THOUGHTS AND FACTS ON BIBLIOMETRIC INDICATORS IN THE LIGHT OF NEW CHALLENGES IN THEIR APPLICATIONS

Wolfgang Glänzel

Centre for R&D Monitoring and Dept MSI, KU Leuven

ECOOM

Structure

Bibliometrics has successfully developed methods and indicators in the course of its about 50 years history. In this presentation I would like to highlight four of the challenges the discipline is faced with.

- The first one refers to *perspective shift* that took place during the last three–four decades.
- The second issue refers to the focus shift away from macro studies down to meso and micro studies.
- The third, more general issue refers to basic requirements that characterise all scientific methods, namely, meaningfulness, validity, replicability and robustness.
- The fourth problem, finally, emerges from the extension of bibliometric studies beyond their original scope, among others to the social sciences, humanities and the web including social networks.

The origin of bibliometrics/scientometrics

- Bibliometrics represented a statistical approach to master the growing flood of scientific information and to analyse and to understand the cognitive characteristics of “big science” by measuring quantitative aspects of communication in science and by providing the results to scientists and users outside the scientific community.
- Describing, modelling and monitoring the process of knowledge production, of dissemination and use of information was originally in the foreground.

Historical background

- First scientometric applications were developed to improve use of bibliographic databases and to extend information services.
- The journal *Impact Factor* (Garfield & Sher, 1963) was originally used to help select journals for the *Science Citation Index* (SCI).
  - Garfield himself recognised the power of the IF for journal evaluation and considered it later also a journal performance indicator.
- The co-citation based *Atlas of Science* developed and issued by the Institute for Scientific Information (ISI) was considered a new kind of "review literature" which is also suited to help students in choice of career in science (Garfield, 1975).

The perspective shift

- A consequence of the growth of knowledge and the evolution from little science to big science: Need for supplementing research evaluation with quantitative methods and of linking funding to performance indicators.
- The application to science policy has brought a new perspective, and resulted in re-interpretation of bibliometric conceptions.
- The ‘science indicators’ movement in the US with the discussion about the possible use of bibliometrics in science policy in the 1970s marked the beginning of a new era in bibliometrics.

