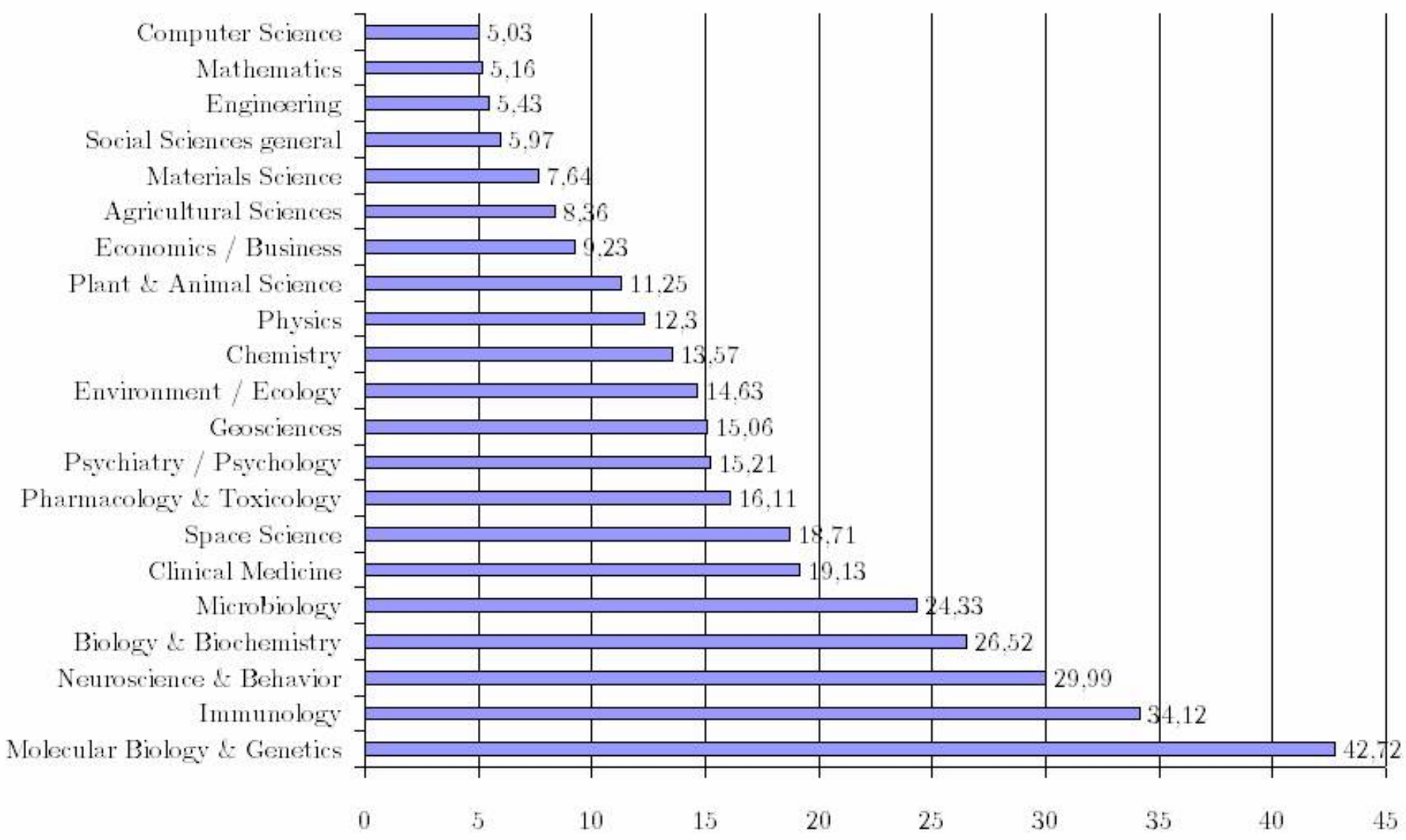

I will not use GS or WoS during evaluation committees
I will not use GS or WoS during evaluation committees
I will not use GS or WoS during evaluation committees
I will not use GS or WoS during evaluation committees
I will not use GS or WoS during evaluation committees
I will not use GS or WoS during evaluation committees
I will not use GS or WoS during evaluation committees
I will not use GS or WoS during evaluation committees
I will not use GS or WoS during evaluation committees
I will not use GS or WoS during evaluation committees
I will not use GS or WoS during evaluation committees

