

Petite encyclopédie de la science ouverte

RÉSUMÉ

Bibliométrie

La science ouverte transforme en profondeur **la bibliométrie**. La disponibilité sans précédent des **métadonnées** bibliographiques, de la **recherche en texte intégral** et d'autres **métriques d'utilisation** crée de nouvelles opportunités pour les **études quantitatives du corpus scientifique**. En outre, elle interroge le focus historique de la bibliométrie sur les données de citations issues d'un noyau de revues présélectionnées. De nouvelles approches cherchent à **étendre le périmètre de la bibliométrie à d'autres usages et publics**, notamment en ce qui concerne l'**impact social** de la recherche.

Historiquement, la bibliométrie a fortement été influencée par la première infrastructure commerciale destinée aux publications universitaires : le Science Citation Index (SCI) qui, dans les années 1960, jeta les bases d'un programme de recherche structuré sur la bibliométrie. Le SCI se concentrait essentiellement sur les données et les réseaux de citations au sein de revues anglophones réputées dans le domaine des sciences, des technologies et de la médecine (STM). Cette approche sélective a permis d'exploiter l'un des premiers moteurs de recherche en dépit des limites de l'informatique des années 1960. Au point de constituer une source d'inspiration directe pour l'algorithme de recherche de Google (le « PageRank »). Cependant, elle a aussi introduit plusieurs **biais performatifs**, car les revues exclues de l'index pour diverses raisons (contenu non anglophone, recherche hors STM) ont alors perdu en visibilité. Les indicateurs bibliométriques étant de plus en plus utilisés pour **évaluer et administrer la recherche**, ils ont fini par pénaliser de nombreuses productions universitaires et par nuire à la **diversité des activités scientifiques**.

Le développement du **Web** a immédiatement métamorphosé la bibliométrie. Malgré certaines analogies avec les données de citations, les hyperliens ne sont pas employés uniquement pour les publications de nature scientifique. Des appellations alternatives comme « webométrie », « infométrie » ou « cybermétrie » ont tenté de redéfinir et d'étendre l'analyse bibliométrique à un plus large éventail de documents et d'objets numériques.

À l'origine, **les relations entre les mouvements de la science ouverte et la bibliométrie étaient mitigées**. Les publications en libre accès devaient prétendument bénéficier d'un avantage en termes de citations du fait de leur visibilité accrue. En outre, les indicateurs bibliométriques étaient couramment utilisés comme arguments en faveur d'une transition vers la science ouverte. Cependant, les indices bibliométriques (comme le facteur d'impact ou l'index h) ont également servi à **consolider la position des grands éditeurs commerciaux**, voire à retarder le passage à la science ouverte. Des approches alternatives (les altermétriques) ont tenté de recadrer les indicateurs bibliométriques pour mieux intégrer les valeurs et les objectifs de la science ouverte, en tenant compte des usages non académiques

à travers des analyses de journaux ou des métriques de données sociales.

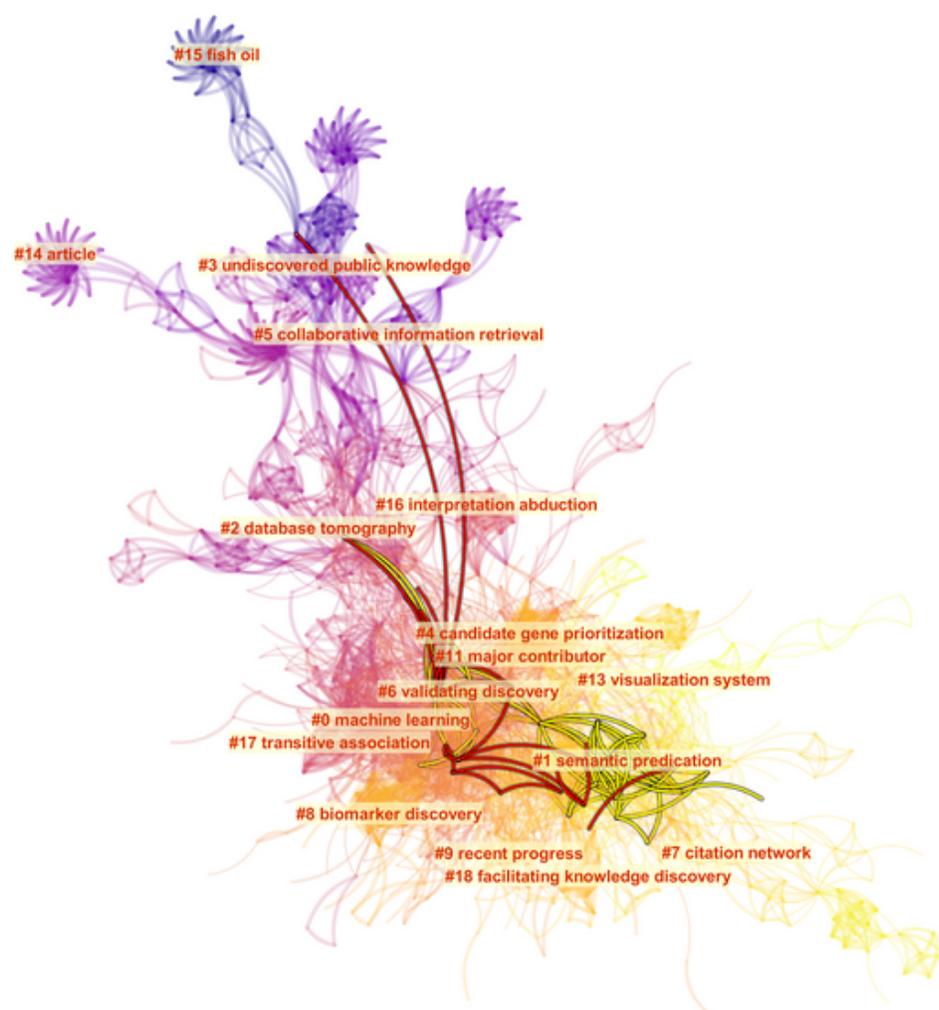
À la fin des années 2020, la bibliométrie a elle-même vécu sa révolution de la science ouverte. Avec l'ouverture des données et métadonnées scientifiques, les éditeurs commerciaux tels que Web of Science ont cédé la place à des initiatives publiques et communautaires, par exemple l'**Initiative for Open Citations** ou **Wikidata**. Les acteurs importants de la bibliométrie ont commencé à repenser leurs méthodes, leur identité de recherche et leurs relations avec la principale base de données commerciale en adoptant la notion plus large d'« études scientifiques quantitatives ».

Cet article est publié sur ce site web et simultanément sous la forme d'un article wikipédia mis à jour de manière indépendante.

Bibliométrie

Langlais, Pierre-Carl CC BY 4.0 publié le 1 juin 2024

ARTICLE



A synthesized document co-citation network of the F dataset along with cluster labels and overlays of main paths of direct citations (red lines) and core references (yellow lines)

CC BY 4.0 , via Wikimedia Commons

Chaomei Chen & Min Song, Visualizing a field of research: A methodology of systematic scientometric reviews (<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0223994>)

La bibliométrie se définit comme l'utilisation de méthodes statistiques pour l'analyse de livres, d'articles et d'autres publications, en particulier dans le domaine scientifique. Ses méthodes sont fréquemment employées au sein

des bibliothèques et dans les sciences de l'information. La bibliométrie est étroitement associée à la scientométrie, qui consiste à analyser des métriques et des indicateurs scientifiques, au point que les deux concepts sont fortement imbriqués.

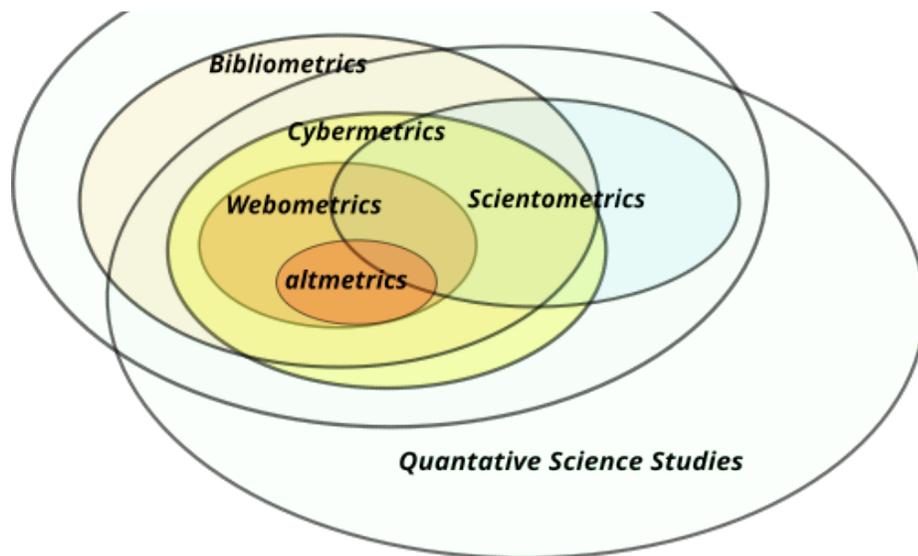
Les premières études bibliométriques apparurent à la fin du XIXe siècle. Elles connurent un développement important après la Seconde Guerre mondiale, mues par un contexte de « crise des périodiques » et par les nouvelles possibilités techniques offertes par l'informatique. Au début des années 1960, le Science Citation Index d'Eugene Garfield et l'analyse des réseaux de citations de Derek John de Solla Price ont jeté les bases d'un programme de recherche structuré sur la bibliométrie.