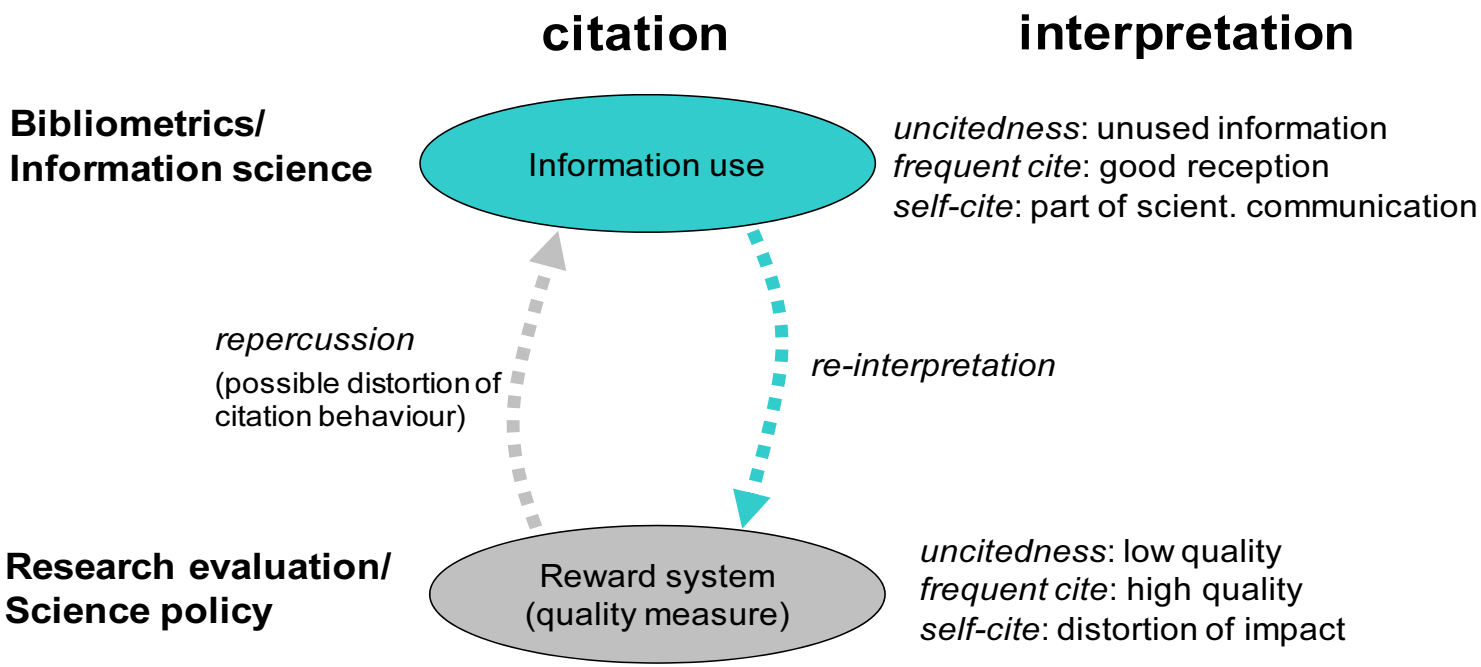
The perspective shift

- Bibliometrics evolved from a sub-discipline of library and information science to an instrument for evaluation and benchmarking. I have called this "perspective shift" in bibliometrics.
- As a consequence of this perspective shift, new fields of applications and challenges opened to bibliometrics.
- Due to the dynamics in evaluation, the availability and the rise of social networking technologies, the focus has shifted away from macro studies towards meso and micro studies of both actors and topics.
  - Many tools were still designed for use in scientific information, information retrieval and libraries.
    - ☛ Those became used in a context for which these were not designed.

The perspective shift

- From the methodological viewpoint, bibliometric applications have two basic sources: *Information science* and *Sociology of science*
  - Sociology of science laid the theoretical groundwork for the paradigmatic perspective shift.
- The scope of bibliometric applications is determined by data availability and ICT related development.
- However, advancement of bibliometric methodology could not always keep pace with the demands and the breathtaking development in data processing.

Re-interpretion of the notion of citation and its consequences

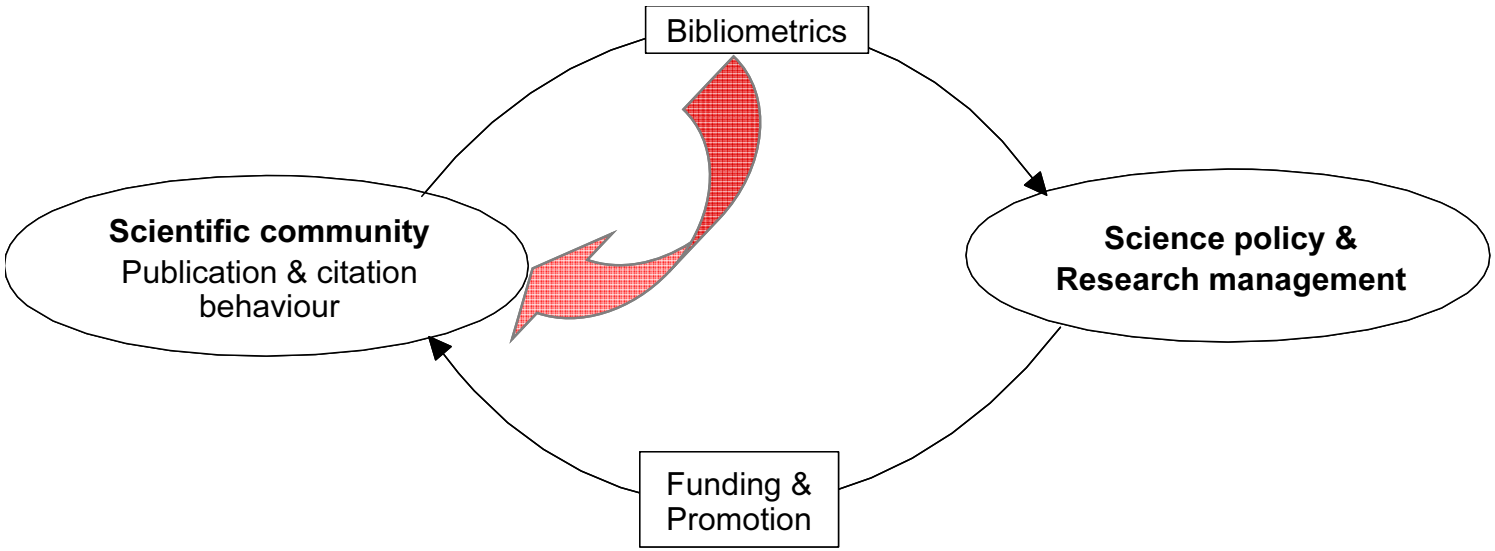


Consequences of the perspective shift

- The development of appropriate methods for different levels of aggregation is necessary.
- Necessary requirements are high quality of data and watertight methodology.
- Data at lower levels of aggregation require highest accuracy and the maximum of precision.
  - Henk Moed raised the question of errors in this context, particularly, of what is an acceptable "error rate" in the assessment process.
- While bibliometric macro and meso data still preserve a certain extent of anonymity, micro-level data expose individual researchers as they are often not treated and applied anonymously.

