Glossary

R. Heller · Date: 2026-05-09 · Last changed: 2026-05-11

Altmetrics. Metrics derived from online activity (social media mentions, news coverage, policy documents) rather than citations. See Chapter 26.

Bibliographic coupling. Two documents are bibliographically coupled when they share one or more references. Strength increases with the number of shared references. See Chapter 16.

Bibliometrics. The quantitative study of scholarly publications, their production, dissemination, and impact. A subfield of scientometrics focused on documents.

Burst detection. A method for identifying terms or topics whose frequency increases sharply in a given time window, signalling emerging research fronts. See Chapter 25.

Citation aging. The temporal pattern of how citations accumulate over time after publication. The citation half-life measures how quickly a paper accrues half its total citations. See Chapter 14.

Citation window. The time period during which citations to a publication are counted. Common choices: 3-year, 5-year, or variable windows.

Co-authorship network. A graph where nodes represent authors and edges connect authors who have co-authored at least one publication. See Chapter 15.

Co-citation. Two documents are co-cited when they are both cited by a third document. Frequently co-cited documents are considered intellectually related. See Chapter 16.

Co-word analysis. Network analysis of the co-occurrence of terms (keywords, title words, or descriptors) within documents. See Chapter 17.

Community detection. Algorithmic identification of densely connected subgroups within a network. Common methods: Leiden, Louvain, modularity optimisation. See Chapter 18.

Disparity filter. A network backbone extraction method that retains statistically significant edges based on local weight distributions. See Chapter 18.

DORA. The San Francisco Declaration on Research Assessment. A set of recommendations for improving research evaluation practices, emphasising that journal-based metrics should not be used as surrogates for individual article quality.

Field normalisation. Adjusting citation counts to account for differences in citation practices across research fields and publication years. See Chapter 13.

Fractional counting. An attribution method where credit for a multi-authored publication is divided equally among contributors (1/n per author or institution), in contrast to full counting. See Chapter 13.

Full counting. An attribution method where each author or institution on a publication receives full credit (count of 1). See Chapter 13.

h-index. A researcher has an h-index of h if h of their papers have each been cited at least h times. See Chapter 10.

Impact factor (JIF). The mean number of citations received in year t by articles published in a journal in years $t-1$ and $t-2$. See Chapter 11.

Leiden Manifesto. Ten principles for responsible research metrics, emphasising that quantitative evaluation should support qualitative, expert assessment rather than replace it. See Chapter 3.

Leiden algorithm. A community detection method that improves on Louvain by guaranteeing well-connected communities. See Chapter 18.

MNCS (Mean Normalised Citation Score). The average of field-normalised citation scores across a set of publications. An MNCS above 1.0 indicates above-world-average impact. See Chapter 13.

Network backbone. A reduced network retaining only statistically significant or structurally important edges, obtained by filtering methods such as the disparity filter. See Chapter 18.

OpenAlex. A free, open catalogue of the global research system, covering works, authors, institutions, sources, concepts, funders, and publishers. The default data source for this book. See Chapter 5.

Overlay map. A visualisation technique where a subset of data (e.g., one institution's publications) is projected onto a pre-computed base map of science. See Chapter 19.

PP(top 10%). The proportion of an entity's publications that rank in the top 10% most cited in their field and year. A percentile-based impact indicator. See Chapter 10.

Price index. The proportion of a paper's references that are no older than 5 years, measuring how current the cited literature is. See Chapter 14.

RAKE. Rapid Automatic Keyword Extraction. An unsupervised method that extracts multi-word keywords based on word co-occurrence within phrases delimited by stopwords. See Chapter 23.

Research front. A cluster of recently published, actively cited papers representing the current edge of a research area.

Retraction. The formal withdrawal of a published paper by its authors or publisher due to errors, misconduct, or irreproducibility. See Chapter 29.

Scientometrics. The quantitative study of science and innovation, encompassing bibliometrics, patent analysis, and the measurement of research systems.

STM (Structural Topic Model). A topic model that allows document-level covariates (e.g., publication year, journal) to affect topic prevalence and content. See Chapter 22.

TF-IDF. Term Frequency–Inverse Document Frequency. A weighting scheme that highlights terms distinctive to individual documents within a corpus. See Chapter 21.

Topic modelling. Unsupervised methods for discovering latent thematic structure in collections of text. Common approaches: LDA, STM, BERTopic. See Chapter 22.

UMAP. Uniform Manifold Approximation and Projection. A dimensionality-reduction technique for visualising high-dimensional data (e.g., document embeddings) in 2D. See Chapter 24.