

Vincent Larivière et Cassidy R. Sugimoto

Mesurer la science



Les Presses de l'Université de Montréal

LIBRE ACCÈS

PDF et ePub gratuits en ligne
www.pum.umontreal.ca

MESURER LA SCIENCE

MESURER LA SCIENCE

Vincent Larivière
et Cassidy R. Sugimoto

*Traduit de l'anglais par
Sophie Chisogne*

Les Presses de l'Université de Montréal

Mise en pages: Yolande Martel

*Catalogage avant publication de Bibliothèque et Archives nationales du Québec
et Bibliothèque et Archives Canada*

Larivière, Vincent

[Measuring research. Français]

Mesurer la science / Vincent Larivière, Cassidy R. Sugimoto.

Traduction de: Measuring research: what everyone needs to know /

Cassidy R. Sugimoto and Vincent Larivière

Comprend des références bibliographiques.

Publié en formats imprimé(s) et électronique(s).

ISBN 978-2-7606-3951-5

ISBN 978-2-7606-3952-2 (PDF)

ISBN 978-2-7606-3953-9 (EPUB)

1. Bibliométrie. 2. Recherche – Évaluation – Méthodes statistiques.

3. Index de citations. 4. Communication savante – Innovations.

I. Sugimoto, Cassidy R, auteur. II. Sugimoto, Cassidy R. Measuring research. Français.

III. Titre. IV. Titre: Measuring research. Français.

Z669.8.S8414 2018

020.72'7

C2018-941604-1

C2018-941605-X

Dépôt légal: 3^e trimestre 2018

Bibliothèque et Archives nationales du Québec

© Les Presses de l'Université de Montréal pour la version française, 2018.

© Oxford University Press pour l'anglais, 2018.

ISBN 978-0-19-064012-5

L'édition originale de ce livre est en anglais. La traduction a été faite en accord avec Oxford University Press. Les Presses de l'Université de Montréal sont les seules responsables de cette traduction et dérogent Oxford University Press de toute responsabilité en cas d'erreur, d'omission ou d'ambiguïté résultant de cette traduction.

Les Presses de l'Université de Montréal remercient de leur soutien financier le Conseil des arts du Canada et la Société de développement des entreprises culturelles du Québec (SODEC).

Financé par le
gouvernement
du Canada

Canada

IMPRIMÉ AU CANADA

*Ce livre est dédié à Noah, Zoé, Anastasia et Madeleine,
qui sont extraordinaires au-delà de toute mesure.*

1

LES BASES

La recherche est une activité sociale complexe. Une pluralité d'acteurs s'y adonne dans des milieux très variés, et elle englobe un grand nombre d'activités distinctes. La production de nouvelles connaissances repose sur un nombre incalculable d'heures de travail en laboratoire et en bibliothèque, de rencontres avec des étudiants et des collaborateurs et d'interactions, officielles ou non, avec d'autres chercheurs ou avec le public. De telles activités sont difficiles, voire impossibles à quantifier directement. La mesure de la recherche, souvent appelée *scientométrie*, a donc comme tâche première de transformer ces activités en unités quantifiables, appelées *indicateurs*. Ce livre a pour objet de décrire ces indicateurs – la manière dont ils sont formulés, leurs forces et leurs faiblesses, de même que l'interprétation et l'usage qu'on devrait en faire.

La mesure de l'activité de recherche concerne trois aspects distincts : ses intrants, ses extrants et ses impacts. Les indicateurs quantifiant les éléments d'entrée tiennent compte des ressources et autres investissements qui alimentent les activités scientifiques. Figurent ainsi parmi les intrants souvent évalués la taille et les diverses caractéristiques de la main-d'œuvre scientifique, ainsi

que le financement accordé à la recherche. Les indicateurs qui quantifient les éléments de sortie s'intéressent plutôt au savoir produit grâce aux intrants : publications savantes et brevets d'invention en sont les exemples les plus courants. Quant aux indicateurs d'impact, ils évaluent les répercussions des travaux de recherche sur la communauté savante et la société en général. Documenter ces différents aspects n'est pas une mince tâche, car beaucoup de leurs effets sont à peine perceptibles ou ne se réalisent que partiellement. Plus encore, nombre de données ne se prêtent pas à une normalisation qui permettrait de les comparer selon différents niveaux d'agrégation. Par conséquent, les mesures actuelles de la recherche reposent sur des activités qui laissent des traces, et des données que l'on peut normaliser et agréger, qui sont contemporaines entre elles. Cette forme de quantification de la recherche, appelée généralement *scientométrie*, regroupe tous les moyens que nous employons pour recueillir et analyser quantitativement la science – en particulier, sur la production et le contenu de documents savants et sur les signes d'une réception des produits issus du travail intellectuel.

Les indicateurs de la science sont bâtis à partir de données provenant de sources très diversifiées. Les intrants sont obtenus le plus souvent au moyen de questionnaires administrés régionalement ou internationalement. Ces données portent le sceau d'agences gouvernementales et scientifiques phares, comme les conseils de recherche nationaux ou les organisations internationales (l'Organisation de coopération et de développement économique [OCDE] ou l'Organisation des Nations Unies pour l'éducation, la science et la culture [UNESCO], p. ex.). Ces données, généralement compilées au moyen de questionnaires, sont souvent enrichies par des informations sociodémographiques, et la stabilité de ces organismes en assure une mise à jour régulière. Les indicateurs de production et d'impact, quant à eux, sont généralement dérivés de banques de données conçues à des fins

de documentation (les index de citations et autres bases de données bibliographiques, p. ex.). Ce sont des sources d'information indirectes, qui ont l'avantage de fournir d'emblée des données uniformes, donc faciles à agréger, plus détaillées que les questionnaires à portée restreinte, et plus stables dans le temps. Malgré des inconvénients liés à l'étendue du contenu (discipline, langue ou pays, p. ex.), les banques de données sur les citations sont devenues la façon la plus commune de mesurer la recherche – une tâche qu'elles effectuent principalement à l'aide des méthodes de la bibliométrie. Le présent ouvrage décrit ces sources de données bibliométriques, de même que les méthodes émergentes en scientométrie.

On accuse souvent la scientométrie de promouvoir la culture de la publication à tout prix (*publish or perish*), de dénaturer la science et d'en fausser grossièrement les objectifs. Nous examinerons ces critiques tour à tour, en soulignant les limites des données et de leur compilation – limites dont il faut tenir compte si l'on entend utiliser correctement les indicateurs. La scientométrie soulève également des inquiétudes sur le plan économique, compte tenu de la concentration de grandes entreprises qui contrôlent la collecte de données (Clarivate Analytics, Elsevier et Google, entre autres) et des coûts élevés de la plupart des outils de mesure. Nous aborderons donc la question de la science entendue comme un bien commun et discuterons des contraintes liées à l'analyse de données non liées appartenant à des entreprises privées. L'élitisme, l'accès difficile aux données, de même que l'interopérabilité de celles-ci seront étudiés dans le contexte des transformations qui perturbent actuellement l'infrastructure traditionnelle.

Depuis quelques dizaines d'années, plusieurs manuels de pointe sur la compilation et l'interprétation des indicateurs bibliométriques ont été publiés – la plupart par des spécialistes du milieu. Il existe aussi un ensemble de manuels techniques publiés par des

organismes internationaux telle l'OCDE qui proposent des normes pour la scientométrie fondée sur les sondages et la formulation d'indicateurs d'intrants technologiques. Cet ouvrage-ci est conçu différemment : il propose de comprendre la scientométrie par une présentation vulgarisée de ses fondements historiques et de ses concepts et termes clés, et par des interprétations, des critiques et des recommandations formulées à son sujet.

Pourquoi mesurer la recherche ?

De nombreux domaines des sciences sociales et humaines ont la science comme objet d'étude. Chacun de ces domaines applique ses propres méthodes à l'étude de la science, que ce soit dans une approche historique, sociologique ou philosophique. La scientométrie nous offre un point de vue exceptionnel sur les éléments relationnels et contextuels de la science. Elle nous permet de mieux comprendre les modes de production et de diffusion des connaissances savantes ainsi que leur évolution au fil du temps, dans les différentes disciplines et en différents endroits du monde. En outre, elle permet de compléter d'autres approches méthodologiques.

La bibliométrie est particulièrement utile quand la quantité de données est trop importante pour qu'un humain puisse les traiter efficacement. Par exemple, si un membre d'un comité de lecture peut produire un jugement sur un document ou un petit ensemble de documents, et si un auteur peut, sans trop de risque de se tromper, établir le nombre de ses publications, la production d'une grande institution ou d'un pays, elle, est plus difficile à évaluer sans données. Quant aux données relationnelles, comme les citations, elles sont presque impossibles à analyser manuellement, même si l'on se limite à la production d'un seul individu. Les mesures de la recherche sont donc le plus utiles à grande échelle,

quand il s'agit de mettre en lumière ce qu'une personne seule ou un petit groupe de chercheurs aurait du mal à observer.

La curiosité – composante de toute activité de recherche – est à l'origine des premiers travaux de recherche fondamentale en scientométrie. Les chercheurs voulaient mieux comprendre le système vaste et complexe de la science en mesurant ses intrants et extrants. Les travaux de Derek de Solla Price et d'autres précurseurs recouraient aux indicateurs de la recherche pour bâtir des théories sur l'activité scientifique, lesquelles servaient également à mettre à l'épreuve certaines hypothèses. Les philosophes, par exemple, s'intéressent depuis longtemps aux consensus et aux révolutions dans les sciences, phénomènes que l'on peut aborder par la quantification des produits de la recherche, l'analyse des thématiques de recherche et l'accumulation de citations. Les historiens, eux, débattent de la réception des idées scientifiques et de l'incidence de découvertes particulières sur les domaines de recherche. Les analyses de la documentation historique classique (archives, entrevues, etc.), renforcées par l'usage d'indicateurs bibliométriques, peuvent témoigner de la réception d'une découverte par les contemporains et de sa postérité. Les sociologues, enfin, ont intégré les mesures de la recherche à leurs analyses afin d'appuyer les théories sur les hiérarchies qui structurent le monde scientifique et les asymétries qui en résultent dans la distribution des récompenses. Pour toutes ces disciplines, le fait de quantifier la recherche, au moyen de la littérature savante principalement, éclaire la science d'un jour nouveau.

Les indicateurs servent par ailleurs d'outil réflexif pour les chercheurs. En appliquant à la science ses propres outils, la scientométrie nous donne les moyens d'observer et de calibrer le système scientifique. La collecte et l'analyse d'informations sur les activités des chercheurs peuvent aider les administrateurs et les responsables des politiques de la recherche à prendre des décisions éclairées. Les mesures permettent de stimuler la recherche,

de distribuer les ressources, de comprendre et de prédire les tendances. Elles peuvent également révéler les iniquités du système. Par exemple, les travaux sur les inégalités entre les sexes dans le domaine scientifique montrent l'asymétrie de la relation entre travail et récompense. Les résultats de ces travaux permettent de formuler des politiques pour améliorer les conditions de travail de tous et élargir la participation en recherche. En somme, ce livre porte sur les outils et les techniques qui permettent de telles observations empiriques, et sur le savoir contextuel nécessaire à leur interprétation.

À qui s'adresse ce livre ?

La ferveur sans précédent pour tout ce qui touche à l'évaluation de la recherche a mené à la multiplication des outils et des sources de données. Pendant longtemps, l'évaluation de la recherche était réservée à des experts, essentiellement des chercheurs actifs en scientométrie et des analystes spécialisés des gouvernements et des universités. Une tendance récente cherche cependant à rendre les données plus accessibles et à confier la mesure de la recherche à des responsables de l'administration et des politiques et à d'autres chercheurs. Bien que cela soit louable à maints égards, cela signifie aussi que ces nouveaux acteurs du système doivent pouvoir compiler et interpréter les indicateurs de recherche de manière pertinente. Or de nombreux outils éloignent l'analyste des données brutes, ce qui augmente les risques liés à la prise de décision fondée sur des algorithmes et autres boîtes noires dans les politiques sur la recherche. L'utilisation adéquate de ces outils exige une connaissance poussée des données disponibles (et de leurs limites), une compréhension solide de la façon dont les indicateurs sont formulés et compilés, et la capacité de les interpréter en fonction de leur contexte. Ceux-ci ne viennent pas dans un format unique applicable en toutes circonstances. Avec ce

livre, nous souhaitons outiller tous ceux qui exercent ou subissent l'évaluation de la recherche de telle sorte qu'ils puissent participer utilement à la mesure de l'activité scientifique et, dans le cas des chercheurs, qu'ils puissent mieux comprendre les évaluations dont ils pourraient être l'objet.

On pourrait défendre l'idée que toute personne qui connaît l'arithmétique devrait être capable de mesurer la recherche, au moins sommairement. En effet, bien que de nombreux indicateurs soient définis au moyen de formules complexes, la majorité de ceux qu'on applique à la recherche se résume à des décomptes de fréquence et à des classements fondés sur ces mesures. Ce livre décrit les aspects plus complexes associés à l'interprétation de telles données, dont certaines viennent d'une compréhension imparfaite de la discipline étudiée ou du champ scientifique dans son ensemble. Les chercheurs à qui l'on confie la tâche de comprendre leur propre travail ou celui de collègues du même domaine traiteront facilement le premier de ces problèmes: bien au fait des enjeux de leur discipline, ils sont suffisamment informés pour comprendre les limites des données et y trouver des solutions. Peu de chercheurs, cependant, sont formés pour comprendre la science dans son ensemble, en tant que système, et pour nuancer les variations de production et d'impact entre les disciplines.

Ce livre traite avant toute chose des particularités des données bibliométriques – de ce qui les distingue d'autres types de données, qui ont toutes leurs propres limites, qu'elles proviennent de sondages, de fichiers d'opérations informatiques ou de fouilles archéologiques. Les données bibliométriques peuvent nous révéler beaucoup d'informations sur le système de la recherche, à condition de les interpréter dans ce contexte particulier. Cela exige que nous comprenions non seulement leurs propriétés techniques et mathématiques, mais également les mécanismes sociologiques et économiques qui régissent le monde de la recherche.

De nombreux scientomètres soutiennent que seuls les spécialistes de ce domaine devraient évaluer la recherche, ce qui soulève deux objections. La première est qu'ils posent une dichotomie artificielle entre les personnes qui possèdent l'expertise bibliométrique et les « non-spécialistes ». Dans la pratique, en effet, la communauté scientifique se gouverne elle-même : l'évaluation par les pairs est l'un des piliers du système, et elle suppose la confiance envers les chercheurs, qui ont une connaissance interne du domaine analysé. Cependant, quand l'évaluation doit se produire en dehors du domaine de l'évaluateur, peu de chercheurs sont assez outillés pour tenir compte des nuances dans les données. L'expertise n'est donc pas absolue, elle s'échelonne par degrés : il n'existe pas de distinction simple entre experts et non-experts quand il s'agit de mesurer la recherche.

La seconde objection est que la scientométrie, tout au long de son histoire, s'est nourrie des contributions de chercheurs de toutes disciplines. Bien que la plupart des premiers scientomètres aient été des scientifiques cherchant à comprendre la structure de la science, la méthode a aussi attiré des scientifiques souhaitant décrire leur propre champ d'études. Cela a donné lieu à la fois à des analyses à grande échelle de l'évolution de la science et à des histoires iconographiques, centrées sur la réussite de grands personnages et sur les découvertes majeures. Les sociologues ont appliqué les théories sociales au système scientifique. Les physiciens ont intégré à ce mélange la science des réseaux et les économistes, les modèles économétriques. À l'heure actuelle, on crée des méthodes qui permettront aux individus de calculer leurs propres indicateurs. Devant la croissance de la demande pour des indicateurs et la diversification des utilisateurs, il devient nécessaire d'offrir un guide descriptif et critique à toutes les personnes qui entendent mesurer la recherche de façon responsable. C'est à elles que ce livre s'adresse.

Quels sont les fondements historiques de la scientométrie ?

Les origines de la scientométrie remontent à différentes activités de documentation – des bibliographies aux enquêtes sur la main-d'œuvre, en passant par les bottins de chercheurs – qui s'étalent dans le temps et l'espace. La scientométrie contemporaine, toutefois, trouve ses racines dans l'Europe et l'Amérique du Nord du XX^e siècle. Le début de ce siècle avait vu se consolider plusieurs des champs de recherche qui allaient fournir des méthodes indispensables à la mesure de la recherche, telles la statistique et la sociologie. Mais ce sont la bibliothéconomie et les sciences qui ont servi de précurseurs à la plupart des méthodes bibliométriques. En effet, afin de déterminer les acquisitions de périodiques qui devaient augmenter leurs collections, les bibliothécaires procédaient au comptage des documents cités par leurs chercheurs. Parce que les données ne pouvaient qu'être compilées manuellement, ces premières études se faisaient à petite échelle, en fonction d'objectifs de gestion des collections. Les méthodes reposaient largement sur les besoins des bibliothèques en recherche documentaire et en gestion des collections. Cette filiation – et la relation inextricable qui unit les publications savantes aux bibliothèques – a rattaché la bibliométrie, avec d'autres approches quantitatives de la mesure de la recherche, au domaine de la bibliothéconomie et des sciences de l'information, pour la plus grande part du siècle dernier.

Si le début du XX^e siècle a fourni bon nombre de méthodes nécessaires à la scientométrie, le milieu du siècle se caractérise quant à lui par l'établissement d'institutions qui allaient motiver et favoriser le développement de cette discipline. L'institutionnalisation du soutien à la recherche scientifique, par l'intermédiaire des conseils de recherche, a remodelé l'économie politique de l'activité savante. Parce que les agences gouvernementales ont commencé à financer des projets de recherche particuliers, celles-ci ont exigé des moyens de suivre les progrès de la recherche scientifique,

perçue comme une composante essentielle de la croissance économique. Les indicateurs scientométriques ont répondu à ce besoin en même temps qu'au regain de la concurrence internationale en matière de progrès scientifique. Parallèlement, la croissance exponentielle du volume des travaux savants appelait de nouvelles façons de cataloguer et de repérer les ouvrages scientifiques. Enfin, la fondation de l'Institute for Scientific Information, aux États-Unis – où fut créé le *Science Citation Index* – a été le dernier pilier de cette institutionnalisation ayant favorisé la conception d'indicateurs pour la science et l'essor d'un domaine consacré à la mesure de la recherche.

Le Britannique Derek de Solla Price, physicien et historien des sciences, a joué un rôle clé dans la cristallisation de la nouvelle discipline au début des années 1960. Il s'inspirait d'un collègue physicien, John Desmond Bernal, qui réclamait l'adoption d'une approche novatrice pour évaluer la production de nouvelles connaissances. Les travaux antérieurs en scientométrie avaient presque toujours été fonctionnels : ils visaient avant tout, on l'a vu, à améliorer pour les bibliothécaires la gestion des collections et la recherche documentaire. Price, lui, invitait la communauté universitaire à « appliquer à la science ses propres instruments¹ », soulignant la relation inextricable entre la science et la société et l'importance d'étudier la science d'un point de vue quantitatif. Price ne négligeait pas les aspects qualitatifs, mais remarquait qu'il faut être « suffisamment scientifique dans l'analyse d'un ensemble de régularités [...] avant d'entreprendre de traiter de façon humaniste les irrégularités qui se produisent parce que les humains sont des humains, et pas des machines² ». Il préconisait l'observation des principales constantes et des modèles de croissance, dans le but de mieux comprendre la nature de l'entreprise

1. Price, *Science et suprascience*. Paris, Fayard, 1972, p. 70.

2. Trad. libre de *Big Science, Little Science*.

scientifique. Son œuvre majeure, *Little Science, Big Science* (en français: *Science et suprascience*), évaluait notamment le volume, la vitesse de circulation et l'interaction de l'information scientifique, ainsi que les propriétés structurales, politiques et sociales du champ scientifique.

L'une de ses observations les plus fructueuses est que la science est essentiellement contemporaine – du fait de son taux de croissance exponentiel, la majorité de ses extrants sont le produit de notre époque. Le constat vaut pour différentes unités d'analyse: nombre de scientifiques, nombre de publications universitaires, quantité de données produites. La scientométrie, par conséquent, s'est avérée extrêmement utile pour explorer ce système scientifique vaste et complexe, permettant d'en analyser quantitativement toutes les catégories, d'étudier les relations entre les occurrences d'une variable et entre différentes variables, de cartographier son passé et de prédire son avenir.

L'Unité de recherche en science de l'information et scientométrie rattachée à la Bibliothèque de l'Académie hongroise des sciences fut l'un des premiers groupes de recherche indépendants en scientométrie. En 1978, le lancement de la revue *Scientometrics* par le chimiste hongrois Tibor Braun allait structurer plus encore le domaine. C'est ce périodique qui a établi l'usage du terme *scientométrie* – dérivé de *naukometriya*, un terme emprunté au mathématicien et philosophe russe V. V. Nalimov – au sens général d'étude quantitative de la science. Mais c'est le sous-titre de la revue – *An International Journal for all Quantitative Aspects of the Science of Science, Communication in Science and Science Policy* – qui révèle l'ampleur du champ qu'on entendait le voir couvrir dans les années 1970, un champ qui englobait la science de la science, la communication scientifique et les politiques sur la science.

Si l'on s'en tient à une interprétation littérale du terme, la scientométrie s'intéresse à la construction des mesures de la science. Celle-ci, bien entendu, n'inclut pas exclusivement des

sciences naturelles, mais comprend toutes les sphères de production du savoir (au sens plus large où l'entendent le latin *scientia* et l'allemand *Wissenschaft*). Bien que les résultats de la scientométrie servent souvent à évaluer la recherche, ils fournissent aussi et, surtout, au système scientifique un excellent moyen de s'observer lui-même. De ce point de vue, la scientométrie est une méta-science, aux côtés de disciplines apparentées telles que l'histoire, la philosophie et la sociologie des sciences, qui prennent toutes la science comme objet d'étude. Dans son ouvrage, Derek de Solla Price parlait d'un « prologue à la science de la science » (nous soulignons). Il ouvrait ainsi un terrain d'investigation où les approches scientifiques et sociales fourniraient des moyens légitimes d'interroger la science et de mesurer la recherche.

Au fil du temps, la science de la science s'est scindée en deux domaines – la scientométrie, d'une part, et l'étude sociale des sciences et des technologies, d'autre part – généralement délimités par des cadres théoriques (approches positivistes *versus* approches constructivistes) et des méthodes (quantitatives *versus* qualitatives, selon la terminologie courante) distincts. En outre, depuis quelques années, la dichotomie entre les données dites massives et les « petits » volumes de données, souvent plus qualitatives, se fait de plus en plus clivante. Mais cette opposition véhicule souvent des suppositions erronées. Les interprétations négatives des études quantitatives les dépeignent comme réductionnistes, décontextualisées et, par conséquent, artificielles. Les études qualitatives sont, pour leur part, perçues comme non généralisables et hautement subjectives. Il manque à chaque camp de reconnaître les distinctions fractales au sein d'un champ, d'un projet ou d'une méthode donnés, en plus de la capacité de chaque spécialisation à fournir un point de vue qui lui est propre sur le système de la recherche (sur le plan « micro » ou « macro », par exemple). Les approches contemporaines de « la science de la science » ont donc intérêt à tenir compte des deux domaines et à

recouper leurs théories et leurs méthodes. Le fait de les considérer comme complémentaires, plutôt que contradictoires, nous donne de meilleurs outils pour comprendre la science.

Quels sont les fondements théoriques de la scientométrie ?

Les premières théories bibliométriques reposaient sur des régularités statistiques observées dans les bibliographies. Dès 1926, Alfred J. Lotka, statisticien à la compagnie d'assurances MetLife, observait qu'une minorité de chercheurs signait la majorité des publications. Inversement, la majorité des chercheurs ne contribuait qu'à une minorité des publications. Appelée loi de Lotka, cette asymétrie dans la production savante s'est révélée constante dans différents ensembles de données et disciplines : 20 % des chercheurs publient 80 % des documents ; les 20 % restants se répartissent entre 80 % des chercheurs. Moins de dix ans après la formulation de la loi de Lotka, Samuel C. Bradford, mathématicien et bibliothécaire au Musée des sciences de Londres, constatait que les références semblaient obéir à un modèle semblable. Il a formulé sa loi de distribution, la loi de Bradford, selon laquelle un petit nombre de périodiques reçoit la majorité des citations, tandis que la grande majorité des périodiques ne reçoit qu'un petit nombre de citations. L'année suivante, en 1935, le linguiste étatsunien George Kingsley Zipf montrait que le même type de loi de puissance était à l'œuvre dans le langage naturel. La loi de Zipf a servi de pierre angulaire à de nombreux systèmes de recherche d'information et s'est avérée utile quand la scientométrie s'est tournée vers l'analyse contextuelle – ou analyse du texte intégral.

L'observation des lois de puissance dans les données d'extrants et d'impact de la recherche a été déterminante pour le développement de ce champ d'études. Ces lois ont montré que les données sur la science ne suivent pas une distribution normale, dont les mesures pointent vers une valeur centrale (moyenne, médiane,

modes). En mettant au jour l'asymétrie prononcée des données bibliométriques, les premières théories ont ouvert la porte à l'idée que ces données ne se comportaient pas comme les autres. Ces observations ont été la clé qui a permis de comprendre la structure et la croissance de la science. Toutefois, la scientométrie est demeurée en dormance, pour ainsi dire, jusqu'au milieu du XX^e siècle, moment où les index de citations à grande échelle ont été mis au point et où l'évaluation de la recherche est devenue plus courante.

L'essor de la scientométrie, rendu possible par la disponibilité croissante des données dans les années 1960 et 1970, s'est doublé d'un intérêt pour l'élaboration de théories capables de rendre compte des extrants et de l'impact de la recherche. Le pendant sociologique des théories statistiques sur l'asymétrie a pris la forme de l'effet saint Matthieu, défini par le sociologue américain Robert K. Merton dans un article publié en 1968 dans la revue *Science*. Merton s'est intéressé à la répartition des marqueurs de reconnaissance des découvertes scientifiques, notamment au cas des découvertes simultanées. Il a montré que la reconnaissance pour une nouvelle découverte était plus susceptible d'être accordée aux chercheurs déjà très reconnus qu'à ceux qui l'étaient moins. Le nom qu'il a donné à ce phénomène s'inspire d'un verset de l'évangile selon Matthieu : « Car à tout homme qui a, l'on donnera et il aura du surplus; mais à celui qui n'a pas, on enlèvera ce qu'il a » (25:29). Suivant Merton, les sociologues (et frères) Jonathan et Stephen Cole ont analysé le processus de stratification sociale de la science. À partir d'un échantillon composé de physiciens, ils ont montré que les indicateurs fondés sur des citations sont liés à d'autres indicateurs de prestige tels que le classement des universités et les prix reçus. Cela met en lumière les inégalités du système de la recherche en tant que milieu de travail et renforce les théories de Merton concernant la répartition disproportionnée des récompenses dans le monde scientifique.

En 1976, Derek de Solla Price a appliqué ces théories aux citations et consolidé ce qu'on savait du « processus des avantages cumulatifs » à l'œuvre dans le processus de citation. Plusieurs autres études ont confirmé le rôle des avantages cumulatifs en science: les chercheurs qui sont affiliés à des établissements prestigieux sont plus susceptibles d'être cités (même lorsque l'on contrôle les autres variables sociodémographiques), les articles qui paraissent dans les revues les plus réputées sont cités davantage que ceux qui paraissent dans les revues moins cotées (en tenant compte encore une fois des autres variables), et les chercheurs (de même que les publications) les plus cités sont plus susceptibles de voir leurs citations encore augmenter de manière non linéaire que les moins cités. Bref, en science comme dans les autres activités sociales, ce sont les riches qui s'enrichissent. Parallèlement, ceux qui sont pauvres en capital ont tendance à s'appauvrir davantage. Surnommé l'effet Matilda, ce phénomène étudié par l'historienne Margaret Rossiter fait référence en particulier à la reconnaissance moins importante que reçoivent les femmes pour un travail équivalent.

Les théories sur les avantages cumulés ont attiré l'attention d'une autre communauté de chercheurs à la fin des années 1990, alors que la science des réseaux s'établissait comme cadre méthodologique de la mesure de la recherche. Les physiciens Albert-László Barabasi et Réka Albert se sont intéressés à l'« attachement préférentiel » dans le contexte de la croissance du Web. Leur théorie de l'attachement préférentiel – une extension de l'effet saint Matthieu – a servi ensuite, avec d'autres théories issues de la science des réseaux, à comprendre la croissance et la structure de l'information scientifique.

On fait souvent référence, en scientométrie, à la notion de capital. La plupart des chercheurs qui l'évoquent se réfèrent aux théories du sociologue français Pierre Bourdieu. Étudiant le système d'enseignement postsecondaire français (à la suite

des manifestations pour la réforme de l'éducation de mai 1968), Bourdieu a constaté que le système universitaire reproduisait les formes de pouvoir de la société en général. Considérant les différentes formes de pouvoir comme des capitaux acquis socialement, politiquement et économiquement, il a émis l'idée que le capital universitaire – dérivé du capital culturel – est également une forme de pouvoir permettant d'établir des hiérarchies. Selon Bourdieu, les individus qui possèdent un capital externe à la science généreront davantage de capital universitaire, et plus rapidement, que ceux qui ne disposent d'aucun capital. C'est par ce processus que le capital et, par conséquent, le pouvoir, se légitiment et se transmettent d'une génération à l'autre. Les théories bourdieusiennes sont essentielles pour comprendre l'échange de biens symboliques sur le marché universitaire, où de multiples formes de capital interagissent. La reconnaissance tacite de l'existence de ce capital et de la valeur des mesures et des évaluations comme indicateurs de cette devise assure la survie du système de la reconnaissance universitaire.

En plus de la statistique et de la sociologie, les scientomètres empruntent certaines théories à la sémiotique – discipline qui étudie les signes et les symboles en tant qu'instruments de communication, et qui s'inspire des travaux de Charles Sanders Peirce, philosophe et logicien étatsunien de la fin du XIX^e siècle. Peirce a étudié la relation entre un objet (le signifié) et le signe qui le représente. Cette approche a des conséquences évidentes pour la mesure de la science, qui utilise les représentations du travail et de l'impact scientifiques (respectivement, le document publié et la citation) comme matériaux de base pour construire des indicateurs. Des bibliomètres ont remis l'œuvre de Peirce au goût du jour à la fin du XX^e siècle et au début du XXI^e, quand ils se sont trouvés aux prises avec les concepts sous-jacents aux citations. Les travaux de Paul Wouters sont particulièrement éclairants à cet égard : il cherchait en effet à distinguer la référence, la citation

(telle qu'elle est représentée dans les index de citations) et la citation telle qu'elle apparaît dans le texte.

La théorie de Henry Small au sujet des « documents cités en tant que symboles conceptuels » est aussi dérivée du courant sémiotique, bien que celui-ci ne soit jamais explicitement mentionné. Small a publié sa théorie des symboles conceptuels en 1978 dans *Social Studies of Science*, en la fondant sur la définition du symbole formulée par le Britannique Edmund Leach, anthropologue des sociétés. Selon Small, quand une référence est effectuée dans le texte, elle devient le symbole d'idées, de méthodes ou de données particulières. Il soutient en outre que le degré de consensus est très élevé dans la communauté scientifique en ce qui concerne cette association du signe et du concept. Il a mis sa théorie à l'épreuve en examinant le contexte des références, et montré qu'un travail donné est généralement évoqué dans un type particulier de contributions (méthodologiques ou théoriques, p. ex.). La théorie des symboles conceptuels de Small est la pierre d'assise des recherches ultérieures sur le sens à donner au fait de citer et sur les motivations qui le sous-tendent. Cela s'avère particulièrement pertinent pour la construction d'indicateurs, qui doivent reposer sur une cohérence forte du signe et du signifié.

Qu'est-ce qu'un indicateur ?

Dans les sciences sociales, la quantification des concepts passe par la formulation d'indicateurs. La force de la relation entre un indicateur et le concept qu'il mesure est d'une importance critique pour sa validité : l'indicateur doit être explicitement lié au concept, et son choix, étayé par des arguments suffisants. Les travaux de Paul Lazarsfeld, sociologue empiriste, ont joué un rôle capital en définissant la relation méthodologique qui unit les concepts aux indicateurs. Lazarsfeld était en quête de variables observables pour représenter, de manière statistiquement valide, des concepts

inobservables. La scientométrie s'inscrit dans cette tradition, puisqu'elle accepte que des variables inobservables (telles que la production et l'impact de travaux scientifiques) puissent être quantifiées par des variables observables et, par conséquent, mesurables (nombre d'articles ou de citations, entre autres). Les indicateurs se concentrent donc sur les phénomènes observables de l'entreprise scientifique en cherchant à rendre manifestes ses aspects inobservables.

La validité d'un indicateur se juge par la taille de l'écart qu'il laisse entre la mesure et le concept – ce que l'historien des sciences Yves Gingras appelle «l'adéquation de la mesure». Gingras fournit deux autres critères pour évaluer les indicateurs, soit leur «respect de l'inertie propre de l'objet» et «l'homogénéité de la mesure». En ce qui concerne l'inertie, Gingras affirme que la vitesse à laquelle un indicateur est compilé devrait correspondre à la vitesse à laquelle les données sont susceptibles d'évoluer. Une université, par exemple, est peu susceptible de subir des changements marqués en un an, ce qui rend les classements annuels inadéquats et les variations annuelles trompeuses. Les indicateurs qui reposent sur des données provenant des médias sociaux, en revanche, varient de façon beaucoup plus dynamique; il est donc acceptable de construire et de diffuser ces indicateurs rapidement, voire en temps réel. Certains indicateurs fondés sur les données des médias sociaux contreviennent au troisième principe, celui de l'homogénéité, parce qu'ils intègrent des mesures multiples et hétérogènes en un même indicateur – en plus, bien souvent, de ne pas correspondre à un concept clairement identifié. C'est le cas de l'*Altmetric Attention Score*, qui amalgame les valeurs obtenues de sources comme Mendeley, Twitter, Facebook et différents journaux (entre autres) en un indicateur composite impossible à interpréter.

Les indicateurs se construisent à partir de variables d'entrée (*input*), de sortie (*output*) et d'impact. Bien entendu, plusieurs questions demandent que l'on compare ces différents types de

variables : le calcul du rendement du capital investi en recherche, par exemple. Mais le manque d'indicateurs d'intrants comparables à l'échelle mondiale ainsi que l'asymétrie de la couverture des index de citations rendent difficile l'usage de tels indicateurs. Disons qu'on veut savoir si l'investissement dans la recherche en Chine rapporte davantage (d'extrants de recherche) que celui fait aux États-Unis. Cette question repose sur plusieurs suppositions : (a) que l'investissement est calculé de la même manière dans les deux pays, (b) que l'étendue de l'analyse des extrants est la même pour les deux pays (ou, du moins, que la mesure tient compte des inégalités), (c) que les deux mesures sont correctement normalisées en fonction de la taille de la production scientifique (qui doit être calculée de la même manière pour les deux pays), et (d) que les extrants sont normalisés en fonction de l'importance accordée aux différentes disciplines dans chaque pays. De tels calculs ne sont pas impossibles, mais ils sont beaucoup plus complexes que celui qui consiste à prendre une variable donnée de la recherche-développement et à la diviser par le nombre de publications provenant du pays étudié. Dans ces conditions, il est impératif que ceux qui formulent et interprètent les indicateurs connaissent parfaitement les forces et les faiblesses des différentes bases de données et qu'ils comprennent précisément comment on construit un indicateur scientifique.

Quelles sont les sources de données employées pour mesurer la recherche ?

Les sources de données utilisées pour quantifier la recherche se divisent en deux grandes catégories, selon qu'elles servent à concevoir des indicateurs d'intrants, d'une part, ou des indicateurs d'extrants et d'impact, d'autre part. La collecte de données qui vont servir de variables d'entrée s'effectue d'habitude localement, au moyen d'enquêtes, par les conseils de recherche ou par

divers organismes internationaux visant à regrouper les données colligées nationalement. C'est ce que font, par exemple, l'OCDE et l'UNESCO depuis le début des années 1960, dans le but d'établir des normes de conceptualisation et d'opérationnalisation pour les données sur la recherche. L'OCDE, par l'intermédiaire du Groupe de travail des experts nationaux sur les indicateurs de la science et de la technologie, a publié plusieurs manuels (appelés communément les manuels de Frascati, de Canberra et d'Oslo) qui fournissent des définitions générales et des lignes directrices pour la compilation d'indicateurs en recherche et développement (R-D). Le *Manuel de Frascati*, sans doute le plus influent des trois, fournit la définition de certains concepts clés, tels que « recherche fondamentale », « recherche appliquée » et « développement expérimental », et il établit la classification des secteurs où s'effectue la recherche. Ces longues définitions très détaillées sont essentielles, puisqu'elles déterminent les types d'activités (ainsi que les dépenses et les ressources humaines qui leur sont associées) qui doivent ou ne doivent pas être comptabilisés.