  $\Rightarrow$ Possible repercussions on the scientists' publication and citation behaviour (Glänzel & Debackere, 2003). Researchers have become more susceptible to the consequences of the use of bibliometrics.

Feedback of policy use of bibliometrics on the scientific community

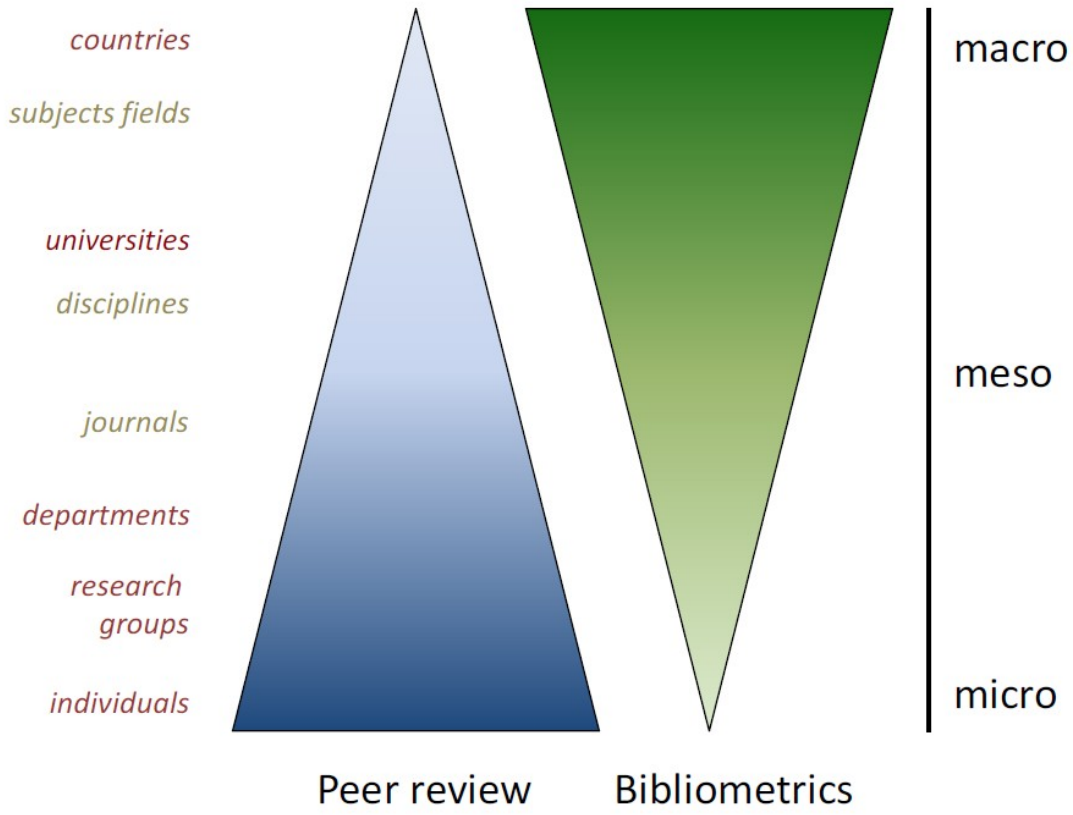


Specific requirements of micro-level bibliometrics

- While macro and meso bibliometrics is still used in traditional contexts, micro-level applications are almost exclusively applied supportive tools in research evaluation.
- Researchers are increasingly concerned by policy use and misuse of bibliometric methods and sometimes they even feel as victims of the evaluation.
- Special caution is always called for at this level and final conclusions should be drawn in combination with "qualitative methods"
- In this context Glänzel & Wouters (2013) have formulated 20 recommendations for bibliometrics ("The dos and don'ts in individual-level bibliometrics").

The challenge of meso and micro bibliometrics

The weight of qualitative (peer evaluation) and quantitative (bibliometrics) methods as function of the aggregation level



Source: Glänzel, 2011

TEN THINGS ONE MUST NOT DO AT THE INDIVIDUAL-LEVEL

1. Don't reduce individual performance to single numbers
2. Don't use IFs as measures of quality
3. Don't apply (hidden) "bibliometric filters" for selection
4. Don't apply arbitrary weights to co-authorship
5. Don't rank scientists according to ONE indicator
6. Don't merge incommensurable measures
7. Don't use flawed statistics
8. Don't blindly trust one-hit wonders
9. Don't compare apples and oranges
10. Don't allow deadlines and workload to compel you to drop good practices