L'analyse des citations est une méthode bibliométrique courante qui repose sur la construction d'un graphe de citations, à savoir une représentation graphique ou en réseau des citations partagées entre documents. De nombreux domaines de recherche utilisent des méthodes bibliométriques pour étudier plusieurs types d'impacts : celui de leur domaine, celui d'un ensemble de chercheurs, celui d'un article particulier ou pour identifier les articles ayant un impact spécifique dans un domaine précis. Les outils bibliométriques sont amplement utilisés en linguistique descriptive, pour développer des thésaurus et pour évaluer l'usage par les lecteurs. Au-delà de l'utilisation scientifique spécialisée, les moteurs de recherche populaires sur le Web, comme l'algorithme PageRank de Google, s'inspirent très fortement des principes et méthodes bibliométriques.

L'émergence du Web et du mouvement de la science ouverte a progressivement transformé la définition et l'objet de la « bibliométrie ». Dans les années 2010, les infrastructures historiques de gestion des données de citations telles que Web of Science ou Scopus commencèrent à subir la concurrence de nouvelles initiatives plus ouvertes. Le Manifeste de Leyde pour la mesure de la recherche (2015) a ouvert un vaste débat sur l'usage et la transparence des métriques.

Définition





Definitions of the different field associated with bibliometrics

CC BY 4.0 , via Wikimedia Commons

Alexander Doria

La première utilisation du terme *bibliométrie* est attribuée à Paul Otlet en 1934.² Il est défini comme « la mesure de tous les aspects liés à la publication et à la lecture de livres et de documents ». ³ La version anglicisée « bibliometrics » fut initialement employée par Alan Pritchard dans un article publié en 1969 intitulé « Statistical Bibliography or Bibliometrics? »⁴, avec la définition suivante : « l'application des mathématiques et des méthodes statistiques aux livres et aux autres moyens de communication ». Le terme « bibliométrie » fut conçu pour remplacer l'expression « bibliographie statistique », principale appellation jusqu'alors employée dans les publications spécialisées : pour Alan Pritchard, « bibliographie statistique » était trop « ambigu » et ne précisait pas clairement l'objet principal de l'étude.⁵

Le concept de bibliométrie « souligne l'aspect matériel de l'entreprise : compter les livres, les articles, les publications, les citations ». ⁶ En théorie, la bibliométrie est un domaine distinct de la scientométrie (du russe « naukometriya »), ⁷ qui repose sur l'analyse d'indicateurs non bibliographiques de l'activité scientifique. En pratique, les études bibliométriques et scientométriques utilisent généralement des sources et des méthodes similaires, car les données de citations sont devenues la principale norme d'évaluation scientifique quantitative au milieu du XXe siècle : « Dans la mesure où les techniques bibliométriques sont appliquées à la littérature

scientifique et technique, les domaines de la scientométrie et de la bibliométrie se superposent considérablement. ». ⁸ Le développement du Web et l'élargissement de l'approche bibliométrique à la production non scientifique ont entraîné l'apparition de dénominations plus larges dans les années 1990 et 2000 : infométrie, webométrie et cybermétrie. ⁹ L'adoption de ces termes est restée marginale, car ils recourent partiellement des pratiques scientifiques existantes, comme la recherche d'informations.

Historique

Selon la définition adoptée, on rencontre des travaux, études et recherches scientifiques à caractère bibliométrique dès le XIIe siècle sous la forme d'index juifs. ¹⁰

Premières expérimentations (1880-1914)

L'analyse bibliométrique est apparue au tournant du XXe siècle. ¹¹ ¹² ¹³

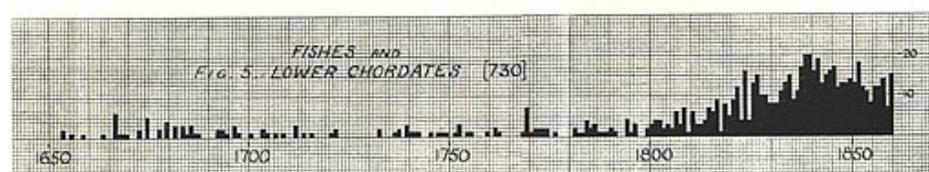
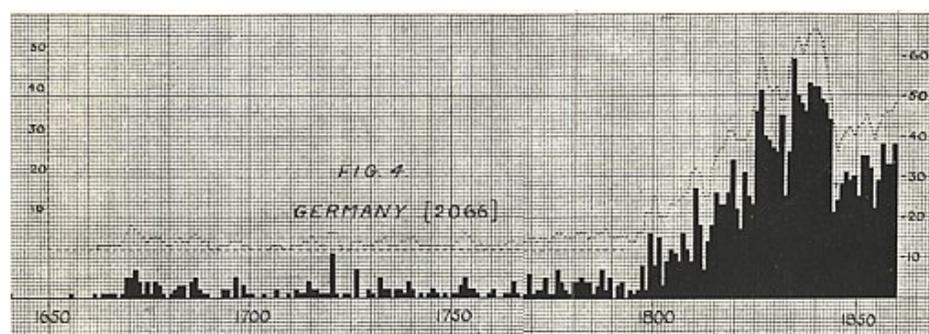
¹⁴ Le phénomène précède de plusieurs décennies la première mention du concept de bibliométrie. D'autres appellations étaient plus répandues : l'expression « statistiques bibliographiques » s'employait couramment après 1920 et resta d'usage jusqu'à la fin des années 1960. ¹⁵ Les premières études statistiques sur les métadonnées scientifiques étaient motivées par la hausse significative des productions de la recherche et le développement parallèle des services d'indexation de bases de données qui facilitaient l'accès à ces informations. ¹⁶ Une ébauche d'index des citations fut appliquée à la jurisprudence dans les années 1860, et le Shepard's Citations de 1873 (l'exemple le plus célèbre) inspirera directement le Science Citation Index un siècle plus tard. ¹⁷

L'émergence des sciences sociales a inspiré de nouvelles recherches spéculatives sur la science de la science et la possibilité d'étudier la science elle-même en tant qu'objet scientifique : « L'idée que les activités sociales, y compris la science, peuvent être réduites à des lois quantitatives, tout comme la trajectoire d'un boulet de canon et les révolutions des corps célestes, remonte à la sociologie positiviste d'Auguste Comte, William Ogburn et Herbert Spencer. ». ¹⁸ L'analyse bibliométrique n'a pas été conçue comme

une étude à part entière, mais comme l'une des méthodes disponibles pour étudier quantitativement l'activité scientifique dans différents domaines de recherche : l'histoire des sciences (Histoire des sciences et des savants depuis deux siècles, d'Alphonse de Candolle en 1885, *The History of Comparative Anatomy, a Statistical Analysis of the Literature*, de Francis Joseph Cole et Nellie B. Eales en 1917), la bibliographie (*The Theory of National and International Bibliography*, de Francis Burbury Campbell en 1896) et la sociologie de la connaissance scientifique (*Statistics of American Psychologists*, de James McKeen Cattell en 1903).

Les premiers travaux de bibliométrie et de scientométrie n'étaient pas simplement descriptifs. Ils exprimaient des points de vue normatifs sur ce que devrait être la science et ses axes possibles de progression. Mesurer les performances des chercheurs, des institutions scientifiques ou de pays entiers était un objectif majeur. ¹⁹ L'analyse statistique de James McKeen Cattell a préparé le chemin à une évaluation à grande échelle des chercheurs américains sur fond d'eugénisme : ce sera *American Men of Science* (1906), « avec son système d'évaluation étonnamment simpliste d'astérisques rattachés à des entrées en proportion de l'éminence estimée du savant désigné ». ²⁰

Développement des statistiques bibliographiques (1910-1945)



An early example of bibliometric analysis of a scientific corpus on anatomy by Francis Joseph Cole and Nellie B. Eales in 1917. with a breakdown by topics and countries.

U.S. Public domain

Francis Joseph Cole, Nellie Barbara Eales, *The History of Comparative Anatomy*, 1917

Après 1910, l'approche bibliométrique, plutôt que d'être une énième méthode quantitative, a pris une place prépondérante dans plusieurs études sur les performances scientifiques.²¹ En 1917, Francis Joseph Cole et Nellie B. Eales militaient pour la valeur statistique primaire des publications, car une publication « est une œuvre isolée et définie, elle est permanente, accessible et peut être jugée ; dans la plupart des cas, il n'est pas difficile de déterminer quand, où et par qui elle a été réalisée et de reporter les résultats sur du papier quadrillé ». ²² Cinq années plus tard, Edward Wyndham Hulme approfondit l'argument en expliquant que les publications pourraient être considérées comme l'étalon de mesure de toute une civilisation : « Si la civilisation n'est que le produit de l'esprit humain opérant sur la plateforme mouvante de son environnement, nous pouvons prétendre que la bibliographie n'est pas seulement un pilier dans la structure de l'édifice, mais qu'elle peut servir à mesurer les forces variables auxquelles cette structure est continuellement soumise. ». ²³ Cette aspiration à prendre en compte les publications n'eut qu'un impact limité : jusque dans les années 1970, l'évaluation nationale et internationale des activités scientifiques « se désintéressait des indicateurs bibliométriques » jugés trop simplistes, au profit de mesures sociologiques et économiques. ²⁴

La nouvelle importance accordée aux publications scientifiques en tant que mesure de la connaissance et les difficultés des bibliothèques à gérer le flux croissant de périodiques académiques ont entraîné le développement des premiers index de citations. ²⁵ En 1927, P. Gross et E. M. Gross compilèrent les 3 633 références citées par le *Journal of the American Chemical Society* au cours de l'année 1926 et classèrent les revues en fonction du niveau de citation. Les deux auteurs ont créé un ensemble d'outils et de méthodes encore utilisés aujourd'hui par les moteurs de recherche universitaires, notamment l'attribution d'un bonus aux citations récentes, car « la tendance actuelle plutôt que les performances passées d'une revue devrait être la première des considérations ». ²⁶ Pourtant, l'environnement scientifique mesuré était sensiblement différent : L'allemand, plutôt que l'anglais, était de loin la langue principale de la chimie, avec plus de 50 % du total des références.