Ces manuels sur les indicateurs d'intrants sont révisés périodiquement afin de refléter les changements dans les pratiques sociales qu'ils visent à mesurer. Cependant, chaque pays a des pratiques spécifiques quant à la collecte et à la présentation des variables d'entrée, ce qui rend les comparaisons difficiles à l'échelle mondiale. Les données sont agrégées dans des rapports assortis de nombreuses notes de bas de page indiquant les particularités de la collecte dans chaque pays et soulignant le manque de cohérence des données, le cas échéant. Les notes contiennent des phrases comme « Ne correspond pas tout à fait aux normes recommandées par l'OCDE », « Discontinuité dans la série avec l'année précédente pour laquelle les données sont disponibles » ou « Surestimé ou fondé sur des données surestimées ». Bien que ces manuels constituent la meilleure source d'intrants disponible, leur incapacité à produire des indicateurs comparables mondialement

demeure problématique. Parallèlement aux grands ensembles de données compilées, généralement par pays, par les organisations internationales, d'autres joueurs, comme des États ou des provinces, des industries et des conseils de recherche – notamment la National Science Foundation (NSF) et les National Institutes of Health (NIH), aux États-Unis – mettent en ligne les données des projets qu'ils financent. Ces données, toutefois, concernent la plupart du temps des subventionnaires particuliers, ou sont agencées en systèmes compilant les données de subventionnaires d'un pays en particulier ou de domaines bien précis, ce qui complique les comparaisons internationales ou multidisciplinaires. Quelques tentatives pour fusionner ces ensembles de données et créer des outils couvrant plusieurs subventionnaires ont vu le jour, mais les résultats sont loin d'être exhaustifs.

Les indicateurs d'intrants compilés par des organismes locaux ou internationaux se focalisent souvent sur les disciplines compatibles avec l'évaluation de la R-D, à savoir la science, la technologie, et la médecine (STM). L'analyse des arts et humanités, ainsi que de certaines sciences sociales, ne figurent pas d'ordinaire dans les rapports de l'OCDE ou des grands organismes nationaux (les indicateurs de la science et de l'ingénierie de la NSF, par exemple). Quand on évalue l'état de la recherche pour un pays donné, par conséquent, on doit comprendre que les indicateurs d'intrants décrivent la force de ce pays dans les sciences naturelles, physiques et médicales. Sur le plan national, il existe aussi des problèmes de couverture : alors que les pays qui investissent beaucoup dans les sciences investissent également dans l'évaluation de leurs variables d'entrée, cela n'est pas vrai des pays moins riches sur le plan scientifique ; c'est ce qui fait que les indicateurs d'intrants ne couvrent pas l'ensemble de la communauté internationale.

Les bases de données bibliographiques construites pour la recherche documentaire sont depuis longtemps les principales sources de données sur les extraits et les impacts, et elles ont

fourni des normes que l'on améliore systématiquement. Il existe plusieurs index de publications fort utiles pour construire des indicateurs de sortie mesurant la production et l'impact de la recherche et la collaboration dans ce domaine. Certains d'entre eux, comme MEDLINE, couvrent des domaines précis dans le monde entier tandis que d'autres, comme SciELO, au Brésil, se concentrent sur les périodiques à l'échelle d'un groupe de pays. Les sources qui offrent des données complètes sur les citations sont relativement peu nombreuses, cependant. Parmi celles-ci, les trois plus connues sont le Web of Science (WoS), Scopus et Google Scholar. Elles se distinguent des bases de données de publications comme MEDLINE en ce que l'information qu'elles fournissent est relationnelle: en plus de contenir les métadonnées des publications, elles font des liens entre les citations. En d'autres mots, elles fournissent, pour chaque article citant, la liste des références auxquelles il renvoie. Une autre caractéristique importante du WoS et de Scopus est qu'ils donnent l'adresse des auteurs, ce qui permet de relier les articles à des établissements et à des pays.

Chacune de ces sources est le produit de son époque. Le *Science Citation Index*, maintenant intégré à la base de données multidisciplinaire Web of Science, a été le premier index de citations détaillé; inspiré des bases de données juridiques, il a été conçu à l'usage des bibliothécaires et des documentalistes, avant l'ère numérique. Ce contexte se reflète clairement à la fois dans ses forces et ses faiblesses. Scopus, lui, a été conçu à l'ère numérique et remédie aux faiblesses du WoS. Il repose sur un vaste corpus de périodiques et, grâce au soutien organisationnel du plus grand éditeur de revues savantes au monde (Elsevier), il a défié le monopole du WoS. Quant à Google Scholar, il est issu lui aussi d'une entreprise existante, et repose sur les capacités d'indexation massives de Google plutôt que sur une sélection précise du contenu. Cette mise en contexte est nécessaire pour comprendre les forces et les faiblesses des sources dont nous traitons ici.

Les ensembles de données par discipline peuvent également servir à quantifier la recherche. MEDLINE, par exemple – produit et mis à jour par la Bibliothèque nationale de médecine des États-Unis (US National Library of Medicine) – couvre presque tous les articles de médecine et son accès est gratuit (on peut aussi le télécharger et il est relativement convivial). Sa couverture remonte jusqu'aux années 1950 et le système de catégorisation thématique qu'il propose, par vedettes MeSH, est l'un des plus précis et des mieux organisés du monde de la recherche. Les vedettes MeSH forment une terminologie hiérarchisée contenant des renseignements détaillés sur les articles recensés – des maladies aux méthodes employées et aux populations étudiées –, ce qui en fait l'outil idéal pour ceux qui s'intéressent au champ thématique du domaine médical. Le principal écueil de MEDLINE, d'un point de vue bibliométrique, est qu'il ne recense pas les citations et que, jusqu'à récemment, il ne donnait l'adresse que du premier (ou principal) établissement; il mesure donc, mais jusqu'à un certain point seulement, la production scientifique, et pas du tout son impact. Par ailleurs, comme il a été conçu pour repérer plutôt que pour compter les articles, il contient moins de métadonnées: les adresses des auteurs, on l'a vu, ne sont données pour tous que depuis 2015. Il est donc impossible de mettre au point des indicateurs mesurant la collaboration ou la production institutionnelle antérieures. En fusionnant MEDLINE et les autres index de citations comme le WoS ou Scopus, on associe les avantages – du premier (les vedettes MeSH, entre autres) à la puissance relationnelle et aux métadonnées des seconds.

Il existe assez peu de sources de données détaillées utiles à la quantification des travaux publiés dans des livres. WorldCat, un réseau de contenu et de services bibliothécaires, est sans doute la plus complète pour quantifier la production d'ouvrages imprimés. Elle donne une indication générale du fonds documentaire des bibliothèques. Les indicateurs d'extrants se calculent en fonction

du nombre de bibliothèques qui possèdent un livre particulier, par année ou par champ disciplinaire. C'est toutefois un outil qui demeure sous-exploité pour la recherche bibliométrique, en grande partie parce que même s'il donne des indications sur la production, il ne fournit aucune information sur les citations. Le *Book Citation Index*, produit par Clarivate Analytics, a été le premier index de citations consacré aux livres. Malheureusement, il répète beaucoup des erreurs de périmètre reprochées au premier Web of Science : il montre un biais favorable aux ouvrages publiés aux États-Unis, par de grands éditeurs, dans le domaine des STM et en langue anglaise. Google Books propose une base de données plus complète. Cette source, toutefois, n'a pas la transparence nécessaire à l'évaluation systématique de la recherche : elle ne fournit aucune indication sur la méthode d'extraction et, en conséquence, aucune indication précise sur la population ou l'échantillon. En outre, comme pour tous les produits de Google, il est assez difficile, voire impossible, d'extraire les données brutes qu'elle contient.

Au cours des dernières années, d'autres outils et sources de données sont apparus qui remettent en question le monopole des index de citations sur la mesure de la production et de l'impact de la recherche. Parmi ceux-ci, on trouve des sources de données qui suivent la trajectoire d'impact des recherches dans le Web social et les mentions d'articles ailleurs que dans la littérature savante, notamment dans les documents de politiques publiques, les médias sociaux et les médias grand public. Elles ont certainement augmenté l'attention potentielle accordée aux travaux savants. En plus des articles savants – le genre dominant dans les index de citations classiques –, l'intérêt ne cesse de croître pour l'impact d'autres formes de savoir, comme les livres ou les données. OCLC, l'entreprise qui possède WorldCat, a conclu en 2013 un partenariat pour intégrer les données de WorldCat dans Plum Analytics, qui compile des données de mesures « alternatives »

pour les bibliothèques et les universités. Une entreprise nommée Altmetric produit des indicateurs pour les livres de certains éditeurs précis. Ces sources demeurent pour l'instant en périphérie de l'exercice de quantification de la recherche. Ce sont toujours les articles et les brevets qui servent, quasi exclusivement, de sources de données pour la mesure des extraits.

2

LES DONNÉES

Qu'est-ce qu'un index de citations ?

Un index de citations est une base de données bibliographiques dans laquelle on peut établir des liens entre les documents indexés. Elle contient, en plus des métadonnées habituelles des documents recensés – auteur, année de publication, titre de l'article et titre du périodique, par exemple –, les références des textes cités par ces documents sources. Le premier index de citations scientifiques s'inspirait du domaine juridique, où l'on se servait depuis 1873, aux États-Unis, d'un système de référence aux cas de la jurisprudence qui avait été mis au point par Frank Shepard. Cet index de citations spécialisé permettait aux juristes de suivre des réseaux de cas imbriqués et d'examiner la provenance et la prépondérance des jugements. Eugene Garfield, chimiste et documentaliste, a vu dans cet usage un modèle qui conviendrait à la littérature scientifique et, dans un article paru en 1955 dans la revue *Science*, il proposait de mettre au point un index des références pour le monde de la science. Le sous-titre de son article, «A new dimension in documentation through association of ideas» [«la documentation acquiert une nouvelle dimension grâce à l'association d'idées»], montre que son but n'était pas l'évaluation de la recherche, comme

on le croit souvent, mais le rapprochement entre des articles en fonction de leurs liens formels. Il présentait son outil comme un « index d'idées liées par association » parce qu'il permettrait aux gens de voyager d'un périodique et d'une discipline à l'autre en suivant les idées, « les unités moléculaires » du travail scientifique. D'après lui, un tel index faciliterait le repérage bibliographique, mais freinerait aussi les comportements inappropriés, en mettant en lumière les usages « non critiques » des références en même temps que tous les précurseurs d'un travail donné. En supposant que les références citées d'une source pertinente soient pertinentes à leur tour, les chercheurs seraient bien outillés pour repérer les travaux apparentés aux leurs.

La proposition de Garfield reçut l'appui d'un certain nombre d'individus et d'organismes. Joshua Lederberg, généticien lauréat d'un prix Nobel, se fit le promoteur du projet, et les National Institutes of Health et la National Science Foundation fournirent le financement initial. Ainsi, vers la fin des années 1950, Garfield faisait l'essai d'un index de citations dans le domaine de la génétique; il se servait de fiches poinçonnées et de rubans magnétiques. Le projet servit de prototype et de validation au *Science Citation Index* (SCI) que l'Institute for Scientific Information, fondé par Garfield en 1960, allait développer. En 1963, le premier SCI était publié, à l'intention des chercheurs et des bibliothécaires; il recensait, dans une série de lourds volumes, les métadonnées des articles de 613 périodiques couvrant toutes les disciplines. L'intention première était de faciliter l'extraction de documents, et non leur évaluation, et le SCI offrait cette nouveauté de permettre des recherches en fonction des relations révélées par les références citées, en plus des recherches classiques par titre de périodique ou nom d'auteur. Chaque document se trouvait donc à s'inscrire dans un réseau de relations qu'on pouvait suivre, au fil d'une idée ou d'un concept, au-delà des frontières habituelles des disciplines. Le caractère multidisciplinaire de ce genre de

repérage documentaire était unique, mais son utilité dépendait de sa taille: pour que son potentiel se déploie pleinement, un index de citations doit couvrir un corpus relativement vaste et varié. Un index de citations qui ne recenserait que quelques périodiques ne contiendrait qu'une petite quantité de références à d'autres documents, ce qui en ferait un outil de repérage inefficace. L'une des caractéristiques primordiales d'un index de citations, on l'aura compris, est le nombre de documents citant qu'il comprend, et qu'on appelle *articles ou documents sources*.

En plus de ces relations citant-cité, les index de citations contiennent des informations géographiques, caractéristique essentielle à la construction d'indicateurs de niveau méso- ou macro. Un index de citations complet inclut, par exemple, l'adresse institutionnelle complète de tous les auteurs, y compris le département, l'université, la ville, la province ou l'État et le pays où ils exercent. Mais la disponibilité de ces données dépend des revues savantes: un grand nombre d'entre elles, en humanités, notamment, ne donnent pas les adresses des auteurs; cela diminue le taux de couverture, soit le périmètre couvert par cette variable dans l'index. D'autres différences existent entre les disciplines en ce qui concerne les métadonnées fournies. Ainsi, dans bien des spécialités de la physique et des mathématiques, on ne mentionne que l'initiale des prénoms des auteurs, ce qui nuit à la désambiguïsation – c'est-à-dire la capacité de rattacher le bon auteur aux bonnes publications dans la base de données – et ainsi aux analyses à l'échelle individuelle. Depuis quelques années, les index de citations intègrent également le contenu de la section «Remerciements» des articles, mais souvent en se limitant à celles qui renseignent sur le financement des travaux de recherche. Cela conduit inévitablement à la surreprésentation des articles provenant des domaines bien subventionnés (comme les sciences naturelles et médicales). En plus des variables explicitement indexées, les analystes peuvent considérer d'autres éléments d'information

grâce aux outils d'analyse superposés : un auteur peut être associé à un sexe, un établissement à un secteur d'activité, un article à un sujet.

Les renseignements sur les sujets traités proviennent traditionnellement de métadonnées comme le titre, le résumé et les mots-clés. Le fait que les documents contemporains soient créés numériquement et que les plus anciens soient numérisés peu à peu ouvre la possibilité d'analyser le texte intégral des articles. Mais une grande part de ce contenu n'est pas dans le domaine public (JSTOR, p. ex.), et si l'expansion des dépôts pré- et postpublication augmente la quantité de données disponibles, dans l'ensemble, l'analyse du texte intégral en vue de construire des indicateurs bibliométriques n'en est qu'à ses balbutiements. Certains ont mis sur pied des dépôts de textes intégraux obtenus par piratage (Sci-Hub en est un exemple), mais leur utilisation se borne à la lecture ; ils ne servent à rien dans l'évaluation de la recherche, et rarement dans sa mesure. D'autres s'efforcent de préciser le repérage en exploitant le texte intégral et les avancées de l'apprentissage automatique : l'outil Semantic Scholar, par exemple, propose plusieurs indicateurs – dont l'un permettant de mesurer les citations très influentes – fondés sur une analyse du texte intégral et du contexte détaillé de chaque citation. Toutefois, ses données se limitent à deux domaines (l'informatique et les neurosciences) et aux articles parus en anglais, et les données sur les articles pour lesquels l'accès est payant ne sont que partielles. Malgré ces contraintes, cette initiative illustre les possibilités qu'offre à la scientométrie la recherche en texte intégral, un aspect des index de citations et de l'évaluation de la recherche susceptible de connaître un essor prochain.

Qu'est-ce que le Web of Science ?

Le Web of Science (WoS) – aussi appelé Web of Knowledge – est la version moderne du SCI qu'Eugene Garfield avait mis au point et publié pour la première fois en 1963. Le WoS est commercialisé à l'heure actuelle par Clarivate Analytics, propriété d'Onex Corporation et de Baring Private Equity Asia, qui l'ont acheté de Thomson Reuters, en 2016, pour 3,55 milliards de dollars américains. Le SCI, premier index de citations consacré à la recherche scientifique, ne couvrait que les sciences naturelles et médicales. En 1973, un index des citations en sciences sociales (le SSCI, pour *Social Science Citation Index*) a vu le jour et en 1975, c'était au tour des arts et des humanités d'avoir le leur (l'AHCI, pour *Arts and Humanities Citation Index*). L'ordre dans lequel cette expansion s'est produite, combiné à l'écart de volume dans la production d'articles entre les sphères scientifiques – au regard d'autres types de documents, comme les monographies ou les actes de colloques –, explique l'asymétrie des périmètres couverts par les différents index du WoS: la couverture des sciences naturelles et médicales est plus étendue que celle des sciences sociales, elle-même plus étendue que celle des humanités.

Eugene Garfield n'a jamais envisagé le SCI comme une base de données exhaustive. Il visait plutôt l'indexation des périodiques les plus importants de chaque discipline, l'importance relative étant déterminée par la loi de dispersion de Bradford. Cette loi énonce que la plus grande partie de la littérature citée dans un domaine est publiée dans un petit nombre de revues et, inversement, que la majorité des revues ne contiennent qu'une petite part des documents cités. En d'autres mots, selon la loi de Bradford, les références suivent une distribution de Pareto, avec une minorité de revues au centre et la majorité en périphérie. Garfield avait pour objectif d'indexer le contenu de ces revues qui se retrouvent au centre parce qu'elles reçoivent le plus de citations

et qu'en conséquence, elles sont les plus utilisées de leur domaine. C'est une approche qui s'est avérée assez fonctionnelle en sciences naturelles et médicales, mais qui demande davantage de critères de sélection en humanités, compte tenu de l'importance relative plus importante des monographies. Garfield n'a jamais caché les limites associées à la couverture – ni les pressions, notamment financières, qu'il subissait au moment de choisir les revues, celles-ci pouvant demander à être répertoriées. Il a répété à plusieurs reprises qu'il gérait une entreprise, pas un organisme de bienfaisance, surtout quand on lui reprochait d'accepter que les revues payent pour être indexées (à condition tout de même de respecter les critères de qualité).

Dans un article récent paru sur le site Web de Clarivate Analytics, et sans doute en réaction à la couverture plus étendue de Scopus, la politique d'indexation du WoS a été clarifiée afin mettre en valeur le processus de sélection. Les critères employés actuellement comprennent les normes éditoriales (relecture par les pairs, régularité des parutions, usage de l'anglais, p. ex.), le contenu rédactionnel et la portée internationale de la revue, en plus des critères citationnels. Une équipe analyse ces critères et décide si la revue mérite d'être répertoriée ou non. La politique d'indexation du WoS mentionne aussi l'importance de tenir compte des questions locales et régionales dans la décision d'inclure ou non une revue dans les index SSCI ou AHCI; cependant, la façon dont ces questions sont intégrées demeure floue.

Le WoS a d'abord été conçu comme un outil de repérage destiné aux bibliothécaires et à d'autres spécialistes de la recherche documentaire; par la suite, les métadonnées de l'index ont facilité certains types d'extraction. Très tôt dans l'histoire du SCI – dès 1973 –, l'établissement d'attache des auteurs figurait dans la base de données, sans doute pour permettre aux chercheurs ou aux bibliothécaires de repérer des articles par établissement d'enseignement, par ville ou par pays. Plus encore, le WoS notait

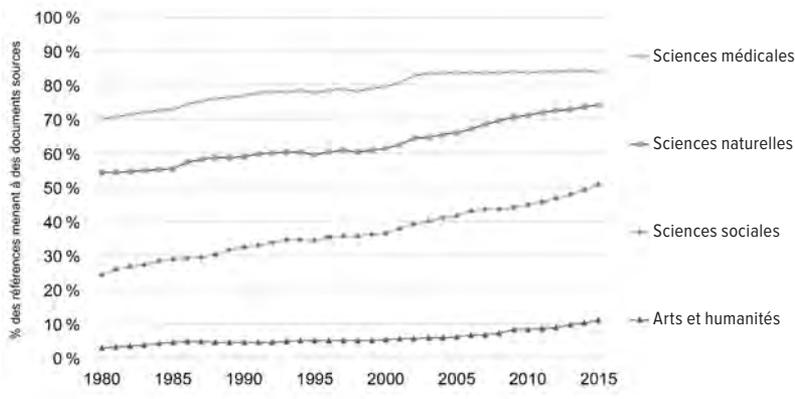
cette information pour chaque auteur, pas seulement l'auteur principal, comme le font d'autres bases de données (MEDLINE, notamment). Cette caractéristique a permis la construction d'indicateurs de collaboration, puisqu'on pouvait repérer rapidement les circuits de coopération entre établissements à de multiples niveaux d'agrégation des données (établissement, ville, pays). Jusqu'en 2008, l'inconvénient était l'absence de lien entre les adresses des établissements et les auteurs particuliers, ce qui nuisait à la désambiguïsation automatique des noms d'auteurs: dans le cas d'une collaboration entre établissements, l'affiliation précise de chacun des auteurs particuliers demeurait inconnue. Ce n'est qu'en 2008 également que les prénoms des auteurs ont été fournis systématiquement. Avant cela, une recherche pour «R. Dupont» dont le but était d'extraire les documents signés par Robert Dupont pointait également vers les articles de Raoul Dupont, Robert Dupont, etc. Enfin, contrairement à Scopus, qui dispose d'un algorithme de désambiguïsation, le WoS n'effectue pas cette tâche automatiquement; il fait intervenir un identifiant unique, ResearcherID, lequel demande l'intervention des auteurs – qui ne l'ont, en très grande majorité, pas adopté. Conséquemment, bien que le WoS soit utile à l'analyse de données cumulées (par sujet, périodique, établissement, ville, État ou province, pays), la bibliométrie à l'échelle individuelle à partir de cette source demeure problématique car elle demande un important travail de nettoyage des données.

Le SCI et ses successeurs ont fait l'objet d'autres critiques. La plus grave concerne les biais de couverture: on observe des disparités entre les revues, les domaines et les genres indexés, ainsi que dans la provenance géographique et la langue du contenu répertorié. Plus encore, on s'inquiète de la qualité disparate de l'indexation: le contenu non anglophone présente une proportion d'erreurs beaucoup plus élevée (ce qui nuit à la fois au repérage et à la désambiguïsation) que le contenu anglophone. La qualité de

l'indexation des références est extrêmement préoccupante aussi, car elle peut mener à la sous-estimation du nombre de citations que reçoivent articles et revues. Mais la qualité de l'index, comme Garfield s'était empressé de le souligner, dépend de la qualité du contenu indexé, c'est-à-dire que l'index ne peut pas corriger les erreurs introduites par les auteurs, les rédacteurs ou les éditeurs dans leurs publications. Le fait que le contenu provienne des éditeurs – et que Clarivate Analytics doit donc attendre qu'on le lui fournisse – entraîne d'ailleurs certains délais d'indexation; un obstacle que des sources comme Google Scholar permettent toutefois de surmonter.

L'autre problème majeur est la couverture des références citées qui, considérée dans son ensemble, donne une indication de l'applicabilité de l'analyse des citations à certaines disciplines. Quand on examine le pourcentage de documents sources dans la liste des références d'un article donné – c'est-à-dire le pourcentage de documents cités dans l'article indexé qui figurent également dans le WoS –, il devient évident que cet index, proche de l'exhaustivité pour les sciences naturelles et médicales, ne fournit qu'une faible couverture des articles sources en sciences sociales, en arts et en humanités. Alors que 85 % environ du contenu cité par les articles publiés en sciences médicales renvoie à des documents recensés par la base de données, cela tombe à moins de 75 % en sciences naturelles et à juste au-dessus de 50 % en sciences sociales (figure 2.1; chiffres de 2015). Dans les arts et les humanités, une petite partie seulement des références citées dans les documents recensés par le WoS fait référence à d'autres documents de cette base de données (environ 10%). La tendance est à la hausse dans toutes les sphères, cependant, ce qui suggère que le périmètre de littérature pertinente couvert par l'index s'étend (ou que les revues répertoriées par le WoS s'entrecroisent de plus en plus). Cela pourrait correspondre aussi à un changement de comportement des chercheurs en matière de publication et de citation:

FIGURE 2.1 Pourcentage des références citant des documents sources (documents indexés par le Web of Science), 1980-2015



Source: Web of Science, données agrégées selon les disciplines de la NSF.

s'ils ont recours, par exemple, à des articles plutôt qu'à des livres ou d'autres types de documents pour étayer leurs arguments.

L'arrivée des cédéroms, dans la première moitié des années 1990, a favorisé l'utilisation des index pour l'évaluation à grande échelle de la recherche et la mise sur pied de groupes de recherche spécialisés en bibliométrie. L'invention d'INCITES – une couche analytique superposée aux données de citation brutes utilisées pour les indicateurs bibliométriques et les analyses comparatives – a élargi le bassin d'utilisateurs potentiels. D'autres entreprises, depuis, ont compris le potentiel commercial de cet ajout: la firme Academic Analytics, entre autres, a mis sur le marché à l'intention des universités et, dans une moindre mesure, des individus, des données servant à produire des indicateurs et des comparaisons entre les institutions.

Dans le but d'étendre le périmètre couvert par les données qu'elle fournit, Clarivate Analytics compile, depuis 1990, un index

de citations pour les actes de colloques (*Conference Proceedings Citation Index*) et, depuis 2008, un autre pour les monographies (*Book Citation Index*). Le premier cherchait à étendre la couverture des domaines comme l'ingénierie et l'informatique, tandis que le second visait à remédier aux carences de la couverture dans les sciences sociales et, surtout, dans les arts et les humanités. En plus de ces index principaux, Clarivate Analytics fédère aussi plusieurs index régionaux, dont l'index de citations brésilien SciELO, un index de citations scientifiques en russe, un index de citations scientifiques en chinois et la base de données des périodiques coréens (KCI), ce qui augmente la couverture des bases de données de Clarivate Analytics de quelques millions de fiches. Dans les faits, comme l'interopérabilité de ces index et du portail principal laisse à désirer, tout comme leur accessibilité, ils sont très peu utilisés. En outre, depuis 2012, en réaction à l'intérêt croissant pour les données en accès libre, Clarivate Analytics produit un index de citations des données (*Data Citation Index*). La médecine clinique est omniprésente dans cet index. En 2013, en effet, quatre dépôts de données fournissaient 75% des fiches: Gene Expression Omnibus, UniProt Knowledgebase, PANGAEA et le Bureau du recensement des États-Unis, avec ses fiches TIGER/Line Shapefiles.

Globalement, en décembre 2016, les index du WoS couvraient les articles publiés dans environ 12 700 revues, 160 000 actes de colloques et 68 000 monographies, pour un total approximatif de 55 millions de documents et 1,1 milliard de citations depuis 1900. Même si tous les documents répertoriés par le WoS comportent des renvois à d'autres documents, on considère qu'il est « exhaustif » – couvrant toutes les disciplines et exerçant une indexation cohérente – à partir de 1980 seulement.

Qu'est-ce que Scopus ?

Le monopole du Web of Science sur l'analyse bibliométrique à grande échelle a pris fin en 2004, quand le plus important éditeur de publications savantes au monde, Elsevier, a lancé un index de citations nommé Scopus. Avant la création de Scopus, les chercheurs fouillaient des bases de données (comme le WoS) pour trouver des articles et en obtenir les renseignements bibliographiques, puis ils récupéraient ces articles sur la plateforme en ligne Science Direct d'Elsevier. On présente souvent la création de Scopus comme un cas remarquable d'intégration verticale, où la firme qui compile l'index de citations est la même que celle qui possède les revues indexées. C'est là une différence fondamentale entre Scopus et le WoS, ce dernier comptant sur les éditeurs pour obtenir les métadonnées nécessaires à la production de ses index. Elsevier, déjà propriétaire des publications, donc des métadonnées, a créé un index de citations recensant les principales revues savantes de son catalogue – en s'abstenant toutefois de les recenser toutes, sans doute pour éviter qu'on lui reproche de manquer de sélectivité. En 2016, environ 10 % des périodiques indexés par Scopus étaient publiés par Elsevier. Cette intégration verticale, très profitable à l'entreprise, soulève des inquiétudes quant au monopole de celle-ci sur les communications savantes : Elsevier est tout à la fois l'un des plus importants éditeurs, indexeurs et fournisseurs d'outils d'évaluation de la recherche.

L'accès limité aux données sources de Scopus rend difficile l'analyse détaillée de sa couverture et de ses limites. Alors que la liste des périodiques dépouillés par le WoS repose depuis toujours sur des données citationnelles, les critères de Scopus sont plus flous – du moins l'étaient-ils au départ. Elsevier a fini par mettre sur pied un comité consultatif pour la sélection du contenu, formé de scientifiques et de bibliothécaires chargés d'évaluer les périodiques en vue de leur inclusion parmi les sources. En plus des

éléments de base (comité de lecture, présence de l'ISSN), le comité de sélection du contenu évalue les revues selon cinq autres critères : politique rédactionnelle, contenu, notoriété (nombre de citations reçues), régularité des parutions et disponibilité en ligne. Elles sont soumises par ailleurs à une réévaluation périodique, et susceptibles de recevoir une mention indiquant qu'elles ont un degré élevé d'autocitation ou qu'elles comptent généralement peu d'articles, de citations ou de clics vers le texte intégral. Elsevier ne révèle pas le nombre de revues qui ont été retirées de sa base de données depuis sa création.

L'indexation des livres est plus obscure. On sait qu'Elsevier tient compte de la réputation de la maison d'édition, de la qualité des ouvrages et des politiques de publication, mais la façon dont elle manie ces concepts n'est pas précisée. L'indexation se fait par maison d'édition plutôt que par titre d'ouvrage. De la même façon, l'indexation des actes de colloques repose sur la qualité du colloque et sur la réputation de l'organisateur ou de l'éditeur. Dans l'ensemble, les critères de sélection produisent une base de données apparemment plus inclusive quant aux types de documents qu'elle contient, au nombre total de titres indexés et à la couverture des revues à portée nationale – des aspects que la communauté des chercheurs et des experts en bibliométrie jugeait annuellement problématiques dans le cas du WoS. Une autre des particularités notables de Scopus est la désambiguïsation automatique des noms d'auteurs, qui permet d'attribuer les articles à des individus précis avec ce qui semble être une marge d'erreur relativement faible – bien que l'algorithme derrière ce travail de désambiguïsation ne soit pas dévoilé.

Dans sa première version, Scopus n'indexait pas les arts et les humanités, domaines bien mal couverts par les périodiques et où de nombreux facteurs limitaient l'applicabilité de l'analyse des citations. Ces disciplines ont été incluses quelques années après le lancement de la base de données. En 2016, Scopus avait indexé

plus de 60 millions de notices extraites d'environ 23 000 revues de toutes les disciplines, ainsi que 6 millions de textes parus dans des actes de colloques et 130 000 monographies. Cette base de données n'en demeure pas moins très « jeune » en ce qui concerne l'étendue de sa couverture temporelle: si son contenu remonte dans certains cas à 1823, ce n'est que depuis 1996 qu'elle l'indexe de façon systématique. Tout cela fait de Scopus une source de grande qualité pour l'analyse de documents contemporains, mais inférieure à WoS pour celle de documents plus anciens.

Qu'est-ce que Google Scholar Citations ?

Un an après Scopus, en 2005, Google Scholar est entré en scène afin d'aider les chercheurs à repérer et à récupérer des documents savants en texte intégral. Quelques années après sa mise en ligne, des indicateurs bibliométriques relatifs aux auteurs et aux revues (assez limités dans ce dernier cas) lui ont été ajoutés, sous la rubrique Mes citations. L'algorithme qui traite le contenu de Google Scholar Citations n'est pas bien documenté: à l'exception d'une note succincte affichée sur le site Web, qui souligne entre autres que l'index recense des articles (ou documents) et non des revues, rien n'est dit au sujet de la méthode qu'il emploie ou de son algorithme de recherche. On suppose que Google Scholar indexe tous les articles se trouvant sur les plateformes de diffusion des revues savantes, déposés sur les dépôts institutionnels des universités et autres dépôts disciplinaires, et tous ceux mis en ligne sur les sites Web des chercheurs. On sait qu'il recense les documents en plusieurs langues, notamment en russe depuis 2016. Cependant, Google Scholar intègre souvent des articles ne provenant pas de périodiques: des présentations PowerPoint aux chapitres de livres, voire des plans de cours – tout ce qui ressemble de près ou de loin à une parution universitaire. Bien qu'il n'existe pas de moyen précis de calculer le nombre total de documents indexés, on l'estime à

160 millions, ce qui en fait le plus gros index de citations actuellement disponible.

Étant donné l'hétérogénéité des sources et le caractère automatique de l'indexation, les données de Google Scholar sont incomplètes et sont en général de faible qualité. Contrairement au WoS et à Scopus, l'index ne précise pas l'établissement d'attache des auteurs. En termes de bibliométrie, les seules mesures qu'on peut obtenir directement à partir de Google Scholar Citations concernent les individus et les périodiques. Le portail donne le nombre total de citations, l'indice h et l'indice i10 (soit le nombre d'articles ayant été cités au moins 10 fois). Tout auteur peut se créer un profil, et Google Scholar Citations lui attribuera automatiquement les documents portant son nom. Ce processus de repérage est dynamique : le dossier d'un chercheur peut donc varier de jour en jour, selon la disponibilité des sources de données. Cette instabilité des dossiers nuit à l'utilisation bibliométrique de la source, rend les analyses moins reproductibles, et complique les comparaisons et les analyses diachroniques. Chaque auteur a la possibilité de trier sa liste de documents, en fusionnant les fiches dupliquées et en supprimant celles qui sont inexactes. Il y a peu d'incitatifs à effectuer ce nettoyage, toutefois, puisqu'en fusionnant et en supprimant des fiches, on diminue les scores de productivité et d'impact associés au dossier. Par conséquent, malgré la mise à jour en temps réel de Google Scholar Citations, fournissant les statistiques les plus à jour, son utilisation dans l'évaluation de la recherche demeure problématique, parce que les taux de production et de citations peuvent varier quotidiennement, dans un sens ou dans l'autre (les diminutions provenant de la suppression des documents attribués par erreur et de la fusion des doublons).

L'indexation automatique et les techniques de désambiguïsation employées par Google Scholar ont comme conséquences un nombre élevé de doublons et de documents non scientifiques, de même que de travaux scientifiques attribués par erreur. Les

profils de Google Scholar Citations contiennent beaucoup d'articles qui n'ont pas été écrits par l'auteur en question (ni par un homonyme) et d'entrées mentionnées à de multiples reprises avec des variantes dans l'une des variables de l'enregistrement. Cette situation est extrêmement problématique, puisqu'elle gonfle grossièrement les mesures de productivité et d'impact. La validité des données indexées – et donc des résultats compilés – par Google Scholar a été mise à l'épreuve dans un certain nombre d'études empiriques. En 2009, Cyril Labbé, un informaticien français, a produit un ensemble de 101 articles générés automatiquement qui se citent entre eux (ils citent aussi un article déjà recensé par Google Scholar, afin de s'assurer de l'indexation par la base de données). À la suite de cette intervention, Ike Antkare (le nom du chercheur fictif « auteur » des 101 articles) a été propulsé tout en haut du palmarès des chercheurs les plus cités de tous les temps, avec un indice h d'environ 100. Une expérience semblable a été menée par le groupe de recherche EC3, de Grenade en Espagne, et a montré à quel point il était facile de gonfler les mesures de citations. Alors que Labbé s'était servi de SciGen pour produire son corpus de faux articles savants, l'équipe espagnole a simplement copié-collé plusieurs fois de suite un court texte dans un document dont la mise en forme s'apparentait à un article savant, et dont toutes les références renvoyaient à des travaux de leur groupe de recherche. Les six articles ainsi créés, attribués au chercheur fictif Marco Alberto Pantani-Contador (en référence à deux cyclistes coupables de dopage), citaient chacun 129 fois les publications du groupe de recherche, augmentant de 774 leurs citations reçues. Mais l'élément le plus intéressant du canular est l'incapacité de Google Scholar à le détecter : il a fallu que les auteurs se dénoncent publiquement, par la publication d'un article et d'un billet de blogue, pour que le profil de Pantani-Contador soit supprimé du portail. Les deux canulars que nous venons de mentionner soulignent le problème fondamental de Google Scholar : son incapacité à

distinguer les vrais articles scientifiques des documents qui n'en ont que l'apparence.

Même si la couverture de Google Scholar est beaucoup plus vaste que celle des deux autres index – ce qui la rapproche davantage du c.v. du chercheur, qu'elle arrive même à dépasser –, ses capacités ne sont pas du tout les mêmes. On ne peut pas extraire ou télécharger les fiches de Google Scholar, qui ne précise pas la taille de son échantillon et ne fournit aucune mesure de référence pouvant servir de base à la comparaison, comme le font le WoS et Scopus. En fait, il existe depuis 2006 un outil complémentaire mis au point dans le but de calculer des indicateurs bibliométriques avancés pour des groupes d'articles indexés par Google Scholar. Anne Wil-Harzing, qui à l'époque était professeure à l'Université de Melbourne, en Australie, a conçu un logiciel, Publish or Perish (PoP), capable d'analyser les données de Google Scholar et de calculer ensuite diverses mesures. Le logiciel ne peut traiter plus de 1000 fiches, et il ne retient par défaut que les résultats les plus pertinents. Cet outil permet de construire des indicateurs plus précis et des analyses plus détaillées que Google Scholar employé seul. Il ne permet cependant pas les analyses à l'échelle des adresses institutionnelles (à l'échelle du monde, des établissements ou des disciplines), d'une part à cause de la taille restreinte des données qu'il permet d'extraire, et d'autre part parce qu'il se fonde sur les données et l'algorithme opaques de Google Scholar. La qualité des métadonnées qu'il permet d'obtenir est plutôt médiocre : les noms d'auteurs ne sont pas toujours fournis, les titres des périodiques ne sont pas homogènes et l'information sur les établissements d'attache est inexistante. Bref, cet outil nous donne une estimation de la productivité et de l'impact d'une personne ou d'un petit groupe de recherche, mais rien de plus qui soit utile à la quantification systématique de la recherche.

Qu'est-ce qui distingue les principaux index de citations ?

Les trois index que nous avons présentés diffèrent par leur taille et leur couverture. En 2014, on estimait que Google Scholar contenait environ 160 millions de documents, comparativement à 60 millions pour Scopus et à 55 millions pour le Web of Science. Mais Google Scholar manque de profondeur historique, puisqu'il ne recense que les articles disponibles en ligne. Les données de qualité de Scopus ne couvrent également que la période contemporaine, puisque les noms d'auteurs et les établissements d'attache ne sont indexés que depuis 1996. C'est le WoS qui a le plus de profondeur historique, grâce à une indexation qui remonte à 1900, même si les données pour le début du XX^e siècle sont clairsemées.