TEN THINGS ONE MIGHT DO WITH INDIVIDUAL-LEVEL BIBLIOMETRICS

1. Also individual-level bibliometrics is statistics
2. Analyse collaboration profiles of researchers
3. Always combine quantitative and qualitative methods
4. Use citation context analysis
5. Analyse subject profiles
6. Make an explicit choice for oeuvre or time-window analysis
7. Combine bibliometrics with career analysis
8. Clean bibliographic data carefully and use external sources
9. Even some "don'ts" are not taboo if properly applied
10. Help users to interpret and apply your results

Deterministic vs. probabilistic approach

- The deterministic approach is the easiest way to process simple counts of raw data and measurements to indicators.
  - Mostly elementary mathematical operations (e.g., shares, averages, ratios) are used.
  - The interpretation of more complex measures and constructs (composite indicators) becomes more problematic.
- In social processes, observations are subject to a variety of influences that are partially not directly measurable.
  - The complexity of social interactions itself yields the effect that is usually interpreted as randomness.
- In bibliometrics, the events seem to be random as they are conditioned by a plethora of superposing actions, processes and effects.
  - Communication, mobility, collaboration, publications and citations all are subject to these effects.

Glänzel & Moed, *Thoughts and facts on bibliometric indicators*, 2012

The stochastic nature of bibliometric indicators

One important consequence of the probabilistic approach is that indicators of units of analysis that take certain values, can be considered derivatives of empirical distributions.

- ☞ This property applies to many statistics like moments, relative frequencies, quantiles, etc.

Another important feature is the option to introduce a time-dependent parameter resulting in stochastic processes that are able to reflect the changes of probabilities, moments and empirical values in time.

Example

Dieks and Chang (1976) introduced a mathematical model describing citing as a stochastic process.

- ☛ Furthermore, the parameter of the Poisson distribution itself could be assumed to be a random variable (according to the subject, age, social status, etc. of the author) at any time so that a compound process is obtained (cf. Burrell, 2005).

☛ Bookstein (1997) distinguishes three "demons to measurement" causing uncertainty in social sciences: *Randomness* (can be coped with using stochastic methods) as well as *Fuzziness* and *Ambiguity*.

Pitfalls and Caveats

- One has to distinguish between samples, in general, and *random* samples, in particular.
- The conditions for the application of statistical standard methods (including tests) are not always met.
- Sometimes it is more important to know if deviations are *not* significant than that they might be significant.
- Regression analysis based on statistical correlation is one of the most popular methods but correlation does not imply causality.
  - Independent variables are uncorrelated but the *reverse statement does not necessarily* hold.

Approximate normality of means and shares

Authorship, publication activities, references, citation histories and other bibliometric phenomena can be expressed by random variables or stochastic processes.

Scientometric distributions are usually extremely skewed, some are discrete and integer-valued but most indicators derived from these are *approximately* normally distributed.

Sample means and shares of uncited papers according to Glänzel & Moed (2013)
(1% of Belgian publications in 2004 with 3 year citation window)

k	n	x	f_0	k	n	x	f_0	k	n	x	f_0
1	126	5.016	20.6%	8	113	5.265	21.2%	15	108	5.028	16.7%
2	94	7.160	23.4%	9	115	5.652	18.3%	16	108	4.102	25.0%
3	133	5.639	20.3%	10	117	5.538	20.5%	17	130	6.362	24.6%
4	122	5.951	19.7%	11	122	5.385	19.7%	18	137	4.569	28.5%
5	126	6.262	22.2%	12	149	6.913	23.5%	19	114	4.456	19.3%
6	112	5.768	17.0%	13	103	4.641	29.1%	20	110	6.473	21.8%
7	128	6.992	22.7%	14	145	7.807	21.4%	Total	2412	5.788	21.8%