REFERENCES. 27

Au cours de la même période, les algorithmes fondamentaux, les métriques et les méthodes de la bibliométrie furent pour la première fois identifiés dans le cadre de projets distincts, ²⁸ la plupart étant liés aux inégalités structurelles de la production scientifique. En 1926, Alfred Lotka présenta sa loi de productivité à partir d'une analyse des publications dans *Chemical Abstracts* et *Geschichtstafeln der Physik* : le nombre d'auteurs produisant n contributions est égal au nombre $1/n^2$ d'auteurs n'ayant produit qu'une seule publication. ²⁹ En 1934, le bibliothécaire en chef du Science Museum de Londres, Samuel Bradford, développa la « loi de la diffusion » à partir de son expérience de l'indexation bibliographique : le rendement de la recherche de références dans les revues scientifiques est exponentiellement décroissant, car il faut consulter de plus en plus d'ouvrages pour trouver des travaux pertinents. D'après leurs contempteurs, les lois de Lotka et Bradford étaient loin d'être universelles et mettaient plutôt en évidence une relation approximative de type loi de puissance obtenue au moyen d'équations à la précision trompeuse. ³⁰

Crise des périodiques, numérisation et index des citations (1945-1960)

Après la Seconde Guerre mondiale, les difficultés croissantes de gestion et d'accès aux publications scientifiques ont abouti à une véritable « crise des périodiques » : les revues ne pouvaient pas suivre l'accélération de la production scientifique stimulée par les projets de Big Science. ³¹ ³² La question est devenue politiquement sensible après le lancement réussi de Spoutnik en 1957 : « La crise du Spoutnik a transformé le problème du contrôle bibliographique rencontré par les bibliothécaires en crise nationale de l'information. ». ³³ Dans un contexte de changements rapides et spectaculaires, le domaine émergent de la bibliométrie était lié à des réformes à grande échelle de l'édition universitaire et à des visions presque utopiques de l'avenir scientifique. En 1934, Paul Otlet s'appuya sur le concept de bibliométrie (ou « bibliologie ») pour lancer un projet ambitieux visant à quantifier l'impact des textes sur la société. Contrairement à la définition restrictive de la bibliométrie qui s'imposera après les années 1960, celle d'Otlet dépassait le cadre de l'édition scientifique, la publication n'y étant même pas fondamentale : elle consistait à « réduire les textes à des éléments

atomiques, ou idées, qu'il situait dans les différents paragraphes (alinéa, verset, articulet) formant un livre ». ³⁴ En 1939, John Desmond Bernal imagina un réseau d'archives scientifiques auquel la Royal Society s'intéressa brièvement en 1948 : « L'article scientifique envoyé à l'office central de publication, après approbation par un comité de lecture, serait microfilmé et une sorte de système d'impression à la demande serait ensuite mis en place. ». ³⁵ Sans qu'il utilise pour autant le concept de bibliométrie, l'influence de J D Bernal fut formatrice pour des grandes figures du domaine telles que Derek John de Solla Price.

Les technologies informatiques naissantes furent immédiatement considérées comme une solution potentielle pour rendre lisibles et consultables un plus grand nombre de productions scientifiques. Au cours des années 1950 et 1960, une vague d'expériences spontanées dans les technologies d'indexation entraîna le développement rapide de concepts majeurs de la recherche informatique. ³⁶ En 1957, Hans Peter Luhn, ingénieur chez IBM, fit autorité avec un paradigme d'analyse statistique de la fréquence des mots, selon lequel « la communication d'idées au moyen de mots repose sur une probabilité statistique ». ³⁷ De plus, la traduction automatique de travaux scientifiques non anglophones a fortement contribué à la recherche fondamentale sur le traitement en langage naturel des références bibliographiques. En effet, à l'époque, un grand nombre de publications scientifiques n'étaient pas encore disponibles en anglais, en particulier celles provenant du bloc soviétique. Des membres influents de la Fondation nationale pour la science tels que Joshua Lederberg plaidèrent pour la création du SCITEL, un « système d'information centralisé » partiellement inspiré des principes de John Desmond Bernal. Dans un premier temps, ce système devait coexister avec les revues imprimées avant de les remplacer progressivement du fait de son efficacité. ³⁸ Dans le plan présenté par Joshua Lederberg à Eugene Garfield en novembre 1961, un référentiel centralisé devait indexer jusqu'à un million d'articles scientifiques par an. Au-delà de la recherche en texte intégral, l'infrastructure devait indexer les citations et d'autres métadonnées, mais aussi effectuer la traduction automatique des articles en langue étrangère. ³⁹

Le premier prototype opérationnel de système de recherche en ligne développé en 1963 par Doug Engelhart et Charles Bourne à l'Institut de

recherche de Stanford prouva la faisabilité du concept, bien que fortement limité par des problèmes de mémoire : il ne pouvait pas indexer plus de 10 000 mots d'un petit nombre de documents. ⁴⁰ Les premières infrastructures informatiques scientifiques se concentraient sur certains domaines de recherche, par exemple MEDLINE pour la médecine, NASA/RECON pour l'ingénierie spatiale ou OCLC Worldcat pour les collections de bibliothèques : « Parmi les anciens systèmes de recherche en ligne, la plupart donnaient accès à une base de données bibliographiques, les autres utilisaient un fichier contenant d'autres types d'informations comme des articles d'encyclopédie, des données d'inventaire ou des composés chimiques ». ⁴¹ La focalisation exclusive sur l'analyse de texte s'est avérée restrictive à mesure de la croissance des collections numérisées : une requête pouvait donner un grand nombre de résultats, et il était difficile d'en évaluer la pertinence et l'exactitude. ⁴²

La crise des périodiques et les limites des technologies de recherche dans les index ont motivé le développement d'outils bibliométriques et de grands index de citations comme le Science Citation Index d'Eugene Garfield. Celui-ci travaillait initialement sur l'analyse automatisée de textes. Contrairement aux investigations en cours largement axées sur les relations sémantiques internes, E Garfield souligna « l'importance du métatexte dans l'analyse du discours », comme les phrases d'introduction et les références bibliographiques. ⁴³ Les formes secondaires de production scientifique, telles que les revues de la littérature et les notes bibliographiques, sont devenues essentielles à la vision d'E Garfield, comme elles l'étaient déjà à la perception des archives scientifiques de John Desmond Bernal. ⁴⁴ En 1953, E Garfield se consacra définitivement à l'analyse des citations : dans une lettre privée à William C. Adair, le vice-président de l'éditeur de l'index Shepard's Citations, « il proposa une solution éprouvée au problème de l'indexation automatique, à savoir « shepardiser » la littérature biomédicale, afin de démêler l'écheveau de son contenu en suivant le fil des liens de citation, comme le faisait Shepard's Citations avec les décisions de justice ».

⁴⁵ En 1955, E Garfield publia son article fondateur « Citation Indexes for Science », qui définit les grandes lignes du Science Citation Index et influença considérablement l'évolution de la bibliométrie. ⁴⁶ L'index général des citations préfiguré par F Garfield était au départ un simple composant de

statistique préfigurée par le Camera étant au départ un simple composant de l'ambitieux projet de Joshua Lederberg visant à informatiser la littérature scientifique. ⁴⁷ Faute de financement, ce projet n'a jamais vu le jour. ⁴⁸ En 1963, Eugene Garfield créa l'ISI (Institute for Scientific Information) afin de rentabiliser les projets initialement esquissés avec Joshua Lederberg.

Réductionnisme bibliométrique, métriques et structuration du champ de recherche (1960-1990)

Le domaine de la bibliométrie s'est constitué parallèlement au développement du Science Citation Index, qui allait devenir son infrastructure fondamentale et sa source de données : ⁴⁹ « Si le début du XXe siècle vit apparaître des méthodes nécessaires à l'évaluation de la recherche, le milieu de ce même siècle fut caractérisé par le développement d'institutions qui ont motivé et facilité cette évaluation. ». ⁵⁰ Parmi les influences significatives de ce domaine naissant, il y a John Desmond Bernal et Paul Otlet, mais aussi Robert K. Merton et sa sociologie des sciences réinterprétée de manière anti-éthique : l'effet Matthieu, c'est-à-dire l'intérêt croissant pour des chercheurs déjà connus, n'est plus considéré comme une dérive mais comme une caractéristique de la science normale. ⁵¹

Un disciple de J D Bernal, l'historien des sciences britannique Derek John de Solla Price, influença ô combien la formation disciplinaire de la bibliométrie : avec « la publication de *Science Since Babylon* (1961), *Little Science, Big Science* (1963) et *Networks of Scientific Papers* (1965) par Derek Price, la scientométrie disposait déjà d'une solide boîte à outils empirique et conceptuelle ». ⁵² D J Price était un partisan du réductionnisme bibliométrique. ⁵³ Comme Francis Joseph Cole et Nellie B. Eales en 1917, il soutenait que la publication est la meilleure norme possible pour articuler une étude quantitative de la science : elle « s'apparente à un assemblage de briques (…) que l'habileté et l'artifice ont transformé pour l'éternité en un édifice intellectuel reposant sur des fondations primitives ». ⁵⁴ D J Price a amplifié cette approche réductionniste en limitant à son tour le vaste ensemble de données bibliographiques existantes aux données de citations. Comme E Garfield, D J Price accepte le postulat d'une inégalité structurelle de la production scientifique. Selon lui, une minorité de chercheurs génère une grande partie des publications et un nombre encore plus réduit possède

un impact réellement mesurable sur les recherches ultérieures (avec seulement 2 % des articles cités au moins 4 fois à l'époque).⁵⁵ Malgré l'essor inédit de la science après-guerre, D J Price pressentait l'existence d'un collège invisible de scientifiques d'élite qui, comme au temps de Robert Boyle, se chargeait des travaux les plus importants.⁵⁶ D J Price avait conscience des relations de pouvoir qui permettent la domination d'une telle élite, mais il subsistait une ambiguïté fondamentale dans les études bibliométriques : elles mettaient en évidence la concentration de l'édition universitaire et du prestige, et pourtant elles créaient des outils, des modèles et des métriques normalisant les inégalités en place.⁵⁷ La position centrale du Scientific Citation Index a amplifié cet effet performatif. À la fin des années 1960, Eugene Garfield formula une loi de concentration qui, en fait, réinterprétait la loi de diffusion de Samuel Bradford, avec une différence majeure : tandis que Bradford s'exprimait dans la perspective d'un projet de recherche spécifique, Garfield généralisait sa loi à l'ensemble des publications scientifiques : « Toutes disciplines confondues, la littérature scientifique représente un noyau d'au plus 1 000, voire 500 revues ». Cette loi justifiait également la limitation pratique de l'index des citations à un sous-ensemble de revues majeures, avec l'hypothèse implicite qu'une extension à des revues de second rang affecterait quantitativement les résultats.⁵⁸