Quant aux titres, presque tous les articles sources du WoS se retrouvent dans Scopus. Plus précisément, la couverture de Scopus dans le domaine des sciences sociales et des arts et humanités est plus vaste que celle du WoS, qui est restée stable, voire a décliné au fil du temps. Après les avoir exclues au début, Scopus s'efforce d'indexer les publications en arts et en humanités, notamment les revues nationales, qui jouent un rôle central dans nombre de disciplines. Les seules disciplines pour lesquelles la couverture du WoS est plus étendue que celle de Scopus sont celles des sciences naturelles, possiblement pour des raisons historiques (par exemple la création d'indices spécifiques pour la chimie). Il est impossible de comparer avec précision les revues indexées par Google Scholar avec celles des deux autres index. Il est cependant probable que Google Scholar, qui ne recense que des articles publiés en ligne, couvre un périmètre moins étendu des arts et humanités, puisque dans ces domaines une proportion importante des revues n'est encore disponible qu'en format imprimé et que les contenus en accès libre sont plus rares.

Le degré de standardisation des données varie lui aussi d'une base de données à l'autre. Le WoS a la réputation de fournir des

données de meilleure qualité, grâce à plus de cinquante années de pratiques éprouvées. En effet, les premières consignes d'indexation ont été formulées avant l'ère numérique, quand le peu d'espace disponible sur les disques exigeait le tri minutieux de renseignements uniformes. L'indexation se faisait sur des documents imprimés, ce qui facilitait la prise en compte de différents aspects comme le type de document, indexé manuellement, et l'adresse institutionnelle, abrégée de manière standardisée. Quant à Scopus, qui dépend beaucoup de métadonnées fournies dans un tout autre but (publier plutôt qu'extraire), cette indexation précise lui a fait défaut, du moins pour les premières années. Aujourd'hui, même si la qualité des données fournies s'est améliorée, elle demeure en deçà de celle du WoS, notamment pour l'adresse institutionnelle des auteurs. Google Scholar ferme la marche, avec les données les moins fiables des trois – aucun nettoyage ni indexation manuels, seulement des algorithmes automatisés.

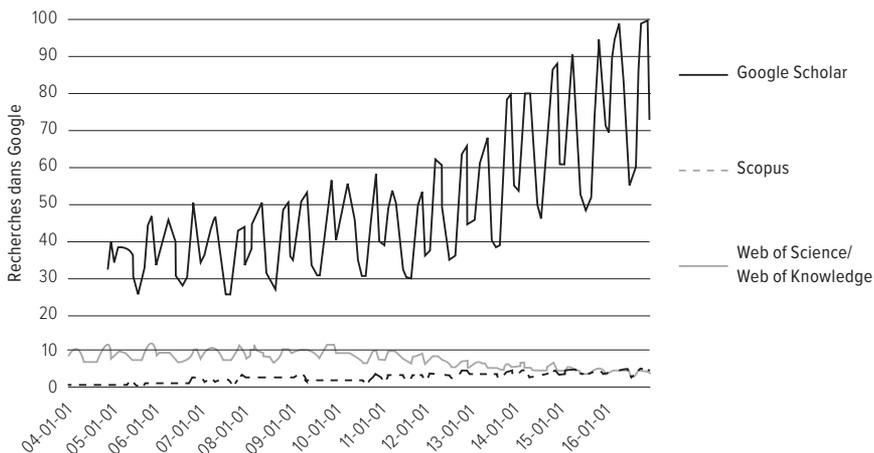
En ce qui concerne la désambiguïsation des noms d'auteurs, Scopus s'avère le plus précis. Les premiers index de citations de Clarivate Analytics n'attribuaient pas d'identifiant unique aux auteurs, et l'indexation des prénoms ne remonte qu'à 2008. Google Scholar attribue d'entrée de jeu des documents aux auteurs, mais ceux-ci doivent trier leur liste manuellement pour éliminer les erreurs. La fiabilité de la désambiguïsation est donc extrêmement variable: l'exactitude des données dépend de la rigueur de l'auteur quant au paramétrage et à l'entretien de son profil. Cette question semble peut-être relever du détail technique. Elle est cruciale, pourtant, du point de vue de la bibliométrie à l'échelle individuelle: sans désambiguïsation efficace des noms d'auteur, les mesures de chacun sont remplies d'erreurs, ce qui a des conséquences importantes sur son rang au sein de la communauté scientifique.

Un autre enjeu est celui des types de documents. Le WoS et Scopus répartissent les textes en catégories: articles originaux,

articles de synthèse, éditoriaux, lettres à l'éditeur, etc. Google Scholar ne fait aucune distinction de ce genre, ce qui ne permet pas de restreindre l'analyse aux documents évalués par des pairs (généralement les articles originaux et les articles de synthèse). Même si les catégories de Scopus et du WoS ne se recoupent pas parfaitement, leurs définitions se rejoignent sur ce qui constitue un article original ou un article de synthèse, les deux types d'articles que l'on considère habituellement comme des contributions à l'avancement des connaissances – originales et évaluées par les pairs.

Compte tenu de ces réserves concernant l'uniformisation des données, on s'étonnera peut-être de l'usage distinctif qui est fait de ces bases de données, et de Google Scholar en particulier. La figure 2.2 illustre le volume de recherches effectuées dans Google

FIGURE 2.2 Recherches dans Google (max = 100) pour les trois principaux index de citations, 2004-2016



Source: Google Trends.

pour chacun des trois index depuis 2004. C'est une limite en soi, puisque la plupart des utilisateurs du WoS et de Scopus emploient probablement un moteur de recherche intégré au portail de leur bibliothèque. Quoi qu'il en soit, il est évident que la popularité de Google Scholar augmente à un rythme qui n'a rien à voir avec celui de Scopus et du WoS. Alors que l'écart entre les trois bases de données demeure relativement stable jusqu'à la fin de 2011, le lancement de la nouvelle version de Google Scholar, qui permet aux chercheurs de créer leur propre profil de citations, provoque une forte hausse de l'intérêt pour cette base, en même temps qu'une lente diminution relative de l'intérêt pour le WoS, qui cède la deuxième place à Scopus en 2015. Bien que les tendances illustrées dans la figure soient fort probablement fonction de l'usage des plateformes pour le repérage, et non pour l'évaluation de la recherche, il n'en demeure pas moins que Google Scholar occupe une part grandissante du marché – Scopus suit le mouvement, mais de loin.

En résumé, il y a plusieurs éléments à considérer quand on utilise une source de données: l'équilibre entre exhaustivité et accessibilité, la qualité des données, et le type d'analyse que l'on souhaite réaliser. Ainsi, Google Scholar ne convient pas aux indicateurs de collaboration ou de performance des établissements; il ne peut qu'être appliqué (avec prudence) à des chercheurs, à de petits groupes, ou à des revues. Scopus a certaines limites pour les analyses historiques, et le WoS employé seul ne donne pas de bons résultats bibliométriques à l'échelle des individus – il faut lui adjoindre un outil de désambiguïsation. On doit considérer également certaines difficultés liées aux types de documents et aux disciplines: les chercheurs des sciences sociales, des arts et humanités sont mieux couverts dans Scopus que dans le WoS, alors que les deux bases de données donnent des résultats semblables pour les chercheurs des sciences naturelles et médicales. La justification la plus courante du choix d'un index est la commodité; mais

les chercheurs ont tout intérêt à connaître les limites inhérentes à chacune des bases de données, et à s'interroger sur la pertinence de celle qu'ils choisissent au regard du but qu'ils veulent atteindre.

Quels sont les biais culturels des sources de données ?

Les sources de données à partir desquelles on construit les indicateurs de la recherche ne sont pas exemptes de biais. Elles sont en fait le produit de contextes historiques, politiques, économiques et sociaux. On le voit très clairement dans les biais importants qui caractérisent leur périmètre linguistique et géographique. Même si l'on admet communément que le WoS n'est pas représentatif de l'ensemble de la recherche publiée, beaucoup de gens croient encore erronément qu'il en représente un échantillon généralisable. Malheureusement, contrairement à ce que l'on aurait obtenu avec un échantillon stratifié ou déterminé par quota, la représentation des auteurs par pays montre une inégalité considérable. Les auteurs des États-Unis comptent pour près du quart de tous les auteurs, tant dans le WoS que dans Scopus. La Chine, le Royaume-Uni, l'Allemagne et le Japon suivent (dans cet ordre) dans la liste des pays les plus productifs, et ce, dans les deux bases de données. Le 6^e rang est occupé par l'Inde dans Scopus, par la France dans le WoS. Cette différence est un artéfact créé par la disparité des sources indexées par chacune des bases de données : Scopus couvre davantage de pays non occidentaux, et le WoS, davantage de pays occidentaux.

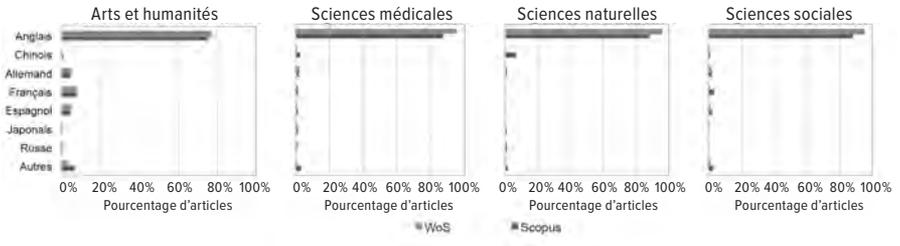
La simple comparaison des sources ne révèle rien, pourtant, de la relation entre l'étendue de la couverture et la production réelle de savoir dans le monde. On évaluera mieux cette relation à partir du répertoire de périodiques Ulrich (*Ulrich's Periodical Directory*), puisque son objectif est l'indexation systématique de tous les périodiques existants, avec les métadonnées relatives à leur langue et à leur pays d'origine. Ce répertoire, même s'il

ne fournit pas d'information sur le volume de production ou les auteurs qui figurent dans les revues, nous donne une idée de la couverture géographique et linguistique des titres qu'offrent les diverses bases de données. Des études montrent que les Pays-Bas, le Royaume-Uni et les États-Unis sont surreprésentés dans tous les domaines, c'est-à-dire que les revues publiées dans ces pays sont recensées d'une façon disproportionnée par les principaux index de citations. Cela s'explique probablement par la concentration de maisons d'édition installées dans ces pays. Il existe cependant des différences entre les disciplines. Dans les sciences sociales, par exemple, l'Australie et l'Allemagne sont surreprésentées au sein du WoS, mais sous-représentées dans Scopus. En arts et humanités, les périodiques français sont davantage représentés dans le WoS, alors que les périodiques chinois, indiens et russes le sont beaucoup plus dans Scopus.

Un examen de la langue de parution des travaux révèle des parallèles. La langue de diffusion du savoir a changé radicalement au cours du siècle dernier. Au début du XX^e siècle, l'allemand, l'anglais et le français se partageaient le marché de la publication scientifique en parts quasi égales. Cela a changé après la Deuxième Guerre mondiale, quand l'anglais et le russe sont devenus les deux langues de publication privilégiées, surtout dans les sciences naturelles et médicales. L'effondrement de l'Union soviétique a conduit à la diminution marquée de l'usage du russe, faisant désormais de l'anglais la *lingua franca* de la science. Si le WoS et Scopus présentent un biais favorable à l'anglais, force est de constater qu'il en va de même pour l'ensemble du monde scientifique.

L'anglais est la langue de 95 % et de 87 %, respectivement, des articles indexés par le WoS et Scopus (figure 2.3). Cette tendance s'observe dans les quatre grandes sphères scientifiques. La place plus grande laissée à la documentation parue en d'autres langues que l'anglais est une conséquence du périmètre plus large couvert

FIGURE 2.3 Pourcentage des articles de Scopus (1996-2015) et du Web of Science (1980-2015), par langue

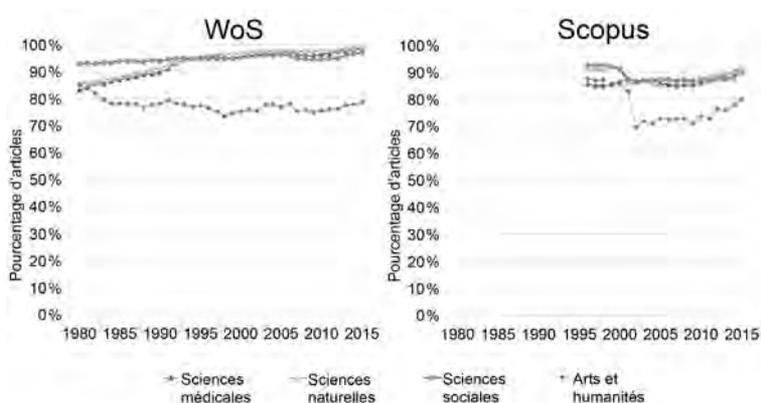


Source: Web of Science et Scopus, données agrégées selon les disciplines de la NSF.

par Scopus, puisque les deux bases de données ont le même noyau de périodiques en anglais. Elles partagent aussi le même bassin de langues les plus fréquemment employées – le chinois, l’anglais, le français, l’allemand, le japonais, le russe et l’espagnol –, bien qu’entre celles-ci, la dominance varie selon la base de données et la discipline auxquelles on s’intéresse. La proportion d’articles écrits en chinois est plus grande dans Scopus que dans le WoS (4,2 % *vs* 0,3 %). Les articles écrits en russe sont mieux couverts dans le WoS que dans Scopus, ce qui est sans doute l’effet de leur couverture historique dans le WoS. La présence du français et de l’allemand est équivalente dans les deux bases de données, mais, pour les arts et humanités, la proportion d’articles dans ces langues est plus grande dans le WoS que dans Scopus. Les langues qui ne font pas partie du groupe des sept premières couvrent 3 % des articles dans Scopus, mais moins de 1 % dans le WoS.

Ces tendances ne sont pas stables chronologiquement (figure 2.4). Depuis 35 ans, la base de données du WoS subit une anglicisation marquée – sauf pour les arts et les humanités –, ce qui semble refléter l’anglicisation générale de la recherche. Alors que l’anglais était la langue de 83 % et 85 % des articles en sciences naturelles et médicales en 1980, ce nombre est passé, en 2015, à 98 % et 99 %

FIGURE 2.4 Pourcentage d'articles en langue anglaise, Web of Science (1980-2015) et Scopus (1996-2015)



Sources: Web of Science et Scopus, données agrégées selon les disciplines de la NSF.

respectivement. La tendance est beaucoup moins prononcée dans Scopus, qui montre une diminution marquée de l'anglais dans ses premières années, suivie d'une hausse depuis quelques années. Enfin, bien qu'il n'existe pas de données comparables pour Google Scholar, les premières analyses montrent un fort biais en faveur de l'anglais, attribuable à l'usage exclusif par cette base de données d'articles publiés en ligne (et à la dominance de l'anglais sur le Web). Dans l'ensemble, ces résultats confirment la prépondérance, dans les index contemporains, de la recherche publiée en langue anglaise.

Comme pour les données sur le pays de publication, la comparaison des index ne reflète pas la représentation des langues dans l'ensemble de la littérature publiée. La comparaison avec le répertoire Ulrich des titres de périodiques montre que Scopus et le WoS surestiment tous deux les publications en langue anglaise dans tous les domaines, de même que le français et l'allemand

dans les sciences naturelles. Dans les sciences médicales, le WoS affiche une importante surreprésentation de la recherche présentée en ukrainien, alors que celle-ci est sous-représentée dans Scopus; et Scopus favorise le français, alors que cette langue est sous-représentée dans le WoS. La couverture de l'anglais est nettement disproportionnée en sciences sociales et en humanités – plus encore qu'en sciences naturelles et médicales. Ces différences montrent donc à la fois les écarts entre les périmètres linguistiques et la nécessité d'interpréter correctement les données en fonction de l'index utilisé.

Malgré les disparités que nous venons de souligner, plusieurs défendent l'idée que les bases de données bibliométriques couvrent simplement les travaux scientifiques les plus cités et, par extension, les plus déterminants. Une telle affirmation néglige la portée des périodiques nationaux, qui publient souvent des travaux qui s'avèrent capitaux à l'échelle nationale. En santé, par exemple, une grande part de la recherche effectuée dans un pays donné porte sur des sujets qui touchent directement sa population. Cela s'explique en partie par le fait que ce domaine de recherche relève par nature de l'intérêt public, et en partie par la disponibilité de sujets pour les expériences: un grand nombre de travaux portant sur des sujets humains se servent d'échantillons de commodité, et ils ont besoin d'une population atteinte assez nombreuse à proximité. Dans ce domaine, par conséquent, il n'est pas rare que les travaux sur des problèmes de santé soient publiés dans des revues à portée nationale. De plus, comme les systèmes sociaux ne sont pas identiques d'un pays à l'autre, plusieurs autres sujets d'intérêt public rejoignent uniquement une population locale. L'éducation, la santé et l'administration publique en sont des exemples: elles diffèrent notablement entre les pays, et la valeur qu'on leur accorde est donc plus élevée à l'intérieur du périmètre national. La conséquence est qu'un grand nombre des références faites par ces études renvoient elles aussi à cet échelon – d'où un

impact moins visible sur le plan international et nettement plus marqué sur le plan national. Voilà qui plaide pour des indicateurs normalisés, capables de rendre compte de tels effets aussi bien que des publications nationales qui rendent visibles ces études. Dans le même ordre d'idées, le fait d'encourager les chercheurs à publier dans des revues internationales a comme effet négatif de limiter la diffusion de la recherche sur des sujets locaux ou nationaux, principalement parce que les périodiques d'envergure mondiale sont très peu susceptibles de publier des articles d'intérêt national – particulièrement lorsque ces nations ne sont pas les États-Unis ou le Royaume-Uni.

Cela explique l'émergence, depuis près de deux décennies, de banques de données ayant pour objectif de rendre accessibles les travaux d'un pays ou d'un domaine en particulier. Au Brésil, SciELO a fort bien élargi le périmètre national, en fournissant aux périodiques scientifiques de ce pays – de l'Amérique latine et de certaines parties de l'Afrique également – une couverture beaucoup plus complète que celle du WoS. Malheureusement, c'est à grande échelle que les index de citations sont le plus efficaces : l'échantillon couvert par le périmètre, en soi, ne suffit pas ; il faut pouvoir recenser tous les documents cités. Dans le WoS, par exemple, environ 75 % des articles cités sont indexés par la base de données (ce sont des documents sources). Par contraste, les index de citations locaux, en raison de leur taille, renvoient beaucoup moins souvent à leurs documents sources. Ce sont des bases de données qui, pour utiles qu'elles soient à l'évaluation de la production nationale, conviennent moins à la formulation d'indicateurs d'impact. En fusionnant les sources de données nationales avec des ensembles plus vastes (comme le WoS), on arrive néanmoins à mettre en relation l'impact national et l'impact international. Certains de ces index nationaux sont en cours d'intégration à des bases de données plus grandes. Clarivate et SciELO, notamment, travaillent ensemble à la fusion de celui-ci dans celle-là.

La symbiose est avantageuse pour les deux parties, puisque le WoS peut faire valoir une couverture internationale élargie et que SciELO réalise sa mission de promouvoir la recherche brésilienne et latino-américaine.

Comment définit-on les disciplines ?

Les disciplines sont des unités primordiales dans la mesure de la recherche. Il existe un énorme corpus de textes qui s'efforcent de définir les disciplines et débattent de ce qui les constitue. Au sens le plus large, une discipline regroupe, autour d'un domaine particulier du savoir, une communauté de chercheurs disposant d'une infrastructure de recherche et de diffusion des connaissances. Les composantes de cette infrastructure comprennent, entre autres, les programmes de doctorat, les canaux de diffusion des connaissances comme les périodiques, les sociétés savantes. Aux fins de l'évaluation de la recherche, ce sont les différents sujets ou thèmes d'une classification donnée qui servent d'équivalents aux « disciplines ». La classification disciplinaire doit contenir des unités – des subdivisions – significatives pour un ensemble donné de chercheurs. La plupart des systèmes de classification sont divisés en unités disciplinaires qui correspondent aux départements des universités ou à la nomenclature des grands colloques ou périodiques d'un domaine. Si l'unité d'analyse est trop petite, elle n'est pas reconnaissable; trop grande, l'évaluation n'aura pas de sens. C'est l'histoire de Boucle d'Or appliquée à la scientométrie: l'unité d'analyse doit être juste de la bonne taille. L'interprétation et la normalisation en dépendent.

Il existe de nombreux systèmes de classification et de catégorisation. Le WoS et Scopus ont chacun le leur pour les périodiques; ils emploient les mots-clés et d'autres ressources fournies par les indexeurs et les éditeurs. La National Science Foundation établit une classification des disciplines utilisée afin de mesurer

différents aspects de l'enseignement supérieur étatsunien. Celle de ProQuest sert à catégoriser les mémoires et les thèses. La taxonomie de la Library of Congress sert à classer les livres. Les vedettes MeSH permettent d'organiser la recherche biomédicale au moyen d'un système hiérarchique de classement par mots-clés. En ce qui concerne l'évaluation de la recherche, la validité de ces systèmes de classification est nécessaire s'ils doivent servir à la normalisation par domaine, aux analyses par discipline, ainsi qu'à l'étalonnage et aux comparaisons. Comme pour tous les indicateurs, leur validité est déterminée par l'exactitude avec laquelle ils représentent le concept qu'ils évaluent. Dans le cas de la bibliométrie, c'est le degré de correspondance entre la classification disciplinaire et la structure du savoir, ainsi que la précision avec laquelle les unités sont rattachées aux disciplines, qui sont déterminants.

La grande majorité des classifications disciplinaires en scientométrie s'effectue à l'échelon des revues, et le WoS comme Scopus rattachent les revues qu'ils indexent à une ou plusieurs disciplines. Le WoS emploie deux modèles différents: l'un par domaine de recherche et l'autre par sujet. Le premier divise le monde de la recherche en cinq grands domaines et 151 disciplines; le second répartit les revues en 252 sujets. On attribue un domaine de recherche à toutes les ressources indexées par le WoS, tandis que les sujets, plus précis, sont réservés aux documents qui constituent la collection centrale du WoS.

Scopus emploie une classification semblable. Il répartit les revues en quatre grandes catégories, les « grappes » (*clusters*), puis en 27 champs d'études principaux qui se redivisent ensuite en plus de 300 champs d'études secondaires. Dans chacun des index, un même périodique peut appartenir à plus d'une catégorie. Bien que les résultats de recherche de Google Scholar ne fournissent aucune information sur la discipline – une caractéristique qu'il a perdue au fil des années –, le portail permet une classification des revues

par domaine, ce qui sert à construire des indicateurs de citations pour les revues. Aucun détail n'est fourni quant à la méthode de classification employée.

Quand, dans un même travail de recherche, on veut comparer des données ou employer plusieurs sources de données, il est souvent nécessaire de fusionner ou de superposer les systèmes de classification. On arrive assez facilement à une interopérabilité des données des revues, grâce à la standardisation des titres ou des identifiants directs qui leur sont attribués (tel l'ISSN), et qui servent donc de mécanismes transversaux. C'est l'un des avantages des systèmes de classification par périodique. L'autre est un relatif immobilisme, propice aux analyses synchroniques et diachroniques.

Mais la classification par discipline à l'échelle des périodiques n'est pas sans inconvénients, le plus problématique étant que la discipline d'une revue ne corresponde pas toujours à la discipline de chacun des articles qu'elle publie. Les revues phares *Science* et *Nature*, par exemple, sont associées à toutes les disciplines dans Scopus et classées parmi les revues multidisciplinaires dans le WoS. Cela noie l'information concernant les articles qu'elles publient. Cette situation peut se produire en outre au sein même des disciplines: le *Journal of the American Chemical Society*, par exemple, publie des articles sur la chimie des polymères, la chimie organique, la géochimie, l'agrochimie, etc. Chacun de ces sous-domaines a des pratiques de citation distinctes, ce qui crée des inégalités entre ces spécialités quand leurs données sont regroupées: certaines spécialisations sont plus productives que d'autres, ou présentent un taux de citations plus élevé.

La normalisation dépend de l'homogénéité: quand les unités sont trop hétérogènes, cela crée de la distorsion et provoque des disparités dans la perception de la performance et du mérite relatif de certains domaines. Prenons l'exemple du Research Excellence Framework, au Royaume-Uni. Imaginons un groupe

de chercheurs très prolifiques dont les travaux sont souvent cités par leurs pairs; ce groupe relève des sciences mathématiques. L'unité que constitue ce groupe paraît très productive et reçoit beaucoup de ressources. Mais un jour, quelqu'un décide que les travaux de certains individus dans le groupe relèvent en fait de l'informatique. Avec la même production et les mêmes citations, ces chercheurs ne paraissent plus que moyennement productifs et influents par rapport à leurs pairs de la nouvelle discipline, et on ne leur accorde plus autant de ressources. On voit ainsi qu'il y a des conséquences directes à la classification par discipline des unités – groupes de recherche ou individus.

Les disciplines doivent aussi demeurer relativement stables dans le temps. Ce n'est pas toujours une évidence, puisque certaines évoluent très vite par rapport à d'autres. Or les mesures qui servent à la comparaison doivent reposer sur des paramètres uniformes. Un trop grand dynamisme, et l'évaluation des tendances devient difficile, tout comme la formulation d'indices de comparaison. Il existe aussi des variations dans les niveaux d'agrégation des domaines, lesquelles sont fonction de la production et du périmètre. Dans la délimitation des domaines d'études du WoS et de Scopus, la physique est divisée en plusieurs sous-domaines, mais pas la sociologie; pourtant, les deux disciplines sont équivalentes dans la structure habituelle des universités (chacune forme un département). Dans l'ensemble des index, la classification par domaine est plus détaillée pour les sciences naturelles et médicales que pour les sciences sociales et humaines, ce qui signifie que la normalisation de celles-ci est beaucoup moins précise.

Comme on l'a souligné, chaque base de données a sa manière de classer les disciplines. Ces différences sont dues au fait qu'on a créé leurs classifications disciplinaires pour faciliter la recherche documentaire, et non pour mesurer la production scientifique. C'est pourquoi il peut s'avérer très difficile de quantifier simultanément la recherche dans différents genres et d'évaluer la cohé-

rence et la validité des systèmes de classification. L'élaboration de politiques adéquates en matière de recherche dépend pourtant de la compréhension non seulement de la production de certaines unités, mais aussi de la relation entre unités (entre les étudiants au doctorat d'un domaine et la production de la recherche dans ce domaine, par exemple).

Il est parfois intéressant, pour les administrateurs des universités, de donner un rang à un individu au sein d'une discipline. Dans ce cas, on se réfère habituellement au département auquel l'individu s'associe ou aux revues dans lesquelles il publie pour lui assigner une discipline. Mais cela peut produire de graves distorsions, en particulier dans le cas des personnes qui cultivent un profil pluridisciplinaire. Les modèles de catégorisation supposent souvent l'exclusivité – chaque unité étant rattachée à une seule discipline. Cela s'avère problématique dans les domaines où la multidisciplinarité est de rigueur, en plus de ne pas offrir une image exacte de la science telle qu'elle se pratique aujourd'hui.

L'un des moyens souvent employés pour résoudre cette tension entre exclusivité et multidisciplinarité consiste à recourir à des classifications reposant sur les articles eux-mêmes. Ces systèmes-là s'appuient sur la cartographie tracée par les références des articles, ou sur l'information fournie par les résumés, les titres et les mots-clés. On extrait ces données dans le but notamment d'améliorer la classification des articles selon les domaines de spécialisation, de prévoir l'évolution des tendances de la science et d'explorer l'interdisciplinarité. Les percées de l'analyse thématique et des algorithmes de regroupement ont augmenté la possibilité de produire des classifications disciplinaires à partir des articles. Celles-ci ont toutefois leurs limites. Elles exigent une grande puissance de traitement informatique, l'accès à des métadonnées de qualité et, dans de nombreux cas, l'accès au texte intégral des articles.

Une telle approche inductive de la classification disciplinaire – qui repose sur les relations, et qui varie par conséquent

avec chaque nouvel ensemble annuel de données – est utile pour repérer des tendances, mais problématique quand il s’agit de quantifier la recherche au fil du temps en fonction de catégories stables. De plus, les résultats en sont souvent difficiles à interpréter : certaines de ces analyses ont divisé la science en plus de 20 000 sous-domaines – bien trop pour qu’on puisse en tirer des nuances statistiques significatives. De plus, comme ces sous-spécialisations sont définies selon l’algorithme en fonction de la similarité entre les documents, l’étiquetage se fait après coup et manuellement. Le nombre de sous-spécialisations rend cet étiquetage non seulement spectaculairement chronophage, mais presque impossible à structurer, puisqu’il requiert une connaissance approfondie de tous les domaines de spécialisation de la science. Comme si cela ne suffisait pas, étant donné que les groupements disciplinaires émergent des données, les groupes d’articles ne sont pas structurés de la même façon pour les champs à densité élevée de citations que pour ceux où les citations sont moins fréquentes : les sciences médicales paraissent fragmentées, tandis que les sciences sociales forment un amas. Cela n’est pas représentatif de ce que nous savons de la recherche, ni des structures organisationnelles de ces domaines. Bref, la cartographie de la littérature scientifique à l’échelle des articles offre une représentation précise des relations entre les documents, techniquement très avancée dans sa construction, elle est souvent structurellement, épistémologiquement ou sociologiquement incohérente. Les classifications effectuées à l’échelle des périodiques proposent des catégories plus cohérentes, qui balisent l’espace du savoir de la même façon que le font spontanément la majorité des chercheurs d’un domaine.

3

LES INDICATEURS

Comment l'autorat d'un travail est-il défini et mesuré ?

La notion d'autorat est au cœur de la bibliométrie : ce qu'elle mesure, ultimement, ce sont les auteurs et les caractéristiques sociodémographiques apparaissant dans leurs publications (institutions, pays, genre, p. ex.). Dans l'histoire, le mot *auteur* a généralement fait référence à ceux qui écrivent. Dans son usage contemporain, toutefois, il recouvre une multitude de contributions à un document savant : rédaction, certes, mais aussi conception, analyse, expérimentation, ou fourniture de ressources. Parfois aussi, certaines personnes qui participent aux travaux ne figurent pas dans la liste d'auteurs. Malgré cette complexité, le fait de signer une publication savante demeure le marqueur essentiel – l'attribut mesurable, par conséquent – d'une contribution à la recherche.

Les pratiques d'attribution du statut d'auteur varient selon les disciplines. La notion classique a toujours cours dans les arts et les humanités et la plupart des sciences sociales : l'auteur, c'est celui qui écrit. La rédaction est la forme de contribution qui permet d'obtenir le statut d'auteur, et les tâches techniques ou autres sont rarement reconnues. Compte tenu du mode d'investigation et des pratiques établies, les listes d'auteurs sont en général très

courtes dans ces disciplines. Des différences s'observent toutefois entre les spécialisations et entre les pays. Si l'on compare, par exemple, les sociologues de la France à ceux des États-Unis, on observe des listes d'auteurs plus courtes et davantage de travaux de nature qualitative dans le premier pays, et un travail plus souvent quantitatif, conduisant à des listes d'auteurs plus longues, dans le second.

Le cas extrême dans le spectre des pratiques d'autorat est la physique des hautes énergies. L'un des laboratoires principaux du domaine, le CERN (Centre européen pour la recherche nucléaire), maintient une liste d'auteurs standardisée, qui inclut tous les chercheurs établis qui consacrent au moins 50 % de leur temps au CERN, ainsi que tous les doctorants et postdoctorants qui y consacrent 100 % du leur. Tous ces noms figurent par défaut dans la liste d'auteurs de toutes les publications, bien qu'un chercheur puisse retirer le sien s'il le souhaite. Ces auteurs demeurent sur la liste encore un an après avoir quitté le CERN. Un élément important à considérer au sujet de ces collaborations à grande échelle est que les noms apparaissent en ordre alphabétique, ce qui rend la contribution relative des individus impossible à déterminer à partir de la liste d'auteurs. On doit cette façon particulière d'attribuer la paternité d'une recherche à la complexité de l'outillage technique, qui demande la contribution de centaines, voire de milliers d'individus. Une publication de 2015 comptant plus de 5000 auteurs illustre parfaitement le propos : des 33 pages de l'article, 24 sont consacrées à la liste des auteurs. Une telle démesure met bien en évidence la complexité du double rôle de la paternité intellectuelle : reconnaître les apports de chacun, mais aussi, en corollaire, engager leur responsabilité à l'égard du projet de recherche.

Les disciplines de laboratoire demandent souvent un travail collaboratif – même si ce n'est pas à l'échelle de la physique des hautes énergies. Dans ces disciplines, l'ordre des auteurs dans la

liste contient souvent des renseignements implicites sur l'importance de leur contribution. Généralement, la première personne mentionnée est celle qui a réalisé l'essentiel du travail (c'est souvent un doctorant ou un postdoctorant), tandis que la dernière est le chef du laboratoire, ou un autre chercheur établi qui a contribué au travail conceptuel sans avoir nécessairement participé aux expérimentations. Ceux dont le nom apparaît au milieu sont la plupart du temps les techniciens, les étudiants ou autres collaborateurs associés à une tâche unique – comme la réalisation des expériences. Le statut et la contribution de chacun s'inscrivent implicitement dans l'ordre de la liste.

On notera que cet ordre, employé comme indicateur de contribution, perd de sa précision à mesure que la liste des contributeurs s'allonge. Cette situation a donné lieu à une nouvelle façon de concevoir et de mesurer les contributions dans laquelle les apports individuels sont rendus explicites – un peu comme le générique d'un film. De nombreuses revues ont adopté la pratique qui consiste à consigner ces informations dans une section désignée de l'article. Mais les données sont souvent difficilement accessibles, ce qui rend à peu près impossibles les analyses systématiques à grande échelle – et, par conséquent, la construction d'indicateurs. Cela dit, un changement est en cours depuis quelques années; on voit de plus en plus de périodiques réunir ces données et les publier avec d'autres métadonnées. Cela permet d'aller au-delà de la mesure du nombre d'articles attribués à un individu, et de mesurer la nature exacte de la contribution d'une personne à un document savant. La mise en place de la taxonomie CRediT et son adoption par les périodiques du site PLOS (Public Library of Science) devraient donner de l'essor à la collecte de ce type de données et à la construction d'indicateurs de contribution. Pour l'instant, comme les grandes bases de données n'indexent pas les contributions (outre l'autorat), le concept n'a pas encore d'indicateur lui correspondant.

Malgré l'existence de normes concernant l'octroi du statut d'auteur, telles celles de l'ICMJE (International Committee of Medical Journal Editors), certaines pratiques non éthiques persistent. Les deux plus connues sont l'autorat fantôme et l'autorat honorifique. Dans le premier cas, une personne contribue au savoir contenu dans un article sans que son nom figure parmi ceux des auteurs: on ne lui donne pas le crédit (ni la responsabilité) qui lui revient. Le second cas est l'exact opposé: un individu figure parmi les auteurs de l'article sans avoir joué de rôle, du moins aucun rôle significatif, dans la production des connaissances publiées. La difficulté d'identifier ces autorats problématiques se double des variantes entre disciplines concernant la reconnaissance du travail intellectuel. Un historien, par exemple, pourrait avoir profité grandement du soutien d'un doctorant pendant ses recherches et, tout en reconnaissant son apport dans la section des remerciements, ne pas lui accorder la place qui lui reviendrait comme auteur dans d'autres disciplines. Avec une contribution semblable à un article de sociologie, par exemple, le doctorant aurait eu une publication à faire valoir à titre d'auteur. D'autres inégalités dans l'établissement et l'application des critères d'autorat touchent le rang qu'occupent les individus dans la hiérarchie scientifique. Un étudiant de doctorat, par exemple, devra fournir beaucoup de travail pour que son nom figure parmi ceux des auteurs, tandis qu'un chercheur établi aura le même résultat avec un apport plus petit. Ces inégalités sont associées aux effets saint Matthieu et Matilda (voir le chapitre 1): production d'avantages cumulés pour les chercheurs d'expérience, et de manques à gagner pour les jeunes chercheurs, les femmes et les chercheurs au statut précaire. La collecte généralisée des données de contribution devrait éliminer ces pratiques contraires à l'éthique, puisqu'elle rend explicite la répartition du travail ayant donné lieu à la production scientifique.

Comment la production en recherche est-elle définie et mesurée ?

En recherche, la production est définie comme la quantité d'extrants – articles, livres et actes de colloque – publiée par différentes unités de recherche (individus, groupes, établissements, notamment). Dans cette perspective, plus le nombre d'extrants d'un chercheur ou d'un groupe de recherche est élevé, plus la capacité perçue de production scientifique est élevée. Lorsqu'elle est quantifiée, la production est exactement ce qu'elle prétend être : une mesure de quantité plutôt que de qualité. Ces deux notions sont malheureusement amalgamées quand la production est valorisée hors de tout contexte. Le fait que les indicateurs de production servent de fondement à l'attribution de ressources et autres marqueurs de reconnaissance encourage la production, parfois au détriment de la qualité. Dans l'ensemble, les indicateurs de production permettent, par exemple, aux universités de connaître les domaines où l'activité de recherche de leurs membres est importante, ou aux gouvernements de vérifier si l'augmentation de l'activité de recherche dans un domaine est le fruit d'investissements ciblés.

Les chercheurs emploient différents canaux de communication pour diffuser leurs recherches. Les livres, les articles de périodiques et les actes de colloques sont depuis longtemps les formes privilégiées. Les lettres à l'éditeur et les articles de synthèse documentent les interactions des membres de la communauté savante. Les chercheurs rendent également accessibles de plus en plus de jeux de données et de sources non textuelles, ce qui augmente la quantité de connaissances disponibles. Pour quantifier la production de la recherche, toutefois, on se limite habituellement aux articles originaux et aux articles de synthèse – les *éléments citables* traditionnels du Web of Science (WoS) –, ainsi qu'aux actes de colloques et aux monographies, quand ces documents sont indexés. C'est là une limite bien connue des indicateurs de

production, qui sont bien souvent fonction des sources de données disponibles.