Data source: Thomson Reuters Web of Knowledge

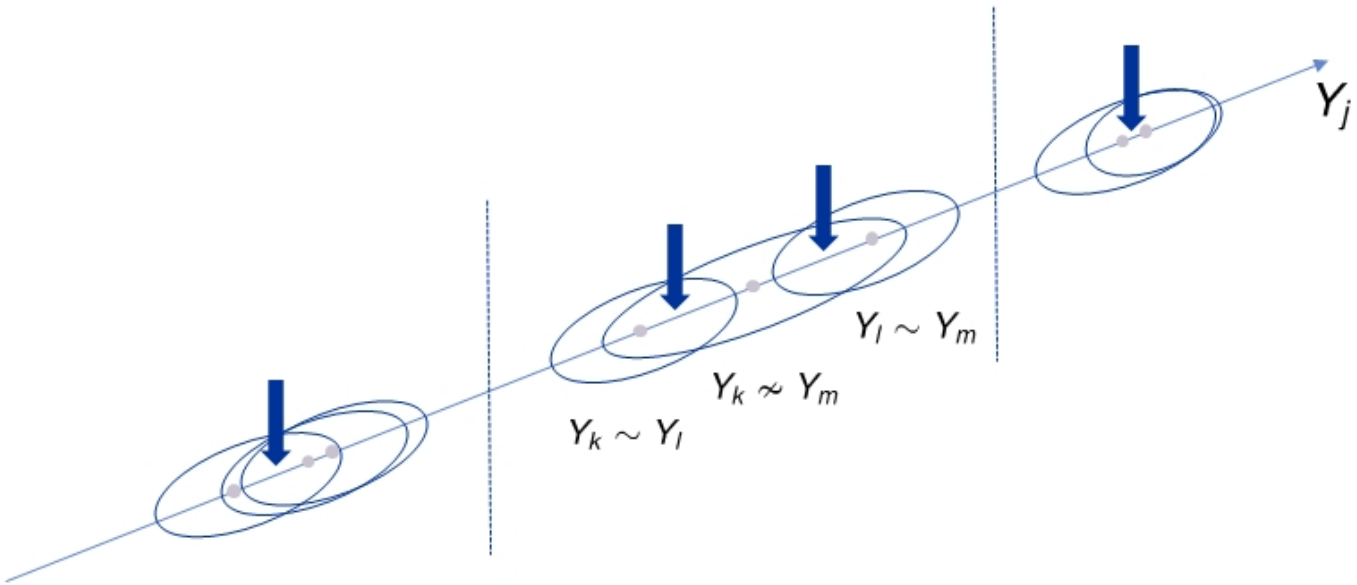
Reliability issues: “predictive bibliometrics”

Predictive bibliometrics refers to the future prediction of productivity and citation impact on the basis of observations. Using the model of a negative-binomial process, Glänzel & Schubert (1995) have shown that the “random error” of predictions mainly depend on three factors.

1. The sample size (– decreases reciprocally with the square root of the sample size)
2. The observation period (– decreases reciprocally with the square root of the mean value in the observation period)
3. The prediction period (– increases proportionally with the square root of the mean value in the prediction period)

☛ Note that predictions do never apply to citation rates of individual papers or productivity of individual authors.

Statistical reliability in ranking
(Pinprick and separation problem – sketch)



Next steps – "Non-linear bibliometrics"

Background:

- Continuous debate about normalisation of bibliometric indicators
- Insufficiency of using a single indicator to depict any quantitative aspects of research performance (e.g., Glänzel, 2009)
- Extreme values might bias scientometric indicators (Waltman et al., 2012)
- The disproportion between standard models of "regular" and "outstanding" performance as reflected by citation impact

☛ A solution is needed that integrates measures of outstanding (and even extreme performance) seamlessly into the standard tools of scientometric performance assessment.

Next steps – “Non-linear bibliometrics”

Desired properties of the solution:

- provide a parameter-free formula for different performance classes
- is applicable to both subject analysis and the combination of different subjects
- is suited for application at different levels of aggregation (macro, meso and micro level)
- replace the conception of “linearly structured” indicators by performance profiles

The “high-end” of performance

Extreme citation rate received by one paper till September 2014

A short history of SHELX
By: Sheldrick, George M.
ACTA CRYSTALLOGRAPHICA SECTION A Volume: 64 Pages: 112-122 Part: 1 Published: JAN 2008

Times Cited: 37,623
(from Web of Science Core Collection)

The phenomenon of this “outlier” publication has been pointed out by Waltman et al., 2012. The extreme citation rate of this paper even biases the mean citation impact of the university.

- ☞ While in other fields statistics might be adjusted, in bibliometrics these values represent the high-end of research performance and deserve special attention.

Parameter-free solutions

An alternative is a "reduction" of the distribution over individual items to a distribution over some essential classes representing specific sections of the original one.

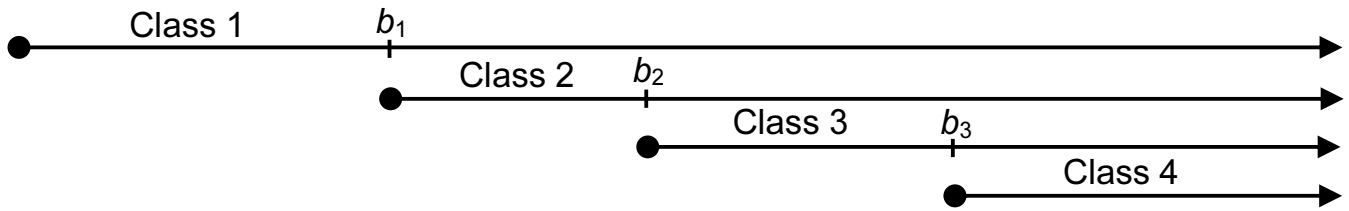
☞ Outliers then become just a "member" of a performance class without distorting indicator values.