A few words of warning

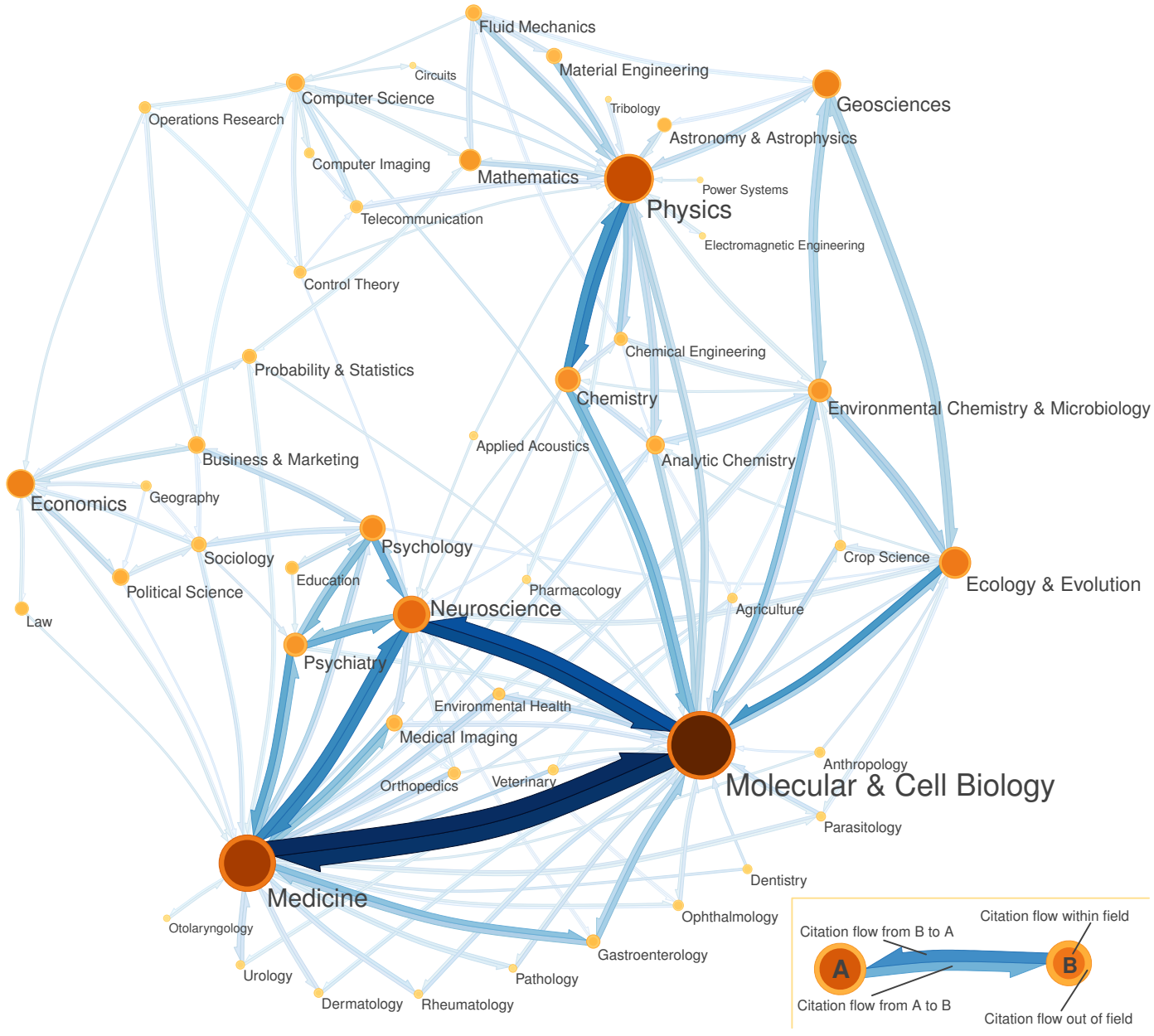
Databases

- cleansing is needed and not easy to do!
  - names: diacritical signs, TEX ligatures, transliteration, homonyms
  - correct affiliations are extremely difficult to determine
  - counting: original articles, letters, notes, erratum, editorials
  - spelling errors + incorrect citations
  - lost citations (up to 30%)
- important differences between fields
  - publication intensity
  - citation intensity & behavior
  - longevity of papers (months vs decades)

Citation intensity for the 21 WoS categories (2000)



Map of scientific fields (PNAS, 2008)



Bibliometric nightmares

- how to deal with multiple authors (sometimes more than 1 000)
- how to deal with multiple affiliations
- how to compare people having different career length
- people react and adapt quickly: perverse effects are pervasive
- how to understand the meaning of a citation (papers on Hydroxychloroquine cure)

Examples of papers with many authors (2011)

Papers with highest numbers of authors, by year, 2002-2011

Year	Paper	Number of authors
2011	ATLAS Collaboration (G. Aad, *et al.*), "Search for quark contact interactions in dijet angular distributions in pp collisions at root s=7 TeV measured with the ATLAS detector," *Phys. Lett. B*, 694(4-5): 327-45, 2011.	3,179
2010	ATLAS Collaboration (G. Aad, *et al.*), "Charged-particle multiplicities in pp interactions at root s=900 GeV measured with the ATLAS detector at the LHC ATLAS Collaboration," *Phys. Lett. B*, 688(1): 21-42, 2010.	3,221
