

BIBLIOMETRICS IN A NUTSHELL

WOLFGANG GLÄNZEL

KU Leuven, ECOOM & Dept MSI, Leuven (Belgium)

Leuven, 2012

1. What is bibliometrics?

- Bibliometrics and scientometrics are sub-disciplines of information science.

Both terms are used almost synonymously for

- “*the application of quantitative methods which are dealing with the analysis of science viewed as an information process*” (NALIMOV & MULCHENKO, 1969),
 - “*the application of mathematical and statistical methods to books and other media of communication*” (PRITCHARD, 1969).
- Bibliometrics is a tool and extension of *scientific information*.
 - Bibliometrics can be used to develop and provide tools to be applied to *research evaluation*, but is **not designed to directly evaluate research performance**, and **does not aim at correcting or replacing qualitative methods** by quantitative approaches.

1. What is bibliometrics?

- Bibliometrics and scientometrics are sub-disciplines of information science.

Both terms are used almost synonymously for

- “*the application of quantitative methods which are dealing with the analysis of science viewed as an information process*” (NALIMOV & MULCHENKO, 1969),
 - “*the application of mathematical and statistical methods to books and other media of communication*” (PRITCHARD, 1969).
- Bibliometrics is a tool and extension of *scientific information*.
 - Bibliometrics can be used to develop and provide tools to be applied to *research evaluation*, but is **not designed to directly evaluate research performance**, and **does not aim at correcting or replacing qualitative methods** by quantitative approaches.

1. What is bibliometrics?

- Bibliometrics and scientometrics are sub-disciplines of information science.

Both terms are used almost synonymously for

- “*the application of quantitative methods which are dealing with the analysis of science viewed as an information process*” (NALIMOV & MULCHENKO, 1969),
 - “*the application of mathematical and statistical methods to books and other media of communication*” (PRITCHARD, 1969).
- Bibliometrics is a tool and extension of *scientific information*.
 - Bibliometrics can be used to develop and provide tools to be applied to *research evaluation*, but is **not designed to directly evaluate research performance**, and **does not aim at correcting or replacing qualitative methods** by quantitative approaches.

2. Why bibliometrics?

Whenever a discipline reaches a stage that requires the support of statistical methods, a metrics emerges from this discipline.

This applies to information science, too: Statistical methods were needed to measure and model specific phenomena of present-day scientific communication.

Particular challenges are:

- the enormous growth of scientific literature
- assistance needed to cope with the information overload
- the increasing complexity of scholarly communication
- the change from “little science” to “big science”
- competitive access to financial and human resources
- crisis of the peer review system

2. Why bibliometrics?

Whenever a discipline reaches a stage that requires the support of statistical methods, a metrics emerges from this discipline.

This applies to information science, too: Statistical methods were needed to measure and model specific phenomena of present-day scientific communication.

Particular challenges are:

- the enormous growth of scientific literature
- assistance needed to cope with the information overload
- the increasing complexity of scholarly communication
- the change from “little science” to “big science”
- competitive access to financial and human resources
- crisis of the peer review system

3. What are the benefits of bibliometrics?

Bibliometrics can provide useful information in three main areas of application, particularly in

- scientific information
 - what is “hot” in my discipline
 - how to get fast access to this information
 - what are the networks in my field
- studies of structure and dynamics of science
 - mapping and visualisation of the structure of science
 - emergence of new trends, research fields and hot topics
- research evaluation and funding
 - development of methods to measure research output and performance
 - bibliometric components in funding formulas
 - comparative studies for research assessment

3. What are the benefits of bibliometrics?

Bibliometrics can provide useful information in three main areas of application, particularly in

- scientific information
 - what is “hot” in my discipline
 - how to get fast access to this information
 - what are the networks in my field
- studies of structure and dynamics of science
 - mapping and visualisation of the structure of science
 - emergence of new trends, research fields and hot topics
- research evaluation and funding
 - development of methods to measure research output and performance
 - bibliometric components in funding formulas
 - comparative studies for research assessment

3. What are the benefits of bibliometrics?

Bibliometrics can provide useful information in three main areas of application, particularly in

- scientific information
 - what is “hot” in my discipline
 - how to get fast access to this information
 - what are the networks in my field
- studies of structure and dynamics of science
 - mapping and visualisation of the structure of science
 - emergence of new trends, research fields and hot topics
- research evaluation and funding
 - development of methods to measure research output and performance
 - bibliometric components in funding formulas
 - comparative studies for research assessment

3. What are the benefits of bibliometrics?

Bibliometrics can provide useful information in three main areas of application, particularly in

- scientific information
 - what is “hot” in my discipline
 - how to get fast access to this information
 - what are the networks in my field
- studies of structure and dynamics of science
 - mapping and visualisation of the structure of science
 - emergence of new trends, research fields and hot topics
- research evaluation and funding
 - development of methods to measure research output and performance
 - bibliometric components in funding formulas
 - comparative studies for research assessment

4. What are the usual approaches in bibliometrics?

The application of bibliometrics requires four steps.