Plutôt que de simplement observer les tendances et les modèles structurels, la bibliométrie tend à les amplifier et à les stratifier davantage : « En poussant leur logique jusqu'au bout, les index de citations d'E Garfield auraient dépeint une littérature scientifique stratifiée, produite par (…) un groupuscule de revues internationales incontournables de haute qualité et détenues par un nombre toujours plus faible de multinationales dominant le marché mondial de l'information. ».⁵⁹

Sous l'impulsion de Garfield et Price, la bibliométrie est devenue à la fois un domaine de recherche et un terrain d'essai pour l'évaluation quantitative de la recherche. Ce dernier point n'était pas un objectif majeur du Science Citation Index, mais il s'est développé progressivement : le célèbre « facteur d'impact » fut imaginé dans les années 1960 par E Garfield et Irving Sher pour sélectionner le noyau dur des revues devant figurer dans Current Contents et le Science Citation Index. Il n'a été publié régulièrement qu'à partir de 1975.⁶⁰ La métrique elle-même est un rapport très simple entre le

nombre total de citations de la revue pendant l'année écoulée et sa productivité lors des deux dernières années, ceci afin de pondérer la prolificité de certaines publications. ⁶¹ Par exemple, en 2017 le facteur d'impact de Nature était de 41,577 : ⁶²

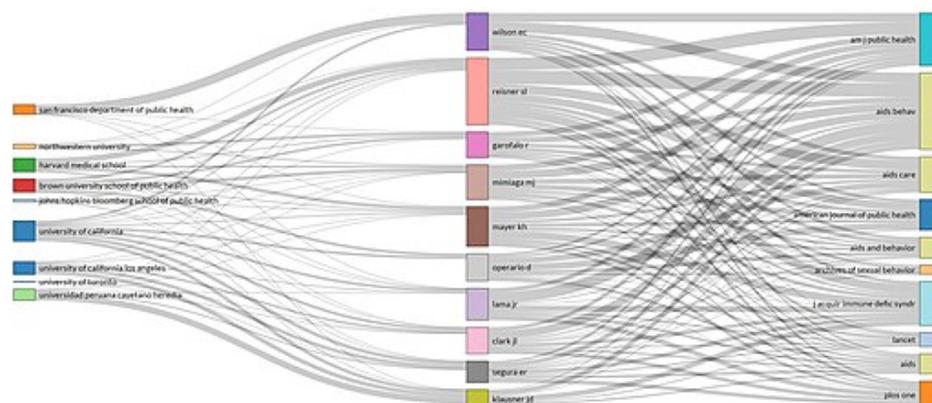
(tableau à insérer)

$$IF_{2017} = \frac{\text{Citations}_{2017}}{\text{Publications}_{2016} + \text{Publications}_{2015}} = \frac{74090}{880 + 902} = 41.577.$$

Il est probable que la simplicité du facteur d'impact ait considérablement favorisé son adoption par les institutions scientifiques, les revues, les bailleurs de fonds et les évaluateurs : « Aucune des versions revisitées ou des substituts du FI de l'ISI n'a dépassé le cadre d'acceptation de ses propres partisans, peut-être parce que les prétendues alternatives étaient bien plus délicates à interpréter que l'original ». ⁶³

En plus de ces mesures simplifiées, E Garfield continua à soutenir et financer la recherche fondamentale en histoire et sociologie des sciences. Sorti en 1964, l'ouvrage *The Use of Citation Data in Writing the History of Science* compile plusieurs études de cas expérimentales s'appuyant sur le réseau de citations du Science Citation Index, avec notamment une reconstitution quantitative de la découverte de l'ADN. ⁶⁴ L'intérêt d'E Garfield pour ce domaine persista bien après le rachat de l'Index par Thomson Reuters : en 2001, il dévoila HistCite, un logiciel d'« historiographie algorithmique » créé en collaboration avec Alexander Pudovkin et Vladimir S. Istomin. ⁶⁵

Le virage du Web (1990-...)



A three-fields plot that shows the relationship of authors between their institutions and cited sources within the

A three-nodes plot that shows the relationship of authors between their institutions and cited sources within the retrieved literature. Created with Biblioshiny - an online bibliometrics data visualisation tool.

CC BY 4.0

Liam Gary Arnall, Anuj Kapilashrami & Margaret Sampson (2021) Visualizing patterns and gaps in transgender sexual and reproductive health: A bibliometric and content analysis of literature (1990–2020), *International Journal of Transgender Health*, DOI: 10.1080/26895269.2021.1997691

Le développement du World Wide Web et la révolution numérique ont eu un impact complexe sur la bibliométrie.

Le Web lui-même et certains de ses composants clés (comme les moteurs de recherche) sont un héritage partiel des théories de la bibliométrie. Dans sa forme originale, il dérivait d'une infrastructure scientifique bibliographique nommée ENQUIRE commandée à Tim Berners-Lee par le CERN pour les besoins spécifiques de la physique des particules. La structure d'ENQUIRE était plus proche d'un réseau interne de données : elle connectait des « nœuds » qui « pouvaient se référer à une personne, un module logiciel, etc. et pouvaient être raccordés entre eux par diverses relations telles que « fait », « inclut », « décrit », etc. ». ⁶⁶ Le partage des données et de leur documentation était l'un des principaux objectifs de la présentation initiale du World Wide Web lorsque le projet fut dévoilé en août 1991 : « Le projet WWW fut lancé pour permettre aux physiciens des particules de partager des données, des informations et de la documentation. L'extension du Web à d'autres domaines et la mise en place de serveurs passerelles pour d'autres données nous intéressent fortement. ». ⁶⁷ Le Web a rapidement supplanté les autres infrastructures en ligne, même lorsqu'elles étaient plus avancées sur le plan informatique. ⁶⁸ La valeur centrale de l'hyperlien dans la conception du Web semble valider les intuitions des figures fondatrices de la bibliométrie : l'apparition du World Wide Web au milieu des années 1990 a donné plus de corps au rêve d'Eugene Garfield sur l'emploi des citations. Dans le réseau mondial de l'hypertexte, non seulement la référence bibliographique est une des formes possibles de l'hyperlien dans la version électronique d'un article scientifique, mais le Web lui-même présente une structure citative, les liens entre les pages Web étant dans la forme similaires aux citations bibliographiques. ⁶⁹ Ainsi, les principales technologies de communication des algorithmes de recherche de Google se sont enrichies de notions bibliométriques : « Le concept de pertinence basé sur les citations appliqué au réseau d'hyperliens entre les pages Web allait révolutionner la façon dont les moteurs de recherche permettent aux utilisateurs de trouver

rapidement des documents utiles dans l'univers anarchique de l'information numérique. ». ⁷⁰

Si le Web a élargi l'influence intellectuelle de la bibliométrie bien au-delà de la recherche scientifique spécialisée, il a aussi fait voler en éclats les principes essentiels du domaine. Contrairement aux visées utopiques de Bernal et Otlet qui l'ont en partie inspiré, le Science Citation Index fut toujours pensé comme une infrastructure fermée, non seulement du point de vue de ses utilisateurs, mais aussi de l'index de collection : la conclusion logique de la théorie du collège invisible de Price et de la loi de concentration de Garfield était de se concentrer exclusivement sur un ensemble limité de revues scientifiques majeures. Avec l'expansion rapide du Web, de nombreuses formes de publications (notamment les préprints ou prépublications), d'activités scientifiques et de communautés sont soudainement devenues visibles et ont souligné par contraste les limites de la bibliométrie appliquée. ⁷¹ L'autre aspect fondamental du réductionnisme bibliométrique, à savoir la focalisation exclusive sur les citations, fut de plus en plus fragilisé par la multiplication des sources de données alternatives et l'accès sans précédent aux corpus de textes intégraux qui ont permis de relancer l'analyse sémantique à grande échelle dont E Garfield avait une vision avant-gardiste au début des années 1950 : « Les liens seuls, tout comme les citations bibliographiques seules, ne semblent donc pas suffire pour cerner les modèles de communication critiques sur le Web. Dans les années à venir, leur analyse statistique suivra probablement le même chemin que l'analyse des citations, s'associant avec réussite à d'autres points de vue qualitatifs et quantitatifs émergents sur le paysage du Web. ». ⁷²