La méthode la plus simple pour quantifier la production de la recherche est le comptage unitaire des publications savantes, qui accorde une « unité d'article » (c'est-à-dire 1 article) à chaque entité (auteur, établissement, pays, etc.) indiquée dans l'article. C'est l'équivalent du comptage des publications dans un c.v., où est énumérée chaque publication à laquelle un chercheur a contribué. La même règle s'applique quand des données sont agrégées à une plus grande échelle : chaque unité de recherche (centre de recherche, université, pays) reçoit une unité pour chaque publication à laquelle elle est associée. Cette méthode de comptage intégral a comme conséquence une exagération de la production réelle : compte tenu des travaux faits en collaboration, la somme des articles de tous les chercheurs (ou de toute autre unité d'agrégation) excédera toujours le nombre de documents distincts réellement produits par ces unités. En 2015, par exemple, le WoS indexait un peu plus de 2 millions de documents. Ceux-ci contiennent plus de 5,2 millions d'adresses institutionnelles et plus de 10 millions de noms d'auteurs, ce qui prouve l'étendue de l'inflation associée au comptage intégral des publications savantes. Cette inflation crée un écart entre le concept de production scientifique et l'indicateur, en surestimant la contribution de chaque unité de recherche à l'unité de savoir produit. Cette méthode de comptage pourrait contribuer, d'ailleurs, aux attributions non éthiques du statut d'auteur (tel l'autorat honorifique), dans la mesure où le fait d'ajouter des auteurs à la liste n'enlève rien aux autres coauteurs. Malgré ces inconvénients, le comptage unitaire est la méthode la plus employée en bibliométrie, surtout à cause de sa simplicité. Les résultats obtenus avec cet indicateur doivent être interprétés comme le nombre d'articles auxquels une unité donnée *a contribué*.

Le comptage fractionné est une solution de rechange au comptage intégral : avec cette méthode, chaque auteur reçoit une fraction de « l'unité article ». Dans un article signé par trois personnes, chacune reçoit le tiers de l'unité article ; le même fractionnement s'applique aux établissements, aux pays ou à tout niveau d'agrégation. L'avantage du comptage fractionné est que le nombre de publications de toutes les unités de recherche recensées est égal à la production réelle. En revanche, l'interprétation des résultats est plus difficile : un score de 33 articles fractionnés pourrait être le résultat de 33 articles publiés en solo, ou de 330 articles pour lesquels il y avait en moyenne 10 auteurs. Chacun de ces cas de figure donne lieu à une interprétation nettement différente de la production scientifique : le comptage fractionné donne une indication de la contribution proportionnelle, mais aucune quant au nombre d'articles publiés. Par conséquent, on aura une compréhension plus fine de la production et, dans une certaine mesure, de la collaboration, en recourant aux deux méthodes de comptage simultanément.

En revanche, ces deux méthodes de comptage ont en commun l'inconvénient de supposer que tous les auteurs contribuent de façon égale à la production du savoir. Une troisième méthode, le comptage harmonique, suppose au contraire que l'ordre de la liste d'auteurs est associé à l'importance de la contribution, la première personne mentionnée étant celle qui a le plus contribué à la production de connaissances. Le comptage harmonique accorde au premier auteur la fraction de productivité la plus grande et aux suivants une fraction de ce que le premier a reçu ; le troisième reçoit une fraction de ce que le deuxième a reçu, etc. Bien que cette méthode reflète un peu mieux la proportion des différentes contributions, elle néglige le rôle prépondérant joué par le dernier auteur mentionné, qui est souvent le directeur du laboratoire et l'auteur de correspondance.

Bon nombre d'indicateurs bibliométriques reconnaissent la prépondérance du premier auteur, du dernier auteur et de l'auteur de correspondance en leur attribuant, et à eux seuls, une part de la production. Ce faisant, ils excluent la contribution de nombreux joueurs importants. Ni le comptage harmonique ni la méthode des auteurs principaux ne fournissent le nombre total d'articles produits par une unité donnée, non plus que sa contribution relative. En outre, l'analyse des auteurs principaux est plutôt un indicateur des rôles de leadership associés à un article donné; elle ne peut servir qu'aux disciplines qui utilisent l'ordre descendant pour marquer l'importance des contributions, et qui réservent la dernière place à l'un des auteurs principaux.

Toutes les méthodes de comptage sont fortement corrélées lorsque les données sont agrégées à un haut niveau, mais cette corrélation est variable sur le plan micro – particulièrement sur le plan individuel. En outre, les différences entre les disciplines sont extrêmement marquées. Les auteurs du domaine de la physique des hautes énergies, par exemple, sont susceptibles, de manière disproportionnée, de contribuer à davantage d'articles que ceux des autres disciplines, et aussi d'avoir davantage de coauteurs. Dans ce groupe, les mesures par comptage unitaire ont tendance à donner des résultats très élevés, tandis que par comptage fractionné, ils sont très bas. On comprend donc que les niveaux de production sont fortement liés aux pratiques d'attribution du statut d'auteur, une caractéristique dont les analyses transdisciplinaires doivent tenir compte. On ne peut pas, à cause de ces différences significatives, comparer les niveaux de production sans les normaliser en fonction de la discipline.

Comme tous les indicateurs bibliométriques, ceux qui mesurent la production de la recherche dépendent des sources de données utilisées. Une analyse fondée sur le WoS conclurait, par exemple, que 27,7% des publications savantes dans le monde proviennent des États-Unis et 14,6% de la Chine. Scopus montrerait

une légère différence, soit 22,5% pour les États-Unis et 15,9% pour la Chine. Il ne faut pas oublier, cependant, que le comptage de ces documents est un indicateur des connaissances indexées, et non le reflet de la production réelle du savoir en Chine. Une part non négligeable de la recherche publiée en Chine n'est pas recensée par les index de citations généralement utilisés. Ceux-ci fournissent en fait des indicateurs de production fondés sur la littérature scientifique recensée, et c'est ainsi qu'on doit les interpréter. S'il est possible de montrer la différence entre valeurs réelles et valeurs mesurées à partir de l'analyse d'un c.v., l'échelon auquel on évalue habituellement la production scientifique demande l'usage d'indicateurs à grande échelle. Mais il est possible également de compiler des indicateurs de production pour une unité donnée à partir de bases de données nationales combinant plusieurs sources, de manière à couvrir presque entièrement la production savante. Certains pays dont la production est sous-estimée par les index bibliométriques dominants, comme les pays scandinaves, ont mis au point de telles bases de données qui sont gérées localement.

On doit distinguer enfin la production de la productivité, qui considère dans ses calculs les intrants du système scientifique. La productivité d'un département universitaire peut se mesurer en divisant le nombre d'articles publiés par ses membres, par le montant des subventions externes obtenues, ou par le nombre de chercheurs au département. Mais la décision de recourir à ces indicateurs doit être soigneusement pesée par les administrateurs et les responsables des politiques, car cela peut avoir certains effets pervers, telle l'augmentation de la productivité sans augmentation correspondante de la qualité, voire une baisse. C'est pourquoi les indicateurs de productivité sont souvent employés de concert avec d'autres indicateurs, afin de donner une image plus complète de la production scientifique, tenant compte non seulement de la quantité de travaux produits, mais aussi de leur qualité.

Comment la collaboration en recherche est-elle définie et mesurée ?

La recherche est une entreprise collective, on en convient de plus en plus. Les interactions entre chercheurs sont variées, tantôt formelles, tantôt informelles. Du côté informel, ils discutent entre eux au cours de conférences et dans les couloirs de l'université; ils font lire les versions préliminaires de leurs articles à des collègues, en consultent d'autres sur des questions de méthode ou d'analyse, co-supervisent des étudiants aux cycles supérieurs et partagent des données et autres résultats de recherche. De manière plus formelle, ils partagent les installations d'un centre de recherche ou les ressources d'une subvention dans le cadre d'un projet conjoint. Toutefois, bon nombre de ces interactions ne laissent pas de traces mesurables et, ainsi, la collaboration en recherche est généralement mesurée à partir des cosignatures apparaissant dans une publication. Un document qui compte plus d'un auteur est considéré comme le produit d'une collaboration; c'est pourquoi les relations de co-autorat servent de fondement aux indicateurs de ce type. Les formes de collaboration qui ne se concluent pas par une copublication demeurent ainsi invisibles du point de vue de la bibliométrie.

La collaboration, telle qu'elle est mesurée par les copublications, augmente dans tous les domaines, y compris les humanités, où la norme jusqu'à récemment était la publication individuelle. Plusieurs explications ont été tentées dans le but de rendre compte de ce phénomène: la complexité et l'interdisciplinarité toujours croissantes de la recherche, la simplification des échanges grâce à la puissance de l'infrastructure technologique et à la plus grande mobilité des chercheurs, les programmes de financement transnationaux qui encouragent le partage du personnel et des ressources au-delà des frontières, et le coût accru des technologies dans certains domaines (la physique des hautes énergies, au cas où

vous n'y pensiez plus). Il est intéressant de se demander, toutefois, si l'accumulation de noms d'auteurs en tête des articles est bien le reflet d'une augmentation absolue de la collaboration – du nombre de personnes, d'établissements et de pays qui auraient travaillé ensemble à produire un nouveau savoir. Il se pourrait bien qu'elle reflète surtout la reconnaissance tardive – et la part de responsabilité – des avatars du « technicien invisible » que décrit Steven Shapin. En recherche biomédicale, par exemple, le nombre d'auteurs a été multiplié par six. On peut admettre que cela soit dû à la complexité croissante des instruments à manipuler et à la nécessité afférente de s'adjoindre des collaborateurs spécialisés et complémentaires. Mais il est fort probable qu'on assiste aussi à un changement dans les pratiques de reconnaissance du travail scientifique et de la responsabilité associées à la contribution à un travail de recherche. Quoi qu'il en soit, ces pratiques varient selon les disciplines, et cela influence les indicateurs de collaboration et l'interprétation qu'on en fait. On s'aperçoit ainsi, en combinant les noms dans la section des remerciements et dans la liste d'auteurs, que les équipes de recherche en sciences sociales et en sciences naturelles – excluant évidemment la physique des hautes énergies – ont à peu près la même taille. Il est donc possible que la réduction du concept de collaboration aux seules données concernant les auteurs laisse de côté certaines contributions.

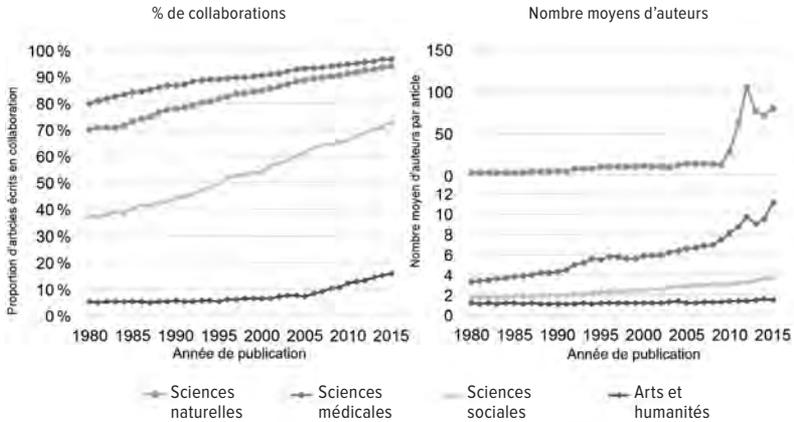
Bien que les indicateurs de collaboration soient rattachés aux individus – ceux qui signent ou cosignent la publication savante –, il est possible d'agréger les indicateurs de collaboration à plusieurs niveaux d'analyse, grâce aux métadonnées associées aux publications. En effet, chaque chercheur peut être affilié à un département et/ou un groupe de recherche, une université ou un centre de recherche, un secteur institutionnel et un pays. Par ses publications, chaque chercheur se rattache à un domaine de spécialisation, à des revues, à des disciplines. Tous ces niveaux d'analyse peuvent servir à construire des indicateurs de collaboration.

Le premier groupe d'indicateurs de collaboration est basé sur la proportion d'articles associés à plus d'une unité – ceux qui nomment plus d'un auteur, plus d'un établissement ou plus d'un pays. Le degré de collaboration varie significativement dans le temps et entre les disciplines. La proportion d'articles en collaboration dans les sciences médicales, naturelles et sociales augmente constamment depuis les années 1980, et frôle aujourd'hui les 100 % dans les deux premiers cas (figure 3.1, panneau de gauche). Dans les arts et les humanités, une augmentation s'observe à partir du milieu des années 2000, mais la collaboration demeure l'exception plutôt que la norme. Dans tous les cas, si les indicateurs de proportion révèlent l'existence de collaborations en recherche, ils ne disent rien de la taille des équipes.

Le deuxième groupe d'indicateurs de collaboration est basé sur des moyennes. Ces indicateurs rendent compte de la taille des équipes en compilant le nombre moyen (ou médian) d'unités qui contribuent à un travail de recherche donné. Calculer le nombre moyen d'auteurs pour un ensemble d'articles nous donne une idée de la taille des équipes, tandis que le nombre moyen de pays par article nous renseigne sur leur caractère international. Comme avec tous les indicateurs bibliométriques, on peut nuancer l'analyse en lui ajoutant différents niveaux d'analyse. Une question importante du point de vue des politiques scientifiques, par exemple, consiste à analyser non seulement le nombre de publications produites en collaboration, mais aussi la proportion de celles résultant d'une collaboration internationale ou multidisciplinaire. La formulation d'un tel indicateur exige des données de qualité, qui permettent de rattacher les auteurs aux variables appropriées (ici, la localisation géographique et la discipline).

Le nombre moyen d'auteurs (par article) est sujet à des asymétries provoquées par des aberrations statistiques. Certaines disciplines et spécialisations suscitent des résultats extrêmes: c'est le cas de la physique des hautes énergies (encore!), spécialité de la

FIGURE 3.1 Pourcentage d'articles et d'articles de synthèse écrits en collaboration (panneau gauche), nombre moyen d'auteurs (panneau droit) par domaine, 1980-2015



Source: Web of Science, données agrégées selon les disciplines de la NSF.

physique où le nombre maximal d'auteurs en 1980 était de moins de 100. En 1989, il avait atteint 500; en 2006, il avait quintuplé, et avait encore doublé en 2015, atteignant les 5000; l'effet de telles aberrations est évident dans le panneau droit de la figure 3.1. Il est donc nécessaire de tenir compte des aberrations statistiques et d'être attentif aux regroupements disciplinaires quand on se sert du nombre moyen d'auteurs comme indicateur de collaboration.

L'une des limites actuelles des indicateurs de collaboration est que, dans bon nombre d'analyses, on ne fait pas de distinction entre collaboration et coaffiliation – le fait pour un chercheur d'appartenir à plus d'un groupe ou établissement. Les indicateurs de collaboration supposent généralement que chaque auteur n'a qu'une affiliation, bien que les chercheurs soient de plus en plus souvent rattachés à plus d'un établissement et, par conséquent, à plus d'une localisation géographique. Dans ce cas, un article signé

par un seul auteur pourrait compter comme une collaboration entre établissements si l'auteur était affilié à deux établissements. Bien que certains considèrent une telle coaffiliation comme une forme de collaboration, cette vue ne fait pas consensus, et davantage de travaux seront nécessaires afin de mieux comprendre ces deux concepts.

Une autre limite est la nature non relationnelle de beaucoup d'indicateurs: s'ils arrivent à décrire le degré de collaboration caractéristique d'une unité, ils ne nous permettent pas de situer cette unité par rapport aux autres dans l'écosystème de la recherche. Certaines avancées dans l'analyse de réseaux permettent de prendre en compte la position et la relation entre les composantes d'un réseau de collaboration. Ces méthodes sont utilisées pour décrire la centralité des collaborateurs dans un réseau, quel que soit le niveau d'agrégation des données. En analyse de réseaux, les techniques de visualisation sont primordiales, et bien que les indicateurs de centralité (p. ex. d'intermédiarité, de proximité, de vecteur propre, etc.) fournissent de l'information signifiante aux statisticiens aguerris, c'est cette capacité de représentation visuelle qui rejoindra sans doute un vaste public et rendra compte de façon plus éloquente des relations entre divers acteurs du système scientifique.

Collaboration et production sont intimement liées, en recherche; les auteurs qui comptent le plus de publications sont généralement des chercheurs qui collaborent beaucoup. La collaboration répartit la charge de travail, ce qui augmente les extrants collectifs; mais cette augmentation doit être comparée à son pendant du côté des intrants. Avec le comptage intégral (ce qu'on voit sur un c.v., par exemple), les chercheurs les plus productifs du système de la recherche d'aujourd'hui le sont davantage que ce qui était observé auparavant, parce qu'ils parviennent à figurer comme auteurs sur un plus grand nombre de documents, souvent à titre d'auteur secondaire ou de directeur d'un groupe de recherche. En

revanche, avec le comptage fractionné, les chercheurs paraissent moins productifs qu'avant. Cette différence s'explique par l'écart de croissance entre le nombre d'auteurs et le nombre d'articles : le nombre d'articles publiés annuellement a connu une augmentation régulière, mais le nombre d'auteurs distincts a connu, lui, une augmentation plus marquée, ce qui a fait baisser le nombre de publications moyen par chercheur à l'échelle de l'ensemble du système. Il se pourrait aussi que ce résultat soit dû à l'augmentation du nombre d'auteurs transitoires – ceux qui ne contribuent qu'à de rares articles dans leur carrière (étudiants ou assistants de recherche, par exemple) –, qu'on ne comptabilisait pas nécessairement dans les listes d'auteurs auparavant. Ce que l'on constate alors ne serait pas un changement dans la productivité, mais dans les pratiques de désignation des auteurs.

Comment l'interdisciplinarité est-elle définie et mesurée ?

Une part toujours plus grande de la production du savoir, aujourd'hui, prend forme aux intersections et dans les interstices des disciplines traditionnelles. On dit de ces lieux qu'ils sont pluridisciplinaires, interdisciplinaires, multidisciplinaires ou transdisciplinaires. Sans que ces termes soient tout à fait synonymes, comme le montrent les discussions substantielles sur ce qui les distingue, on les emploie souvent de façon interchangeable, avec interdisciplinarité comme terme principal. Les indicateurs de l'interdisciplinarité reposent sur les classifications disciplinaires — il faut des disciplines pour mesurer l'interdisciplinarité — qui sont analysées de multiples façons afin d'apprécier l'interdisciplinarité d'un domaine d'études, d'un individu ou de toute autre unité de recherche. Les indicateurs de l'interdisciplinarité généralement employés sont fondés sur la collaboration entre unités disciplinaires ou sur les références et les citations.

Lorsque l'interdisciplinarité est mesurée par les pratiques de collaboration, chaque auteur est associé à une discipline, et on quantifie la présence, dans les publications co-écrites, de chercheurs provenant de différentes disciplines. Ce calcul est toutefois rendu compliqué par les difficultés liées à l'attribution d'une discipline à un chercheur, laquelle peut se faire par affiliation départementale ou par la discipline du doctorat (lorsque l'information est accessible, souvent dans des études à plus petite échelle). On peut également utiliser les disciplines dans lesquelles les chercheurs publient comme indicateur de leur appartenance disciplinaire, et mesurer les collaborations entre chercheurs qui publient dans des disciplines différentes. Mais, comme beaucoup de chercheurs publient dans un éventail de champs disciplinaires, très peu d'analyses ont été réalisées à ce niveau. Dans l'ensemble, compte tenu des difficultés d'assigner une discipline à un chercheur, la plupart des études faites sur la base de tels indicateurs de l'interdisciplinarité ont été effectuées à petite échelle, à partir de listes composées manuellement en associant les individus aux disciplines.

Les analyses de l'interdisciplinarité fondées sur les références et les citations sont, pour leur part, assez communes — les classifications disciplinaires des références citées étant elles-mêmes très répandues. Ces indicateurs de l'interdisciplinarité ont les documents plutôt que les individus comme unités d'analyse. La multidisciplinarité d'un document se mesure en fonction de l'usage qu'il fait de documents publiés dans d'autres disciplines (références), ou de l'impact qu'il a sur d'autres disciplines (citations). En comparant ainsi la discipline du document (ou périodique) citant et du document (ou périodique) cité, on parvient à mesurer diverses dimensions de l'interdisciplinarité. Au cœur de ces indicateurs se trouve la notion de diversité, généralement prise en compte de trois façons : (a) l'équilibre (différentes disciplines sont citées de manière égale), (b) la variété (la quantité de disciplines

citées) et (c) la disparité ou la similarité (le degré de différence entre les disciplines qui sont citées — par exemple physique et sociologie). Toutes ces mesures, soulignons-le, ne concernent que les documents sources, puisque c'est à ce sous-ensemble d'articles que l'on peut associer une discipline et une spécialité. En conséquence, les domaines où une part importante des références citent des articles recensés par la base de données – telles les sciences et la médecine – permettent de compiler des indicateurs plus précis, tandis que dans ceux où ce n'est pas le cas – telles les sciences sociales et les humanités –, les indicateurs sont beaucoup moins représentatifs des pratiques de recherche interdisciplinaires des chercheurs. Bien qu'il y ait de plus en plus d'intérêt pour les indicateurs de l'interdisciplinarité, la complexité du concept qu'ils visent à mesurer ainsi que leur relative indisponibilité sur les plateformes bibliométriques rendent leur utilisation plutôt limitée.

Comment l'impact est-il défini et mesuré ?

L'impact est un concept fondamental en bibliométrie et le Graal des responsables des politiques scientifiques, des administrations universitaires et des organismes subventionnaires. C'est sans doute l'aspect de la recherche le plus difficile à définir et à mettre en application. Depuis quelques dizaines d'années, on définit principalement l'impact de la recherche comme ce qui a un effet sur la communauté scientifique, effet que l'on mesure grâce aux citations. Cette définition repose sur un raisonnement relativement simple selon lequel, si un article en cite un autre, cela prouve une forme d'interaction et d'arrimage par lesquels l'article citant s'appuie sur l'article cité. Mais nombre de travaux s'intéressant à ce qui motive les citations montrent qu'il existe une variété de raisons pour lesquelles les auteurs font référence à d'autres travaux. Une citation peut servir, entre autres, à mettre un travail de recherche en contexte, à comparer des résultats, à justifier

l'emploi d'une méthode, à étayer un argument ou, au contraire, à le réfuter. Ces différences fonctionnelles ont conduit certains à défendre l'idée qu'une citation considérée seule est indicatrice d'une utilisation, et non d'un impact.

À petite échelle, l'interprétation des mesures de citations est extrêmement variable. Par contre, à grande échelle, on en tire une information précieuse. Pour le dire simplement, si l'on croit en la nature cumulative de la science – à tout le moins, en période de « science normale » –, alors l'accroissement des citations permet de représenter la croissance du savoir. La bibliographie d'un article savant contient le socle des connaissances existantes sur lesquelles s'appuient les nouvelles connaissances qu'il contient; on peut ainsi supposer que les documents qui sont plus souvent mentionnés dans les bibliographies ont un impact plus important sur la production de nouvelles connaissances que ceux qui le sont plus rarement, voire jamais. Lorsqu'elles sont agrégées, les citations permettent de montrer les effets d'un travail sur le paysage de la science. Plus encore, les citations révèlent à la fois la force de l'impact (par le nombre de citations) et sa nature (par les disciplines ou les pays qui citent le travail). Ces arguments tiennent mieux dans les disciplines dont la nature est cumulative que dans celles qui ne le sont pas, ce qui nous amène à formuler toutes sortes d'avertissements lorsque nous appliquons l'analyse des citations aux arts et aux humanités, ainsi qu'à certaines disciplines des sciences sociales.

La mesure de l'impact de la science sur la société et l'économie est au centre de nombreux débats et travaux chez les spécialistes de ces disciplines. Toutefois, les indicateurs de ces formes d'impact sont moins standardisés – ils sont plus difficiles à mesurer, surtout à grande échelle – et, en conséquence, sont demeurés en périphérie du champ de la scientométrie. Certains soutiennent que les références faites à des articles savants dans les demandes de brevet sont révélatrices de l'impact économique de la recherche

fondamentale, mais cet indicateur, techniquement difficile à compiler, présente un biais favorable aux disciplines hautement technologiques. Pour quantifier l'impact sociétal, on se sert des citations de travaux scientifiques dans les documents de politiques publiques ou les journaux. Depuis peu, l'omniprésence des médias sociaux – Twitter, Facebook, etc. – a mené à la recherche, dans ces plateformes, de signes d'un impact de la recherche sur la société qu'on peut mesurer rapidement. Regroupées sous le terme générique d'*altmetrics* – soit de mesures « alternatives » aux citations –, les mentions d'articles scientifiques sur ces plateformes seraient, selon certains, révélatrices de l'incidence de la recherche dans un bassin plus large d'intervenants. Toutefois, leur sens exact ne fait pas encore l'unanimité, principalement en raison de leur nature volatile et virale. On n'a pas encore tranché quant à leur capacité à rendre compte de l'impact des travaux scientifiques sur la société – du moins de la façon dont l'entendent habituellement les sociologues, les économistes et les responsables des politiques scientifiques. Ceux-ci cherchent des indicateurs permettant de différencier les travaux ayant une valeur scientifique élevée (les percées majeures, p. ex.) des travaux ayant des conséquences favorables pour la société (les innovations qui réduisent le coût des soins de santé, p. ex.), et capables en même temps de stimuler l'intérêt des chercheurs pour ces derniers. D'après les premières analyses, ces indicateurs hétérogènes fournissent dans certains cas des résultats faiblement corrélés avec les citations (Twitter, p. ex.), alors que dans d'autres, ils sont plus fortement corrélés avec les citations (lecteurs sur Mendeley, p. ex.), et permettent même de les prédire légèrement, compte tenu de la plus grande rapidité avec lesquels ils s'accumulent.

De nombreux autres concepts sont souvent amalgamés à celui d'impact, tant dans la construction que dans l'interprétation des indicateurs. En ce qui concerne les citations, on confond souvent l'impact avec la qualité. Si l'on peut considérer sans crainte qu'un

article cité plus de cent fois a plus d'effet sur la communauté scientifique qu'un article jamais cité, on se gardera bien d'affirmer qu'il est de meilleure *qualité*. Plusieurs raisons peuvent expliquer qu'un travail de recherche soit fréquemment cité, y compris l'actualité du sujet et l'utilité pour un vaste public. Un classement établi sur la foi d'un indicateur d'impact ne signifie pas que telle unité est de meilleure qualité qu'une autre, mais plutôt qu'elle a reçu davantage d'attention et que ses résultats ont été davantage utilisés – et, conséquemment, qu'elle a eu davantage d'impact sur la production de nouvelles connaissances.

La mesure de l'impact social soulève des inquiétudes semblables compte tenu de l'amalgame entre attention et impact. Un article peut être relayé un grand nombre de fois dans les médias sociaux à cause de son contenu humoristique ou populaire, sans nécessairement avoir d'impact social ou scientifique durable. En 2016, l'article ayant attiré le plus d'attention, d'après l'Altmetric Attention Score, était une communication spéciale parue dans le *Journal of the American Medical Association* et relayée par Twitter essentiellement à cause de la notoriété de son auteur : le président des États-Unis, Barack Obama. Bien qu'il y ait un relatif consensus sur l'interprétation des mesures de l'impact de la science sur elle-même, son impact sur la société, elle, demeure imprécis.

Quelle est la différence entre références et citations ?

Citations et références sont les deux revers d'une même médaille ; ils sont le même événement, soit l'acte de citer. Quand un article fait *référence* à un autre travail, ce dernier est *cité* ; en d'autres termes, la référence est l'acte de citer effectué dans le document citant, et la citation est le produit de cet acte pour le document cité. Cette distinction toute conceptuelle est très importante en bibliométrie. D'abord, bien que presque tous les travaux de recherche contiennent des références, tous ne sont pas cités. Ensuite, chronologiquement,

les références sont figées dans le passé : elles sont statiques, et ne vont ni croître ni changer au fil du temps. Les citations reçues, en revanche, se produisent dans le futur, sont fonction des nouveaux textes citant publiés, et sont un ensemble ouvert et dynamique.

Bien qu'elles soient le produit du même acte de citer, références et citations renvoient, comme indicateurs, à des concepts différents. Le recours à la quantification des citations pour mesurer l'impact des travaux scientifiques est très répandu, alors que l'usage des références l'est moins. Celles-ci, pourtant, peuvent nous apprendre beaucoup de choses utiles sur le monde de la recherche. Elles indiquent, entre autres, le type de travaux ayant servi à produire un ensemble donné d'articles, une information riche d'applications potentielles : les bibliothécaires pourraient décider de choisir leurs acquisitions d'après les travaux mentionnés par les professeurs de leur université, par exemple. La date de parution des références joue un rôle important dans la description des aspects structuraux de la science : on pourrait ainsi étudier les corpus utilisés par les chercheurs dans la production de nouvelles connaissances, et examiner, par exemple, les différences entre les populations, les disciplines, les établissements et les pays. De telles observations ont des applications directes en bibliométrie : la normalisation de certains indicateurs d'impact repose sur les références faites, qui permettent d'une certaine façon de cerner le domaine de recherche d'un article citant. Mais pour l'essentiel, ce sont les citations, et non les références, qui servent d'unité de base à la compilation d'indicateurs – principalement en termes d'impact de la recherche.

Pourquoi cite-t-on les travaux de recherche ?

La quête d'une théorie des citations a commencé en même temps que le déploiement des grands index les comptabilisant. L'un de ses objectifs consistait à dresser une typologie de leurs rôles et fonc-

tions dans le processus de diffusion des connaissances. Plusieurs ont été développées au fil des années, mais la plus durable est celle proposée dans les années 1970 par Michael J. Moravcsik et Poovanalingam Murugesan à partir d'articles sur la physique des hautes énergies. Pour décrire les fonctions des citations, ils ont eu recours à quatre dichotomies non exclusives. La première distingue les contributions conceptuelles et opérationnelles, à savoir si l'article est cité pour son apport théorique ou méthodologique. La deuxième distingue la nature organique ou superficielle du lien entre la référence et l'article citant, à savoir si la citation est nécessaire à l'article qui la cite, ou si elle ne sert qu'à orienter le lecteur. Les deux dernières catégories sont fortement reliées : le caractère évolutionnaire ou juxtaposé décrit le rapport entre texte cité et texte citant, à savoir si celui-ci poursuit la réflexion de celui-là ou s'il s'en détache pour proposer autre chose ; et l'axe tendu entre la conformité et la négation décrit la façon dont le document citant confirme ou infirme le document cité.

Aux premiers temps de la bibliométrie, la présence de citations négatives était préoccupante pour plusieurs. Certains chercheurs voyaient les références négatives comme un problème minant l'utilisation des citations comme mesures d'impact. Selon eux, les auteurs ne devraient pas accumuler de citations – c'est-à-dire du capital universitaire – si leurs travaux sont convoqués dans le seul but de les réfuter ou de les contredire. On a prétendu qu'il fallait identifier les citations négatives et les retirer des données des indicateurs. Deux arguments récurrents s'élèvent cependant contre cette proposition. Le premier est le faible taux de citations négatives : des études empiriques ont montré que ce taux oscille entre 1 % et 15 % selon la discipline et la période étudiées. Le second est que, dans la mesure où le concept envisagé est l'impact plutôt que la qualité, l'opinion à l'égard de la citation importe assez peu. Si un texte est cité très fréquemment, même pour des raisons négatives,

alors on ne peut ignorer son impact sur le domaine d'études: il fait avancer la science, ne serait-ce que par réfutation.

Les typologies comme celle de Moravcsik et Murugesan ont été étudiées «manuellement», en général: un examen attentif permettait d'inférer la fonction de la citation à partir de divers signaux repérés dans le texte. L'un des plus forts est l'endroit du texte où apparaît la citation. En sciences naturelles et médicales, par exemple, la plupart des articles sont organisés sur le modèle introduction – méthodes – résultats – discussion. Suivant cela, le caractère conceptuel ou opérationnel d'une citation est prévisible grâce à la position de la citation: dans le premier cas, elle sera dans l'introduction, et dans le deuxième, dans la section méthodes. Dans le même esprit, les articles cités dans l'introduction sont plus susceptibles d'avoir un lien plus superficiel au contenu, alors que les références organiquement liées au contenu devraient normalement figurer dans la discussion. Mais la position seule ne rendra pas compte du caractère évolutionnaire ou juxtaposé, ni conforme ou négateur du rapport de citation. Pour ceux-ci, il faut consulter directement le contexte immédiat de la citation. Plusieurs méthodes fondées sur l'apprentissage-machine s'efforcent d'analyser le contexte de citation sur de grandes quantités de données, en se servant à la fois de la taxonomie de Moravcsik et Murugesan, des méthodes de la linguistique informatique et de la disponibilité toujours plus grande des données du texte intégral.

Les indicateurs capables de tenir compte du contexte de la citation (et de ce qui la motive) fourniront des interprétations plus nuancées de l'impact d'un article. Mais cet aspect de la bibliométrie n'en est qu'à ses balbutiements. Pour repousser les limites actuelles et mieux comprendre la nature d'une publication à partir de la façon dont elle est citée, des études plus poussées du contexte de citation s'imposent; études qui sont de plus en plus possibles grâce à l'accessibilité des grandes bases contenant le plein texte des documents savants.

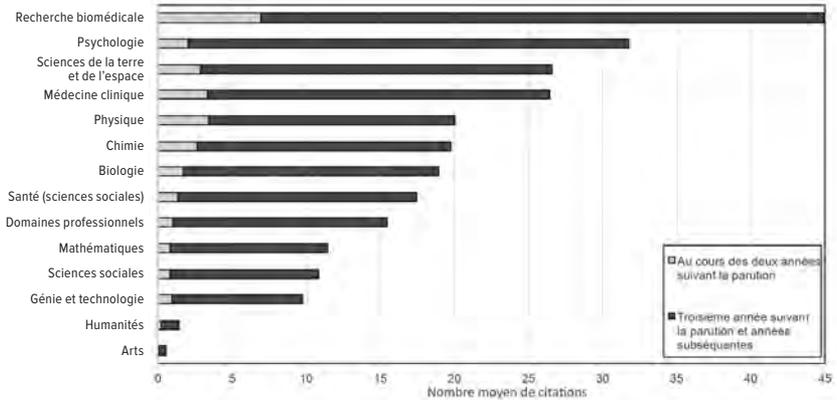
Comment les taux de citation varient-ils dans le temps et par discipline ?

Les différents taux de citation que reçoivent les documents savants sont le fait des différences structurelles entre les disciplines et les spécialités, le temps écoulé depuis leur parution et l'année où ils ont été publiés, entre autres. Un document publié dans une discipline où les publications et les citations sont nombreuses est plus susceptible d'être cité qu'un article publié dans une discipline où la densité des publications et des citations est plus faible. Plus précisément, plus la quantité de documents produits dans une discipline est grande, plus les occasions d'être cité sont nombreuses. À l'opposé, les documents publiés dans des domaines surspécialisés comptant peu de chercheurs ou ayant un rythme de publication plus lent ont moins d'occasions d'être cités. La nature cumulative de la science joue également un rôle important dans le fait d'être cité ou non. Dans les domaines où l'on s'appuie plus directement sur les travaux antérieurs, les citations s'accumuleront plus vite (et en plus grand nombre) que dans les domaines que l'on considère comme non cumulatifs.

La figure 3.2 illustre l'ampleur des différences entre disciplines selon le nombre de citations, à partir de l'ensemble des articles publiés en 1990 et recensés dans le Web of Science (WoS). Deux ans seulement après la publication, des différences se manifestent dans le taux de citation, et l'écart se creuse avec le temps. Dans la sphère des arts et des humanités, les articles reçoivent moins d'une et moins de deux citations, respectivement, pendant la période. Dans la recherche biomédicale, en revanche, les articles reçoivent en moyenne 7 citations au cours des deux premières années et 45 au cours des 26 années suivantes.

Certains facteurs extérieurs au contenu des articles et des revues influencent aussi les taux de citation. La langue de publication en est un, puisqu'elle détermine qui peut lire l'article et,

FIGURE 3.2 Nombre moyen de citations reçues par article publié en 1990, pour les deux années suivant la parution et les années subséquentes, par discipline

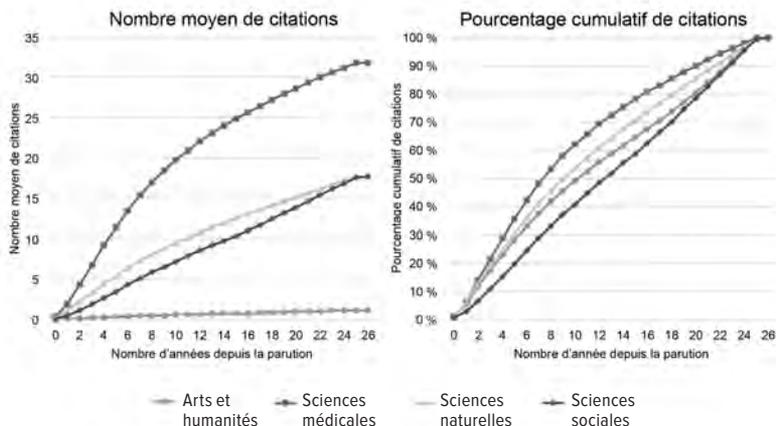


Source: Web of Science, disciplines de la NSF.

conséquemment, le citer. La notoriété et le prestige de la revue ou de la maison d'édition influencent aussi le nombre de citations reçues. Les documents qui sont indexés par les grandes banques de données bibliographiques ou publiés en libre accès obtiennent généralement des taux de citation plus élevés. La popularité du sujet, sa nouveauté ou son intérêt multidisciplinaire augmentent aussi le nombre de citations reçues. L'article qui ouvre un nouveau champ d'investigation est plus susceptible d'être cité que celui qui résout un problème déjà connu. Même le type de document (article de synthèse, article méthodologique ou article plus théorique) influence le nombre de citations potentielles: les articles de synthèse reçoivent en moyenne davantage de citations que les articles de recherche qui font connaître des travaux originaux.