2009	LIGO Sci. Collaboration, Virgo Collaboration (B.P Abbott, *et al.*), "An upper limit on the stochastic gravitational-wave background of cosmological origin," *Nature*, 460(7258): 990-4, 2009.	657
2008	CMS Collaboration (S. Chatrchyan, *et al.*), "The CMS experiment at the CERN LHC," *J. Instrumentation*, 3: No. S08004, 2008.	3,101
2007	CMS Collaboration (G.L. Bayatian, *et al.*), "CMS physic technical design report, volume II: Physics performance," *J. Phys. G.-Nucl. Part. Phys.*	2,011
2006	ALEPH, DELPHI, L3, OPAL, and SLD Collaborations (S. Schael, *et al.*), "Precision electroweak measurements on the Z resonance," *Phys. Reports*, 427(5-6): 257-454, 2006.	2,517
2005	Antiretroviral Therapy Cohort Collaboration (D. Costagliola, *et al.*), "Incidence of tuberculosis among HIV-infected patients receiving highly active antiretroviral therapy in Europe and North America," *Clin. Infect. Diseases*, 41(12): 1772-82, 2005.	859
2004	MEGA Study Group (H. Nakamura, *et al.*), "Design and baseline characteristics of a study of primary prevention of coronary events with pravastatin among Japanese with mildly elevated cholesterol levels," *Circulation J.*, 68(9): 860-7, 2004.	2,459
2003	D. Acosta, *et al.* (CDF II Collaboration), "Measurement of the mass difference M(D(s)(+))-m(D(+)) at CDF II," *Phys. Rev. D*, 68(7): No 072004, 2003.	818
2002	B. Aubert, *et al.* (BABAR Collaboration), "The BABAR detector," *Nucl. Instr. Meth. Phys. Res. Sect. A*, 479(1): 1-116, 2002.	824

Bibliometric indices

Hypotheses

- all above problems have been taken care of
- you have a good, verified, and cleaned database
- otherwise, do not use evaluative bibliometrics!