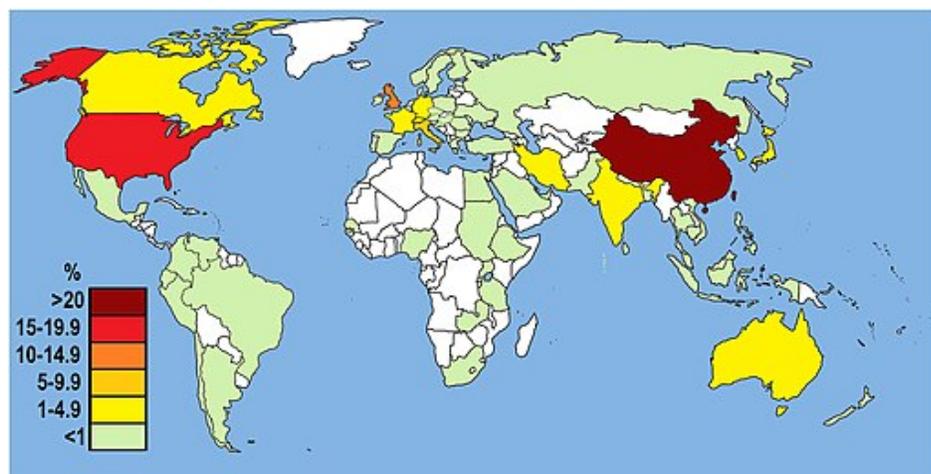
Les liens étroits entre la bibliométrie et les fournisseurs commerciaux d'indicateurs et de données de citations se sont distendus depuis les années 1990. Les principaux éditeurs scientifiques ont diversifié leurs activités au-delà de la publication et sont passés « d'une activité de fourniture de contenu à une activité d'analyse de données ». ⁷³ En 2019, Elsevier a acquis ou construit un vaste portefeuille de plateformes, d'outils, de bases de données et d'indicateurs couvrant en totalité les aspects et les étapes de la recherche scientifique : « Le plus grand fournisseur de revues académiques est également chargé d'évaluer et de valider la qualité et l'impact de la recherche (Pure Plum Analytics Sci Val) d'identifier les experts universitaires pour le

« ... et, dans certains cas, de rassembler les experts universitaires pour le compte des employeurs potentiels (Expert Lookup), de gérer les plateformes collaboratives de recherche (SSRN, Hivebench, Mendeley) et les outils de localisation des financements (PlumX, Mendeley, Sci Val), mais aussi de contrôler les plateformes permettant d'analyser et de stocker les données des chercheurs (Hivebench, Mendeley). ».⁷⁴ Les métriques et les indicateurs sont des éléments clés de cette intégration verticale : « L'évolution d'Elsevier vers un service d'aide à la décision basé sur des métriques est aussi un moyen d'accroître son influence sur l'ensemble du processus de production de connaissances et de monétiser encore plus son accumulation disproportionnée de contenus. ».⁷⁵ Les marchés émergents de la publication et des données scientifiques ont été comparés au modèle économique des réseaux sociaux, des moteurs de recherche et d'autres formes de capitalisme de plateforme.^{76 77 78} L'accès au contenu est certes gratuit, mais il est indirectement rémunéré par l'extraction et la surveillance des données.⁷⁹ En 2020, Rafael Ball prévoyait un avenir sombre pour les bibliomètres, dont les recherches allaient contribuer à l'instauration d'une forme très invasive d'« économie de la surveillance » : les scientifiques « se verraient attribuer une série de notes qui non seulement donnerait une image plus complète de leurs performances académiques, mais aussi de leur perception, de leur comportement, de leur attitude, de leur apparence et de leur crédibilité (subjective) »⁸⁰ En Chine, ce type d'analyse des données personnelles est déjà mis en œuvre et utilisé simultanément comme levier d'incitation et de sanction. ».⁸⁰

Le Manifeste de Leyde pour la mesure de la recherche (2015) a mis en évidence le fossé croissant entre les fournisseurs commerciaux de métriques scientifiques et les communautés bibliométriques. Les signataires ont souligné les dommages sociaux potentiels d'une évaluation et d'une surveillance sans contrôle basées sur les métriques : « En tant que scientomètres, spécialistes des sciences sociales et administrateurs de la recherche, nous avons observé avec une inquiétude croissante le détournement généralisé des indicateurs d'évaluation des performances scientifiques ».⁸¹ Plusieurs réformes structurelles de la recherche bibliométrique et de l'évaluation de la recherche sont proposées, notamment un recours accru à l'évaluation qualitative et l'adoption d'une approche « ouverte, transparente et simple » de la collecte de données. ».⁸²

« ouverte, transparente et simple » de la collecte de données. ⁸² Le Manifeste de Leyde a suscité un important débat parmi les acteurs de la bibliométrie, de la scientométrie et de l'infométrie, certaines critiques estimant que l'élaboration de métriques quantitatives ne rend pas responsable des utilisations abusives au sein des plateformes commerciales et de l'évaluation de la recherche. ⁸³

Bibliométrie et science ouverte



Distribution of corresponding authors of scholarly articles on SARS-CoV-2 and COVID-19 between January and March 2020

CC BY 4.0

Nowakowska, Joanna; Sobocińska, Joanna; Lewicki, Mateusz; Lemańska, Żaneta; Rzymiski, Piotr (2020). "When science goes viral: The research response during three months of the COVID-19 outbreak". *Biomedicine & Pharmacotherapy*: 110451. doi:10.1016/j.biopha.2020.110451.

Le mouvement de la science ouverte a été reconnu comme la plus importante révolution vécue par la bibliométrie depuis son apparition dans les années 1960. ⁸⁴ ⁸⁵ Le libre partage sur le Web d'un grand nombre de productions scientifiques a affecté la bibliométrie à tous les niveaux : la définition et la collecte des données, l'infrastructure et les métriques.

Avant sa cristallisation autour du Science Citation Index et des théories réductionnistes de Derek John de Solla Price, la bibliométrie avait été largement influencée par des projets utopiques d'amélioration du partage de la connaissance au-delà des communautés académiques spécialisées. Les réseaux scientifiques imaginés par Paul Otlet ou John Desmond Bernal ont connu un regain de pertinence avec le développement du Web :

« L'inspiration philosophique des pionniers dans la poursuite des pistes de

recherche susmentionnées s'est toutefois progressivement estompée (…) Alors qu'en définitive l'apport de Bernal trouverait un prolongement idéal dans le mouvement de libre accès, la machine à citations mise en branle par Garfield et Small a suscité la prolifération d'études sectorielles de nature fondamentalement empirique. ».⁸⁶

Des altermétriques aux métriques ouvertes

Voir aussi : Altermétriques

À ses débuts, le mouvement de la science ouverte a partiellement adopté les outils standards de la bibliométrie et de l'évaluation quantitative : « L'absence de référence aux métadonnées dans les principales déclarations sur le libre accès (Budapest, Berlin, Bethesda) a entraîné une situation paradoxale (…) c'est par l'utilisation de Web of Science que les défenseurs du libre accès ont souhaité montrer à quel point l'accessibilité était plus avantageuse que les articles payants en matière de citations. ».⁸⁷ Après 2000, une importante littérature bibliométrique a été consacrée au bénéfice des publications en libre accès pour générer des citations.⁸⁸

À la fin des années 2000, certains indicateurs comme le facteur d'impact et d'autres furent estimés de plus en plus responsables d'un verrouillage systémique de sources prestigieuses non accessibles. Des figures clés du mouvement de la science ouverte tels que Stevan Harnad appelèrent à la création d'une « scientométrie du libre accès » qui « tirerait parti de la richesse des indicateurs d'usage et d'impact résultant de la multiplication des archives numériques en ligne, en texte intégral et en libre accès ».⁸⁹ La science ouverte s'étant popularisée au-delà des cercles universitaires, de nouveaux indicateurs devraient chercher à « mesurer les impacts sociétaux à plus grande échelle de la recherche scientifique ».⁹⁰

Le concept d'altermétrique fut introduit en 2009 par Cameron Neylon et Shirly Wu en tant qu'indicateur au niveau de l'article.⁹¹ Contrairement aux principaux indicateurs centrés sur les revues (facteur d'impact) ou, plus récemment, sur le chercheur lui-même (indice h), les indicateurs au niveau de l'article autorisent un suivi individuel des publications en circulation : « (Un) article autrefois rangé sur une étagère se trouve désormais dans Mendeley, CiteULike ou Zotero, où il est possible de le voir et de le comptabiliser ».⁹² Il est ainsi davantage compatible avec la diversité des stratégies de publication

est un avantage compatible avec la diversité des stratégies de publication, qui caractérise la science ouverte : les prépublications, les rapports et même les résultats non textuels tels que les référentiels ou les logiciels peuvent aussi disposer d'indicateurs associés. ⁹³ Dans leur proposition de recherche initiale, Neylon et Wu favorisaient l'utilisation de données provenant de logiciels de gestion de références tels que Zotero ou Mendeley. ⁹⁴ Le concept d'altérométrie a évolué pour venir englober les données extraites « d'applications de médias sociaux, comme les blogues, Twitter, ResearchGate et Mendeley ». ⁹⁵ Les sources des médias sociaux montrèrent davantage de fiabilité à long terme, alors que des outils universitaires spécialisés tels que Mendeley finirent par intégrer l'écosystème développé en propre par les plus grands éditeurs scientifiques. Les principales altérométries apparues dans les années 2010 sont Altmetric.com, PlumX et ImpactStory.