- A solution using six classes has been suggested by Leydesdorff et al. (2011).
- According to their model, a pre-set number of six rank percentages is calculated on the basis of the reference distribution.
- Individual observations are then scored according to the percentage the publications in question belong to.

☛ Two particular problems arise from this approach, namely the arbitrariness of pre-set percentiles and the ties in both the reference distribution and the observations.

Characteristic scores are obtained from iteratively truncating samples at their mean value and recalculating the mean of the truncated sample. The procedure is repeated until a given number of scores k is reached.

Visualisation of characteristic scores and scales for four classes (Glänzel, 2007)



$[b_0, b_1)$ is the class of 'poorly cited' papers,
$[b_1, b_2)$ contains 'fairly cited' papers,
$[b_2, b_3)$ contains 'remarkably cited' papers and
$[b_3, \infty)$ is the class of 'outstandingly cited' papers.

The values $k = 2$ and $k = 3$ are often used to identify *highly cited* papers.

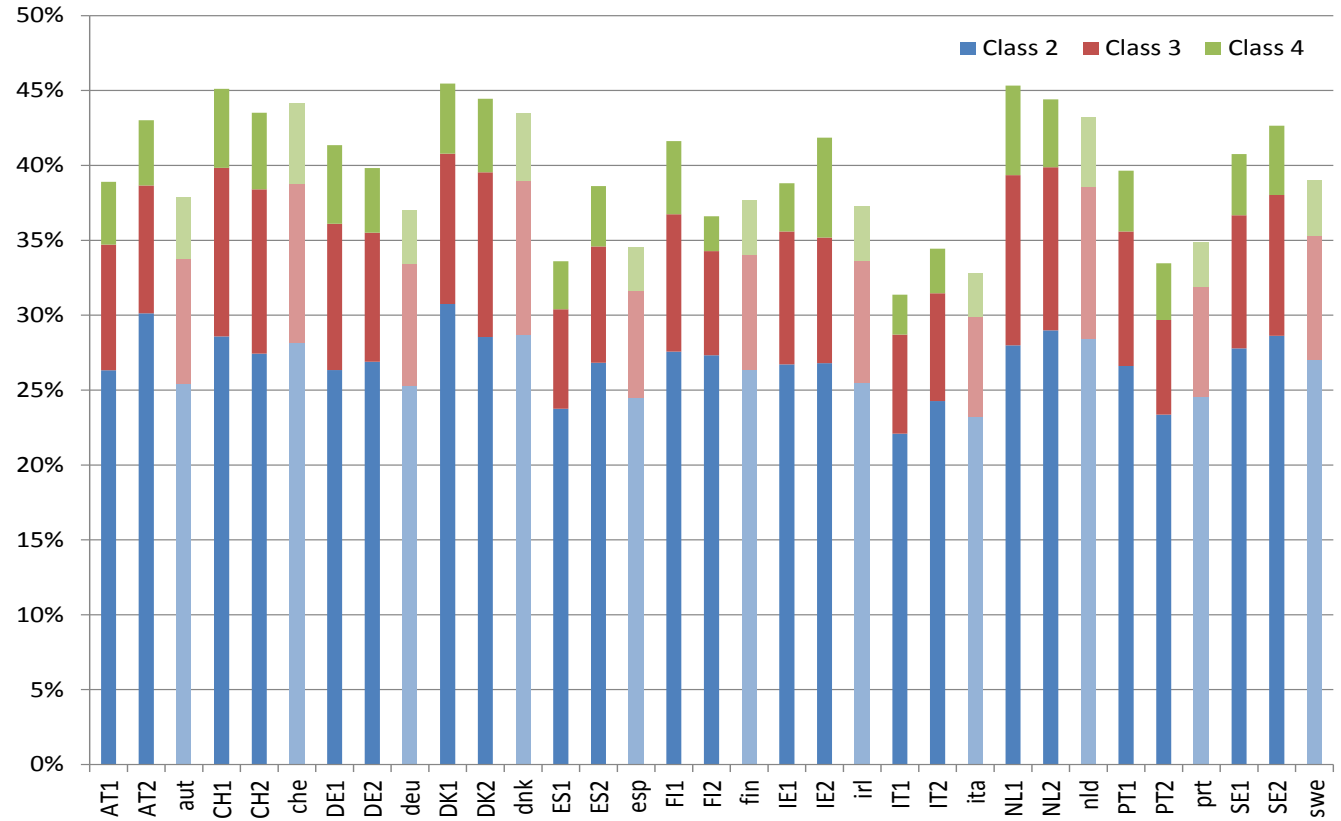
CSS classes in all fields combined in 2007 and 2009