Bibliometric indices

Hypotheses

- all above problems have been taken care of
- you have a good, verified, and cleaned database
- otherwise, do not use evaluative bibliometrics!

Many possible indices

- counting of papers
- counting of citations
- sum of Impact Factors
- Markovian indices (e.g., PageRank-like)
- h-index

Bibliometric indices

Hypotheses

- all above problems have been taken care of
- you have a good, verified, and cleaned database
- otherwise, do not use evaluative bibliometrics!

Many possible indices

- counting of papers
- counting of citations
- sum of Impact Factors
- Markovian indices (e.g., PageRank-like)
- h-index

Bibliometric Indices

- what properties?
- how to compare (combine, use) them?

Potential problems with the h-index (1/2)

h-index, J. Hirsch, PNAS, 2005 (6 199 citations on WoS, Sept. 2022)

- the h-index of an author is x if this author has x papers having at least x citations each (and her other papers have at most x citations each)

Potential problems with the h-index (1/2)

h-index, J. Hirsch, PNAS, 2005 (6 199 citations on WoS, Sept. 2022)

- the h-index of an author is x if this author has x papers having at least x citations each (and her other papers have at most x citations each)

- author f: 4 papers with 4 citations each $(4 \cdot \mathbf{1}_4)$
- author g: 3 papers with 6 citations each $(3 \cdot \mathbf{1}_6)$
- $i_h(f) = 4 > i_h(g) = 3$

Potential problems with the h-index (1/2)

h-index, J. Hirsch, PNAS, 2005 (6 199 citations on WoS, Sept. 2022)

- the h-index of an author is x if this author has x papers having at least x citations each (and her other papers have at most x citations each)

- author f: 4 papers with 4 citations each $(4 \cdot \mathbf{1}_4)$
- author g: 3 papers with 6 citations each $(3 \cdot \mathbf{1}_6)$
- $i_h(f) = 4 > i_h(g) = 3$

- both authors publish a new paper with 6 citations $(\mathbf{1}_6)$
- $i_h(f^*) = 4 = i_h(g^*) = 4 \qquad (f^* = f + \mathbf{1}_6 \quad g^* = g + \mathbf{1}_6)$

Potential problems with the h-index (1/2)

h-index, J. Hirsch, PNAS, 2005 (6 199 citations on WoS, Sept. 2022)

- the h-index of an author is x if this author has x papers having at least x citations each (and her other papers have at most x citations each)

- author f: 4 papers with 4 citations each $(4 \cdot \mathbf{1}_4)$
- author g: 3 papers with 6 citations each $(3 \cdot \mathbf{1}_6)$
- $i_h(f) = 4 > i_h(g) = 3$

- both authors publish a new paper with 6 citations $(\mathbf{1}_6)$
- $i_h(f^*) = 4 = i_h(g^*) = 4 \qquad (f^* = f + \mathbf{1}_6 \quad g^* = g + \mathbf{1}_6)$

- both authors publish a new paper with 6 citations $(\mathbf{1}_6)$
- $i_h(f^{**}) = 4 < i_h(g^{**}) = 5 \qquad (f^{**} = f^* + \mathbf{1}_6 \quad g^{**} = g^* + \mathbf{1}_6)$

Potential problems with the h-index (1/2)

h-index, J. Hirsch, PNAS, 2005 (6 199 citations on WoS, Sept. 2022)

- the h-index of an author is x if this author has x papers having at least x citations each (and her other papers have at most x citations each)

- author f: 4 papers with 4 citations each $(4 \cdot \mathbf{1}_4)$
- author g: 3 papers with 6 citations each $(3 \cdot \mathbf{1}_6)$
- $i_h(f) = 4 > i_h(g) = 3$