Avec le glissement de sens des altérométries, le débat sur l'impact positif des indicateurs a évolué pour les redéfinir dans le cadre d'un écosystème de science ouverte : « Les discussions sur l'usage abusif des indicateurs et leur interprétation viennent les placer au centre des pratiques de la science ouverte ». ⁹⁶ Alors que les altérométries furent initialement conçues pour les publications scientifiques ouvertes et pour faciliter leur circulation bien au-delà des cercles universitaires, leur compatibilité avec les nouvelles exigences en matière d'indicateurs ouverts a été remise en question : les données des réseaux sociaux, en particulier, sont loin d'être transparentes et facilement accessibles. ⁹⁷ ⁹⁸ En 2016, Ulrich Herb a publié une évaluation systématique des indicateurs de publication clés selon les principes de la science ouverte. Il en a conclu que « ni les indicateurs d'impact basés sur les citations ni les altérométries ne peuvent être qualifiés d'ouverts. Tous manquent de fondement scientifique, de transparence et de vérifiabilité. ». ⁹⁹

| Metric | Provider | Sources | Free access | Data access | Open Data | Open Software |
|------------------------------|------------------|-------------------------------------|--------------------|--------------------|------------------|----------------------|
| <u>Journal Impact Factor</u> | <u>Clarivate</u> | Citations (<u>Web of Science</u>) | No | No | No | No |
| <u>SCImago</u> | | | | | | |

| | | | | | | |
|-------------------------------|-----------------------|----------------------------|---------|-----|------------------------------------|-----|
| <u>Journal Rank</u> | <u>Elsevier</u> | Citations (Scopus) | Yes | Yes | No | No |
| <u>SNIP</u> | Elsevier | Citations (Scopus) | Yes | Yes | No | No |
| <u>Eigenfactor</u> | Clarivate | Citations (Web of Science) | Yes | No | No | No |
| <u>Google Journal Ranking</u> | <u>Google Scholar</u> | Citations (Google) | Yes | No | No | No |
| <u>h-index</u> | Clarivate | Citations (Web of Science) | No | No | No | No |
| h-index | Elsevier | Citations (Scopus) | No | No | No | No |
| h-index | Google Scholar | Citations (Google) | Yes | No | No | No |
| <u>Altmetrics</u> | <u>PLUM Analytics</u> | Varied sources | No | No | No | No |
| Altmetrics | Altmetric (Macmillan) | Varied sources | Partial | No | No | No |
| Altmetrics | <u>PLOS</u> | Varied sources | Yes | Yes | Partial (include proprietary data) | Yes |
| Altmetrics | <u>ImpactStory</u> | Varied sources | Yes | Yes | Partial (include proprietary data) | Yes |

| | | | | | | |
|--------------------|--------------------------------------|----------------|-----|-----|-----|-----|
| Open Citation Data | OpenCitations Corpus | Varied sources | Yes | Yes | Yes | Yes |
|--------------------|--------------------------------------|----------------|-----|-----|-----|-----|

Évaluation des principales métriques et altermétriques de publication au regard des principes de la science ouverte ¹⁰⁰

Ulrich Herb a établi un programme alternatif pour les métriques ouvertes dont le développement reste à venir. ¹⁰¹ ¹⁰² Les principaux critères étaient :

- Une large sélection d'éléments de publication (articles de revues, livres, jeux de données, logiciels) correspondant aux pratiques d'écriture et de lecture des communautés scientifiques. ¹⁰³
- Des sources de données pleinement documentées. ¹⁰⁴
- Un processus transparent et reproductible pour le calcul des métriques et des autres indices. ¹⁰⁵
- Des logiciels ouverts. ¹⁰⁶
- La promotion d'usages réflexifs et interprétatifs des métriques pour éviter leur utilisation abusive lors des évaluations quantitatives. ¹⁰⁷

Cette définition a été mise en œuvre dans certains programmes de recherche, par exemple le projet ROSI (Reference Implementation For Open Scientometric Indicators). ¹⁰⁸ En 2017, le groupe d'experts de la Commission européenne sur les altermétriques a élargi le programme de métriques ouvertes d'Ulrich Herb dans le cadre d'un nouveau concept, les Métriques de nouvelle génération. Ces métriques devraient être gérées par des « infrastructures de données ouvertes, transparentes et liées ». ¹⁰⁹ Le groupe d'experts souligne que tout n'est pas à mesurer et que toutes les métriques ne sont pas pertinentes : « Mesurons ce qui compte : la prochaine génération de métriques devrait commencer par les qualités et les impacts que les sociétés européennes valorisent le plus et pour lesquels elles ont le plus besoin d'indices, plutôt que par les plus faciles à collecter et à mesurer. ». ¹¹⁰

Infrastructures pour les données de citations ouvertes

Voir aussi : Initiative for Open Citations

Jusqu'aux années 2010, l'impact du mouvement de la science ouverte était fortement limité aux publications scientifiques : il « a plutôt négligé l'importance des structures sociales et des contraintes systémiques dans la

conception de nouvelles formes d'infrastructures du savoir ». ¹¹¹ En 1997, Robert D. Cameron appela au développement d'une base de données de citations ouvertes qui révolutionnerait les conditions de la communication scientifique : « Imaginez une base universelle de données bibliographiques et de citations reliant tous les travaux scientifiques jamais écrits – quel que soit leur mode de publication – à tous les travaux qui les citent. Imaginez une telle base de données de citations librement accessible sur Internet et quotidiennement mise à jour avec tous les nouveaux travaux publiés le jour même, y compris les articles sortis dans des revues traditionnelles et électroniques, les documents de conférence, les thèses, les rapports techniques, les documents de travail et les prépublications. ». ¹¹² Malgré le développement d'index spécifiquement dédiés aux œuvres en libre accès, comme CiteSeer, aucune grande alternative au Science Citation Index n'a vu le jour en format ouvert. La collecte des données de citations est restée dominée par de grandes structures commerciales telles que Web of Science, le descendant direct du Scientific Citation Index. L'écosystème émergent des ressources ouvertes est ainsi resté en périphérie des réseaux universitaires : « Le patrimoine commun des ressources n'est pas régi ou géré par l'initiative actuelle des communs universitaires. Il n'existe pas d'infrastructure matérielle dédiée et, malgré l'émergence d'une communauté, il n'y a pas d'adhésion formelle. ». ¹¹³

Depuis 2015, les infrastructures, plateformes et revues de science ouverte ont convergé vers la création de communs académiques numériques, de plus en plus structurés autour d'un écosystème de services partagés, et des normes ont émergé à travers les interdépendances entre les infrastructures. Ce mouvement découle d'une attitude de plus en plus critique à l'égard des principales bases de données privées. En 2012, la Déclaration de San-Francisco sur l'évaluation de la recherche (DORA) appelait à « cesser d'utiliser les facteurs d'impact des revues dans les décisions de financement, de recrutement et de promotion ». ¹¹⁴ Le Manifeste de Leyde pour la mesure de la recherche (2015) a encouragé le développement de collections de données « ouvertes, transparentes et simples ». ¹¹⁵

La collaboration entre les acteurs universitaires et non universitaires engagés collectivement dans la création et l'administration des communs de la connaissance s'est avérée déterminante dans la mise en place d'une nouvelle

infrastructure consacrée aux données de citations ouvertes. Depuis 2010, des chercheurs ont rassemblé un jeu de données de citations ouvertes, l'Open Citation Corpus, à partir de différentes sources en libre accès (dont PLOS et Pubmed).¹¹⁶ Cette collection formait la charpente de l'Initiative for Open Citations (I4OC), lancée en 2017 à la suite de problèmes d'accessibilité des données rencontrés par Wikidata, un projet Wikimedia. Une conférence de Dario Taraborelli, responsable de la recherche à la Fondation Wikimedia, a montré que seulement 1 % des articles figurant dans Crossref possédaient des métadonnées de citations librement accessibles et que les références stockées dans Wikidata ne permettaient pas d'inclure le très vaste segment des données non libres. Quand Elsevier a finalement rejoint l'initiative en janvier 2021, plus de la moitié des articles enregistrés sont devenus directement accessibles.¹¹⁷

Depuis 2021, le catalogue OpenAlex constitue une infrastructure ouverte majeure pour les métadonnées scientifiques. Initialement créé pour remplacer la base de données MAG (Microsoft Academic Graph) qui cessait d'être mise à jour, OpenAlex indexait en 2022 quelque 209 millions de travaux universitaires provenant de 213 millions d'auteurs ainsi que la liste des institutions, lieux et concepts associés dans un graphe de connaissances intégré au Web sémantique (et à Wikidata).¹¹⁸ En raison de sa couverture étendue et de l'important volume de données récupéré dans MAG, OpenAlex « semble être au moins aussi pertinent que MAG pour l'analyse bibliométrique des années de publication antérieures à 2021 ». ¹¹⁹ En 2023, une étude sur la couverture des revues de données dans les index scientifiques a révélé que, tout comme Dimensions, OpenAlex « jouit d'un avantage conséquent par rapport aux deux bases de données plus traditionnelles, WoS et Scopus » ¹²⁰ et qu'il est, dans l'ensemble, particulièrement performant pour indexer les publications hors revues, par exemple les livres ¹²¹ et les productions des chercheurs non occidentaux. ¹²² L'ouverture des données scientifiques fut un sujet de débat majeur dans la communauté de la bibliométrie et de la scientométrie, avec des conséquences sociales et intellectuelles très variées. En 2019, le comité scientifique du Journal of Infometrics a démissionné en bloc et créé une nouvelle revue en libre accès, Quantitative Science Studies. La revue était publiée par Elsevier depuis 2007, et les membres du comité se montraient de

plus en plus critiques vis-à-vis du manque d'avancées dans le libre partage des données de citations ouvertes : « Notre domaine dépend de métadonnées scientifiques de haute qualité. Pour que notre science soit plus robuste et reproductible, ces données doivent être aussi ouvertes que possible. C'est pourquoi notre comité éditorial a très mal vécu le refus d'Elsevier de participer à l'Initiative for Open Citations (I4OC). ».¹²³