Il faut du temps pour que les citations s'accablent. Dans le calcul des taux de citation, il est donc très important de tenir compte du temps écoulé depuis la parution d'un document. Le

FIGURE 3.3 Nombre moyen de citations reçues (panneau gauche) et pourcentage cumulé de citations (panneau droit) par nombre d'années suivant la parution et par domaine, pour les articles, les notes et les revues de littérature publiés en 1990

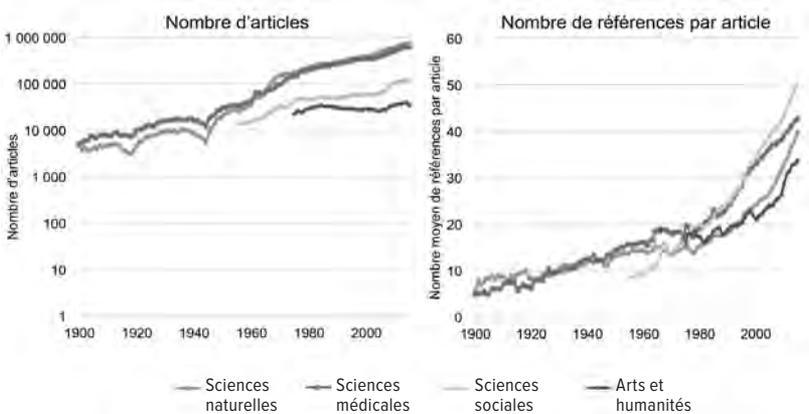


Source: Web of Science, données cumulées par disciplines de la NSF.

jour de leur publication, la plupart des articles ont un compte de citations nul (quoique l'usage croissant des prépublications tende à changer cela). Après un an, le taux d'accumulation diffère nettement selon les domaines, et en quelques années, celui des travaux parus en sciences médicales atteint le double de celui des travaux parus en sciences naturelles et sociales (figure 3.3). Ces éléments sont donc essentiels à la construction d'indicateurs de citations: on doit toujours garder à l'esprit le temps écoulé entre la parution de l'article et l'évaluation bibliométrique, ainsi que les différences entre les disciplines.

En plus du temps écoulé depuis la publication, on tiendra compte de l'année exacte où le document a été publié. Les taux de citation ont augmenté régulièrement tout au long du XX^e siècle, principalement à cause de l'augmentation du nombre de publi-

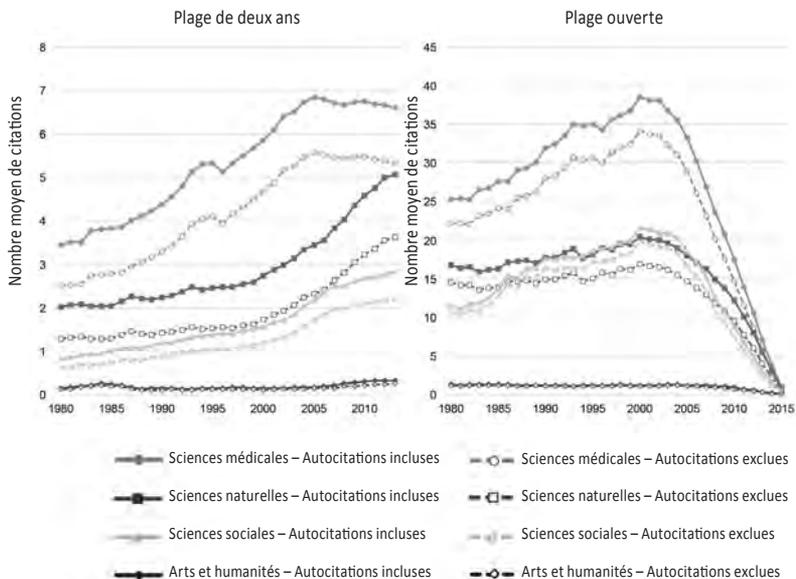
FIGURE 3.4 Nombre de documents (articles, notes et revues de littérature) publiés (panneau gauche) et nombre moyen de références par document (panneau droit), par domaine, de 1900 à 2015



Source: Web of Science, données cumulées par disciplines de la NSF.

cations et de la longueur des bibliographies – deux facteurs qui accroissent le nombre de citations potentielles (figure 3.4). On trouve des preuves de cette croissance en analysant la littérature indexée par le WoS, par discipline. À l'exception d'une diminution pendant la Première puis la Seconde Guerre mondiale, le nombre de documents indexés en sciences naturelles et médicales a crû constamment pendant le siècle dernier; pour les sciences sociales, ce nombre croît aussi depuis qu'elles sont recensées par le WoS. On observe un changement marqué dans le nombre de références par document, et ce, dans tous les domaines. Le cas des sciences sociales est particulièrement frappant: le nombre moyen de références est passé de moins de dix dans les années 1950 à plus de 50 en 2015. De telles variations structurales modulent les attentes quant au nombre de citations, et l'on doit en tenir compte quand on applique les indicateurs diachroniquement.

FIGURE 3.5 Nombre moyen de citations reçues par article (calculé avec, puis sans les autocitations), note et article de synthèse pendant les deux années suivant la parution (panneau gauche), puis jusqu'en 2015 (panneau droit), de 1980 à 2015



Source: Web of Science, données cumulées par disciplines de la NSF.

Ces différences structurales ont conduit, au fil des années, à une croissance du nombre moyen de citations des articles (figure 3.5). Pour les articles publiés en 1980 et indexés par le WoS, par exemple, le nombre moyen de citations reçues au cours des deux années suivant l'année de publication est de 3,5 en sciences médicales et de 2 en sciences naturelles; il est passé à 6,6 et 5, respectivement, 34 ans plus tard. On observe des variations similaires pour les sciences sociales, et les arts et humanités, même si, dans ce dernier cas, le nombre de citations demeure peu élevé (avec une moyenne de 0,32 citation par article pour les parutions

de 2013). Les tendances sont les mêmes quand on considère toutes les citations reçues – c'est-à-dire sans limiter la fenêtre de citation à 2 ans –, à la différence que l'écart entre citations et autocitations diminue. Voilà qui souligne l'importance de l'autocitation dans les quelques années qui suivent immédiatement la publication. La construction d'indicateurs de citations comparant des années entre elles demande donc une attention particulière, puisqu'il faut tenir compte des autocitations et de la discipline.

Qu'est-ce qui n'est pas cité ?

En 1990, la revue *Science* publiait un article au sous-titre accablant : «New evidence raises the possibility that a majority of scientific papers make negligible contributions to new knowledge» («De nouvelles données suggèrent que l'apport de la majorité des articles scientifiques au savoir est négligeable»). Cet article rapportait que la majorité (55 %) des articles publiés entre 1981 et 1985 dans les périodiques indexés par le WoS n'avait reçu aucune citation pendant les cinq années suivant leur parution. L'auteur soulignait qu'il s'agit pourtant des revues phares de la science – donc, implicitement, que le corpus des connaissances non indexées était probablement moins cité encore. Extrapolant les données d'une étude antérieure, il estimait que 80 % des articles publiés pendant cette période n'avaient jamais été cités plus d'une fois et, de là, il affirmait que «plus de la moitié – peut-être même plus des trois quarts – de la littérature scientifique est essentiellement inutile». Les nombreux aspects problématiques de cet article illustrent clairement les idées erronées et omniprésentes qui circulent au sujet de l'absence de citations.

Le premier problème est le présupposé selon lequel la valeur scientifique et le fait d'être cité sont intrinsèquement liés, c'est-à-dire que le but du travail scientifique serait d'être cité et que la valeur proviendrait de la citation. Dans cette perspective, un

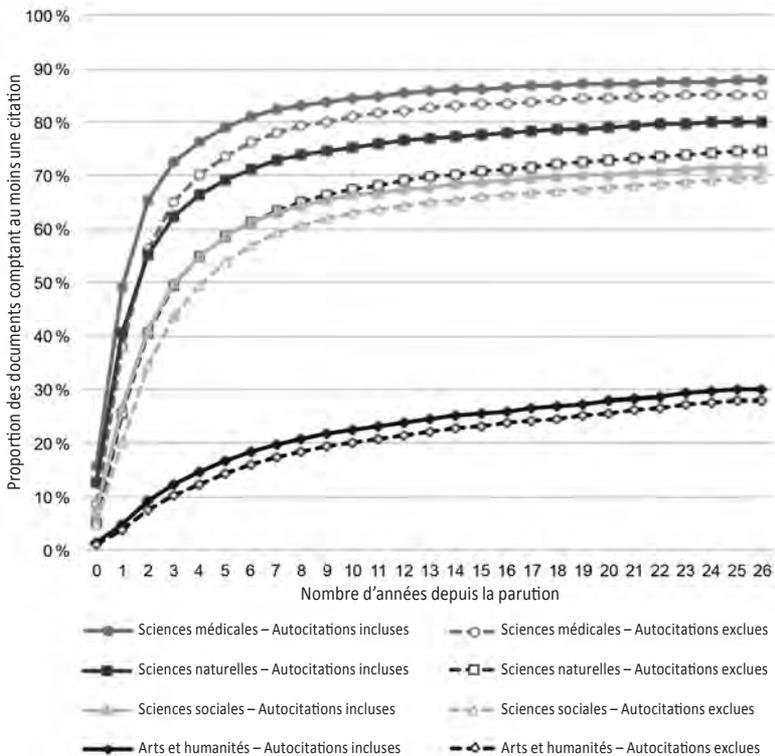
article non cité ou peu cité est, comme le suggère l'auteur, « essentiellement inutile ». Mais – c'est curieux tout de même de devoir le préciser – le but du travail scientifique n'est pas d'être cité, et la citation seule ne détermine pas la valeur du travail. La communauté scientifique devrait cesser de percevoir négativement l'absence de citation et entreprendre de la décrire afin de la comprendre.

Le deuxième problème est que l'article néglige, dans son analyse, le facteur temporel. Le fait d'être cité est fortement associé à la longueur de fenêtre de citation, et cette association varie selon les domaines. Pour qu'un article soit cité, quelqu'un doit l'avoir repéré et intégré à ses propres travaux, dont le produit doit ensuite être soumis à une revue, qui doit l'évaluer, le réviser, le corriger et le publier. Pour de nombreux périodiques, le processus d'évaluation dure plus de 6 mois (voire quelques années) et la file de numéros en attente de publication prolonge encore les délais. Par ailleurs, le temps requis pour produire un nouvel élément de savoir varie selon les disciplines. On ne s'attend pas, par exemple, à ce qu'un article en histoire soit abondamment cité dès l'année suivant sa parution.

La réalité, aujourd'hui, est que la plupart des articles en sciences médicales, naturelles et sociales sont cités quelques années après leur parution. Les citations prennent plus de temps à s'accumuler dans les arts et les humanités, mais environ le tiers des documents auront reçu des citations dans les 25 années suivant leur publication (une estimation conservatrice, étant donné le faible taux de couverture des articles sources dans ces sphères du savoir). Les autocitations influencent beaucoup les taux de citation des articles dans les années suivant leur publication, en particulier dans les sciences naturelles. En effet, comme le montre la figure 3.6 fondée sur les articles publiés en 1990, on voit que 55 % des articles sont cités au moins une fois pendant les deux années suivant la publication, quand on tient compte des autocitations ; ce nombre tombe

à 41% quand on les exclut du calcul. Ces observations suggèrent que les fenêtres de citation doivent être adaptées aux disciplines (plus courtes en sciences médicales, plus longues en sciences sociales), mais aussi que l'analyse des citations en arts et humanités, à cause de leur faible densité, demeure très problématique – tout particulièrement quand elle vise à évaluer la recherche.

FIGURE 3.6 Proportion d'articles cités (calculée avec, puis sans les autocitations) par nombre d'années suivant la parution et par domaine, pour les articles et les revues de littérature publiés en 1990



Source: Web of Science, données cumulées par disciplines de la NSF.

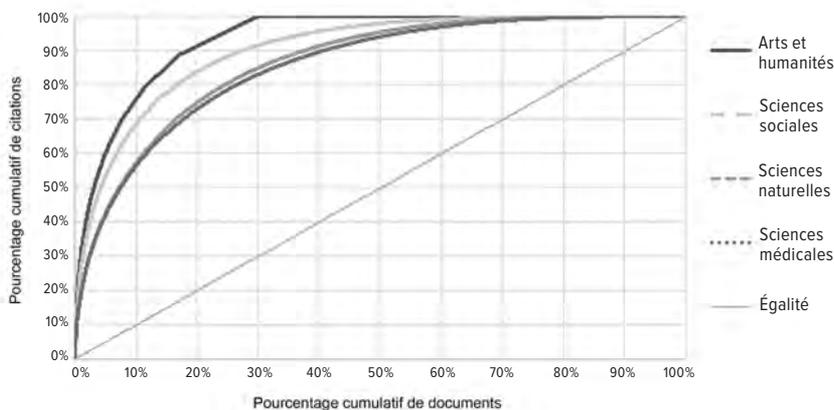
Le fait, pour un article, de ne recevoir aucune citation est devenu de plus en plus rare tout au long du XX^e siècle. Alors que plus de 80 % et 60 % des articles en sciences médicales et naturelles parus au début des années 1900 n'avaient reçu aucune citation dix ans après leur parution, une minorité d'articles publiés dans les dernières années est dans la même situation. Les documents sont en outre cités plus rapidement, une proportion plus élevée d'entre eux étant citée pour la première fois quelques années seulement après leur parution. Par ailleurs, le fait de ne pas être cité est souvent lié à la nature même de la recherche des différentes disciplines. En ingénierie, par exemple, la proportion d'articles non cités est plus élevée qu'en physique, une différence qui s'explique principalement par le caractère appliqué des disciplines du génie. En effet, dans les sciences appliquées, un grand nombre de travaux paraissent sous des formats qui ne sont pas recensés par les principaux index de citations (communications de colloque, revues professionnelles, rapports de recherche, etc.). Raison de plus pour distinguer soigneusement les notions de couverture et de taux de citation : les taux de citation dépendent de la couverture des articles citant par les bases de données servant à l'analyse. Dans les sciences médicales, la plupart des références faites pointent vers des publications indexées par la base de données. Résultat : des pourcentages d'articles non cités moins élevés que dans d'autres disciplines. Les pourcentages d'articles non cités plus élevés observés dans les arts et les humanités s'expliquent par la couverture moins étendue des revues nationales, des livres et d'autres types de documents non recensés par les index. L'invisibilité de ces documents entraîne une sous-estimation de leurs taux de citation. Dans l'ensemble, non seulement les documents sont de plus en plus cités, mais le caractère non exhaustif des index de citations font qu'ils le sont sans doute encore plus qu'ils ne le semblent.

Quel est le degré de concentration des citations ?

Les taux de citation sont asymétriques à l'extrême, on l'a vu : quelques documents seulement reçoivent la grande majorité des citations. La figure 3.7 montre, pour les quatre grands domaines scientifiques, le pourcentage des documents qui reçoivent un pourcentage précis de citations (courbes de concentration ou courbes de Lorenz). On voit, pour chacun des domaines, qu'une minorité de documents reçoit la majorité des citations ; on voit aussi qu'il y a des écarts entre les domaines. L'un des moyens bien connus pour étudier la concentration est la règle du 80-20 (ou loi de Pareto) qui, lorsqu'elle est appliquée aux citations, montre que 80 % des citations devraient provenir de 20 % des articles. Cela reflète assez bien les tendances observées, même si le chiffre exact varie selon la discipline : 80 % des citations renvoient à 12 % des documents en arts et en humanités, à 16 % en sciences sociales, à 24 % en sciences naturelles et à 26 % en sciences médicales. Une partie de ces différences entre disciplines s'explique par les articles non cités, qui représentent une proportion considérable des publications en arts et en humanités, et une petite proportion seulement dans les autres domaines.

Le degré de concentration des citations a varié énormément au cours du siècle dernier, accompagné d'une diversification croissante des articles cités et, subséquemment, d'une diminution de la concentration des citations. En 1940, par exemple, 20 % des articles recevaient 80 % des citations ; en 2005, ce 80 % se répartissait entre 40 % des articles. La dispersion s'observe aussi à l'échelon des revues : les travaux les plus cités paraissent maintenant dans un plus grand nombre de revues que par le passé. Alors que les *Proceedings of the National Academy of Sciences* publiaient presque 9 % du 1 % des articles les plus cités au monde en 1986, leur part baissait à moins de 3 % en 2010. On observe des tendances semblables pour tous les périodiques phares (*Nature*, *Science*,

FIGURE 3.7 Distribution cumulative de Pareto (courbes de Lorenz) des documents et des citations, par domaine, pour les documents publiés en 1990. La diagonale représente l'égalité parfaite.



Source: Web of Science, données cumulées par disciplines de la NSF.

Cell, p. ex.), tendances que vient corroborer, à l'échelle mondiale, la diminution du pouvoir prédictif du « facteur d'impact » des revues, ou FI, un indicateur sur lequel nous reviendrons plus loin. On observe également une diminution de la concentration géographique des citations reçues, ce qui est dû surtout à la diversification des pays et des établissements actifs en recherche et à une couverture améliorée de leurs extrants. La part de citations reçues par les articles provenant des États-Unis diminue, tandis que celle des articles provenant de la Chine, de l'Inde et d'autres pays émergents augmente. La concentration géographique des citations, par conséquent, diminue également.

Comment les citations sont-elles comptées ?

Plusieurs méthodes permettent de compiler des indicateurs fondés sur les citations, de la simple addition aux calculs plus complexes. Avant d'entrer plus avant dans le détail de chacune, précisons que les citations, du moins dans les principales bases de données, sont compilées de façon binaire : peu importe le nombre de fois où un document est cité par un document citant, il ne reçoit qu'une seule citation. Bien qu'il existe de nouveaux outils – Semantic Scholar en est un – qui prennent en compte le nombre de fois où un document est cité à l'intérieur d'une même source, l'approche binaire (cité/non cité) demeure la norme. Tandis que les indicateurs de production et de collaboration s'appuient, règle générale, sur des métadonnées (auteurs, établissements, revues, etc.), les indicateurs de citations exigent en plus certaines données relationnelles : les références dans les articles recensés doivent être liées aux articles cités correspondants. C'est ainsi que MEDLINE, qui ne recense pas les références faites par les articles qu'elle contient et qui n'est donc pas un index de citations, peut servir à calculer des indicateurs de production, mais pas de citations.

L'indicateur de citations le plus simple est le décompte, soit la somme du nombre de fois où un document donné figure dans la liste de références de documents citant. Le nombre de citations peut s'agréger à n'importe quelle échelle : périodique, université, pays... La question clé est de savoir comment agréger les citations comptées. Si la distribution asymétrique caractéristique des citations suggère de recourir à la médiane (la valeur qui partage la distribution en deux parties égales) plutôt qu'à la moyenne (la somme des citations obtenues, divisée par le nombre total d'observations), la plupart des analyses fondées sur les citations ont recours à la moyenne plutôt qu'à la médiane, entre autres parce que la valeur modale est souvent 1 ou 0, ce qui renseigne peu sur la distribution. Au-delà de la simple commodité (il est plus facile de

compiler des moyennes que des médianes), l'usage de la moyenne a l'avantage de tenir compte de tous les documents de la distribution, pas seulement de quelques-uns. Les médianes présentent aussi l'inconvénient d'avoir souvent une valeur très faible, proche de zéro. Les moyennes, en revanche, sont très sensibles aux valeurs extrêmes : dans les calculs fondés sur elles, un article cité 1000 fois aura autant de poids que 1000 articles cités une fois.

Une façon de remédier à l'asymétrie des citations consiste à ne considérer qu'une partie de la distribution et à présenter les données obtenues comme une proportion des articles les plus cités – 10 %, 5 % ou 1 % des articles les plus cités, par exemple. L'avantage de cette méthode est qu'elle ignore la longue traîne de la distribution, ce qui a comme effet d'encourager les chercheurs, plutôt qu'à miser sur la quantité, à publier moins d'articles, mais qui offrent une forte probabilité de citation. Cette méthode de calcul a pris son essor et elle est désormais intégrée aux principaux indicateurs, tel le classement universités que fait l'université de Leiden aux Pays-Bas.

Une approche semblable consiste à transformer la distribution entière en rangs centiles, au lieu de ne s'intéresser qu'à certaines zones de la distribution. Mais, étant donné la difficulté d'attribuer un centile distinct (de 1 à 100) à chaque article en fonction de son taux de citation (difficulté liée en particulier à la quantité d'articles faiblement cités dans certaines disciplines), les rangs percentiles sont habituellement cumulés en groupes. On trouve ainsi le groupe des documents qui sont sous le 50^e percentile, suivi du groupe qui s'étend du 50^e au 75^e percentile, puis du 75^e au 90^e, du 90^e au 99^e, et enfin la tranche supérieure équivalant à 1 % des documents les plus cités. Les comparaisons entre individus, établissements et pays se fondent sur l'examen des différences de proportion entre ces catégories. On peut ainsi mesurer, par exemple, la proportion de la production d'un groupe de chercheurs qui figure dans le 10 % supérieur des documents les plus cités. On

peut appliquer ce calcul à tous les niveaux d'agrégation. Les catégories par centile combinent ainsi les données sur les articles et les citations de manière nuancée, ce qui permet d'examiner le niveau de citation des articles et de le comparer au taux d'une discipline, d'un établissement ou d'un pays.

On se sert également des logarithmes des taux de citation, ce qui permet de rééchelonner les citations comptées sur une échelle moins asymétrique. Toutefois, même si cette façon de procéder semble fournir une solution simple à l'asymétrie des distributions des citations, elle n'est pas sans soulever certains problèmes concernant les articles non cités, puisque leur valeur de zéro ne peut pas être transformée logarithmiquement. Certains ont proposé d'autres solutions, comme la suppression des articles non cités du calcul ou l'ajout systématique d'une citation à chaque article, mais aucune n'a été adoptée à grande échelle. Les transformations logarithmiques ont par ailleurs comme conséquence de diminuer l'importance des articles très cités, qui sont, contrairement à ce qu'on croit spontanément, les travaux qu'on cherche à mettre en évidence et à encourager par la production d'un indicateur de citations.

Aux plus hauts niveaux d'agrégation des données, ces diverses méthodes de calcul sont fortement corrélées. Aux niveaux les plus bas, toutefois – ceux des chercheurs pris individuellement ou des petits regroupements –, les différentes méthodes de comptage produisent différents résultats. Ce constat vient renchérir sur l'importance de trianguler les indicateurs, c'est-à-dire d'en utiliser plusieurs afin de consolider l'analyse.

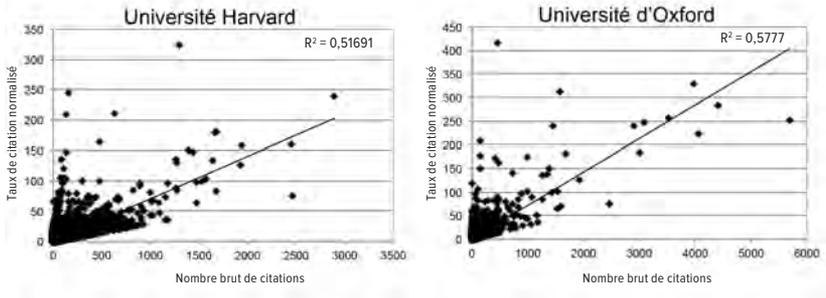
Comme nous l'avons déjà souligné, la discipline à laquelle un document est associé et la date de publication sont deux facteurs qui ont des effets non négligeables sur les taux de citation. Les techniques de normalisation servent à moduler ces différences. Dans le cas des indicateurs basés sur la moyenne, la procédure de normalisation habituelle consiste à diviser le nombre de citations

de chaque document par le nombre moyen de citations reçues par les documents publiés dans la même discipline ou spécialisation pendant la même année, ce qui produit des scores de citations qui s'étendent entre 0 et ∞ , avec une moyenne de 1 pour l'ensemble des articles analysés. Ces scores, appelés « indicateurs relatifs » ou « indicateurs normalisés par discipline », permettent de comparer des ensembles de documents publiés dans des disciplines différentes et à des moments différents: par exemple, l'ensemble des documents produits par les universités ou les pays. Des procédés semblables s'appliquent également aux indicateurs de citations basés sur des centiles: les citations sont réparties en rangs centiles en fonction de leur distribution dans une discipline et une année données, ce qui permet ensuite de comparer les rangs centiles où se classent des groupes de documents publiés à différents moments dans des disciplines différentes.

La figure 3.8 montre que les taux de citation normalisés par discipline et par année de publication demeurent néanmoins corrélés aux taux de citation absolus – c'est particulièrement vrai pour les documents qui obtiennent un nombre élevé de citations. La corrélation est plus faible, cependant, pour le sous-ensemble de documents recevant un nombre moins élevé de citations; à l'échelle de la discipline, cette variabilité pourrait être la conséquence d'une densité de citations plus faible; à l'échelle des articles, ce pourrait être un effet de la récence de la parution. Dans de tels cas, le dénominateur (la valeur attendue) est parfois si bas que le fait de recevoir ne serait-ce qu'une citation se traduit par un taux de citation normalisé très élevé.

Les citations s'analysent aussi relationnellement, ce qui produit des données permettant de dégager des réseaux de relations entre les documents. La cocitation et le couplage bibliographique sont les deux indicateurs fondamentaux de la relation entre deux documents. Des documents sont cocités quand ils figurent ensemble dans le groupe des références citées par un document

FIGURE 3.8 Relation entre le taux de citation normalisé, par domaine et année de parution, et le nombre absolu de citations, pour les documents de 2010 à 2015 de l'Université Harvard et de l'Université d'Oxford



Source: Web of Science. Normalisation effectuée par disciplines ou spécialisations de la NSF et année de parutions

citant: si l'article A cite les articles B et C, alors les articles B et C sont cocités. Le couplage bibliographique, quant à lui, repose sur la relation entre documents citant: des documents sont bibliographiquement couplés s'ils ont une référence en commun. Si les documents A et B font tous deux référence au document C, alors A et B sont bibliographiquement couplés. Ces mesures de relation se compilent à plusieurs niveaux d'agrégation, mais, en règle générale, les articles, les auteurs et les revues – ainsi que les disciplines auxquelles on les associe – sont les unités pour lesquelles on peut compiler des indicateurs utiles. Ces mesures sont intégrées aux systèmes de repérage d'information afin de suggérer de la documentation pertinente aux utilisateurs – on reconnaît la fonction de suggestion d'Amazon, par exemple –, mais aussi pour quantifier et illustrer les relations entre auteurs, établissements et disciplines. Les mesures de cette sorte peuvent indiquer, entre autres, la centralité d'un auteur ou d'un établissement, ou l'importance du rôle de relais d'une discipline par rapport à une autre.

Qu'est-ce qu'une autocitation ?

Autocitation et autoréférence sont des termes qui renvoient au fait, pour une entité de recherche (auteur, établissement, pays), de faire référence à un document en provenance de cette même entité ou d'en recevoir une citation. Imaginons qu'un chercheur écrit un article. Quelques années plus tard, il écrit d'autres articles où il se réfère à cette publication antérieure. Celui-ci reçoit par ailleurs des citations provenant d'autres articles écrits par d'autres personnes. Au total, 50 % des citations qu'il reçoit proviennent des articles qu'il a lui-même écrits. On appelle ces citations des autocitations. Ce concept (et les indicateurs qui lui sont associés) est distinct – c'est important de le souligner – de l'autoréférence. L'autoréférence est l'action de l'auteur en train de s'autociter, par définition dans le texte citant : l'auteur rédige un article et y insère une référence à l'une de ses publications antérieures. La citation que reçoit de ce fait le document cité est une autocitation : c'est le produit de l'autoréférence, vu dans le texte cité. Supposons que dix des trente références qui figurent dans un article renvoient aux publications de l'auteur de l'article. Dans ce cas, le tiers des références de cet article sont des autoréférences. Du point de vue mathématique, bien que les deux indicateurs aient le même numérateur – soit le nombre de documents inclus dans l'intersection de l'ensemble des auteurs citant et de l'ensemble des auteurs cités –, le dénominateur varie selon que l'on considère les références ou les citations. Le taux d'autoréférencement est la part des références qui renvoie à des articles publiés par l'auteur de l'article citant. Le taux d'autocitation est la part de citations reçues par un article donné qui provient de documents signés par l'auteur même de l'article.

Bien qu'en général, le terme « autocitation » désigne indifféremment l'une ou l'autre de ces réalités, le fait d'amalgamer autoréférence et autocitation mène à des conclusions inexactes.

L'autoréférence décrit la façon dont un auteur (ou un groupe d'auteurs) s'appuie sur ses travaux antérieurs pour faire avancer sa recherche actuelle. Cela suppose un bagage d'articles antérieurs pertinents, qui dépend lui-même de l'avancement de sa carrière, de la production et d'un programme de recherche bien focalisé. L'autocitation, elle, est compilée à partir de l'impact d'un travail sur la communauté scientifique. Un taux élevé d'autocitation suggère que la publication nourrit avant tout les travaux subséquents de l'auteur et qu'elle a peu d'incidence sur le reste de la communauté. En somme, autocitations et autoréférences sont deux types très différents de mesures et de concepts.

Il existe plusieurs types d'autocitations et d'autoréférences. Dans le cas des publications à un seul auteur, l'autocitation est facile à définir: c'est, tout simplement, quand le nom de l'auteur cité et de l'auteur qui cite est identique. Mais la recherche, ne l'oublions pas, est de plus en plus collaborative, et le calcul des autocitations dans les articles écrits en collaboration est plus complexe. Il est possible de compter une autocitation dès qu'il y a chevauchement d'un seul nom entre la liste des auteurs cités et celle des auteurs citant. Il est possible aussi de limiter l'autocitation aux cas où le premier auteur de l'article citant est le même que le premier auteur de l'article cité. Plusieurs autres correspondances (entre le premier auteur citant et n'importe quel auteur cité, p. ex.) et de multiples niveaux d'agrégation des données (par pays, par établissement ou par revue) sont également possibles.

À l'échelle individuelle, un nombre élevé de publications correspond souvent à un taux élevé d'autoréférence. C'est ce qui découle du fait d'avoir de nombreuses sources potentielles à exploiter. Le phénomène fonctionne à rebours quand on se sert de la discipline comme unité d'analyse: plus elle est petite, plus le nombre d'autocitations et d'autoréférences y est élevé. En effet, la spécialisation tend à faire augmenter la proportion d'autocitations, puisque le bassin de publications à citer est limité. Le même

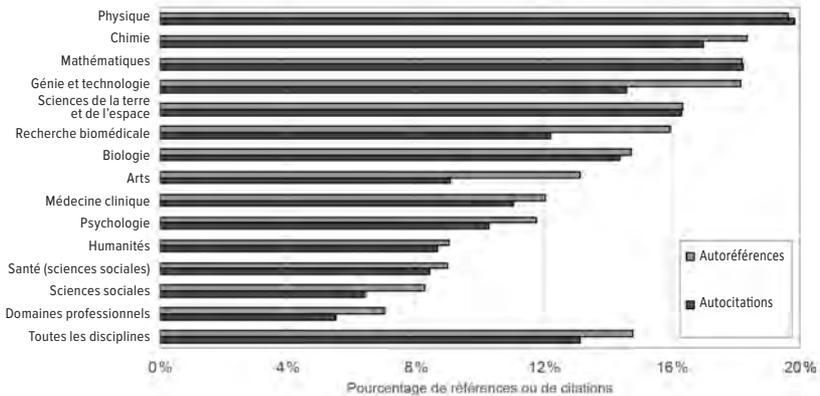
raisonnement s'applique aux revues très spécialisées. Du point de vue de la simple probabilité, plus un domaine de spécialisation est pointu et moins il compte d'individus, plus le nombre d'autocitations et d'autoréférences qu'on s'attend à y observer est élevé.

Tant l'autocitation que l'autoréférence ont augmenté au fil du temps, suivant l'augmentation du nombre moyen d'auteurs par document. Pour les documents publiés en 2000, tous domaines confondus, elles représentent environ 15 % des références et 13 % des citations (figure 3.9), quand on inclut toutes les différentes formes de chevauchement des noms d'auteurs cités et citant. Évidemment, les chiffres varient avec les disciplines : les proportions d'autocitations et d'autoréférences sont plus faibles dans les sciences sociales, les arts et les humanités, et plus élevées dans les sciences naturelles et médicales. Ces différences s'expliquent avant tout par le plus grand nombre d'auteurs par document dans ces derniers groupes de disciplines.

Autocitations et autoréférences varient également selon le lieu d'origine des auteurs et des citants : à l'échelle mondiale, un pays est susceptible de citer davantage de documents provenant du pays même que d'autres pays (figure 3.9). Ainsi, alors que les États-Unis sont responsables de 21,8 % de tous les documents indexés pendant la période de 2010 à 2015, 58 % des références des articles de ce pays renvoient à des articles de même provenance. Ce chiffre est encore plus élevé pour les petits pays : la Suède, qui a publié 1 % des documents (2010-2015), consacre 20 % de ses références à des documents publiés par des Suédois. Cela s'explique en partie par la proximité géographique, mais aussi par la pertinence des recherches à l'échelle nationale. Ce constat se vérifie en particulier en sciences sociales et en humanités, où les thèmes abordés par les travaux de recherche trouvent leur plus grande pertinence localement.

L'autocitation est au cœur de la critique liée à l'usage des citations comme indicateurs d'impact. Ses détracteurs affirment

FIGURE 3.9 Pourcentage d'autocitations et d'autoréférences des documents parus en 2000, par discipline. Les autocitations et autoréférences (de même que leurs dénominateurs respectifs) ne tiennent compte que des documents sources.



Source: Web of Science, données cumulées par disciplines de la NSF.

qu'elle gonfle artificiellement le compte des citations et qu'elle est fondamentalement différente de la citation reçue d'un auteur qui n'a aucune affiliation avec le document cité. Certains soutiennent même que les autocitations devraient être exclues des indicateurs, notamment quand on cherche à mesurer l'impact intellectuel de la recherche. Pourtant, même s'il existe en effet des chercheurs qui s'autocitent de manière complètement disproportionnée, le fait de citer son propre travail ne devrait pas être perçu comme intrinsèquement mauvais. En fait, étant donné la nature cumulative du travail scientifique, de même que les trajectoires professionnelles, on devrait même s'attendre à ce qu'un chercheur fasse avancer ses recherches en s'appuyant sur leurs phases antérieures. Pour le

chercheur qui a fait part d'une découverte ou qui a proposé une méthode dans un précédent article, c'est tout naturel de se référer à celui-ci dans des publications subséquentes. En outre, le chercheur est sans doute celui ou celle qui a la meilleure compréhension de son propre travail, ce qui le rend d'autant plus susceptible de citer son corpus personnel. Combiné avec la parenté thématique incontournable d'un parcours de recherche individuel, cela rend très probable, quoique non nécessaire, le phénomène de l'autocitation à l'échelle individuelle.

Cela dit, il est très possible de miser sur l'autocitation pour faire gonfler le nombre de citations reçues. Dans le contexte où les indicateurs de citations sont de plus en plus employés pour évaluer les chercheurs et revues savantes, certains auteurs et certains éditeurs ont trouvé des façons créatives de propulser artificiellement leurs taux de citation vers les sommets. L'autocitation abusive est une pratique assez facile à détecter, et l'argument a déjà servi à faire retirer certaines données des indicateurs fondés sur les citations, comme le facteur d'impact. Dans l'édition de 2016 du *Journal Citation Reports* (JCR), par exemple, les statistiques de 13 périodiques ont été supprimées, parce que l'analyse de leurs citations montrait des anomalies, comme l'échange de citations entre revues. Celui-ci se produit quand, afin de déjouer les mesures simplement fondées sur les autocitations, une revue en cite une autre de manière disproportionnée. Cela peut arriver de bonne foi, à cause de l'étroite parenté thématique des revues. Les pratiques malhonnêtes sont celles où une tricherie volontaire vise à faire augmenter le taux de citation des diverses revues. Tout comme l'autocitation abusive, la citation frauduleuse entre deux revues est assez évidente. Elle devient plus pernicieuse quand elle relève d'une entente entre plus de deux joueurs – on parle alors d'un « cartel de citations ». En 2011, on s'est aperçu que cinq revues brésiliennes s'adonnaient à cette pratique, mais elle demeure difficile à détecter.

Comment l'obsolescence est-elle mesurée ?

La vie utile d'un document savant n'est pas éternelle. On appelle *obsolescence* la mesure de la décroissance des citations accordées à un document. Malgré l'augmentation constante du nombre de références par document depuis le début du XX^e siècle et la relative homogénéité de ce phénomène dans toutes les sphères scientifiques, les documents des diverses disciplines ne sont pas cités au même rythme une fois qu'ils sont publiés, et la période pendant laquelle on les cite n'est pas la même non plus. L'année même de leur parution, la plupart des articles ne sont pas cités, et cela prend ensuite quelques années avant qu'ils atteignent le nombre annuel maximal de citations qu'ils recevront. Cette durée est plus courte en sciences naturelles et médicales qu'en sciences humaines et sociales. Inversement, les articles du premier groupe de disciplines deviennent obsolètes plus rapidement que ceux du deuxième, qui sont cités pendant plus longtemps.