- both authors publish a new paper with 6 citations $(\mathbf{1}_6)$
- $i_h(f^*) = 4 = i_h(g^*) = 4 \qquad (f^* = f + \mathbf{1}_6 \quad g^* = g + \mathbf{1}_6)$

- both authors publish a new paper with 6 citations $(\mathbf{1}_6)$
- $i_h(f^{**}) = 4 < i_h(g^{**}) = 5 \qquad (f^{**} = f^* + \mathbf{1}_6 \quad g^{**} = g^* + \mathbf{1}_6)$

Independence is violated

Potential problems with the h-index (2/2)

Evaluation of authors and departments

- the h-index of a department is x if this department has x papers having at least x citations each (and its other papers have at most x citations each)

Potential problems with the h-index (2/2)

Evaluation of authors and departments

- the h-index of a department is x if this department has x papers having at least x citations each (and its other papers have at most x citations each)

Department $F = (f_1, f_2)$

- author $f_1 = 4 \cdot \mathbf{1}_4$
- author $f_2 = 4 \cdot \mathbf{1}_4$
  - h-index of both authors is 4
  - h-index of the department is 4

Potential problems with the h-index (2/2)

Evaluation of authors and departments

- the h-index of a department is x if this department has x papers having at least x citations each (and its other papers have at most x citations each)

Department $F = (f_1, f_2)$

- author $f_1 = 4 \cdot \mathbf{1}_4$
- author $f_2 = 4 \cdot \mathbf{1}_4$
  - h-index of both authors is 4
  - h-index of the department is 4

Department $G = (g_1, g_2)$

- author $g_1 = 3 \cdot \mathbf{1}_6$
- author $g_2 = 3 \cdot \mathbf{1}_6$
  - h-index of both authors is 3
  - h-index of the department is 6

Potential problems with the h-index (2/2)

Evaluation of authors and departments

- the h-index of a department is x if this department has x papers having at least x citations each (and its other papers have at most x citations each)

Department $F = (f_1, f_2)$

- author $f_1 = 4 \cdot \mathbf{1}_4$
- author $f_2 = 4 \cdot \mathbf{1}_4$
  - h-index of both authors is 4
  - h-index of the department is 4

Department $G = (g_1, g_2)$

- author $g_1 = 3 \cdot \mathbf{1}_6$
- author $g_2 = 3 \cdot \mathbf{1}_6$
  - h-index of both authors is 3
  - h-index of the department is 6

Consistency is violated

- the “best” department contains the “worst” authors!

Bart knows!


I will not use the h-index anymore
I will not use the h-index anymore
I will not use the h-index anymore
I will not use the h-index anymore
I will not use the h-index anymore
I will not use the h-index anymore
I will not use the h-index anymore
I will not use the h-index anymore
I will not use the h-index anymore
I will not use the h-index anymore
I will not use the h-index anymore

Outline

Model of Authors

Authors

- an author is a function f from $\mathbb{N}$ to $\mathbb{N}$
- $f(x)$ is the number of papers by this author having received x citations

Model of Authors

Authors

- an author is a function f from $\mathbb{N}$ to $\mathbb{N}$
- $f(x)$ is the number of papers by this author having received x citations

x	0	1	2	3	4	5	6	7	8	...
f	1	2	2	1	0	0	0	0	0	...

6 papers, 9 citations

Model of Authors

Authors

- an author is a function f from $\mathbb{N}$ to $\mathbb{N}$
- $f(x)$ is the number of papers by this author having received x citations

x	0	1	2	3	4	5	6	7	8	...
f	1	2	2	1	0	0	0	0	0	...

6 papers, 9 citations

Objective

- build a binary relation $\succsim$ on $\mathscr{A}$
- $f \succsim g$ if "given their publication/citation record", scientist f is at least as good as scientist g

Model of Authors

Authors

- an author is a function f from $\mathbb{N}$ to $\mathbb{N}$
- $f(x)$ is the number of papers by this author having received x citations

x	0	1	2	3	4	5	6	7	8	...
f	1	2	2	1	0	0	0	0	0	...