Bibliométrie sans évaluation : passage aux études scientifiques quantitatives

La disponibilité sans précédent d'un large éventail de productions scientifiques (publications, données, logiciels, conférences, évaluations, etc.) a donné un tournant plus radical au projet bibliométrique. Pour les nouveaux travaux alternatifs fidèles aux principes de la science ouverte, la bibliométrie telle que définie par Garfield et Price dans les années 1960 doit être repensée. La présélection d'un corpus limité de revues majeures ne semble ni nécessaire ni appropriée. En 2019, les promoteurs du projet Matilda « ne [voulai]ent pas simplement « ouvrir » l'information fermée existante, mais redonner leur juste place à tous les contenus académiques jusqu'alors exclus de ces outils en appliquant la philosophie « tous les textes naissent égaux » ». ¹²⁴ Ils espéraient « redéfinir les outils bibliométriques en tant que technologie » en se focalisant sur l'exploration et la cartographie du corpus scientifique. ¹²⁵

Les problématiques d'inclusivité et d'approche critique des inégalités structurelles dans la science sont aujourd'hui plus présentes dans la scientométrie et la bibliométrie, en particulier la question des disparités entre les sexes. ¹²⁶ ¹²⁷ ¹²⁸ Après 2020, l'un des débats les plus vifs dans ce domaine ¹²⁹ concernait la réception d'une étude sur la parité hommes-femmes en physique fondamentale. ¹³⁰

À cause de l'évolution structurelle des définitions de la bibliométrie, de la scientométrie et de l'infométrie, il a fallu trouver d'autres appellations. Le concept d'études scientifiques quantitatives fut introduit à la fin des années 2000 en vue de permettre une nouvelle évaluation critique des analyses bibliométriques traditionnelles. ¹³¹ Il a pris une importance plus marquée à la fin des années 2010. Après avoir quitté Elsevier, les rédacteurs de *Journal of Infometrics* ont adopté cet intitulé comme titre de leur revue (*Quantitative*

inométrie ont adopté cet intitulé comme titre de leur revue (Quantitative Science Studies). Le premier éditorial abolissait toute référence aux métriques et indiquait vouloir inclure plus largement les recherches quantitative et qualitative sur la science de la science :

Nous espérons que les personnes qui se reconnaissent dans des appellations telles que la scientométrie, la science de la science et les métasciences se sentiront chez elles dans QSS. Nous reconnaissons également la diversité des disciplines pour lesquelles la science est un objet d'étude : nous nous réjouissons d'accueillir au sein de notre revue les historiens, les philosophes et les sociologues des sciences. Et malgré notre référence aux études quantitatives, nous sommes ouverts à toutes les perspectives épistémologiques. Les études scientifiques quantitatives ne peuvent pas évoluer en vase clos : il n'y a pas de travail empirique solide sans intégration des théories et connaissances venues de toutes les métasciences. ¹³²

Plan

Définition

Historique

- Premières expérimentations (1880-1914)
- Développement des statistiques bibliographiques (1910-1945)
- Crise des périodiques, numérisation et index des citations (1945-1960)
- Réductionnisme bibliométrique, métriques et structuration du champ de recherche (1960-1990)
- Le virage du Web (1990-...)

Bibliométrie et science ouverte

- Des altermétriques aux métriques ouvertes
- Infrastructures pour les données de citations ouvertes
- Bibliométrie sans évaluation : passage aux études scientifiques quantitatives

Notes

1. Hutchins et al. 2019.
2. Otlet 1934.
3. Rousseau 2014.
4. Pritchard 1969.
5. Hertzfel 2003, p. 288.
6. Bellis 2009, p. 3.
7. Bellis 2009, p. 12.
8. Bellis 2009, p. 3.
9. Bellis 2009, p. 4.
10. Jovanovic 2012.
11. Bellis 2009, p. 2.
12. Godin 2006.
13. Danesh & Mardani-Nejad 2020.
14. Hertzfel 2003, p. 292.
15. Hertzfel 2003, p. 292.
16. Bellis 2009, p. 6.
17. Bellis 2009, p. 23.
18. Bellis 2009, p. 1.
19. Bellis 2009, p. 6.
20. Bellis 2009, p. 2.
21. Bellis 2009, p. 7.
22. Cole & Eales 1917, p. 578.
23. Hulme 1923, p. 43.

24. Bellis 2009, p. 14.
25. Bellis 2009, p. 9.
26. Gross & Gross 1927, p. 387.
27. Gross & Gross 1927, p. 388.
28. Bellis 2009, p. 75.
29. Bellis 2009, p. 92.
30. Bellis 2009, p. 99.
31. Wouters 1999, p. 61.
32. Bellis 2009, p. 12.
33. Wouters 1999, p. 62.
34. Bellis 2009, p. 10.
35. Bellis 2009, p. 52.
36. Bellis 2009, p. 27.
37. Luhn 1957.
38. Wouters 1999, p. 60.
39. Wouters 1999, p. 64.
40. Bourne & Hahn 2003, p. 16.
41. Bourne & Hahn 2003, p. 12.
42. Bellis 2009, p. 30.
43. Bellis 2009, p. 34.
44. Bellis 2009, p. 53.
45. Bellis 2009, p. 35.
46. Bellis 2009, p. 35.
47. Bellis 2009, p. 36.
48. Bellis 2009, p. 37.
49. Bellis 2009, p. 49.
50. Suaimoto & Larivière 2018 n 8

50. Sugimoto & Larivière 2016, p. 9.

51. Bellis 2009, p. 57.

52. Bellis 2009, p. 49.

53. Bellis 2009, p. 62.

54. Price 1975, p. 162.

55. Bellis 2009, p. 65.

56. Bellis 2009, p. 67.

57. Bellis 2009, p. 67.

58. Bellis 2009, p. 103.

59. Bellis 2009, p. 104.

60. Bellis 2009, p. 187.

61. Bellis 2009, p. 186.

62. Nature. 2017 Journal Citation Reports (PDF) (rapport). Web of Science (Science ed.). Thomson Reuters. 2018.(À vérifier).

63. Bellis 2009, p. 194.

64. Bellis 2009, p. 153.

65. Bellis 2009, p. 173.

66. Hogan 2014, p. 20.

67. Tim Berners-Lee, « Qualifiers on Hypertext Links », courriel envoyé le 6 août 1991 au groupe alt.hypertext.

68. Star & Ruhleder 1996, p. 131.

69. Bellis 2009, p. 285.

70. Bellis 2009, pp. 31–32.

71. Bellis 2009, p. 289.

72. Bellis 2009, p. 322.

73. Aspesi et al. 2019, p. 5.

74. Chen et al. 2019, par. 25.

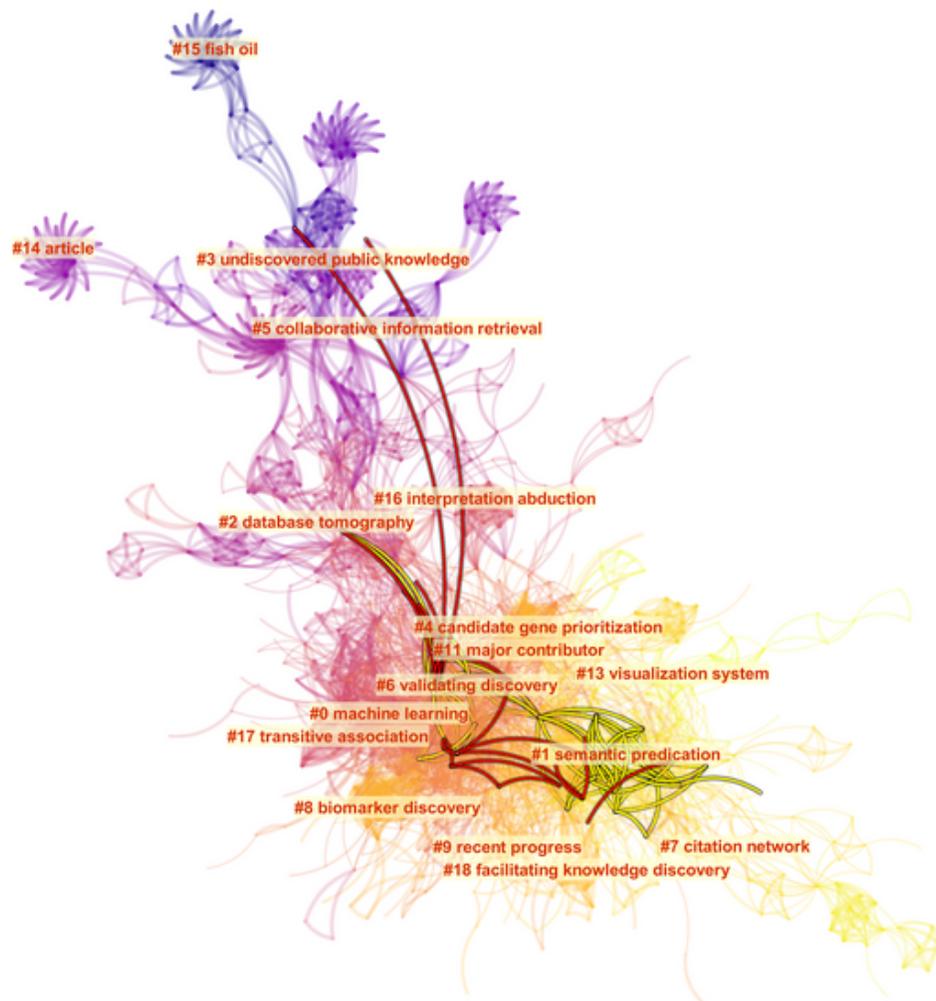
75. Chen et al. 2019. par. 29.

75. Chen et al. 2019, p. 201.
76. Moore 2019, p. 156.
77. Chen et al. 2019.
78. Wainwright & Bervejillo 2021.
79. Wainwright & Bervejillo 2021, p. 211.
80. Ball 2020, p. 504.
81. Hicks et al. 2015, p. 430.
82. Hicks et al. 2015, p. 430.
83. David & Frangopol 2015.
84. Bellis 2009, p. 288 sq.
85. Heck 2020.
86. Bellis 2009, p. 336.
87. Torny, Capelli & Danjean 2019, p. 1.
88. Sugimoto & Larivière 2018, p. 70.
89. Bellis 2009, p. 300.
90. Wilsdon et al. 2017, p. 9.
91. Neylon & Wu 2009.
92. Priem et al. 2011, p. 3.
93. Wilsdon et al. 2017, p. 9.
94. Neylon & Wu 2009.
95. Wilsdon et al. 2017, p. 9.
96. Heck 2020, p. 513.
97. Bornmann & Haunschild 2016.
98. Tunger & Meier 2020.
99. Herb 2016, p. 60.
100. Herb 2016.
101. Herb 2012, p. 29.