Les mesures synchrones de l'obsolescence quantifient les citations reçues après l'année de parution. On mesurera, par exemple, le temps nécessaire pour que les articles publiés dans une discipline donnée à une année donnée obtiennent la moitié des citations qu'ils recevront dans toute leur vie. Parce que les citations sont cumulatives, la plage de temps considérée est extrêmement importante. En sciences humaines, une plage de cinq ans ne saisirait probablement pas le profil de citation de l'article, et la médiane tirée de ces données serait très imprécise. Par conséquent, pour que l'analyse synchronique soit signifiante, la plage envisagée en sciences humaines doit être extrêmement longue – ce qui signifie, en corollaire, que les propriétés de vieillissement des documents de ce domaine ne se comprennent que par l'étude de publications nettement plus vieilles.

L'obsolescence se mesure aussi diachroniquement, en compilant l'âge moyen ou médian des *références* citées pendant une

année donnée, cette dernière étant également appelée la *demi-vie des citations*. Lorsqu'une fenêtre de citation de 100 ans est utilisée, l'âge moyen des documents cités est de 5 à 6 ans en sciences médicales, de 7 ans environ en sciences naturelles et en ingénierie, et de 8 ans environ en sciences sociales. L'âge moyen des documents cités dans le domaine des arts et des humanités est de plus de 14 ans, un écart qu'on doit surtout à l'usage d'archives comme sources primaires, au caractère non cumulatif de ces disciplines et à un recours moindre aux données quantitatives (ce qui rend les articles moins vulnérables à l'obsolescence factuelle).

Plusieurs raisons expliquent que le taux de citation d'un article aille en s'amenuisant. L'une est l'utilité réelle ou perçue du savoir qu'il produit : les données ou les descriptions qu'il emploie sont remplacées par de nouvelles, plus précises ; les problèmes scientifiques qu'il expose ont été résolus et les travaux de recherche ont évolué vers d'autres thématiques ; ou son contenu est jugé inexact ou insuffisant. Une autre est que ses thèses se sont consolidées au fil du temps et font partie des connaissances incontournables de la discipline. Ce phénomène, appelé oblitération par incorporation, cause lui aussi une diminution des citations. On donne souvent l'exemple d'Albert Einstein, rarement cité par les physiciens contemporains tellement on tient pour acquis que son œuvre est connue de tous. Dans ce cas-ci, le fait de ne pas être cité ne correspond pas à un manque d'utilité, mais plutôt à une *hyperutilité* : les travaux en question sont devenus fondamentaux pour ce domaine de la science.

Contrairement à la croyance fort répandue voulant que les connaissances s'accumulent et se périment plus rapidement qu'avant – en réaction aux croissances concurrentes de la main-d'œuvre scientifique et des technologies numériques, principalement –, les travaux savants sont cités pendant plus longtemps qu'auparavant. Autrement dit, l'âge des documents cités dans les articles publiés récemment est plus élevé que dans les articles

plus anciens. Cette tendance n'est pas tout à fait linéaire : l'âge des références citées a augmenté pendant la Deuxième Guerre mondiale, diminué entre 1945 et 1975 et, depuis, il augmente de nouveau. L'augmentation observée pendant la guerre témoigne de l'un des facteurs d'obsolescence, soit la quantité de documentation disponible. Il s'est publié moins de littérature scientifique pendant les années de guerre, et les articles publiés durant cette période avaient donc moins de travaux contemporains à citer. Inversement, la croissance exponentielle observée après-guerre a fait baisser l'âge moyen des travaux cités pendant cette période, car une quantité massive de littérature récente était constamment accessible. Le fait que les chercheurs actuels s'appuient de plus en plus sur des recherches plus vieilles s'explique par le passage d'une croissance exponentielle à une croissance linéaire, ainsi que par la numérisation rétrospective des documents plus anciens, qui les rend plus accessibles. L'arrivée de bases de données bibliographiques offrant une indexation rétrospective, comme Google Scholar, est particulièrement importante à cet égard. C'est l'une des explications plausibles du phénomène de la reconnaissance tardive, que l'on observe dans certains articles scientifiques qui chamboulent les modèles habituels, et qu'on cite pour la première fois de longues années après leur publication.

Les mesures d'obsolescence viennent renforcer la nécessité d'établir soigneusement la plage temporelle des indicateurs basés sur les citations. S'il est possible de calculer des taux de citation pour les articles récents dans la plupart des disciplines (en gardant à l'esprit que ces taux reflètent l'impact des documents à un moment précis, et que leur croissance ou décroissance relative pourrait être plus rapide ou plus lente que la moyenne), dans le cas des arts et des humanités, les taux sont tellement bas et la demi-vie tellement longue que l'analyse des citations, par consensus, y est rarement utilisée.

Qu'est-ce que le facteur d'impact ?

Le facteur d'impact des revues – ou simplement le facteur d'impact (FI) – est l'un des indicateurs scientométriques les plus employés, débattus et... décriés. Il a fait son apparition dans les années 1960, peu après les premières versions du *Science Citation Index*, et a soulevé depuis, au sein de la communauté scientifique, un intérêt et un mépris aussi énormes l'un que l'autre. Plus de 7000 articles lui ont été consacrés depuis 50 ans, suivant une courbe d'intensité croissante. Le sujet n'est pas la chasse gardée des scientomètres: la majorité des articles à son sujet sont publiés dans des revues de science et de médecine, ce qui met en relief l'importance de cet indicateur pour l'ensemble de la communauté des chercheurs.

Le FI est intégré à la base de données du Web of Science (WoS): il a été mis au point à partir de cet index et se nourrit de ses données. Conséquemment, ce « facteur d'impact des périodiques » n'est compilé que pour les revues savantes recensées dans le *Science Citation Index Expanded* et le *Social Science Citation Index*. Étant donné la longue demi-vie des citations (et des références) en arts et en humanités, on ne compile pas de FI pour les périodiques recensés dans l'*Arts and Humanities Citation Index*.

Le FI a été conçu pour aider, d'une part, Garfield et ses collègues de l'Institute for Scientific Information à choisir les revues qui allaient figurer dans leur base de données bibliographiques, et, d'autre part, les bibliothécaires à décider des acquisitions qui allaient augmenter leurs collections. Au fil du temps, toutefois, il en est venu à mesurer la valeur du chercheur plutôt que de la revue: plus le facteur d'impact du périodique est élevé, plus le capital universitaire du chercheur qui y publie s'accroît. L'usage du FI est très répandu, non sans certains effets pervers: on a prouvé qu'il orientait les pratiques de publication des chercheurs — qui pour beaucoup visent simplement à publier dans la revue au plus haut FI possible – et entraînait un déplacement des objectifs

(c'est-à-dire chercher pour publier plutôt que de chercher pour trouver). Cela a provoqué beaucoup de mécontentement dans la communauté scientifique et certains ont proposé de remplacer voire d'éliminer l'indicateur.

À priori, pourtant, le FI semble inoffensif : il calcule le nombre total de citations reçues, pendant une année de référence, par les articles parus dans un périodique donné au cours des deux années antérieures ; puis il divise ce nombre par le nombre d'articles publiés pendant ces deux années. Ce calcul nous renseigne donc sur « l'impact moyen » des articles de ce périodique. Mais l'indicateur ainsi conçu présente plusieurs failles. La première est que les revues publient divers types de documents. Certains types sont cités fréquemment (les articles de recherche et les articles de synthèse, qui reçoivent dans le WoS l'étiquette de documents *citables*), d'autres moins (lettres à l'éditeur, éditoriaux, nouvelles, etc., que le WoS catégorise comme *non citables*). Or, il existe une asymétrie concernant l'intégration au calcul des éléments citables et non citables : le numérateur inclut tous les types de documents, tandis que le dénominateur n'inclut que les articles de recherche et les revues de littérature (documents citables). L'effet de cette asymétrie est de gonfler artificiellement le FI d'une revue quand elle publie des éléments étiquetés non citables, mais, dans les faits, souvent cités – ce qui est le cas des éditoriaux et des lettres à l'éditeur, entre autres. On voit un bon exemple de cet effet dans les revues *Science* et *Nature*, dont le FI pour 2009 passe, respectivement, de 34 480 à 22 770 et de 29 747 à 20 902 une fois qu'on a exclu du calcul les éléments non citables. Cela a donné lieu à des manipulations, et l'on a vu des éditeurs changer le type de certains documents de « citable » à « non citable » afin d'augmenter leur FI en réduisant le dénominateur.

Le calcul est vulnérable également aux « douces persuasions » des éditeurs de revues par exemple aux suggestions des rédacteurs ou des évaluateurs d'intégrer plus de références au périodique où

l'auteur souhaite publier ou, dans un scénario plus complexe, à l'un ou l'autre des périodiques avec lesquels ils auront formé un cartel de citations, comme on l'a vu précédemment. Les éditeurs des revues savantes sont parfois directement coupables de tels procédés, quand ils se servent de leur rôle éditorial afin de suggérer des références à leur revue, afin de faire grossir leur FI. La réaction de Clarivate Analytics a été de surveiller et de punir les abus flagrants de telles autocitations. Malgré cela, le scepticisme survit dans la communauté scientifique quant à l'intégrité de l'indice.

Une autre de ses failles importantes concerne la comparabilité entre les disciplines. Les pratiques en matière de références – comme le nombre de références par document et l'âge moyen de celles-ci – varient énormément d'une discipline à l'autre. Cela produit des FI généralement plus élevés en recherche biomédicale, disons, qu'en chimie, en physique ou dans la plupart des sciences sociales. On observe aussi des différences à l'intérieur même d'une discipline, selon les modalités de référence et de publication de ses différents domaines de spécialisation. On ne peut donc pas comparer les FI des disciplines et des spécialités directement, c'est-à-dire sans tenir compte de ces différences.

La fenêtre de citation relativement courte de l'indice est un autre de ses aspects problématiques. En comptant les citations sur une période de deux ans, on se limite à une petite proportion de toutes les citations reçues. Les sciences sociales, où les références citées sont en moyenne plus anciennes et les citations plus lentes à s'accumuler, s'en sortent nettement mieux quand on prolonge la plage d'analyse. Même en recherche biomédicale, où les citations s'accumulent à un rythme beaucoup plus rapide que dans d'autres disciplines, l'analyse d'articles publiés en 1990 montre que les citations reçues en 1991 et 1992 ne représentent que 16 % environ des citations qui seront reçues dans les 20 années suivant la publication. Pour rendre compte de telles différences, le *Journal Citation Report* fournit, depuis 2007, un FI fondé sur une fenêtre

de citation de 5 ans. Mais le FI calculé sur deux ans demeure la mesure la plus utilisée, et c'est lui qui est mis en évidence dans le matériel promotionnel des périodiques.

On se préoccupe également du fait que le FI calcule une moyenne, ce qui est difficile à interpréter à l'échelle des articles. La moyenne indique le taux de citation « typique » ou « normal » d'un article paru dans la revue. Or, comme on l'a vu, les citations ne suivent pas une distribution normale. Elles sont au contraire fortement asymétriques : une minorité de publications reçoit une très grande part des citations, et la majorité se partage le reste. En termes scientifiques, on dit que la distribution des citations est non paramétrique, alors que l'interprétation de mesures à tendance centrale suppose une distribution paramétrique. En conséquence, on ne saurait considérer le FI comme un indicateur adéquat de l'impact scientifique typique des articles individuels parus dans la revue.

Plusieurs analyses empiriques ont cherché à comprendre le pouvoir prédictif de l'indicateur, à savoir si le FI d'une revue permet de prédire le succès d'un article en particulier. Ils ont montré que si le FI est bel et bien corrélé au taux de citation des articles, son efficacité prédictive a diminué depuis 20 ans. C'est dire qu'il est risqué d'extrapoler la valeur (ou le potentiel) d'un article ou d'un individu à partir de l'indice attribué à un périodique. On fera bien, alors, de réserver l'utilisation de cet indicateur aux analyses des seules revues.

Qu'est-ce que le score Eigenfactor ?

L'indicateur Eigenfactor a été mis au point par un expert en sciences des données, Jevin West, et un biologiste, Carl Bergstrom, tous deux de l'Université de Washington. Pour déterminer l'importance d'une revue scientifique, l'algorithme de l'Eigenfactor – qui est aussi à la base de l'*Article Influence Score* (AIS), un indicateur

apparenté – exploite la puissance du réseau de citations, au lieu de se fier uniquement aux citations reçues. Contrairement aux algorithmes des autres indicateurs de citations, celui de l'Eigenfactor accorde plus de poids à une citation provenant d'une source centrale (on parle ici de centralité de réseau) qu'à une citation provenant d'une source plus périphérique. L'idée de cet algorithme est dérivée des travaux du sociologue Phillip Bonacich, qui, en 1972, a introduit la notion de centralité de vecteur propre (*eigenvector*, en anglais) pour repérer les individus influents au sein d'un réseau de communication. Dans cette approche, on détermine les individus clés en fonction de leur centralité dans le réseau. L'algorithme PageRank, qui est la base du moteur de recherche de Google, est une autre application bien connue de la centralité de vecteur propre.

Dans leur présentation du concept, West et Bergstrom invitent le lecteur à imaginer un chercheur errant sans fin parmi les rayons d'une bibliothèque, passant d'un article scientifique à l'autre au hasard des références. Le score Eigenfactor est censé représenter, en pourcentage, le nombre de fois où le parcours du chercheur l'orienterait vers une revue donnée. Un score de 3, par exemple, indiquerait qu'il consacrerait 3% de son temps à la revue X, au sein du réseau des revues étudiées : ainsi, plus le score Eigenfactor est élevé, plus l'importance de la revue dans le réseau est élevée. Puisque l'Eigenfactor est un pourcentage, la somme est de 100 si l'on considère toutes les revues d'un réseau. L'indicateur est donc additif : pour mesurer le réseau, il faut additionner les scores individuels de ses éléments. Mais la démarche n'est pas naturelle, quand on sait que le score de la plupart des revues est de moins de 0,01 et qu'il est difficile à interpréter dans l'absolu. Le score Eigenfactor normalisé vient corriger cette situation par rééchelonnement, en accordant un score de 1 à la revue moyenne. L'Eigenfactor est calculé selon une plage de 5 ans, ce qui le distingue du FI et de sa plage habituelle de 2 ans.

Exploitant l'algorithme de l'Eigenfactor, l'*Article Influence Score* (AIS) se compare plus directement au facteur d'impact, puisqu'il repose sur la force des citations reçues par article – mais il est censé offrir de nombreux avantages par rapport au FI. En plus de la plage de 5 ans, mieux adaptée à de nombreuses disciplines, l'algorithme réduit les distorsions entre disciplines en calculant des proportions, plutôt qu'un nombre absolu de citations. En outre, il exclut les autocitations : son approche itérative ne permet pas à un périodique de renvoyer à lui-même.

La firme Thomson Reuters a ajouté l'Eigenfactor et l'AIS au *Journal Citation Report* (JCR) en 2009. En 2013, l'Eigenfactor a été adapté au classement non seulement des revues savantes, mais aussi de la production des auteurs, des départements, des universités et des pays (à partir des données des citations obtenues à l'échelon des auteurs).

Qu'est-ce que le SNIP ?

L'indicateur d'impact des articles selon la source normalisée (ou SNIP, pour *Source Normalized Impact per Paper*) a été mis au point par le scientomètre Henk Moed, alors qu'il travaillait au Centre for Science and Technology Studies de l'université de Leiden. Le SNIP est un indicateur de citations des périodiques qui prend en compte le contexte citationnel d'une revue. Il est centré sur le concept du « potentiel de citation », c'est-à-dire qu'il adopte la perspective de l'article citant plutôt que celle de l'article cité, examinant le nombre moyen de références citées par document dans un domaine donné. Plutôt que de rattacher un périodique à un domaine en fonction d'une liste préétablie, on définit le domaine comme la collection des documents qui citent le périodique. Le domaine se construit donc à partir d'un faisceau de citations provenant d'articles plutôt que de revues. De plus, on affine le potentiel de citation en examinant le contexte de l'index de citations, c'est-à-dire qu'on tient

compte non seulement du nombre total de références citées, mais aussi du nombre de ces références parues dans les périodiques indexés (items sources) par la base de données (Scopus, dans ce cas-ci). Enfin, le calcul de l'indicateur tient compte de la vitesse d'accumulation des citations au sein du domaine. Comme c'est le cas pour le score Eigenfactor et le *SCImago Journal Rank* (SJR; voir ci-dessous), les éléments non citables sont exclus. Cela atténue certains partis pris, mais ne représente pas les périodiques qui contiennent une forte proportion d'articles de synthèse (et dont le score SNIP est généralement élevé). Les avantages de l'indicateur SNIP corrigent amplement les problèmes liés à la classification par domaine, les différences entre les domaines et à l'intérieur de ceux-ci concernant le périmètre et les pratiques de citation, de même que les risques d'abus ou de dérapage liés à l'inclusion dans les calculs des éléments non citables. Comme le SJR, le SNIP tient compte d'une fenêtre de citation de trois ans. Mais contrairement au score Eigenfactor, il inclut toutes les autocitations, ce qui peut provoquer certaines distorsions.

Qu'est-ce que le *SCImago Journal Rank* ?

Le classement des périodiques SCImago (SJR, pour *SCImago Journal Rank*) a été conçu par Félix de Moya-Anegón, du Conseil supérieur de recherche scientifique de l'Espagne (CSIC), et Vicente Guerrero-Bote, de l'Université de l'Estrémadure. C'est maintenant l'Université de Grenade qui l'exploite et le met à jour. Comme le score Eigenfactor, ce classement a recours à la centralité de vecteur propre de Bonacich pour évaluer le prestige d'un périodique. Mais pour ce faire, il mesure les liens de proximité entre périodiques, déterminés par leurs relations de cocitations. L'hypothèse est que deux revues étroitement liées – présentant un potentiel de transfert important – sont susceptibles d'avoir des sujets en commun. L'algorithme du SJR accorde donc davantage de poids aux

citations provenant d'une revue voisine qu'aux citations provenant d'une revue périphérique. La justification avancée pour soutenir ce choix méthodologique est que les revues dites périphériques ont moins d'autorité sur le sujet traité dans la revue citée (mais cette affirmation est contestée, ce qui limite potentiellement la puissance de l'indicateur). Les autocitations sont incluses dans le calcul, jusqu'à concurrence du tiers de toutes les citations reçues par le périodique. SCImago permet aux utilisateurs d'examiner le rang de chaque revue indexée par Scopus, à partir d'un portail Web accessible au public. En 2012, des améliorations ont été apportées au SJR original. Le SJR2 repose sur une plage de 3 ans, il normalise le calcul du prestige en tenant compte de la proportion de documents citables, et fixe un seuil au transfert de prestige d'une revue à l'autre, cela dans le but de réduire la distorsion causée par les manipulations et par les revues très spécialisées contenant peu de références.

Qu'est-ce que CiteScore ?

À la fin de l'année 2016, Elsevier a ajouté à Scopus un nouvel indicateur d'impact pour les revues : CiteScore. Exploitant la base de données de Scopus, qui couvre, on s'en souvient, un plus grand périmètre que celle du WoS, CiteScore s'applique donc à davantage de périodiques et s'efforce de répondre, de manière transparente, à certaines des critiques formulées à l'endroit du FI. Contrairement à de nombreux autres indicateurs, dont l'Eigenfactor et le SJR, le calcul de CiteScore est simple, en ce sens qu'il ne corrige ni l'asymétrie de la distribution ni la différence de densité entre disciplines en ce qui a trait aux citations. On obtient le résultat CiteScore d'une revue pour 2015, disons, en additionnant le nombre total de citations reçues en 2015 par tous les documents publiés dans cette revue entre 2012-2014, et en divisant ce nombre par celui de tous les documents publiés dans la revue pendant la même période. La

plage de parution est donc plus longue d'une année que celle du FI (3 ans au lieu de 2). Sa simplicité rend l'indicateur facile à comprendre et à interpréter : un CiteScore de 4 signifie que les articles publiés dans la revue au cours des 3 années précédentes ont reçu en moyenne 4 citations au cours de l'année actuelle.

L'indice CiteScore répond à l'un des reproches les plus récurrents adressés au FI, soit l'asymétrie entre numérateur et dénominateur. Cet indice ne distingue pas les documents selon leur type ; il inclut dans le numérateur les citations reçues par tous les types de documents et, symétriquement, additionne dans le dénominateur tous les types de documents publiés. Par conséquent, tandis que les revues comptant une forte proportion d'éditoriaux, de nouvelles, de lettres à l'éditeur, de notices nécrologiques, etc. – comme *Science* et *Nature* – bénéficient de citations « gratuites » dans le calcul du FI (parce que le dénominateur ne contient que les articles originaux et les articles de synthèse), elles obtiennent un résultat CiteScore moins élevé, parce que ces mêmes documents sont inclus dans le dénominateur (et que leur taux de citations est tout de même moins élevé que celui des autres types de documents). Les revues sont ainsi encouragées à publier peu de documents autres que les articles, ce qui favorise, comme l'ont souligné les critiques, les périodiques d'Elsevier, qui en publient proportionnellement moins que les autres éditeurs. Cela révèle un important conflit d'intérêts : avec la création de CiteScore, Elsevier figure désormais parmi les évaluateurs des revues, alors même qu'elle en publie.

Qu'est-ce que l'indice h ?

L'indice h a été mis au point en 2005 par le physicien Jorge Hirsch, qui voulait construire un indicateur composite reflétant, à l'échelle du chercheur individuel, à la fois la productivité et l'impact. Le calcul de l'indice h est relativement simple : un chercheur obtient

un indice de h quand il a publié h articles, et que chacun a été cité au moins h fois. Un auteur a un indice h de 14, par exemple, si au moins 14 de ses articles ont reçu au moins 14 citations. L'attrait de la simplicité a joué dans l'adoption rapide de l'indice h dans beaucoup de disciplines, notamment en sciences naturelles et en génie, où l'on a cru qu'il pouvait mesurer la qualité individuelle des scientifiques. Une « bulle h » s'est ensuivie, quand certains scientomètres et autres chercheurs se sont empressés de décrire l'indice h en fonction de différentes populations et d'en construire de nouvelles variantes. On l'a intégré également aux bases de données bibliométriques, ce qui a démocratisé son usage.

Or, alors que l'indice h cherche à combiner les publications et les citations, en réalité il fait de la publication la variable dominante du calcul, puisqu'elle constitue la borne supérieure de l'indice: celui-ci ne sera jamais plus élevé que le nombre de publications du chercheur. Par ailleurs, dans ce calcul, chaque citation supplémentaire – qui fait que le nombre de citations dépasse le nombre de publications du chercheur – est perdue. Prenons, par exemple, ces deux collègues. Le premier a publié 10 articles et chacun a été cité 10 fois; il a donc un indice h de 10. Le second a publié 5 articles, et chacun a été cité 100 fois. Il a donc un indice h de 5, la moitié de celui de son collègue, bien que chacune de ses publications soit citée 10 fois plus que les siennes. Cet exemple montre que l'indicateur est sujet à des distorsions, puisqu'il favorise la quantité de publications (laquelle est influencée lourdement par l'ancienneté, les collaborations et le champ disciplinaire).

Plusieurs dizaines de variantes de l'indice h ont été proposées, et nombre d'entre elles cherchaient à corriger ses nombreux points faibles. Mais le défaut le plus accablant de cet indicateur est qu'il n'a pas de concept sous-jacent: il intègre (et par le fait même, il amalgame) deux concepts, la productivité et l'impact. Mais on ne saisit pas clairement celui des deux qui est représenté par cet indicateur composite. Toutes les variantes qui en ont été dérivées

souffrent donc du même défaut fondamental que l'original : elles tentent de combiner deux concepts en un seul indicateur.

Que sont les mesures alternatives ?

On reproche depuis longtemps aux citations de n'offrir qu'un portrait partiel de l'impact – d'illustrer avant tout l'impact de la science sur d'autres documents scientifiques, en se fondant exclusivement, ou presque, sur l'interaction entre les articles publiés dans les périodiques. En 2010, le quasi-monopole des citations sur les indicateurs d'impact a été remis en question par de nouvelles façons de quantifier ce concept : les mesures dites alternatives, destinées à compléter les mesures plus classiques, voire à les remplacer. L'expression anglaise, *altmetrics*, est de Jason Priem, qui était alors doctorant à l'Université de la Caroline du Nord à Chapel Hill. Priem et d'autres qui pensaient comme lui ont invité la communauté scientifique à diversifier ses manières de quantifier l'impact de la recherche.

Les données les plus accessibles viennent des réseaux sociaux, grâce auxquels on peut étudier les interactions, en ligne, des documents scientifiques dans des gestionnaires de références (Zotero, Mendeley, CiteULike...), des plateformes de médias sociaux (Twitter, Facebook...) et d'autres dépôts ou services d'archivage (Figshare, SlideShare, GitHub...). Certaines plateformes éditoriales, comme celle de la Public Library of Science (PLOS), ont adopté ces idées et innové dans le domaine, en récoltant des données, pour la plupart non citationnelles, qu'elles présentent pour chaque article (souvent appelées les Article-Level-Metrics). Des mesures antérieures à la création des médias sociaux ont été rattachées à cette catégorie, telles que les données de téléchargement des revues (les statistiques de COUNTER, p. ex.). Cependant, même si ce genre de mesures existait avant 2010, c'est la formulation de l'expression, parallèlement à l'essor du rôle des médias sociaux dans la création, la diffusion

et la discussion des contenus scientifiques qui a fourni le terreau fertile où le concept de mesures alternatives s'est épanoui. En quelques années à peine, les mesures alternatives en sont venues à constituer un ensemble hétérogène d'indicateurs ayant comme trait commun de ne pas employer les données citationnelles des index classiques.

Rapidement, des entreprises se sont créées pour profiter de l'intérêt croissant envers ces mesures alternatives. Trois des mieux établies sont Altmetric, ImpactStory et Plum Analytics. Altmetric a été fondée en 2011 par Euan Adie, grâce aux fonds de démarrage obtenus à la suite du concours « Apps for Science » organisé par Elsevier. En 2012, afin de poursuivre ses activités, elle obtenait des fonds de Digital Science – exploitée, celle-ci, par le Holtzbrinck Publishing Group (propriétaire de *Nature*, Macmillan, Springer et d'autres maisons d'édition). L'objectif initial d'Altmetric était d'agrèger les données des mesures alternatives pour chaque article, c'est-à-dire repérer les mentions d'un article dans diverses sources – énoncés de politiques, médias grand public, blogues, plateformes d'évaluation par les pairs postpublication, médias sociaux et plans de cours. La marque d'Altmetric est son fameux « beigne arc-en-ciel » : une roue de tracteur dont les stries représentent, pour un article donné, les sources qui mentionnent cet article et qui contribuent, ce faisant, à son « score d'attention Altmetric » (*Altmetric Attention Score*), un indicateur numérique inscrit dans le trou du beigne et qui quantifie l'attention que l'article a reçue.

ImpactStory a été fondée par Heather Piwowar et Jason Priem. Le projet, baptisé *total-impact*, a vu le jour lors de l'atelier *Beyond Impact* de 2011 : il s'agissait alors d'un marathon de programmation. L'Open Society Foundation a fourni les fonds de démarrage, et la National Science Foundation et la fondation Alfred P. Sloan, des subventions subséquentes. ImpactStory se sert d'Altmetric pour fournir toutes les données liées à chaque contribution

savante, données qu'elle complète avec celles de Mendeley et de Twitter. ImpactStory a ceci d'original qu'elle fournit des données liées au chercheur en particulier, plutôt qu'à un article ou à un établissement. L'attribution des publications aux individus se fait au moyen d'ORCID, un service de gestion spécialisée de l'identité des chercheurs. C'est d'ailleurs l'une des limites de ce service, puisque les auteurs doivent à la fois posséder un identifiant ORCID et tenir à jour leur dossier de publications dans ImpactStory. Crossref fournit des métadonnées supplémentaires sur les articles. Les réalisations – figurer dans le 10 % ou 25 % supérieur des chercheurs pour certains éléments, par exemple – se dégagent de comparaisons effectuées entre plus de trente mille chercheurs possédant un profil ORCID. Toutefois, on ne peut pas non plus examiner les profils des autres (comme avec Google Scholar). Chaque personne n'a accès qu'à son propre profil, une fois qu'elle a lancé une session. Les profils sont gratuits (bien que cela n'ait pas toujours été le cas) par l'intermédiaire d'un compte Twitter.

Plum Analytics, fondée en 2012 par Andrea Michalek, fournit des données cumulatives provenant de plus de 30 sources. Membre de la famille d'entreprises EBSCO (jusqu'en février 2017), PlumX Analytics exploite les données d'utilisation et de téléchargement des bibliothèques, ainsi que les données sur leurs fonds documentaires provenant de WorldCat. Les données tirées de Mendeley, de GitHub, des blogues, du site Goodreads (pour les livres), des articles de Wikipédia et des vidéos de Vimeo fournissent ce que l'entreprise appelle des « captures » ou des « mentions ». Les plateformes de médias sociaux comme Reddit, Twitter, Facebook et Figshare sont toutes recensées. En plus de ces ressources dites alternatives, PlumX Analytics emploie aussi les données citationnelles de Crossref, de MEDLINE et de Scopus, les données cliniques de Dynamed Plus, et les données des brevets déposés auprès du Bureau des brevets des États-Unis (United States Patent and Trademark Office, USPTO). L'entreprise propose plusieurs

produits, destinés avant tout aux bibliothécaires et aux administrateurs. Vendus à Elsevier en février 2017, les outils de mesures alternatives de Plum Analytics ont été intégrés aux outils d'analyse d'Elsevier, tels SciVal et Scopus.

Chacune de ces entreprises fournit des indicateurs légèrement différents – par exemple, Altmetric inclut les références aux articles scientifiques provenant des énoncés de politiques publiques et Plum Analytics, les fonds documentaires des bibliothèques. Elles se distinguent aussi par leurs niveaux d'analyse: globalement, les indicateurs d'Altmetric quantifient la recherche au niveau des articles, ceux de Plum Analytics, au niveau des établissements et ceux d'ImpactStory, au niveau des individus. Ces variations des données et des modèles de catégorisation entraînent des modes de présentation différents des données; Altmetric, par exemple, propose des représentations géospatiales du profil d'impact de chaque article, tandis qu'ImpactStory informe les chercheurs de la durée de la popularité de leurs travaux dans les médias sociaux.

Les sources diffèrent également par la disponibilité de leurs données. Altmetric rend toutes ses données accessibles aux chercheurs; elles servent de base, par conséquent, à la plupart des études empiriques publiées sur les mesures alternatives. Les chercheurs peuvent également intégrer l'outil Bookmarklet d'Altmetric à leur moteur de recherche, gratuitement. Cet applet permet à l'utilisateur de générer un profil pour n'importe quel document grâce à un *Digital Object Identifier* (DOI). Le profil contient le score d'attention d'Altmetric pour ce document, ainsi que le résumé et les détails de toute l'attention qu'il reçoit. ImpactStory, par contraste, ne permet qu'aux utilisateurs de consulter leur profil, et seulement s'ils ouvrent une session sur le portail. Quant à Plum Analytics, seuls ses abonnés peuvent en obtenir des rapports. Plum Analytics et ImpactStory ne publient pas leurs données,

qui ne peuvent donc pas servir à des analyses externes à grande échelle par les scientomètres.

L'une des caractéristiques communes à toutes ces sources est leur dépendance envers les DOI. Les DOI sont très répandus sur toutes les plateformes, même si plusieurs autres identifiants sont employés (par exemple, Altmetric se sert de l'ISBN des livres et recense les identifiants propres à certaines plateformes, comme ceux de MEDLINE, d'arXiv, d'ADS et du SSRN; ImpactStory dépend d'ORCID pour la désambiguïsation des noms d'auteurs). De nombreux autres types de documents sont donc ignorés, ce qui limite la portée des mesures d'impact à celles qui réservent une place à ces identifiants particuliers. En se fiant à ces identifiants uniques, on occulte en particulier une grande part des échanges qui se déroulent dans les médias sociaux au sujet des travaux de recherche sans être rattachés à un identifiant (DOI ou autre), de même que les débats sur la recherche dans les grands journaux. Bien que les organisations produisant des mesures alternatives essaient d'étendre leur couverture de ces activités en ligne, à ce jour, les indicateurs dépendant encore largement des articles de périodiques et de leurs DOI. L'interprétation des indicateurs de mesures alternatives doit donc tenir compte de cette contrainte.

On trouve dans la littérature quelques tentatives pour formuler des indicateurs fondés sur les mesures alternatives. Les plus élémentaires représentent le nombre et la moyenne des scores obtenus par les articles sur ces plateformes, et concernent la couverture, la densité et l'intensité. La couverture est le pourcentage de documents ayant au moins une mention (ou « événement ») sur une plateforme donnée. La densité est le nombre moyen d'événements par document. Étant donné l'étroitesse du périmètre de couverture, la densité est souvent biaisée par le grand nombre de valeurs égales à zéro. C'est pourquoi on a introduit la notion d'intensité, qui est le nombre moyen d'événements pour un ensemble de documents, excluant ceux qui n'ont aucune mention. Chacun

de ces indicateurs est compilé pour une plateforme ou un genre précis (on étudie seulement, par exemple, les mentions sur Twitter, le nombre de lecteurs sur Mendeley, ou le nombre de fois où un document est couvert dans les médias). Des chercheurs se sont efforcés à quelques reprises de formuler des indicateurs composites, mais la communauté scientifique ne les a pas adoptés – à l’exception du score d’attention mis au point par Altmetric.

Ce score est composite parce que les différentes couleurs des stries qui composent «le beigne», qui correspondent au type d’attention reçue par le document analysé, représentent les différentes sources recensées par Altmetric (médias sociaux, blogues, énoncés de politiques publiques, évaluation par les pairs postpublication, nouvelles, vidéos). Le score composite est l’indicateur pondéré de l’attention obtenue selon les différents scores, qui sont ajustés selon le volume, l’importance de la source et les attributs de l’auteur de la source. Le volume représente le nombre de mentions provenant d’un individu et d’une source unique – sur Twitter, par exemple, il ne tient pas compte des mentions multiples, par la même personne, d’un travail donné. Le volume augmente avec chaque mention pour la plupart des sources. L’exception est Wikipédia, où le comptage est binaire: le score reste le même, que le travail soit mentionné dans un article ou dans cent. Toutes les sources n’ont pas le même poids: à chacune correspond un compte particulier. Une mention dans Twitter vaut 1; dans un blogue, elle vaut 5, et 8 si l’attention provient d’un site de nouvelles. Certaines mentions valent moins qu’un point – sur Facebook, YouTube, Reddit ou Pinterest, par exemple. Toutefois, comme le score d’attention est représenté par un nombre entier, le compte de chaque source est arrondi au nombre entier suivant. Il existe aussi une pondération à l’intérieur des sources, en fonction de la *portée*, de la *proximité* et de la notion de *biais*. La portée est mesurée en fonction du public réel ou imaginé du document. Prenons Twitter: sur cette plateforme, le nombre d’abonnés est explicite et le coefficient de

pondération grossit avec le nombre d'abonnés. La taille de l'ensemble des abonnés d'une source comme le *New York Times* justifie la surpondération. Altmetric examine également le comportement antérieur de la source de l'attention. Un pseudonyme Twitter qui retransmet constamment des articles scientifiques aura moins de valeur qu'un pseudonyme qui retransmet moins souvent (promiscuité). Le score d'attention d'Altmetric corrige les biais en modifiant le résultat en fonction des activités autopromotionnelles, des comportements qui révèlent l'utilisation de robots et des aberrations extrêmes. Malgré cette structure de pondération détaillée, le score sur Twitter détermine très souvent le score global, à cause de son périmètre très étendu. Mendeley provoque un effet semblable, même si Altmetric calcule le lectorat d'un document seulement si celui-ci a été mentionné dans au moins une autre source.

Les mesures alternatives ont d'abord été présentées comme des indicateurs d'impact, dans la même catégorie que les citations, et les chercheurs ont voulu tout d'abord examiner le rôle de ces mesures en supposant qu'elles étaient révélatrices de l'impact général des travaux savants. Mais les premières analyses tendent à montrer que le concept qui sous-tend ces mesures se rapproche plus de l'attention que de l'impact. S'il existe bel et bien des traces d'interaction, elles sont d'un genre qui demande beaucoup moins d'investissement et de filtrage que les citations. Les entreprises qui formulent les indicateurs – Altmetric, en particulier – reconnaissent volontiers que leurs données se rattachent à de nombreux concepts, y compris l'attention, la dissémination et l'influence. Altmetric a résolu ce problème en ramenant tout au plus petit dénominateur commun : en 2016, en réaction aux critiques qui lui reprochaient de présenter indûment sa mesure comme un indicateur d'impact, elle lui a ajouté le mot « attention » – « *Altmetric Attention Score* » – afin de rendre explicite le fait qu'il s'agit d'une mesure d'attention et non d'impact. Malgré cela, la nature composite de l'indicateur continue de poser problème : quand les

sources de données d'un indicateur correspondent à plus d'un concept, l'indice obtenu devient impossible à interpréter. Disons, en résumé, que l'hétérogénéité des sources et des concepts est trop grande pour que les mesures alternatives puissent servir d'indicateurs.