1. quantification
 - data reduction (extracting essential information)
 - data cleaning and data disambiguation
2. measuring
 - What *can* and *should* we measure ?
 - What are appropriate measures for the purpose?
3. benchmarking
 - putting data into context and considering limitations
 - comparison (like with like)
4. presentation and interpretation
 - preparing charts and maps
 - providing interpretation of observations
 - linking bibliometric indicators to results of other methods

4. What are the usual approaches in bibliometrics?

The application of bibliometrics requires four steps.

1. quantification

- data reduction (extracting essential information)
- data cleaning and data disambiguation

2. measuring

- What *can* and *should* we measure ?
- What are appropriate measures for the purpose?

3. benchmarking

- putting data into context and considering limitations
- comparison (like with like)

4. presentation and interpretation

- preparing charts and maps
- providing interpretation of observations
- linking bibliometric indicators to results of other methods

4. What are the usual approaches in bibliometrics?

The application of bibliometrics requires four steps.

1. quantification

- data reduction (extracting essential information)
- data cleaning and data disambiguation

2. measuring

- What *can* and *should* we measure ?
- What are appropriate measures for the purpose?

3. benchmarking

- putting data into context and considering limitations
- comparison (like with like)

4. presentation and interpretation

- preparing charts and maps
- providing interpretation of observations
- linking bibliometric indicators to results of other methods

4. What are the usual approaches in bibliometrics?

The application of bibliometrics requires four steps.

1. quantification
 - data reduction (extracting essential information)
 - data cleaning and data disambiguation
2. measuring
 - What *can* and *should* we measure ?
 - What are appropriate measures for the purpose?
3. benchmarking
 - putting data into context and considering limitations
 - comparison (like with like)
4. presentation and interpretation
 - preparing charts and maps
 - providing interpretation of observations
 - linking bibliometric indicators to results of other methods

4. What are the usual approaches in bibliometrics?

The application of bibliometrics requires four steps.

1. quantification
 - data reduction (extracting essential information)
 - data cleaning and data disambiguation
2. measuring
 - What *can* and *should* we measure ?
 - What are appropriate measures for the purpose?
3. benchmarking
 - putting data into context and considering limitations
 - comparison (like with like)
4. presentation and interpretation
 - preparing charts and maps
 - providing interpretation of observations
 - linking bibliometric indicators to results of other methods

5. Bibliometrics at different levels of aggregation

Analogously to economics, bibliometrics distinguishes between different levels of aggregation.

- micro level (e.g., individual scientists or research teams)
- meso level (e.g., research institutions, journals, sub-fields)
- macro level (supra-national regions, countries or major science fields)

☞ Each level needs its specific standards.

☞ Importance and reliability of bibliometric methods grows with increasing level of aggregation.

5. Bibliometrics at different levels of aggregation

Analogously to economics, bibliometrics distinguishes between different levels of aggregation.

- micro level (e.g., individual scientists or research teams)
 - meso level (e.g., research institutions, journals, sub-fields)
 - macro level (supra-national regions, countries or major science fields)
- ➡ Each level needs its specific standards.
- ➡ Importance and reliability of bibliometric methods grows with increasing level of aggregation.

5. Bibliometrics at different levels of aggregation

Analogously to economics, bibliometrics distinguishes between different levels of aggregation.

- micro level (e.g., individual scientists or research teams)
 - meso level (e.g., research institutions, journals, sub-fields)
 - macro level (supra-national regions, countries or major science fields)
-
- Each level needs its specific standards.
 - Importance and reliability of bibliometric methods grows with increasing level of aggregation.

6. The requirements for bibliometric methods

There are five basic requirements for bibliometric methods mirroring basic criteria for any science.

1. *Validity* – measuring what is intended to measure
2. *Relevance* – results must be meaningful in the context of what they are produced for
3. *Reliability* – assessability of correctness of measures and errors
4. *Reproducibility* – Under identical conditions results should be reproducible
5. *Robustness* – results may not be sensitive to marginal fluctuations or minor corrections in the underlying data sources

6. The requirements for bibliometric methods

There are five basic requirements for bibliometric methods mirroring basic criteria for any science.

1. *Validity* – measuring what is intended to measure
2. *Relevance* – results must be meaningful in the context of what they are produced for
3. *Reliability* – assessability of correctness of measures and errors
4. *Reproducibility* – Under identical conditions results should be reproducible
5. *Robustness* – results may not be sensitive to marginal fluctuations or minor corrections in the underlying data sources

6. The requirements for bibliometric methods

There are five basic requirements for bibliometric methods mirroring basic criteria for any science.