102. Herb 2016, p. 70.
103. Herb 2016, p. 70.
104. Herb 2016, p. 70.
105. Herb 2016, p. 70.
106. Herb 2016, p. 70.
107. Herb 2016, p. 70.
108. Hauschke et al. 2018.
109. Wilsdon et al. 2017, p. 15.
110. Wilsdon et al. 2017, p. 15.
111. Okune et al. 2018, p. 13.
112. Cameron 1997.
113. Bosman et al. 2018, p. 19.
114. Wilsdon et al. 2017, p. 7.
115. Hicks et al. 2015, p. 430.
116. Peroni et al. 2015.
117. Waltman, Ludo (22 décembre 2020). QA about Elsevier's decision to open its citations. Leiden Madrics. Universiteit Leiden. Consultation le 11 juin 2021.
118. Priem, Piwowar & Orr 2022, p. 1-2.
119. Scheidsteger & Haunschild 2022, p. 10.
120. Jiao, Li & Fang 2023, p. 14.
121. Laakso 2023, p. 166.
122. Akbaritabar, Theile & Zagheni 2023.
123. Waltman et al. 2020, p. 1.
124. Torny, Capelli & Danjean 2019, p. 2.
125. Torny, Capelli & Danjean 2019, p. 7.
126. Larivière et al. 2013.

127. Iorny, Capelli & Danjean 2019.
128. Chary et al. 2021.
129. Gingras 2022.
130. Strumia 2021.
131. Glänzel 2008.
132. Waltman et al. 2020.

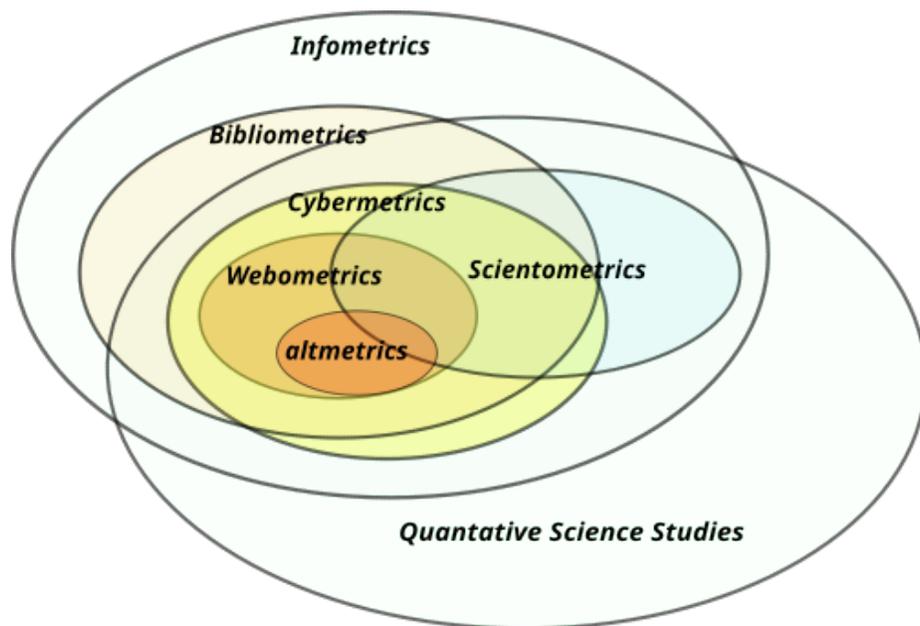
Ressources



A synthesized document co-citation network of the F dataset along with cluster labels and overlays of main paths of direct citations (red lines) and core references (yellow lines)

CC BY 4.0 , via Wikimedia Commons

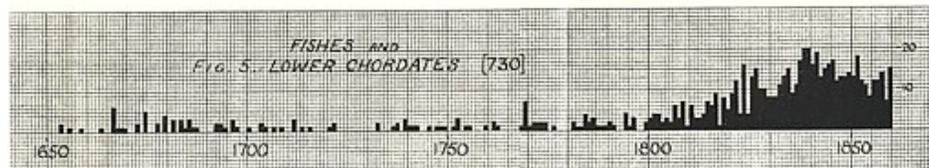
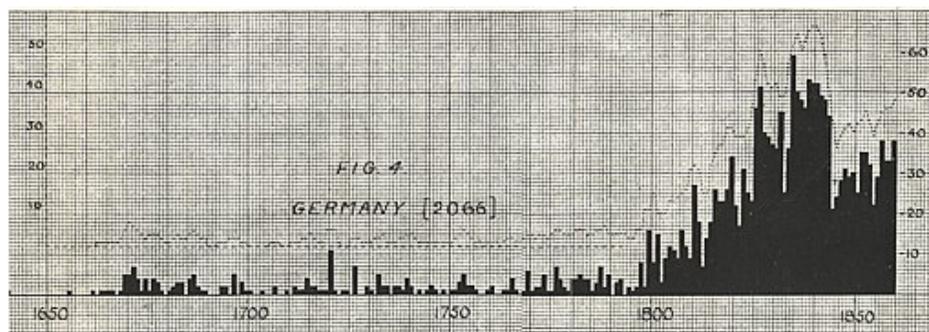
Chaomei Chen & Min Song, Visualizing a field of research: A methodology of systematic scientometric reviews (<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0223994>)



Definitions of the different field associated with bibliometrics

CC BY 4.0 , via Wikimedia Commons

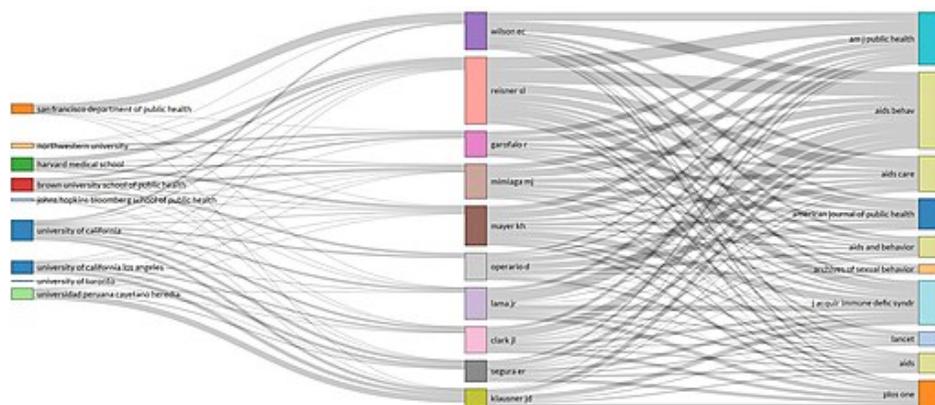
Alexander Doria



An early example of bibliometric analysis of a scientific corpus on anatomy by Francis Joseph Cole and Nellie B. Eales in 1917, with a breakdown by topics and countries.

U.S. Public domain

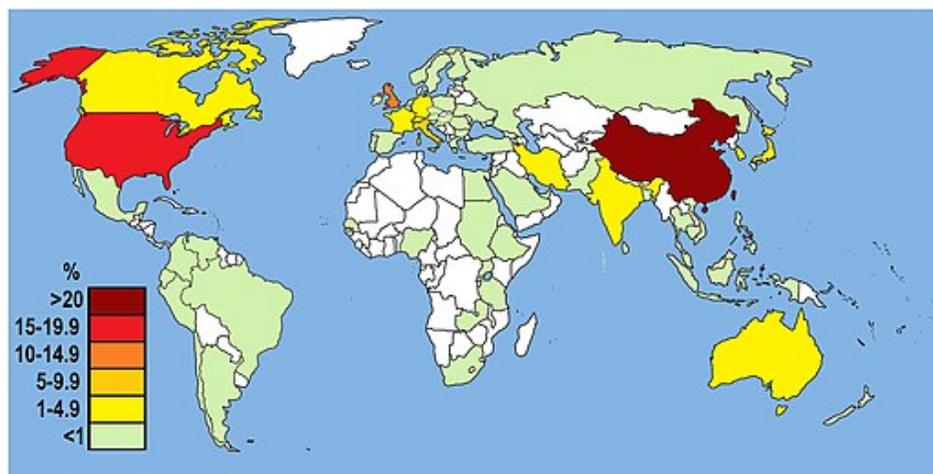
Francis Joseph Cole, Nellie Barbara Eales, The History of Comparative Anatomy, 1917



A three-fields plot that shows the relationship of authors between their institutions and cited sources within the retrieved literature. Created with Biblioshiny - an online bibliometrics data visualisation tool.

CC BY 4.0

Liam Gary Arnall, Anuj Kapilashrami & Margaret Sampson (2021) Visualizing patterns and gaps in transgender sexual and reproductive health: A bibliometric and content analysis of literature (1990–2020), *International Journal of Transgender Health*, DOI: 10.1080/26895269.2021.1997691



Distribution of corresponding authors of scholarly articles on SARS-CoV-2 and COVID-19 between January and March 2020

CC BY 4.0

Nowakowska, Joanna; Sobocińska, Joanna; Lewicki, Mateusz; Lemańska, Żaneta; Rzymiski, Piotr (2020). "When science goes viral: The research response during three months of the COVID-19 outbreak". *Biomedicine & Pharmacotherapy*: 110451. doi:10.1016/j.biopha.2020.110451.