6 papers, 9 citations

Objective

- build a binary relation $\succsim$ on $\mathscr{A}$
- $f \succsim g$ if "given their publication/citation record", scientist f is at least as good as scientist g

Important Limitation

- coauthors are ignored in this talk

Model of Departments

Departments

- a department of size k is an element of $\mathscr{A}^k$: $(f_1, f_2, \ldots, f_k)$

Model of Departments

Departments

- a department of size k is an element of $\mathscr{A}^k$: $(f_1, f_2, \ldots, f_k)$

Objective

- build a binary relation $\trianglerighteq$ on $\mathscr{D}$
- $F \trianglerighteq G$ if "given their publication/citation record of the scientists in departments F and G", department F is at least as good as department G"

Model of Departments

Departments

- a department of size k is an element of $\mathscr{A}^k$: $(f_1, f_2, \ldots, f_k)$

Objective

- build a binary relation $\trianglerighteq$ on $\mathscr{D}$
- $F \trianglerighteq G$ if "given their publication/citation record of the scientists in departments F and G", department F is at least as good as department G"

Important limitations

- multiple affiliations are ignored
- field normalization is ignored

Axioms

Build $\succsim$ and $\trianglerighteq$ satisfying

- Consistency
  - seen above
- Transfer
  - if a member of a department publishes a new paper I do not care about who in the department is doing so
- Homogeneity
  - duplicating all authors in a department leaves unchanged the position of the department
- Archimedean
  - any two citation profiles are commensurate

Axioms

Build $\succsim$ and $\trianglerighteq$ satisfying

- Consistency
  - seen above
- Transfer
  - if a member of a department publishes a new paper I do not care about who in the department is doing so
- Homogeneity
  - duplicating all authors in a department leaves unchanged the position of the department
- Archimedean
  - any two citation profiles are commensurate

Independence is implied

Consistency

$F = (f_1, f_2, \ldots, f_k)$ and $G = (g_1, g_2, \ldots, g_k)$: departments of size k.

If $f_i \succsim g_i$, for all i then $F \trianglerighteq G$

If $f_i \succsim g_i$, for all i and if $f_j \succ g_j$, for some j then $F \triangleright G$

Transfer

$$(f_1, \ldots, f_i + \mathbf{1}_x, \ldots, f_k) \triangleq (f_1, \ldots, f_j + \mathbf{1}_x, \ldots, f_k)$$

Homogeneity

$$(f_1, f_2, \ldots, f_k) \triangleq (\underbrace{f_1, f_1, \ldots, f_1}_{n}, \underbrace{f_2, f_2, \ldots, f_2}_{n}, \ldots, \underbrace{f_k, f_k, \ldots, f_k}_{n})$$

Archimedeanness

$$f \succ g \Rightarrow \exists n \in \mathbb{N} \text{ s.t. } f' + (n \cdot f) \succsim g' + (n \cdot g)$$

Scoring rules for scientists

Definition

$\succsim$ is a scoring rule for scientists (s-scoring rule) if there is a real valued function u on $\mathbb{N}$ such that

$$f \succsim g \Leftrightarrow \sum_{x \in \mathbb{N}} f(x)u(x) \geq \sum_{x \in \mathbb{N}} g(x)u(x)$$

- $u(x)$ gives the worth of one publication with x citations
- many bibliometric indices are scoring rules (but not the h-index)
- all scoring rules satisfy independence

Scoring rules for scientists

Definition

$\succsim$ is a scoring rule for scientists (s-scoring rule) if there is a real valued function u on $\mathbb{N}$ such that

$$f \succsim g \Leftrightarrow \sum_{x \in \mathbb{N}} f(x)u(x) \geq \sum_{x \in \mathbb{N}} g(x)u(x)$$