Class	Share (in %)	
	2007 (5-year cites)	2009 (3-year cites)
1	69.8	69.7
2	21.5	21.4
3	6.3	6.4
4	2.4	2.5

Source: Thomson Reuters – Web of Science Core Collection

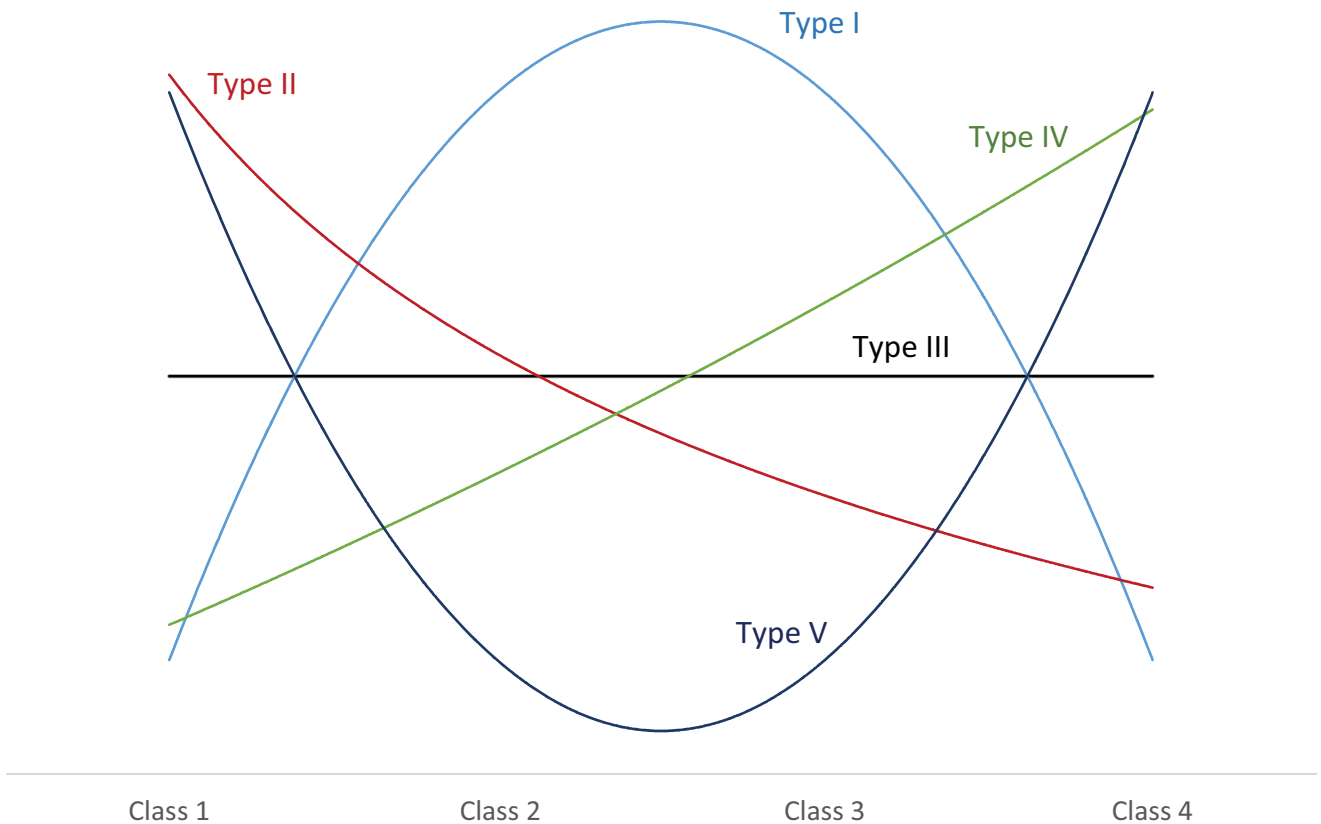
CSS performance classes in institutional research assessment

Shares of publications of selected universities and countries in the upper three CSS classes in all fields combined in 2007 (5-year citation window)



Source: Thomson Reuters – Web of Science Core Collection

Five possible profiles according to a sample's deviation from the reference standard



☞ Profiles are more pronounced at lower levels of aggregation.