1. *Validity* – measuring what is intended to measure
2. *Relevance* – results must be meaningful in the context of what they are produced for
3. *Reliability* – assessability of correctness of measures and errors
4. *Reproducibility* – Under identical conditions results should be reproducible
5. *Robustness* – results may not be sensitive to marginal fluctuations or minor corrections in the underlying data sources

6. The requirements for bibliometric methods

There are five basic requirements for bibliometric methods mirroring basic criteria for any science.

1. *Validity* – measuring what is intended to measure
2. *Relevance* – results must be meaningful in the context of what they are produced for
3. *Reliability* – assessability of correctness of measures and errors
4. *Reproducibility* – Under identical conditions results should be reproducible
5. *Robustness* – results may not be sensitive to marginal fluctuations or minor corrections in the underlying data sources

6. The requirements for bibliometric methods

There are five basic requirements for bibliometric methods mirroring basic criteria for any science.

1. *Validity* – measuring what is intended to measure
2. *Relevance* – results must be meaningful in the context of what they are produced for
3. *Reliability* – assessability of correctness of measures and errors
4. *Reproducibility* – Under identical conditions results should be reproducible
5. *Robustness* – results may not be sensitive to marginal fluctuations or minor corrections in the underlying data sources

6. The requirements for bibliometric methods

There are five basic requirements for bibliometric methods mirroring basic criteria for any science.

1. *Validity* – measuring what is intended to measure
2. *Relevance* – results must be meaningful in the context of what they are produced for
3. *Reliability* – assessability of correctness of measures and errors
4. *Reproducibility* – Under identical conditions results should be reproducible
5. *Robustness* – results may not be sensitive to marginal fluctuations or minor corrections in the underlying data sources

7. Data sources for bibliometrics

These are the main data sources of bibliometrics.

- sufficiently large bibliographies
- bibliographic databases (subject specific or multidisciplinary)
- citation and abstract databases
- full-text databases (limited coverage but enables advanced text mining techniques and its combination with citation-based methods)

7. Data sources for bibliometrics

These are the main data sources of bibliometrics.

- sufficiently large bibliographies
- bibliographic databases (subject specific or multidisciplinary)
- citation and abstract databases
- full-text databases (limited coverage but enables advanced text mining techniques and its combination with citation-based methods)

7. Data sources for bibliometrics

These are the main data sources of bibliometrics.

- sufficiently large bibliographies
- bibliographic databases (subject specific or multidisciplinary)
- citation and abstract databases
- full-text databases (limited coverage but enables advanced text mining techniques and its combination with citation-based methods)

7. Data sources for bibliometrics

These are the main data sources of bibliometrics.

- sufficiently large bibliographies
- bibliographic databases (subject specific or multidisciplinary)
- citation and abstract databases
- full-text databases (limited coverage but enables advanced text mining techniques and its combination with citation-based methods)

7. Data sources for bibliometrics

These are the main data sources of bibliometrics.

- sufficiently large bibliographies
- bibliographic databases (subject specific or multidisciplinary)
- citation and abstract databases
- full-text databases (limited coverage but enables advanced text mining techniques and its combination with citation-based methods)

8. Quantifiable items in bibliometrics

Quantifiable items have to be cleaned, disambiguated and correctly assigned to the unit of analysis before processing.

- ➡ Assignment is mostly not unique.

Quantifiable items:

- publications (different document types, e.g. books, research articles, review articles, proceedings papers, letters, notes, etc.)
- citations (= citing papers)
- references (= cited papers)
- (co-)authors
- affiliations (corporate addresses of authors)
- source (journals, serials, books, etc.)

8. Quantifiable items in bibliometrics

Quantifiable items have to be cleaned, disambiguated and correctly assigned to the unit of analysis before processing.

➡ Assignment is mostly not unique.

Quantifiable items:

- publications (different document types, e.g. books, research articles, review articles, proceedings papers, letters, notes, etc.)
- citations (= citing papers)
- references (= cited papers)
- (co-)authors
- affiliations (corporate addresses of authors)
- source (journals, serials, books, etc.)

9. Challenges in bibliometrics

The development and application of the bibliometric methods raises further principle issues of validity than those that are directly linked to bibliometric measures.

Relevant issues are as follows.

- data completeness vs. representativity and relevance
- varying data coverage of different databases
- different coverage of research fields/disciplines in the same database
- identification problems of authors, institutions, journals (due to homonyms, reorganisation, title mergers, etc.)
- subject-classification problems (in databases where journals are used for subject assignment)

10. New perspectives in bibliometrics

Bibliometrics was originally developed for application to the basic sciences. During the last decades bibliometrics have undergone substantial extensions.

- extension to applied fields, engineering and technology
- links between scientific and patent literature as indicators for science-technology linkage and knowledge transfer (e.g., patent citations, co-activity of inventors/authors)
- appropriate indicators for the assessments of the social sciences, humanities and arts
- analysis of Web visibility and Weblinks (webometrics)