- $u(x)$ gives the worth of one publication with x citations
- many bibliometric indices are scoring rules (but not the h-index)
- all scoring rules satisfy independence

Examples

- $u(x) = x$: number of citations
- $u(x) = 1$: number of publications
- $u(x) = 1$ if $x \geq \alpha$: number of highly cited publications

Rules for departments

Definition

$\trianglerighteq$ is an averaging rule for departments (d-averaging rule) if there is a real valued function v on $\mathbb{N}$ such that

$$(f_1, f_2, \ldots, f_k) \trianglerighteq (g_1, g_2, \ldots, g_\ell) \Leftrightarrow \frac{1}{k}\sum_{i=1}^{k}\sum_{x\in\mathbb{N}} f_i(x)v(x) \geq \frac{1}{\ell}\sum_{i=1}^{\ell}\sum_{x\in\mathbb{N}} g_i(x)v(x)$$

Sample result

Theorem (B & Marchant, 2011)

The relations $\succsim$ and $\trianglerighteq$ are linked by Consistency, $\trianglerighteq$ satisfies Transfer and Homogeneity, $\succsim$ satisfies Archimedeanness
if and only if
$\succsim$ is an s-scoring rule and $\trianglerighteq$ is a d-averaging rule with $u = v$

The function u is unique up to the multiplication by a positive constant

Extensions

Extensions

- add additional conditions to restrict the shape of u
  - u is nondecreasing
  - u is constant
  - u is linear
- characterize indices instead of rankings

Easy!

Extensions

Extensions

- add additional conditions to restrict the shape of u
  - u is nondecreasing
  - u is constant
  - u is linear
- characterize indices instead of rankings

Easy!

Extensions

- coauthors
- multiple affiliations
- field normalization
- length of career ("age")

Difficult!

Outline

Messages

Bibliometrics

- bibliometrics is not limited to evaluative bibliometrics
- (evaluative) bibliometrics is an interesting field of study

Messages

Bibliometrics

- bibliometrics is not limited to evaluative bibliometrics
- (evaluative) bibliometrics is an interesting field of study

Evaluative bibliometrics in practice

- it should be used with much care
- it should not be in the hands of laypersons
- it should not be entrenched in formal rules
- it should always be used as a complement to careful and impartial peer review
  - there is no substitute to reading the papers!
  - there is no substitute to open and public debate!

More Messages

Warning

- there are quite bad indices
- beware of scientists giving their h-index on their Web page or CV!
- beware of comparisons of Universities using bibliometric indices

More Messages

Warning

- there are quite bad indices
- beware of scientists giving their h-index on their Web page or CV!
- beware of comparisons of Universities using bibliometric indices

(Informal) Proposition on Evaluative Bibliometrics

If

- trained bibliometricians have prepared a clean database
- used to compare people of the "same age" and working in the same field
- using scoring rules

then (and only then)

Evaluative Bibliometrics may be of some help

Are you excellent?

Excellence

- excellence is another word for outliers
  - not everyone can be excellent!
  - what should we do with people that are not excellent?
  - is the mantra of excellence a good motivating tool?