CSS for indicator normalisation

The deviation of the two scores $(\beta_2 - \beta_1)$ can be a proxy for the scale parameter of the underlying distribution, where β_i are the empirical values of the corresponding characteristic scores b_i.
The transformation suggested by Schubert et al. (1989) can then be applied for scale normalisation.

$$u^* = \frac{x}{\beta_2 - \beta_1},$$

where x represents the actual citation statistic of the sample distribution.

Example: The citation distribution of a WoS subject category serves as the reference distribution, x stands for the impact factors of the journals assigned to this category.

A posteriori normalisation of journal impact measures

Threshold for "highly" cited papers (β_3) based on *Characteristic Scores and Scales* and their (a posteriori) normalised versions (β_3^*) according to Glänzel (2011)
[B1: biochemistry/biophysics/molecular biology; H1: applied mathematics]

	B1		H1	
	1980-2000	2006-2008	1980-2000	2006-2008
β_3	196.55	22.69	49.66	4.47
β_3^*	3.51	3.53	3.28	3.46

Data source: Thomson Reuters Web of Knowledge

The challenge of network indicators

In the classical, linear model, publications are considered separate entities, and citations regarded as separate events.

- Both citing and cited articles or authors may have all kinds of relationships that need to be taken into account when assessing citation impact of a unit of assessment.

☞ Thus units entering the network might influence indicator values of those units with which they are *not directly linked*.

Network based indicators such as *Scimago Journal Rank indicator*, Thomson Reuters' *Eigenfactor/Article Influence Score* and other Page-Rank based measures can reflect much more complex relationships, but are subject to structural changes in the complete system.

- ☛ Validity requirement concerning network indicators should therefore be considered from a completely different perspective as is usually adopted in the case of traditional statistical functions.

The challenge of new data sources

Bibliometrics was originally developed for and applied to journal literature in basic research in the sciences.

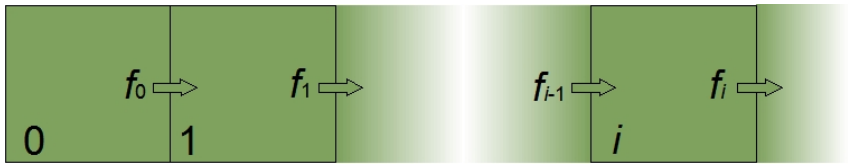
- In the 1990s, its scope was gradually extended to the technical sciences, the social sciences and humanities.
- More recently the web including social networks have substantially extended both tools and application fields of bibliometrics.
  - The extension of data sources resulted in an open, dynamic and almost unlimited universe.
  - The opening of information *sources* and *targets* are changing the types of metrics.
  - New publication and *output* types appear.
    - ☞ Authored and edited books, monographs, e-papers, annotated corpora, scientific data; excavation, artwork etc. with implementations for ageing (time windows), subject classification, version (new, improved editions, translation and languages).

These new challenges also result in conceptual changes.

- The notion of ageing of literature and obsolescence needs to be reconsidered. Different models might be used simultaneously.
- The expansion of publication types and other types of research output raises the question of redundancy of information (including that of its use, e.g., citations).
- Different output types might require different granularities of classification – a challenge to traditional subject classification schemes.
- The development of new models will be necessary.

Web-based bibliometrics – Example

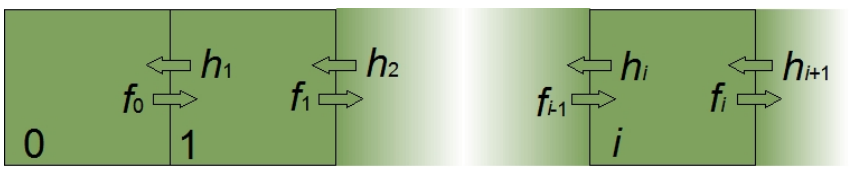
The citation process as a simple birth process in traditional bibliometrics



$$f_i = (a + b \cdot i)x_i \; ; \quad (a > 0; b \geq 0)$$

The solution for any t is a Negative Binomial distribution.

The "Web citation process" in open networks with possible removal of links as a simple birth-and-death process ('ideal case' if sources are not removed)



$$f_i = (a + b \cdot i)x_i \, , \quad h_i = (c + d \cdot i)x_i \; ; \quad (a, c > 0; b, d \geq 0)$$

The general solution is more complicated and analytical expressions are known only for some special cases.

Some consequences

- Birth processes proved to be suitable models for citation processes. Appropriate 'obsolescence' measures can readily be defined through these models.
- Birth-and-death processes are models more appropriate to describe link-based processes in open, dynamic networks. They reflect more complex demographic patterns than the 'one-dimensional' ageing characterising citation processes
- The change of the traditional notion of publication and citation results in non-cumulative metrics.

Thank you very much for your attention.