L'idée de se servir d'indicateurs alternatifs pour mesurer l'impact social est également problématique. Les travaux sur le sujet révèlent qu'une grande partie de ceux qui génèrent les données de ces mesures alternatives (en pointant vers un article sur Twitter ou en sauvegardant un document sur Mendeley, p. ex.) sont en fait d'autres chercheurs, plutôt que le « grand public ». Ce n'est peut-être pas surprenant, étant donné le recours aux identifiants d'objets numériques : les chercheurs sont plus susceptibles de connaître cet usage et de faire des liens vers ces objets dans leurs échanges sur des sujets scientifiques. Le grand public est peut-être plus susceptible de cliquer sur des sources secondaires (billets de blogues ou nouvelles, p. ex.), ou de mentionner des résultats de recherche sans fournir le lien correspondant. Tant qu'ils reposent sur les données des DOI, c'est l'un des obstacles majeurs à l'utilisation des indicateurs alternatifs pour mesurer l'impact social.

Les mesures alternatives sont extrêmement dynamiques : les outils existants changent et disparaissent et de nouvelles sources émergent constamment, ce qui génère de nouveaux indicateurs. De plus, un même indicateur peut varier énormément selon le degré d'accessibilité des données : certaines entreprises obtiennent leurs données directement d'une plateforme, par exemple, alors que d'autres se servent d'une interface de programmation d'application (API). Les éléments téléchargeables changent aussi rapidement, ce qui rend les études fondées sur des mesures alternatives difficiles à reproduire. Dans un index solidement fondé, le nombre de citations ne peut qu'augmenter (Google Scholar est le seul où l'on observe à l'occasion une diminution du nombre de citations). Les mesures alternatives sont nettement moins stables. Le score

d'attention d'Altmetric, par exemple, peut varier à la hausse ou à la baisse en une seule journée. Plusieurs raisons peuvent le faire chuter : la suppression d'une mention par un auteur, la désignation d'un billet comme inapproprié et donc supprimé, la repondération d'un algorithme ou encore, le changement de catégorie de certaines données qui, de publiques, deviennent privées (les profils Facebook, p. ex.). Une hausse soudaine, en contrepartie, correspond souvent à l'intégration de nouvelles sources. L'instabilité de cet indice remet en question l'usage de scores semblables dans les exercices d'évaluation à grande échelle, et empêche les analyses longitudinales.

Cela dit, les mesures alternatives ont quand même des qualités. Elles proposent un nouveau moyen de quantifier l'impact, en trouvant leurs données ailleurs que dans les seuls articles scientifiques. Elles rendent compte des éléments plus vastes liés à la fois à la production et à l'usage de la science, à savoir la diversité de la production et de la réception. Elles ont aussi un avantage chronologique, puisque les données auxquelles elles s'intéressent s'accumulent dès la publication et sont actualisées chaque jour, voire plusieurs fois par jour. Cela ouvre de nouveaux horizons : des méthodes inusitées assurent le suivi des travaux scientifiques, et des marques d'interaction auparavant invisibles deviennent manifestes. Mais tout ce qui est dénombrable ne doit pas nécessairement être dénombré. En formulant des indices qui reposent sur les activités en ligne (gazouillis, blogues, publications Facebook), on suscite des comportements particuliers et la marchandisation potentielle des données sur les interactions par les sites où elles se produisent. Quand les décideurs publics intègrent les données alternatives aux évaluations, ils favorisent implicitement certains types de travaux. En réaction, les chercheurs pourraient choisir de s'intéresser à des sujets « hautement twittables », et consacrer une part de leur temps à l'entretien de leur profil en ligne plutôt qu'à leurs travaux. Les mesures alternatives sont aptes à créer un

milieu où les chercheurs assouissent plus volontiers leur besoin d'attention que leur goût pour la science.

L'expansion qu'on leur doit, par conséquent, ne concerne pas juste les sources de données et les nouveaux indicateurs, mais aussi les concepts qu'elles mesurent – pas seulement la qualité ou l'utilité, mais aussi la popularité ou l'intérêt aux yeux du grand public. Ceux qui évaluent la science doivent rester vigilants et veiller, toujours, à ce que les indicateurs soient associés à un concept bien précis et interprété selon ce concept. La légitimation de ces indicateurs dépend de leur réception par le marché universitaire. À l'heure actuelle, les mesures alternatives ont un statut marginal : reconnues par certains chercheurs, elles ne sont pas explicitement intégrées aux lignes directrices concernant les promotions et la permanence ni aux autres initiatives des décideurs politiques. Toutefois, comme les organismes subventionnaires et les établissements sont en quête de moyens de responsabiliser les chercheurs quant à l'impact général de leurs travaux sur la société, il est fort possible qu'on les intègre davantage à l'entreprise d'évaluation de la recherche.

Comment le financement de la recherche est-il mesuré ?

Les indicateurs du financement de la recherche sont disponibles pour les unités présentes aux deux extrêmes du spectre de la recherche, soit les projets et les pays. Très peu sont conçus pour les niveaux d'agrégation intermédiaires. Au niveau individuel, la plupart des pays possèdent des bases de données publiquement accessibles sur les projets subventionnés par leurs organismes subventionnaires. Mais ces données sont dispersées entre de nombreux organismes, et l'intégration des sources pour un pays, un ensemble de pays ou une région du monde est très rare. Aux États-Unis, on peut télécharger les données brutes de chacun des grands organismes subventionnaires (celles des National Science

Foundation et National Institutes of Health, notamment). Ces données sont cumulées dans une source nommée Federal RePORTER, produite par le projet STARMETRICS et reliée à de nombreuses agences et sources fédérales (p. ex. MEDLINE et PubMed Central). La qualité et l'exhaustivité des données, toutefois, varient énormément d'une source à l'autre. En outre, ces données rendent compte du financement fédéral d'un seul pays, négligeant le financement fourni par les universités ou d'autres organismes, notamment les grandes fondations privées (Sloan ou Mellon, p. ex.).

Certains organismes ont entrepris d'agréger les données de différents pays et types de financement, et fournissent de l'information sur les sommes reçues et le nombre de projets financés. Digital Science, par exemple, a mis au point un outil, Dimensions for Funders, qui propose aux subventionnaires des analyses comparatives de leur portefeuille de projets. Un autre outil, Academic Analytics, se sert des différents ensembles de données disponibles pour compiler les dossiers de subventions des professeurs d'université aux États-Unis. Quoi qu'il en soit de ces initiatives (dont certaines exigent un abonnement), la quantification du financement de la science est encore très localisée, à l'heure actuelle. La création d'un ensemble de données mondialement comparables demeure entravée par le caractère insulaire des données de ce genre, le manque d'uniformité dans leur consignation et la difficulté d'en arriver à des équivalences en fonction des différentes devises.

Le financement souffre des mêmes biais structureaux que les autres indicateurs de la recherche: variable à l'extrême selon les domaines, très asymétrique, concentré pour la plus grande part sur quelques chercheurs d'élite. Un peu comme les mesures de productivité, les données qui se rattachent au financement souffrent également de l'attribution trop précise des fonds. Certaines agences, par exemple, n'inscrivent que le chercheur principal comme bénéficiaire, même s'il est entouré de plusieurs cochercheurs. Quand

la collaboration s'étend au-delà des frontières, il devient particulièrement difficile de comprendre (à partir des métadonnées, à tout le moins) à qui revient l'ultime responsabilité du travail produit et qui devrait figurer comme récipiendaire du financement. Bref, les données sur le financement associent habituellement les ressources (ou le mérite) au chercheur principal, négligeant ainsi le caractère de plus en plus collaboratif de la recherche subventionnée.

Plusieurs autres obstacles nuisent à l'utilisation des données sur le financement pour quantifier l'activité scientifique. L'un d'entre eux est la très grande variabilité de la concurrence pour les différentes subventions. Bien que les taux de succès varient de façon spectaculaire – et qu'ils aient chuté abruptement dans la plupart des pays au cours des dernières décennies –, le financement figure généralement sous la forme d'un unique nombre monétaire, et souvent sans indication de la source. Il existe aussi de grandes différences, pour un pays donné, entre sa capacité d'investissement en science et le nombre de candidats suffisamment qualifiés pour l'exploiter. Associé aux différences entre disciplines concernant leurs besoins en subventions et la disponibilité de celles-ci, ce facteur explique la grande incertitude et l'hétérogénéité du financement, ce qui rend son usage risqué en tant que mesure de l'excellence des chercheurs et de la recherche.

L'OCDE a historiquement considéré le financement comme un indicateur de l'activité scientifique à l'échelle des pays. De ce point de vue, le financement n'est pas un indicateur d'extrant, mais plutôt un indicateur d'intrant qui renseigne sur les investissements consacrés à la recherche. Même si l'on s'attend à ce qu'il existe une relation entre l'intrant « financement » et les extrants tels que la production et l'impact, la mesure, toutefois, s'est transformée depuis quelques années en indicateur de qualité de la recherche sur le plan individuel: le fait de se voir accorder une subvention est interprété comme une mesure d'extrant signalant la qualité de l'individu ou du laboratoire. C'est le cas depuis longtemps en

sciences, et cela le devient aussi en sciences sociales et humaines, domaines où l'obtention de la permanence, les promotions et les autres formes d'évaluation sont de plus en plus liées au fait de savoir attirer les subventions.

Cette inversion des variables d'entrée et de sortie peut se lire sous un autre angle, d'où l'on voit la bibliométrie servir de plus en plus de variable d'entrée dans le processus d'évaluation des demandes de subvention. Pour le dire simplement, un chercheur abondamment publié et cité est plus susceptible de recevoir des subventions qu'un autre dont le compte de citations est moins élevé. Les citations, par le fait même, servent à prouver la crédibilité d'une personne et à réduire le risque de lui confier des fonds. C'est là une inversion de l'approche consistant à utiliser le financement comme indicateur d'intrant et à mesurer ses effets – c'est-à-dire, le rendement du capital investi – par le décompte des citations reçues. L'ajout du financement au rang des indicateurs de qualité – le fait de le considérer comme une mesure d'extrant plutôt que d'intrant – a de nombreux effets négatifs, liés au fait que de nombreux chercheurs pour avancer leurs travaux – par exemple dans les domaines des sciences sociales et humaines – n'ayant aucun besoin de subventions cherchent à en obtenir afin de se distinguer.

Quels sont les indicateurs de la recherche appliquée ?

L'extrant typique de la recherche fondamentale est la publication dans des revues avec comité de lecture. La recherche appliquée et le développement technologique, eux, concrétisent généralement leurs résultats par des brevets. Comme l'article scientifique, le brevet est un document écrit : il décrit une invention, et octroie à la personne ou à l'entreprise qui en détient la propriété intellectuelle le monopole de sa commercialisation dans un pays donné. C'est dire que l'inventeur qui fait connaître publiquement la recette de son invention gagne l'exclusivité de sa commercialisation.

Accordés par les pays, les brevets sont évalués par des examinateurs qui, contrairement à ceux qui interviennent dans la révision par les pairs des autres lieux de la recherche – périodiques et agences subventionnaires –, sont des employés des agences de brevets. Les critères d’admissibilité varient d’un pays à l’autre, mais sont souvent centrés sur : a) la nouveauté (la création ne doit pas déjà exister ou avoir été décrite), b) l’utilité (ou l’application industrielle, en Europe), c) l’activité inventive (ou la non-évidence, aux États-Unis, ou l’inventivité, en Europe). Le monopole (ou la protection conférée par le brevet) est généralement d’un maximum de 20 ans ; les frais annuels, de quelques centaines à quelques milliers de dollars, servent à garantir la protection du brevet.

L’Organisation mondiale de la propriété intellectuelle (OMPI) est un regroupement des bureaux de brevets du monde entier – on en comptait plus de deux cents en 2016. Pour recenser les brevets déposés dans le monde, il existe trois grandes bases de données : celle de l’United States Patent and Trademark Office (USPTO), celle de l’Office européen des brevets (OEB) et celle de l’Office des brevets du Japon (OBJ). Ces trois organismes, qui forment la Coopération tripartite des brevets, font la promotion de normes et d’infrastructures communes. Il existe une hiérarchie entre les bureaux des brevets, car les dépositaires souhaitent maximiser la durée de leur protection et, conséquemment, de leur potentiel commercial. Les inventeurs canadiens, par exemple, font breveter deux fois plus d’inventions aux États-Unis que dans leur propre pays. Les brevets qui ont été déposés auprès des trois grands offices (USPTO, OEB, OBJ) portent le nom de « brevets triadiques » : on tient pour acquis qu’ils contiennent des inventions ayant un potentiel commercial important. L’importance d’un brevet se mesure aussi par le nombre (et la longueur) de ses revendications, qui donnent une idée de l’étendue de la protection juridique demandée pour le brevet.

Les trois bases de données sont accessibles gratuitement sur le Web, afin de faciliter le processus de brevetage; les inventeurs et leurs représentants doivent fouiller le registre pour prouver la nouveauté de l'invention. Étape clé du processus, le dépôt public de l'invention donne en plus aux personnes qui étudient l'innovation l'occasion de consulter les données – celles de l'USPTO sont aisément téléchargeables. Il est important de se rappeler que les brevets inscrits dans ces bases de données sont à des stades différents du processus; l'OEB et l'OBJ contiennent des *demandes* de brevets, c'est-à-dire que les données obtenues concernent des brevets qui n'ont pas encore été accordés – et qui ne le seront peut-être jamais. Les données de l'USPTO concernent à la fois les demandes et les brevets accordés. C'est un facteur à considérer quand on étudie les brevets, notamment si l'on effectue une analyse comparative. Il existe par ailleurs des bases de données cumulatives sur les brevets; PATSTAT, une gamme de produits de l'OEB axés sur la fouille, l'analyse et la visualisation des données des brevets, en est une. Elle couvre plus de 100 millions de brevets et renseigne sur le statut juridique de 200 millions de dossiers depuis le milieu du XIX^e siècle jusqu'à aujourd'hui. Elle cumule les données de 45 organismes de propriété intellectuelle européens et sa mise à jour est bisannuelle. Ses données brutes s'achètent pour une somme assez raisonnable: quelques milliers d'euros, soit une fraction de ce qu'il en coûte pour acheter des bases de données bibliométriques comme le Web of Science ou Scopus.

Les métadonnées des brevets ressemblent beaucoup à celles des articles scientifiques. Les notions liées à la propriété intellectuelle sont toutefois un peu plus difficiles à démêler. En plus des inventeurs, qui sont les individus à qui l'on doit la conception de l'invention, les brevets ont aussi des cessionnaires, qui sont les personnes ou les organisations qui détiennent la propriété intellectuelle du brevet. Ces différents ensembles de noms et d'organismes – ainsi que les villes, provinces, pays, etc., qui leur sont

rattachés – renvoient à deux concepts distincts: celui d'activité inventive (des inventeurs) et celui de propriété intellectuelle (des cessionnaires). Les deux sont liés, bien sûr (la plupart des brevets appartiennent à des entités ayant leur adresse dans la même province ou le même pays que l'inventeur), mais la différence conceptuelle est importante pour la construction des indicateurs. L'activité en recherche appliquée sera ainsi mesurée selon les inventeurs, tandis que la propriété intellectuelle du brevet ou du portefeuille de brevets devrait l'être à partir des variables associées aux cessionnaires. Soulignons également que si les variables associées aux inventeurs demeurent les mêmes après l'obtention d'un brevet (les inventeurs ne changent pas), la propriété des brevets, elle, peut changer quand ils sont vendus (ou que les organismes cessionnaires le sont).

La documentation liée aux brevets contient des références, ce qui signifie qu'on peut compiler le nombre citations reçues par les brevets. Des études ont montré qu'il existe une relation entre un taux de citation élevé et une valeur économique élevée. Cependant, le processus de citation n'est pas le même dans les brevets que dans les articles scientifiques. Une différence importante vient de la contribution des examinateurs ou des juristes promoteurs du brevet qui, en cherchant à étayer la revendication de nouveauté, peuvent ajouter des références. Les brevets citent par ailleurs des sources très hétérogènes, y compris d'autres brevets, des articles scientifiques, des rapports techniques, des catalogues d'objets technologiques et des sites Web.

Les données des brevets peuvent servir à construire des indicateurs analogues à ceux des articles scientifiques. Par exemple, on peut déterminer le nombre de brevets par la méthode du comptage intégral ou fractionné; compiler des indicateurs de copropriété ou de collaboration entre inventeurs; ou quantifier l'impact d'un brevet – toutes ces opérations pouvant se faire selon la classification technologique ou l'année. On peut ensuite agréger ces données

à différents niveaux, comme pour les publications savantes, soit par individus (les inventeurs, habituellement, mais parfois les cessionnaires), par villes, par provinces ou États, par pays ou par domaines. De tels indicateurs se fondent le plus souvent sur les brevets d'utilité et excluent, par conséquent, les brevets de design et les brevets d'obtention végétale, associés à la création de nouvelles variétés de plantes.

Les indicateurs des brevets sont quelque peu limités quant à leur capacité à évaluer la recherche appliquée et l'évolution technologique – certaines innovations n'étant simplement pas brevetables. En Europe, par exemple, les logiciels ne sont pas admissibles au dépôt de brevet, et aux États-Unis, il n'y a pas de voie directe pour les faire breveter. D'autres innovations sont brevetables, mais les inventeurs et les cessionnaires préfèrent ne pas dévoiler le brevet. Alors que la diffusion du savoir est essentielle à la production de capital universitaire, ce n'est pas toujours le cas du brevetage, qui relève d'un autre type de capital – celui-là économique. De nombreuses entreprises préfèrent garder le secret de leur invention ou de leur procédé industriel, quand cela génère davantage de capital économique que l'obtention du brevet. C'est le calcul que doivent faire les entreprises, entre le brevetage et la confidentialité: alors que le brevet promet un monopole d'au maximum 20 ans – mais une concurrence accrue après ce délai –, le secret, lui, pourrait durer plus longtemps. Le brevetage, par ailleurs, n'est pas toujours un indicateur de commercialisation. Certaines entités détiennent des brevets pour des inventions qu'elles ne souhaitent pas commercialiser; elles cherchent plutôt à empêcher la concurrence de le faire. Dans de tels cas, le brevetage nuit à l'innovation au lieu de la susciter. Enfin, soulignons que les brevets ont une couverture disciplinaire extrêmement étroite. Tandis que la quantification de la recherche au moyen des articles savants s'applique à tous les domaines – avec les réserves habituelles au sujet de la couverture des données en sciences sociales

et humaines –, les brevets ne concernent qu'un tout petit nombre de disciplines : la grande majorité des brevets sont accordés dans deux domaines essentiellement, l'électronique et la biotechnologie.

Quelle est la relation entre les indicateurs scientifiques et la révision par les pairs ?

La révision par les pairs est au cœur du système d'évaluation de la recherche. Elle fait référence au processus qui consiste, pour les chercheurs, à soumettre leurs travaux réalisés ou projetés à l'examen de spécialistes de leur domaine de recherche. La méthode touche plusieurs dimensions du système scientifique : elle sert à évaluer le contenu des articles en vue de la publication, les demandes de bourse ou de subvention, et les candidatures à des postes ou à des promotions. Elle est le moyen privilégié pour juger de la qualité des chercheurs et de la recherche depuis le milieu du XX^e siècle. On ne s'étonnera donc pas que les premiers bibliomètres s'en soient servis comme mesure de comparaison des résultats.

Les études réalisées au fil du temps et des disciplines montrent que le nombre de citations et le nombre de publications sont toujours fortement corrélés, le premier plus que le second, à d'autres mesures de l'estime des pairs, bien que ces corrélations varient selon les disciplines – plus fortes en sciences naturelles et médicales qu'en sciences sociales, et nettement moins marquées dans les arts et les humanités. L'écart laisse croire que les deux mesures évaluent des dimensions différentes du savoir, dans certaines disciplines en particulier.

Malgré ces variations, les indicateurs bibliométriques tendent de plus en plus à compléter, voire à supplanter l'évaluation par les pairs. L'une des raisons qu'on évoque pour justifier cette tendance est le coût. Pour 2005-2006, par exemple, on estime à 196 millions d'euros le coût de l'évaluation par les pairs pour les conseils de recherche du Royaume-Uni, ce qui représente 6 % des fonds

accordés cette année-là. D'autres coûts scientifiques s'ajoutent à cela, qui sont liés au fait de libérer plusieurs chercheurs de leurs tâches de recherche et d'enseignement afin qu'ils réalisent des évaluations.

La neutralité de l'évaluation par les pairs fait aussi fait l'objet de débats : certaines études montrent à l'évidence que la subjectivité et les partis pris sont à l'œuvre dans l'évaluation par les pairs. Malgré les idéaux mertonien d'une évaluation rigoureusement fondée sur la qualité des travaux ou du chercheur, des biais explicites et implicites influencent l'évaluation de toutes sortes de façons. Certains affirment que, malgré les inconvénients observés, la production d'un corpus approuvé par les pairs illustre la capacité de fonctionner dans ce système, et que les citations constituent des jugements « intersubjectifs » ayant un impact ultérieur. Si l'on peut débattre longuement des mérites relatifs de chaque approche (et on le fait!), il n'en est pas moins évident que chacune présente des inconvénients à certains niveaux d'analyse. Dans les domaines relativement homogènes, par exemple, l'évaluation par les pairs demeure l'indicateur le plus puissant pour mesurer la valeur des individus et des petits groupes de recherche. Mais quand il s'agit d'évaluer un grand nombre de chercheurs ou des groupes très hétérogènes, la capacité d'analyse à grande échelle des indicateurs scientifiques se révèle plus utile.

Soulignons par ailleurs qu'il n'y a jamais eu d'indépendance totale entre les indicateurs et la révision par les pairs. Les évaluateurs ne sont pas immunisés contre les indicateurs et, même s'ils reçoivent la consigne de n'évaluer que la « qualité » des dossiers qu'ils reçoivent, ils survoleront le c.v. ou le profil Google Scholar du candidat pour connaître sa production, le FI des revues qui le publient et les différents taux de citation. Même quand ces éléments n'ont pas de place officielle dans l'évaluation, ils sont souvent implicitement pris en compte. Les corrélations entre évaluation par les pairs et évaluation par les indicateurs sont

donc inévitables. N'en demeure pas moins que chaque méthode a ses forces et ses faiblesses, et qu'on est sans doute bien avisé de les utiliser en conjonction – une troisième approche qu'on appelle «l'évaluation par les pairs éclairée par des indicateurs».

4

VUE D'ENSEMBLE

Qui contrôle l'évaluation de la recherche ?

Le savoir est un bien public, dit-on : un produit dont la possession exclusive est onéreuse, et dont la valeur ne diminue pas avec le partage. Mais cette définition ne passe pas le test du réel, quand on sait que des éditeurs scientifiques à but lucratif détiennent en quasi-exclusivité le savoir – du moins dans sa forme transmissible – et l'exploitent. La perversion de ce système, où les chercheurs et les établissements auxquels ils sont rattachés doivent payer (plusieurs fois, même, à cause des abonnements qui se chevauchent) pour avoir accès aux travaux qu'ils produisent et fournissent gratuitement, atteint également l'évaluation de la recherche. La communauté scientifique produit les données qui la gouvernent, mais elle ne les possède pas, et les décisions concernant ces données sont souvent motivées par l'intérêt des entreprises, et non de la science.

C'est à grande échelle que les indicateurs de la science déploient pleinement leur efficacité. Deux types d'organisations sont en mesure, plus que les autres, d'exploiter cette amplitude : les gouvernements et les grandes entreprises. Les gouvernements s'intéressent surtout aux données concernant les chercheurs et les

ressources, par des sondages sur l'éducation, la main-d'œuvre scientifique et les investissements en recherche et développement (R-D). Sauf quelques remarquables exceptions (le portail éditorial SciELO, au Brésil, en est une), les gouvernements ne colligent pas directement de données de publication et de citation sur leurs chercheurs. Le Web of Science (WoS) et, plus tard, Scopus, ont été utilisés afin de combler ce manque. Or, le monopole des entreprises sur les données concernant la production et l'impact de la recherche a des répercussions importantes sur la mesure de la recherche. Chaque entreprise a des stratégies et des priorités qui se répercutent sur l'étendue et la disponibilité des données et la fonctionnalité de son portail. Leurs revenus dépendent des abonnements institutionnels – ceux des bibliothèques – qui achètent leurs outils, davantage pour l'extraction documentaire que pour l'évaluation de la recherche. Mais le statut dominant de ces outils parmi les moteurs de recherche est menacé par l'essor de Google Scholar et d'autres sources gratuites offertes sur la Toile. L'incertitude concernant la clientèle de base pourrait expliquer que Thomson Reuters ait vendu sa division Intellectual Property & Science – propriétaire du WoS, maintenant commercialisé sous le nom de Clarivate Analytics – à Onex Corporation (Canada) et Baring Private Equity Asia (Chine) pour 3,55 milliards de dollars américains, en 2016. Dans ce contexte de forte concurrence, Elsevier a plutôt choisi d'acquérir d'autres plateformes, telles Mendeley et Plum Analytics, et d'intégrer leurs fonctionnalités et leurs indicateurs à ses propres outils d'analyse, de manière à diversifier son portefeuille et à élargir la gamme de ses indicateurs de recherche.

Des arguments économiques influencent ainsi les variables indexées par les entreprises et les indicateurs qu'elles construisent. En retour, ces décisions établissent les normes auxquelles la communauté des bibliomètres devra se plier. Que les experts s'adaptent aux outils, plutôt que l'inverse, est contraire à ce qu'on attend d'un

regroupement d'experts, mais c'est pourtant ce qui se produit. La compilation et la promotion du *Journal Impact Factor*, le fameux facteur d'impact (FI), l'illustrent bien : malgré le consensus du milieu de la recherche sur ses failles, très peu de modifications ont été apportées à cet indicateur. Parfois, ce sont des groupes d'intervenants qui pèsent lourd dans la balance : l'indice h, dont la communauté bibliométrique a bien montré les limites, figure en bonne place sur les principaux portails, à cause de l'attention qu'il a reçue du milieu de la recherche et des demandes subséquentes des chercheurs et des administrateurs qui le réclament, bien souvent, en ignorant tout de ses lacunes inhérentes. Dans d'autres cas, des fournisseurs de données deviennent producteurs d'indicateurs, ce qui soulève de toute évidence un conflit d'intérêts. Des études ont montré par exemple que CiteScore, un indice d'impact des revues produit par Elsevier et censé concurrencer le FI, favorise les périodiques d'Elsevier aux dépens de ceux de son concurrent, Springer-Nature. Dans le même esprit, l'indexation récente des remerciements par le WoS se limite aux articles où le financement est dévoilé (au lieu de s'appliquer à tous ceux qui contiennent des remerciements), parce que cette information monnayable est utile aux organismes subventionnaires.

Malgré la mainmise des entreprises sur les données brutes et l'infrastructure, un certain nombre de groupes de recherche indépendants se procurent les données dans le but de réaliser des analyses bibliométriques poussées. Ces groupes font aussi bien de la recherche fondamentale que de la recherche appliquée. Même s'ils financent leurs activités en acceptant des mandats d'organismes gouvernementaux, ils se servent aussi des données pour mettre au point des indicateurs et étudier la structure et le développement de la science. Parce qu'ils ont accès à toute la gamme des données bibliométriques, ils ont été les principaux moteurs de l'innovation et de la critique en évaluation de la recherche. Cela, toutefois, crée de l'inégalité dans le domaine, car la recherche se concentre chez

ceux qui ont les moyens d'acheter les données et de les traiter. En outre, comme ces groupes utilisent les données sous licence (ils achètent le droit d'utiliser les données, mais n'en sont pas propriétaires), ils sont vulnérables aux changements de priorités de l'entreprise. Par exemple, bien que le WoS permette encore à certains groupes de chercheurs de réutiliser leurs données dans des projets de recherche appliquée, Elsevier a décidé de limiter ces usages de manière à privilégier le recours à ses propres outils d'analyse. C'est pourquoi les projets de recherche en bibliométrie appliquée fondés sur les données de Scopus se font essentiellement par l'intermédiaire du portail de recherche et de services d'Elsevier. Ce genre d'entraves affecte la transparence des ensembles de données et, par la suite, limite les possibilités d'analyse.

On pourrait objecter que la production d'outils, quand elle est motivée par les forces du marché, préserve une saine concurrence dans le système et stimule l'innovation. Le WoS a détenu un monopole pendant un demi-siècle et, à l'heure actuelle, de nombreux produits faisant leur entrée sur le marché en bousculent la structure. Certains appartiennent à des mégaentreprises (Microsoft ou Google, p. ex.), d'autres à de petits organismes de recherche (Semantic Scholar, p. ex.). Ces produits pourraient constituer des sources de données plus inclusives et abaisser les barrières d'entrée pour ceux dont le but est de mesurer la recherche. Mais comme ils souffrent actuellement de lacunes importantes quant à la couverture et à la qualité des données, l'avenir de la bibliométrie dépend encore lourdement de la prospérité de quelques entreprises et de leur volonté de donner accès aux données qu'elles possèdent.

Quelles sont les responsabilités des intervenants ?

Plusieurs groupes d'intervenants se partagent la responsabilité de mesurer la recherche, chacun ayant une expertise et un champ

d'action légèrement différents. Les données sont, de bien des façons, la base de cette évaluation. Les fournisseurs de données doivent être tenus responsables de l'exactitude et de la transparence de leurs données, et expliciter leurs critères de sélection et la couverture de leurs indices. Mais il n'existe aucun mécanisme pour les contraindre à assumer cette responsabilité. Dans l'état actuel des choses, ce sont les consommateurs de données qui doivent connaître les particularités des celles-ci et s'en servir judicieusement, à l'intérieur des limites connues. Ils doivent se rappeler avant tout qu'aucune des sources de données utilisées en bibliométrie n'est exhaustive. Les indicateurs, comme leur nom l'indique, pointent vers un résultat, ils ne démontrent rien. On croit souvent que la couverture des indicateurs de production les plus élémentaires est complète. Pourtant, si l'ensemble de périodiques recensés par le WoS, par exemple, peut convenir à l'étude de certaines problématiques, il ne convient pas à toutes. Chacune des bases de données n'inclut qu'une fraction des articles publiés – jamais la totalité, ni même un échantillon représentatif. Les utilisateurs doivent comprendre les aspects fonctionnels des outils pour arriver à interpréter correctement les données obtenues. Ils se souviendront, par exemple, que la plupart des portails qui fournissent des données bibliométriques ont été mis sur pied à des fins de recherche d'information. Or, il y a des contraintes inhérentes au fait d'utiliser un outil de recherche d'information comme un outil d'évaluation, car les caractéristiques attendues de celui-ci (couverture exhaustive, homogénéité des données) ne sont pas nécessairement celles que présente celui-là (taux de rappel, p. ex.). De nombreuses entreprises qui achètent ou compilent des données bibliométriques et qui fournissent des services d'analyse ont vu le jour pour répondre aux besoins de comparaison des établissements. Leur essor soulève le risque d'éloigner les décideurs des données qui fondent les décisions. Ces entreprises, comme Academic Analytics, ont donc la responsabilité d'être

transparentes quant au périmètre couvert par leurs outils ; elles doivent aussi permettre aux individus d'avoir accès à leurs données et, en cas d'erreur, prévoir des mécanismes de correction. Il est possible, en outre, que même des données impeccables ne représentent pas adéquatement l'état d'un domaine – c'est le cas de certaines disciplines qui fusionnent des domaines distincts comme l'astronomie et la physique, qui ont des pratiques de publication et de citation particulières. La qualité des extraits varie alors énormément en fonction du niveau d'analyse, et elle exige une désambiguïsation rigoureuse. En l'absence de données nettes et de déclarations explicites de leurs limites, les rapports obtenus par ces outils présentent une distorsion parfois grave de la réalité du système qu'ils prétendent mesurer. Les administrateurs de la recherche, qui sont les premiers utilisateurs de ces outils, doivent être tenus responsables de l'usage adéquat des données.

Les responsables des politiques publiques jouent un rôle crucial, car ils exploitent les résultats des évaluations, d'une part, tout en consacrant les indicateurs qui produisent ces résultats, d'autre part. Les décisions stratégiques, en matière de financement autant qu'en matière de politiques, s'appuient souvent sur des chiffres. Les politiques, en retour, forment la structure incitative dans laquelle évoluent les chercheurs. Elles subissent aussi l'épreuve des chiffres. Dans l'ensemble, la quantification de la recherche devrait permettre aux décideurs politiques de comprendre le système qu'ils s'efforcent de gouverner et les aider à déterminer les interventions les plus appropriées pour atteindre les objectifs des politiques. La comparaison et la standardisation sont d'une importance critique pour les intervenants de ce groupe. Ils doivent demeurer prudents et savoir reconnaître que tel changement observé est un effet de la politique, et non le signe de l'atteinte réelle de l'objectif. Ainsi, le fait de fonder les évaluations sur la collaboration peut faire augmenter les pratiques d'autorat honorifique, au lieu du nombre de collaborations authentiques. Dans

ce contexte, la responsabilité des décideurs est de rendre transparents les objectifs et l'évaluation.

Les maisons d'édition et les rédacteurs des revues scientifiques sont en quelque sorte les garants des indicateurs. Les éditeurs, en effet, se servent souvent d'indicateurs pour faire la promotion de leurs revues, avec pour conséquence la consécration de ces indicateurs comme mesure d'excellence (des revues et des éditeurs qui les publient). Le FI nous offre l'exemple flagrant d'un usage abusif ayant des effets pervers sur le milieu de la recherche. L'absence de normalisation par domaine, la fenêtre de citation trop courte pour presque toutes les disciplines, les asymétries non corrigées : tout cela montre que l'indicateur vedette est mal conçu, même pour l'usage auquel il était destiné. Mais les effets sociologiques sont encore plus graves : la promotion des indicateurs et les incitatifs qu'on déploie à cause d'eux agissent sur les comportements des auteurs et des éditeurs savants. Ces derniers sont les gardiens de la diffusion des connaissances. C'est un honneur qui vient avec une responsabilité, celle de veiller à ne pas promouvoir les indicateurs ayant un effet néfaste sur la communauté scientifique. De plus, ils sont les mieux placés pour défendre la communauté savante devant les éditeurs commerciaux, dont les objectifs sont plus souvent financiers que scientifiques.

La communauté scientifique forme le groupe social le plus important en ce qui concerne la mesure de la recherche. Elle est à la fois l'objet de l'évaluation et la première consommatrice des données. Il est donc vital que les personnes qui subissent l'évaluation soient conscientes de ce qu'elle suppose et qu'elles y participent d'une façon active. Quand des données sont réunies au sujet d'un chercheur, celui-ci devrait y avoir accès et pouvoir les corriger au besoin. À l'heure actuelle, cependant, les données bibliométriques sont difficiles à désambiguïser au niveau des individus et contiennent beaucoup d'erreurs. Les personnes qui produisent le savoir sont les mieux à même de juger de l'exactitude

de leur dossier. Les chercheurs doivent être très renseignés sur les jeux de données qui servent à évaluer leurs travaux, et cultiver le dialogue avec les personnes qui se servent de ces données pour prendre des décisions. Ce sont souvent les chercheurs eux-mêmes, dans leur rôle d'évaluateurs, qui prennent ces décisions; ils doivent donc impérativement comprendre les réserves émises au sujet des données.

Les étudiants des cycles supérieurs ne sont pas à l'abri de la culture de l'évaluation. De fait, la génération actuelle est sans doute plus sensibilisée aux indicateurs de la recherche qu'aucune autre avant elle. La responsabilité des éducateurs est de veiller à ce que ces étudiants n'apprennent pas à se fier aveuglément aux indicateurs pour mesurer leur réussite. On devrait les former sur la construction et l'interprétation correctes des mesures afin qu'ils ne propagent pas les mythes et les mauvais usages des indicateurs. C'est un fait connu que la concurrence est féroce en matière d'emploi – à l'université plus encore qu'ailleurs –, et qu'on a souvent recours à des indicateurs pour discriminer les candidats à un poste. Malgré tout, les étudiants devraient apprendre à se focaliser d'abord et avant tout sur les intrants du système du savoir, pas sur l'évaluation des extrants: ils doivent prioriser la science, plutôt que de penser leur programme de recherche personnel en fonction des indicateurs.

Les responsabilités de chaque groupe d'intervenants se déclinent en nuances particulières, mais tous les groupes sont concernés par cinq questions fondamentales touchant (1) le temps, (2) la qualité, (3) la normalisation, (4) la couverture, (5) l'adéquation. En ce qui concerne le temps, la fenêtre des publications et des citations étudiées doit correspondre à ce qu'on veut analyser. Souvenons-nous que la production d'un savoir de haut niveau demande du temps, tout comme l'accumulation de citations. Surtout, les taux de production et de citation varient selon les disciplines et ils changent au fil du temps. Les analyses

qui négligent le facteur temporel sont susceptibles de conduire à des conclusions erronées. La mauvaise qualité des données peut également mener à des résultats inexacts. La plupart des ensembles de données servant à la bibliométrie présentent des problèmes de qualité, en particulier en ce qui concerne la standardisation des noms d'auteurs et d'établissements. Un nettoyage minutieux des données doit précéder l'analyse. La couverture et la normalisation influencent aussi sérieusement l'analyse: les données employées doivent couvrir correctement le sujet étudié et, si l'analyse rejoint plusieurs disciplines, elles doivent être normalisées. Le dernier enjeu, celui de l'adéquation, est la clé de toutes les autres. La compilation d'un indicateur bibliométrique doit répondre à un objectif clairement défini, et la sélection des données doit respecter des critères compatibles avec cet objectif. Trop souvent, des analyses sont réalisées sans que les données, les indicateurs et l'objectif de la recherche soient alignés. Ceux qui sont en position d'évaluer et de prendre des décisions basées sur l'évaluation doivent s'assurer que les données et les indicateurs correspondent à l'objectif global de l'évaluation. Ultimement, il revient à l'ensemble de la communauté scientifique d'interpréter les résultats obtenus par l'intermédiaire de ce que Merton appelle «le scepticisme organisé». Trop souvent, on accorde aux résultats quantitatifs une aura de précision et une autorité qu'ils ne méritent pas. Le concept de signification statistique, par exemple, qui a servi, des dizaines d'années durant, de pierre d'assise à la recherche quantitative, est de plus en plus remis en question. Il est souvent et à la fois mal appliqué et mal interprété. C'est le cas tout particulièrement en bibliométrie, où l'on confond les concepts d'échantillon et de population. L'échantillonnage adéquat et la normalité de la distribution sont deux exigences des statistiques auxquelles les données bibliométriques, souvent, passent outre. Celles qu'on obtient du WoS ou de Scopus sont celles d'une population d'articles. Mais cette population est formée des articles

recensés par ces deux index, et non d'un échantillon aléatoire de tous les articles publiés. Par ailleurs, étant donné la taille des ensembles de données dans les exercices d'évaluation à l'échelle mondiale, la probabilité d'obtenir des résultats statistiquement significatifs est élevée, même quand il n'existe aucune différence valide à interpréter. En bibliométrie, comme dans tous les secteurs de la science, la communauté scientifique a donc la responsabilité d'interpréter les résultats en fonction de leur contexte.