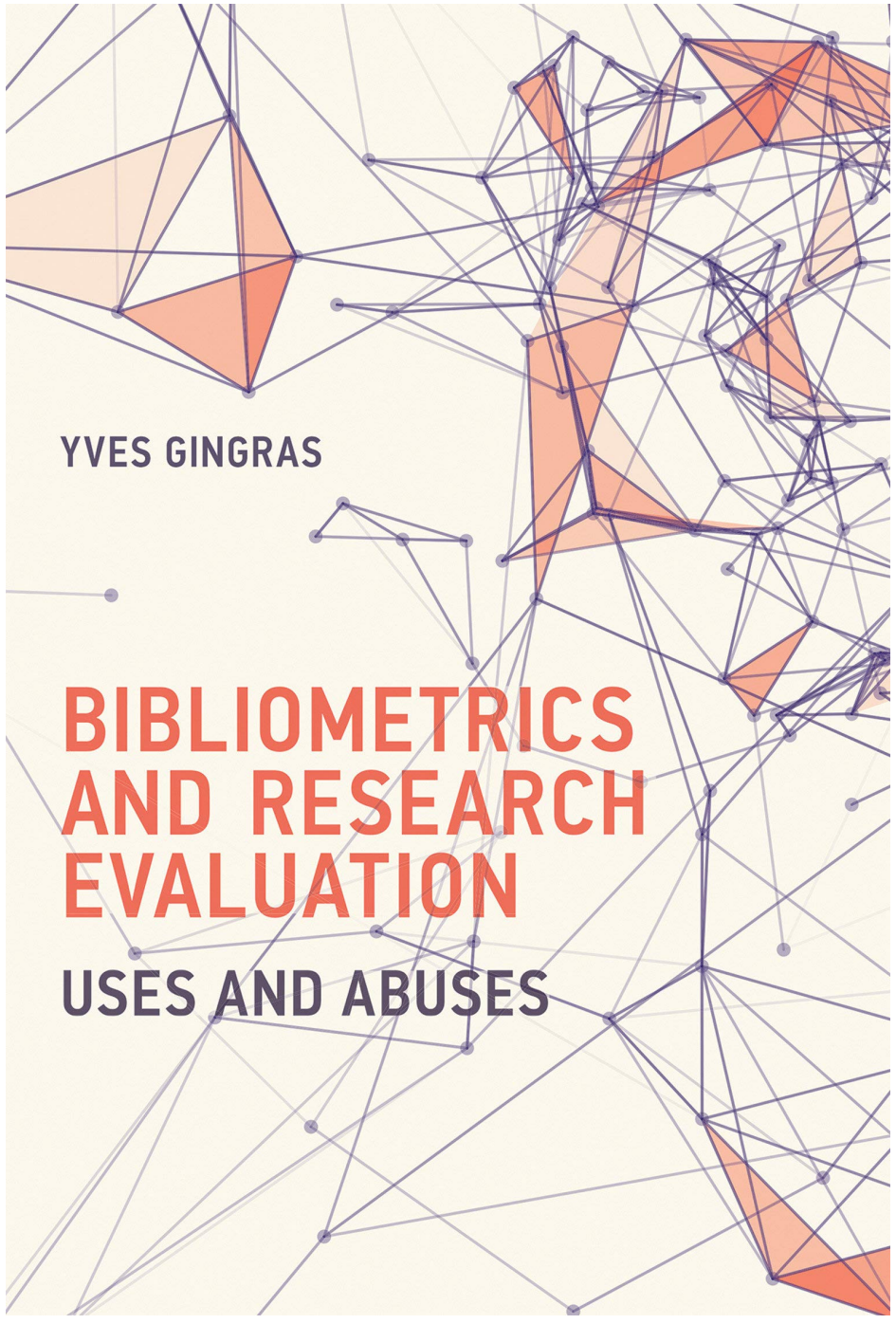
References

Adler, R., Ewing, J., Taylor, P. (2009)
Citation statistics
Statistical Science, **24** (1), 1–14

Bouyssou, D., Marchant, T. (2011)
Ranking scientists and departments in a consistent manner
Journal of the American Society for Information Science and Technology, **62** (9), 1761–1769

Bouyssou, D., Marchant, T. (2014)
An axiomatic approach to bibliometric rankings and indices
Journal of Informetrics, **8** (1), 449–477

Bouyssou, D., Marchant, T. (2016)
Ranking authors using fractional counting of citations: An axiomatic approach
Journal of Informetrics, **10** (1), 183–199


YVES GINGRAS
BIBLIOMETRICS AND RESEARCH EVALUATION
USES AND ABUSES


C'EST
TOUT
POUR
AUJOURD'
HUI.

Questions?