Quels sont les effets négatifs de la quantification ?

Les indicateurs ont des effets sur les comportements. Dans un environnement dépourvu de jugements de valeur, la mesure des extrants ne devrait pas influencer les intrants, ou très peu. Mais les chercheurs, comme tous les êtres qui vivent en société, réagissent au fait d'être observés. À mesure que les indicateurs de la recherche s'institutionnalisent, les chercheurs adaptent leur comportement aux mesures de l'évaluation. De telles adaptations peuvent renverser le système scientifique, en rendant le capital qu'il génère plus important que la science elle-même. Le déplacement d'objectifs est un problème grave en évaluation de la recherche, parce qu'il menace d'une distorsion fondamentale le processus qu'il surveille (un phénomène décrit en 1976 par le psychosociologue Donald T. Campbell et résumé par une loi qui porte son nom). Suivant la formule des économistes suisses Bruno Frey et Margit Osterloh, la quantification de la recherche pourrait avoir incité les chercheurs à cultiver « le goût du classement » davantage que « le goût de la science ».

On reproche souvent aux politiques scientifiques de niveler les comportements, par des mandats explicites ou la simple surveillance. On observe en effet un lien direct entre les régimes de financement et la production : quand la priorité est accordée au nombre de publications, la production augmente ; quand on

privilégie les travaux ayant un fort impact, ce sont les publications dans des revues à taux de citation élevé qui augmentent; et quand on valorise la collaboration, ce sont les travaux réalisés conjointement qui augmentent. Cela illustre sans doute l'efficacité des politiques scientifiques, mais non sans avoir également des conséquences imprévues. Ainsi, quand l'Australie a adopté un nouveau schème d'évaluation récompensant la publication dans les revues recensées par l'Institute for Scientific Information, elle a atteint son but. Mais, tandis que le milieu australien de la recherche augmentait sa part des travaux indexés, l'impact moyen de ces travaux diminuait. Inciter les scientifiques à produire augmente peut-être la production, mais génère en parallèle des stratégies peu recommandables – comme le saucissonnage, c'est-à-dire la répartition des résultats d'une seule expérience en plusieurs articles afin d'augmenter le nombre de publications – et des travaux de moindre qualité.

Les formules qui privilégient les citations ont aussi des effets négatifs. La Chine, par exemple, offre des stimulants pécuniaires aux auteurs qui publient dans des revues ayant un facteur d'impact (FI) donné. Cela augmente radicalement le nombre d'articles soumis à des revues comme *Science*, sans donner lieu à une hausse correspondante des manuscrits acceptés. La communauté scientifique doit donc assumer le fardeau qui consiste à examiner des travaux soumis à une revue qui n'est peut-être pas la plus indiquée, ni par le champ thématique ni par la portée, parce que la proposition est motivée par des objectifs financiers plutôt que scientifiques. Un lien a été établi entre ce genre d'exigences et la fraude, ainsi que d'autres formes d'abus du système scientifique.

Par ailleurs, même dans les cas où les mesures incitatives sont bonnes en soi, elles ne conviennent pas nécessairement à toutes les disciplines. Malgré la valorisation du travail d'équipe pendant les études, l'essor des entreprises transnationales et d'autres projets à vocation mondialisante qui tendent à promouvoir la

collaboration, cette approche ne s'applique pas nécessairement à tous les domaines d'études. De la même façon, si la valorisation de l'impact de la recherche sur la société procède d'une intention louable, elle pourrait aussi favoriser la recherche appliquée au détriment de la recherche fondamentale. On en conclura que les politiques scientifiques doivent être sensibles à la diversité de la science. Les politiques monolithiques qui ne reconnaissent pas l'hétérogénéité patente du paysage de la recherche ont toutes les chances d'avoir des contrecoups néfastes.

Les données et les instruments de mesure utilisés pour l'évaluation peuvent aussi causer des effets non souhaités. Le périmètre des bases de données peut provoquer des distorsions dans les modes de production, avec des effets variables selon les disciplines. La plupart des bases de données bibliométriques, par exemple, sont centrées sur des articles de périodiques. Les indicateurs qui exploitent ces bases de données se trouvent ainsi à discriminer les disciplines dont la production adopte d'autres formats. C'est l'un des facteurs qui pourraient expliquer l'augmentation du recours aux articles – et la diminution correspondante du recours aux livres – comme véhicules de transmission du savoir, en particulier en sciences sociales et humaines. Ces changements dans le mode de production touchent aussi le genre de recherches effectuées: les articles conviennent bien à des travaux brefs et empiriques, car ils présentent des analyses quantitatives ponctuelles plus souvent que des travaux théoriques. Par ailleurs, un travail ne peut être comptabilisé que s'il est indexé. Les écarts de couverture pour les différentes langues amènent les chercheurs à publier en anglais préférablement. Or, selon le domaine d'études, cette langue n'est pas toujours la plus utile.

L'une des failles majeures de la bibliométrie est l'usage des indicateurs à des niveaux d'analyse pour lesquels ils ne sont pas conçus. Il importe de distinguer entre l'évaluation d'un milieu scientifique (à partir de données cumulées par disciplines, par

établissements et par pays) et l'évaluation des individus. La bibliométrie est une méthode dont l'application la plus pertinente se fait à l'ensemble du système scientifique, dans le but de comprendre la structure et la croissance de la recherche fondamentale. Les microanalyses sont sujettes à des erreurs importantes: les problèmes de couverture et les petites inexactitudes peuvent avoir de gros effets. Comme la loi des grands nombres nous invite à le penser, les mesures bibliométriques sont plus stables et leur validité plus assurée aux niveaux supérieurs d'agrégation des données. Au niveau des individus, elles peuvent servir de complément d'analyse, mais ne devraient pas remplacer les autres méthodes d'évaluation: les observations sont tout simplement insuffisantes, et la variance beaucoup trop grande, pour en tirer de l'information significative. Les mesures chiffrées suscitent aussi une sorte de cercle dans l'évaluation: la corrélation est très élevée entre les personnes qui publient beaucoup, qui sont souvent citées et qui reçoivent une grande part des ressources. L'attribution du financement sur la foi de mesures fortement corrélées peut exacerber l'effet saint Matthieu, et empêcher certains groupes sous-représentés de recevoir la reconnaissance et les ressources qu'ils méritent.

De toute évidence, le milieu de la recherche a intériorisé les indicateurs. La culture de l'évaluation, dans les universités, a suscité une grande ferveur pour le comptage et les classements. Ces mesures peuvent en effet fournir de l'information importante pour les comparaisons et la distribution des ressources. On se méfiera pourtant de la fausse précision de la bibliométrie en matière d'évaluation. Le caractère quantitatif des mesures obtenues évoque une valeur de vérité; or il n'existe pas de valeur de référence pour des concepts comme la production ou l'impact du savoir. Les indicateurs servant à mesurer ces concepts doivent nécessairement être compris dans leur contexte et en toute connaissance des réserves que soulèvent les données et les outils. Hors contexte, ces indicateurs sont, au mieux, vides de sens et, au pire, dangereux. C'est

ce qui a poussé de nombreux chercheurs éminents à s'opposer ouvertement aux mesures d'évaluation et à désavouer les revues et les éditeurs qui font la promotion de certains outils. Quoi qu'il en soit, le pouvoir de créer les indicateurs, comme celui de les rejeter, demeure un privilège. L'élite à qui les indicateurs sont profitables jouit d'un avantage cumulatif qui lui ouvre toutes les portes. Ce sont le plus souvent les personnes marginalisées ou déplacées par les indicateurs qui profiteraient le plus de mesures nouvelles, mais qui ont le moins de moyens d'en produire. La communauté scientifique a la responsabilité de trouver une solution aux inégalités créées et perpétuées par les indicateurs.

Quel avenir pour la bibliométrie ?

Le *Science Citation Index*, en son temps, fut une entreprise assez audacieuse. Quelques heures passées à survoler les versions papier de cet index, et l'on se demande quel avenir Eugene Garfield imaginait pour lui. Aujourd'hui, les avancées de l'informatique et la publication numérique des documents savants multiplient les sources actuelles et rendent les sources anciennes rapidement consultables. Les outils bibliométriques continueront à se multiplier, avec quelques différences quant à leurs utilisateurs, à leur longévité et à leur couverture. En ce moment, le marché est favorable aux consommateurs, car il existe des indicateurs pour toutes les dimensions concevables de la recherche. Les personnes qui plaident en faveur de nouveaux indicateurs évoquent souvent l'argument de l'inclusion. Mais la multiplication forcenée des indicateurs ne garantit pas une augmentation de l'équité. Nature Publishing Group, par exemple, a créé un index qui classe les pays et les établissements en fonction du nombre d'articles qu'ils ont publiés dans les revues du groupe. C'est un mécanisme qui sert à générer du capital – économique et scientifique – à partir d'un produit savant. En créant cet index (une démarche que le

capital universitaire de *Nature* rend crédible et souhaitable), la maison d'édition stimule la volonté de publier dans les pages de ses propres périodiques, ce qui génère un capital économique non négligeable. La réciprocité du capital économique et du capital universitaire est un élément central dans l'évaluation des indicateurs. Alors que les classements biaisés sont facilement rejetés quand ils sont le fait de groupes marginaux, les éditeurs, les universités et les individus qui disposent d'un énorme capital jouissent d'un pouvoir de persuasion à l'avenant sur le marché de la science. Ce phénomène ne se limite pas au secteur à but lucratif. Aux États-Unis, les National Institutes of Health, chefs de file en matière de financement des sciences biomédicales, ont produit un nouvel indicateur, le « ratio de citations relatif » (*Relative Citation Ratio*) devant servir aux évaluations internes et l'ont intégré à leur outil en ligne (iCite). L'indicateur a essuyé les critiques sévères de la communauté bibliométrique, mais n'a subi aucune véritable modification depuis sa parution tout en profitant d'une abondante visibilité attribuable à la notoriété des NIH. Bref, le capital engendre le capital, et la mesure selon laquelle les nouveaux outils d'évaluation sont monnayés ou acquièrent le statut de bien public modifiera certainement le monde de la recherche. En outre, bien que l'hétérogénéité en hausse du nombre de mesures possibles présente de nombreux avantages, leur prolifération ouvrirait la porte à un particularisme outrancier, apte à priver les indicateurs de leur capacité à fournir des équivalences à l'échelle mondiale et de faire communiquer les disciplines. La nécessité de la standardisation et de l'interopérabilité a donné lieu à plusieurs projets visant à faciliter le passage entre les ensembles de données grâce à des identifiants uniques. L'ORCID (pour *Open Research Contributor Identification*), par exemple, est intégré aux différents portails bibliométriques depuis 2010 dans le but de fournir à chaque chercheur un identifiant unique. L'ORCID simplifie la désambiguïsation des auteurs en associant à chacun un code relié aux extraits

qui lui correspondent, dans tous les genres et sur tous les portails. À la fin de 2016, l'ORCID comptait près de 2,7 millions d'abonnés et des centaines de membres organisationnels, y compris les principales universités et maisons d'édition. Un autre exemple de standardisation réussie est celui de Crossref, un registre d'identifiants numériques, les *Digital Object Identifiers* (DOI) lancés en 2000 par la Publishers International Linking Association (un organisme à but non lucratif). Le but de ce registre est de faciliter le couplage de documents entre les revues et les portails. Sans offrir lui-même de contenu scientifique, mais en reliant les citations, il sert de pont entre les contenus scientifiques hébergés par différents sites. Crossref fournit des DOI à plusieurs types de documents, y compris des articles en prépublication, des articles publiés, des livres, des actes de colloques, des thèses et des ensembles de données; à la fin de 2016, il en avait attribué 80 millions.

D'autres outils ont été mis au point dans le but non seulement de relier des documents scientifiques, mais aussi de les dépouiller. La disponibilité toujours plus grande de documents en texte intégral, grâce aux services d'autoarchivage (ce qu'on appelle «la voie "verte" du libre accès»), a donné lieu à quelques projets d'envergure visant à les indexer et à s'en servir pour construire des indicateurs. Semantic Scholar, par exemple, un produit de l'Allen Institute for Artificial Intelligence, recense tous les articles d'arXiv et de la Digital Bibliography & Library Project (DBLP) – entre autres bases de données –, ce qui représente une proportion importante de tous les travaux publiés dans les domaines de la physique, des mathématiques et de l'informatique. À partir des données obtenues, Semantic Scholar compile des indices de citations à l'échelon des auteurs et des documents. De plus, il se sert des techniques de l'apprentissage-machine pour déterminer la pertinence d'une référence par rapport au document qui la cite. Cependant, tout comme Google Scholar, sa puissance bibliométrique ne s'exerce que sur le contenu de son portail. De

plus, il n'offre aucune interface de programmation automatique (API) permettant de fouiller et d'extraire son contenu, de sorte qu'il demeure fermé à l'évaluation externe. Quoi qu'il en soit, dans la mesure où la formule du libre accès est en expansion, les initiatives de ce genre devraient prendre leur essor, d'autant plus qu'elles permettent la construction d'indicateurs reposant sur l'analyse contextuelle des citations.

L'avenir de la bibliométrie dépend en grande partie d'outils qui sauront se superposer aux données existantes, qui fonctionneront sur tous les portails et qu'on pourra adapter à des besoins particuliers. Les chercheurs, les administrateurs et les décideurs exigeront des données couvrant un périmètre toujours plus grand et qu'ils pourront moduler au gré de leurs besoins. Pour ce faire, on devra leur fournir de grands ensembles de données hétérogènes et la capacité de les homogénéiser, de les normaliser et de les contextualiser rapidement. Beaucoup des nouveaux utilisateurs de ces données ne seront pas des spécialistes. Les concepteurs d'outils et les analystes devront donc veiller à préserver un fragile équilibre dans la construction des indicateurs. Les plus simples sont faciles à comprendre, mais peuvent présenter de graves biais; les plus complexes sont souvent plus précis, mais aussi plus difficiles à interpréter. Par ailleurs, beaucoup de ces producteurs de nouvelles mesures ne sont pas formés à l'étude des sciences. Involontairement, ils pourraient formuler des indicateurs qui ne respectent pas les critères de base de l'analyse bibliométrique, en employant des plages inadéquates, en négligeant de préparer et de désambiguïser les données, en employant une couverture trop étroite et en omettant de normaliser les données. Des indicateurs ainsi formulés induisent facilement en erreur et se retrouvent dissociés du concept qu'ils sont censés quantifier. C'est donc à la communauté scientifique de rester vigilante et de surveiller les nouvelles données et les nouveaux outils, tout en formant les utilisateurs à l'application pertinente des indicateurs.

Les spécialistes du domaine ont remarqué depuis longtemps la complexité des questions politiques, économiques et scientifiques qui s'entremêlent dans le système de la science, mais ils ont toujours défendu la neutralité éthique des mesures bibliométriques. Pourtant, on voit désormais que la manière de quantifier la recherche a des effets spectaculaires sur ce qui est évalué. Ces effets sont donc la première chose à prendre en considération quand on se sert d'indicateurs pour évaluer la recherche : on doit s'assurer que les indicateurs choisis ne sont pas nuisibles à la communauté des chercheurs. Les évaluateurs feront bien d'adopter le précepte de la profession médicale : avant tout, ne pas nuire. C'est à l'ensemble du milieu scientifique de veiller à évaluer la recherche d'une manière productive, et de s'assurer que l'évaluation renforce le système scientifique qu'elle observe au lieu de le détruire. C'est notre souhait que cet ouvrage, en proposant une introduction accessible aux fondements historiques de la bibliométrie, en montrant les forces et les faiblesses des données et des indicateurs et en donnant des pistes pour les interpréter lucidement, représente un pas dans cette direction.

SUGGESTIONS DE LECTURE

Ce livre présente une synthèse des travaux réalisés depuis plus de cent ans sur la mesure de la science. Parce qu'il s'agit d'un ouvrage de vulgarisation, nous avons décidé d'omettre dans les chapitres les références détaillées des études citées et de les regrouper ci-dessous, dans une liste où nous mentionnons également d'autres textes importants qui nous ont inspirés et qui seront utiles à qui veut comprendre les bases de l'évaluation de la recherche. Cette liste ne prétend pas à l'exhaustivité, et ne représente pas non plus une bibliographie des classiques de notre domaine. C'est plutôt, pour nous, un moyen de reconnaître notre dette intellectuelle, tout en fournissant aux lecteurs intéressés un point de départ pour approfondir leur connaissance de la bibliométrie.

Chapitre 1

- Barabási, A. L. et R. Albert, « Emergence of scaling in random networks », *Science*, vol. 286, n° 5439, 1999, p. 509-512.
- Bourdieu, P., *Homo academicus*, Paris, Les éditions de Minuit, coll. « Le sens commun », 1992 (1984).
- Bourdieu, P., *Science de la science et réflexivité. Cours du Collège de France, 2000-2001*, Paris, Raisons d'agir, coll. « Cours et travaux », 2001.
- Cole, J. et S. Cole, *Social stratification in science*. Chicago, University of Chicago Press, 1973.
- Cronin, B. et C. R. Sugimoto (dir.), *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact*, Cambridge, MIT Press, 2014.
- De Bellis, N., *Bibliometrics and citation analysis: From the Science Citation Index to cybermetrics*, Lanham, Scarecrow Press, 2009.

- de Solla Price, D. J., *Science et suprascience* [*Little Science, Big Science*], trad. par G. Lévy, Paris, Fayard, 1972 [1963].
- Elkana, Y. *et al.*, *Towards a metric of science: The advent of science indicators*, New York, Wiley, 1978.
- Gingras, Y., *Bibliometrics and research evaluation: Uses and abuses*, Cambridge, MIT Press, 2016.
- Godin, B., *Measurement and statistics on science and technology: 1920 to the Present*, London, Routledge, 2005.
- Lazarsfeld, P. F. et R. Boudon, *On social research and its language*, Chicago, University of Chicago Press, 1993.
- Merton, R. K., «The Matthew effect in science», *Science*, vol. 159, n° 3810, 1968, p. 56-63.
- Merton, R. K., *The sociology of science: Theoretical and empirical investigation*, Chicago, University of Chicago Press, 1973.
- Moed, H. F., *Citation analysis in research evaluation*, Dordrecht, Springer, 2004.
- Moed, H. F., W. Glänzel et U. Schmoch (dir.), *Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems*, Dordrecht, Kluwer Academic Publishers, 2005.
- Rossiter, M. W., «The Matthew Matilda effect in science», *Social Studies of Science*, vol. 23, n° 2, 1993, p. 325-341.
- Small, H. G., «Cited documents as concept symbols», *Social Studies of Science*, vol. 8, n° 3, 1978, p. 327-340.
- Sugimoto, C. R. (dir.), *Theories of informetrics and scholarly communication*, Berlin, De Gruyter Mouton, 2016.
- Todeschini, R. et A. Baccini, *Handbook of bibliometric indicators: Quantitative tools for studying and evaluating research*, Weinheim, Wiley-VCH, 2016.
- van Raan, A. F. J., *Handbook of quantitative studies of science and technology*, Amsterdam, Elsevier, 1988.
- Wouters, P., *The citation culture*, thèse de doctorat, Université d'Amsterdam, 1999.

Chapitre 2

- Aguillo, I., «Is Google Scholar useful for bibliometrics? A webometric analysis», *Scientometrics*, vol. 91, n° 2, 2012, p. 343-351.
- Archambault, É. *et al.*, «Comparing bibliometric statistics obtained from the Web of Science and Scopus», *Journal of the Association for Information Science and Technology*, vol. 60, n° 7, 2009, p. 1320-1326.

- Delgado López-Cózar, E., N. Robinson-García et D. Torres-Salinas, «The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators», *Journal of the Association for Information Science and Technology*, vol. 65, n° 3, 2014, p. 446-454.
- Elsevier, «Scopus Content Coverage Guide», 2017. En ligne : <https://www.elsevier.com/>
- Garfield, E., «Citation indexes for science. A new dimension in documentation through association of ideas», *Science*, vol. 122, n° 3159, 1955, p. 108-111.
- Garfield, E., «Genetics citation index: Experimental citation indexes to genetics with special emphasis on human genetics», *Essays of an Information Scientist*, vol. 7, 1984, p. 515-522.
- Garfield, E., «How ISI selects journals for coverage: Quantitative and qualitative considerations», *Current Contents*, vol. 13, n° 22, 1990, p. 185-193.
- Gordin, M. D., *Scientific Babel: How science was done before and after global English*, Chicago, University of Chicago Press, 2015.
- Jacsó, P., «Google Scholar duped and deduped—the aura of “robometrics”», *Online Information Review*, vol. 35, n° 1, 2011, p. 154-160.
- Kawashima, H. et H. Tomizawa, «Accuracy evaluation of Scopus Author ID based on the largest funding database in Japan», *Scientometrics*, vol. 103, n° 3, 2015, p. 1061-1071.
- Labbé, C., «Ike Antkare one of the great stars in the scientific firmament», *ISSI Newsletter*, vol. 6, n° 2, 2010, p. 48-52.
- Mongeon, P. et A. Paul-Hus, «The journal coverage of Web of Science and Scopus: A comparative analysis», *Scientometrics*, vol. 106, n° 1, 2016, p. 213-228.
- Orduna-Malea, E. *et al.*, «Methods for estimating the size of Google Scholar», *Scientometrics*, vol. 104, n° 3, 2015, p. 931-949.
- Robinson-García, N., E. Jiménez-Contreras et D. Torres-Salinas, «Analyzing data citation practices using the data citation index», *Journal of the Association for Information Science and Technology*, vol. 67, n° 12, 2015, p. 2964-2975.
- Sugimoto, C. R. et S. Weingart, «The kaleidoscope of disciplinarity», *Journal of Documentation*, vol. 71, n° 4, 2015, p. 775-794.
- Testa, J., «Journal Selection Process», 2016. En ligne : <https://clarivate.com/>
- Torres-Salinas, D. *et al.*, «Mapping citation patterns of book chapters in the Book Citation Index», *Journal of Informetrics*, vol. 7, n° 2, 2013, p. 412-424.

Chapitre 3

- Acharya, A. *et al.*, «Rise of the rest: The growing impact of non-elite journals», arXiv preprint, arXiv:1410.2217, 2014.

- Archambault, É. et V. Larivière, «History of journal impact factor: Contingencies and consequences», *Scientometrics*, vol. 79, n° 3, 2009, p. 639-653.
- Bergstrom, C. T. et J. D. West, «Comparing impact factor and Scopus CiteScore», 2016. En ligne : <http://eigenfactor.org/>
- Bergstrom, C. T., J. D. West et M. A. Wiseman, «The Eigenfactor™ metrics», *Journal of Neuroscience*, vol. 28, n° 45, 2008, p. 11433-11434.
- Bertin, M. *et al.*, «The linguistic patterns and rhetorical structure of citation context: An approach using n-grams», *Scientometrics*, vol. 109, n° 3, 2016, 1417-1434.
- Birnholtz, J. P., «What does it mean to be an author? The intersection of credit, contribution, and collaboration in science», *Journal of the American Society for Information Science and Technology*, vol. 57, n° 13, 2006, p. 1758-1770.
- Bollen, J. *et al.*, «A principal component analysis of 39 scientific impact measures», *PLoS ONE*, vol. 4, n° 6, 2009, e6022.
- Bornmann, L., «Scientific peer review», *Annual Review of Information Science and Technology*, vol. 45, n° 1, 2011, p. 197-245.
- Bornmann, L. et H. D. Daniel, «What do citation counts measure? A review of studies on citing behavior», *Journal of Documentation*, vol. 64, n° 1, 2008, p. 45-80.
- Catalini, C., N. Lacetera et A. Oettl, «The incidence and role of negative citations in science», *Proceedings of the National Academy of Sciences*, vol. 112, n° 45, 2015, p. 13823-13826.
- Cole, S. et J. R. Cole, «Scientific output and recognition: A study in the operation of the reward system in science», *American Sociological Review*, vol. 32, n° 3, 1967, p. 377-390.
- Galison, P. et M. Biagioli (dir.), *Scientific authorship: Credit and intellectual property in science*, New York, Routledge, 2003.
- Garfield, E., «Is citation analysis a legitimate evaluation tool?», *Scientometrics*, vol. 1, n° 4, 1979, p. 359-375.
- Gauffriau, M. et P. O. Larsen, «Counting methods are decisive for rankings based on publication and citation studies», *Scientometrics*, vol. 64, n° 1, 2005, p. 85-93.
- Glänzel, W. et H. F. Moed, «Journal impact measures in bibliometric research», *Scientometrics*, vol. 53, n° 2, 2002, p. 171-193.
- Glänzel, W. et U. Schoepflin, «A bibliometric study on ageing and reception processes of scientific literature», *Journal of Information Science*, vol. 21, 1995, p. 37-54.

- González-Pereira, B., V. P. Guerrero-Bote et F. Moya-Anegón, «A new approach to the metric of journals' scientific prestige: The SJR indicator», *Journal of Informetrics*, vol. 4, n° 3, 2010, p. 379-391.
- Griliches, Z. (dir.), *R&D, patents and productivity*, Chicago, University of Chicago Press, 2007.
- Grossetti, M. *et al.*, «Cities and the geographical deconcentration of scientific activity: A multilevel analysis of publications (1987-2007)», *Urban Studies*, vol. 51, n° 10, 2014, p. 2219-2234.
- Hagen, N. T., «Harmonic publication and citation counting: Sharing authorship credit equitably—not equally, geometrically or arithmetically», *Scientometrics*, vol. 84, n° 3, 2010, p. 785-793.
- Hamilton, D. P., «Publishing by—and for?—the numbers», *Science*, vol. 250, n° 4986, 1990, p. 1331-1332.
- Hirsch, J. E. «An index to quantify an individual's scientific research output», *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, n° 46, 2005, p. 16569-16572.
- ICMJE. «Recommendations», 2016. En ligne: <http://www.icmje.org/recommendations/>
- Jaffe, A. B. et M. Trajtenberg, *Patents, citations, and innovations: A window on the knowledge economy*, Cambridge, MIT Press, 2002.
- Katz, J. S. et B. R. Martin, «What is research collaboration?», *Research Policy*, vol. 26, n° 1, 1997, p. 1-18.
- Larivière, V., É. Archambault et Y. Gingras, «Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900-2004)», *Journal of the American Society for Information Science and Technology*, vol. 59, n° 2, 2008, p. 288-296.
- Larivière, V. *et al.*, «Contributorship and division of labor in knowledge production», *Social Studies of Science*, vol. 46, n° 3, 2016, p. 417-435.
- Larivière, V., Y. Gingras et É. Archambault, «The decline in the concentration of citations, 1900-2007», *Journal of the American Society for Information Science and Technology*, vol. 60, n° 4, 2009, p. 858-862.
- Larivière, V. *et al.*, «A simple proposal for the publication of journal citation distributions», *Biorxiv*, 062109, 2016.
- Larivière, V., G. A. Lozano et Y. Gingras, «Are elite journals declining?», *Journal of the Association for Information Science and Technology*, vol. 65, n° 4, 2014, p. 649-655.
- Lee, C. J. *et al.*, «Bias in peer review», *Journal of the American Society for Information Science and Technology*, vol. 64, n° 1, 2013, p. 2-17.

- Leydesdorff, L. *et al.*, «Turning the tables on citation analysis one more time: Principles for comparing sets of documents», *Journal of the American Society for Information Science and Technology*, vol. 62, n° 7, 2011, p. 1370-1381.
- Li, D. et L. Agha, «Big names or big ideas: Do peer-review panels select the best science proposals?», *Science*, vol. 348, n° 6233, 2015, p. 434-438.
- Lozano, G. A., V. Larivière et Y. Gingras, «The weakening relationship between the Impact Factor and papers' citations in the digital age», *Journal of the American Society for Information Science and Technology*, vol. 63, n° 11, 2012, p. 2140-2145.
- McCain, K. W., «Obliteration by incorporation», dans B. Cronin et C. R. Sugimoto (dir.), *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact*, p. 129-149. Cambridge, MIT Press, 2014.
- Moed, H. F., «Measuring contextual citation impact of scientific journals», *Journal of Informetrics*, vol. 4, n° 3, 2010, p. 265-277.
- Moed, H. F. et T. N. Van Leeuwen, «Improving the accuracy of Institute for Scientific Information's journal impact factors», *Journal of American Society of Information Science*, vol. 46, 1995, p. 461-467.
- Moravcsik, M. J. et P. Murugesan, «Some results on the function and quality of citations», *Social Studies of Science*, vol. 5, n° 1, 1975, p. 86-92.
- OCDE, *Manuel de Frascati 2015, Lignes directrices pour le recueil et la communication des données sur la recherche et le développement expérimental*, Paris, Éditions OCDE, 2015. En ligne : <http://dx.doi.org/10.1787/9789264257252-fr>.
- Paul-Hus, A. *et al.*, «The sum of it all: Revealing collaboration patterns by combining authorship and acknowledgements», *Journal of Informetrics*, vol. 11, n° 1, 2017, p. 80-87.
- Pontille, D., «Authorship practices and institutional contexts in sociology: Elements for a comparison of the United States and France», *Science, Technology & Human Values*, vol. 28, n° 2, 2003, p. 217-243.
- Priem, J., «Altmetrics», dans B. Cronin et C. R. Sugimoto (dir.), *Beyond bibliometrics: Harnessing multi-dimensional indicators of performance*, Cambridge, MIT Press, 2014, p. 263-287.
- Rousseau, R., C. García-Zorita et E. Sanz-Casado, «The h-bubble», *Journal of Informetrics*, vol. 7, n° 2, 2013, p. 294-300.
- Seglen, P. O., «The skewness of science», *Journal of the American Society for Information Science*, vol. 43, n° 9, 1992, p. 628-638.
- Shapin, S., «The invisible technician», *American Scientist*, vol. 77, n° 6, 1989, p. 554-563.

- Small, H., « Co-citation in the scientific literature: A new measure of the relationship between two documents », *Journal of the Association for Information Science and Technology*, vol. 24, n° 4, 1973, p. 265-269.
- Sugimoto, C. R., « “Attention is not impact” and other challenges for altmetrics », *blogue Wiley Exchanges*, 24 juin 2015.
- Sugimoto, C. R. *et al.*, « Scholarly use of social media and altmetrics: A review of the literature », arXiv:1608.08112, 2016.
- Van Leeuwen, T., « Testing the validity of the Hirsch-index for research assessment purposes », *Research Evaluation*, vol. 17, n° 2, 2008, p. 157-160.
- Verstak, A. *et al.*, « On the shoulders of giants: The growing impact of older articles », arXiv preprint, 2014, arXiv:1411.0275.
- Wagner, C. S. *et al.*, « Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature », *Journal of Informetrics*, vol. 5, n° 1, 2011, p. 14-26.
- Wallace, M. L., V. Larivière et Y. Gingras, « Modeling a century of citation distributions », *Journal of Informetrics*, vol. 3, n° 4, 2009, p. 296-303.
- Waltman, L. *et al.*, « The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation », *Journal of the American Society for Information Science and Technology*, vol. 63, n° 12, 2012, p. 2419-2432, arXiv:1202.3941.
- Waltman, L. et N. J. Van Eck, « The inconsistency of the h-index », *Journal of the American Society for Information Science and Technology*, vol. 63, p. 2, 2012, p. 406-415.
- Waltman, L. *et al.*, « Towards a new crown indicator: An empirical analysis », *Scientometrics*, vol. 87, n° 3, 2011, p. 467-481.
- Waltman, L. *et al.*, « Towards a new crown indicator: Some theoretical considerations », *Journal of Informetrics*, vol. 5, n° 1, 2011, p. 37-47.
- Wilsdon, J. *et al.*, *The metric tide: Report of the independent review of the role of metrics in research assessment and management*, London, SAGE, 2015.
- Wouters, P. et R. Costas, « *Users, narcissism and control-tracking the impact of scholarly publications in the 21st century* », Utrecht, SURF Foundation, 2012.

Chapitre 4

- Butler, L., « Explaining Australia's increased share of ISI publications—The effects of a funding formula based on publication counts », *Research Policy*, vol. 32, n° 1, 2003, p. 143-155.
- Hicks, D. *et al.*, « The Leiden Manifesto for research metrics », *Nature*, vol. 520, n° 7548, 2015, p. 429-431.

- Larivière, V., S. Haustein et P. Mongeon, « The oligopoly of academic publishers in the digital era », *PLoS ONE*, vol. 10, n° 6, 2015, e0127502.
- Osterloh, M. et B. S. Frey, « Ranking games », *Evaluation Review*, vol. 39, n° 1, 2015, p. 102-129.
- Schneider, J. W., « Caveats for using statistical significance tests in research assessments », *Journal of Informetrics*, vol. 7, n° 1, 2013, p. 50-62.
- Stephan, P., « Research efficiency: Perverse incentives », *Nature*, vol. 484, n° 7392, 2012, p. 29-31.
- Welp, I. M. *et al.* (dir.), *Incentives and performance: Governance of knowledge-intensive organizations*. Cham, Springer International Publishing, 2015.
- Wouters, P. *et al.*, « The dilemmas of performance indicators of individual researchers—An urgent debate in bibliometrics », *ISSI Newsletter*, vol. 9, n° 3, 2013, p. 48-53.

TABLE DES MATIÈRES

1. Les bases	9
2. Les données	35
3. Les indicateurs	67
4. Vue d'ensemble	145
Suggestions de lecture	163

Autres titres en libre accès aux Presses de l'Université de Montréal

Aménagement du paysage urbain. École d'architecture du paysage, 2003-2013 • Nicole Valois (dir.)

L'économie circulaire. Une transition incontournable • Sébastien Sauvé, Daniel Normandin et Mélanie McDonald

Guy Rocher. Le savant et le politique • Violaine Lemay et Karim Benyekhlef (dir.)

L'idée d'université. Anthologie des débats sur l'enseignement supérieur au Québec de 1770 à 1970 • Claude Corbo avec la collaboration de Marie Ouellon

L'interculturel au Québec. Rencontres historiques et enjeux politiques • Lomomba Emongo et Bob W. White (dir.)

Montréal en paysages • Philippe Poullaouec-Gonidec et Sylvain Paquette

Monuments intellectuels de la Nouvelle-France et du Québec ancien. Aux origines d'une tradition culturelle • Claude Corbo (dir.)

Un nouvel ordre mondial made in China? • Mamoudou Gazibo et Roromme Chantal

Petit guide de survie des étudiants • Marie Lambert-Chan

Pratiques de l'édition numérique • Marcello Vitali-Rosati et Michaël Eberle-Sinatra (dir.)

Le Québec en jeu. Comprendre les grands défis • Gérard Daigle et Guy Rocher (dir.)

Le savoir des livres • Benoît Melançon (dir.)

Sciences, technologies et sociétés de A à Z • Julien Prud'homme, Pierre Doray et Frédéric Bouchard (dir.)

Les scientifiques et la paix. La communauté scientifique internationale au cours des années 20 • Brigitte Schroeder-Gudehus

Les visages de la police. Pratiques et perceptions • Jean-Paul Brodeur

L'ensemble de la communauté scientifique réclame depuis plusieurs années des indicateurs fiables permettant de mesurer les répercussions de la recherche. La ferveur inégalée autour de la mesure de l'influence de la recherche, combinée avec les nouveaux modes de diffusion des connaissances à l'ère numérique, a révolutionné le domaine de la scientométrie. Il s'agit là d'une discipline qui comprend toutes les façons dont nous collectons les documents savants et analysons quantitativement leur production ainsi que leurs usages, des citations aux tweets. Les données et les indicateurs ainsi recueillis sont utilisés pour comprendre la science, stimuler la recherche ou distribuer les ressources.

Curieusement, il n'existe aucun ouvrage qui explique les fondements historiques, les concepts et les sources de la scientométrie, ou qui en fournirait une critique éclairée ou même qui formulerait des recommandations pour un usage optimal. D'où l'importance de celui-ci. À sa façon, chacun est un acteur de la société du savoir et devrait se soucier des outils qui aident à guider son évolution : c'est pourquoi ce livre s'adresse à tous, savants comme profanes.

Vincent Larivière est titulaire de la Chaire de recherche du Canada sur les transformations de la communication savante à l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal. Il est également directeur scientifique de la plateforme Érudit et directeur scientifique adjoint de l'Observatoire des sciences et des technologies.

Cassidy R. Sugimoto est professeure agrégée à l'Université de l'Indiana à Bloomington, où elle étudie les modes de production, diffusion, et usages de connaissances scientifiques. Elle préside, depuis 2015, l'International Society for Scientometrics and Informetrics.

22,95 \$ • 20 €

Illustration: gravure extraite de l'*Introductio geographica* de Petrus Apianus, 1532.

Disponible en version numérique
www.pum.umontreal.ca

ISBN 978-2-7606-3951-5

