

Défense et illustration de la bibliométrie

YVES-FRANÇOIS LE COADIC

Conservatoire national des arts et métiers
yvesfrançois.lecoadic@wanadoo.fr

Professeur honoraire de sciences de l'information au Conservatoire national des arts et métiers de Paris (Cnam), Yves-François Le Coadic est ingénieur diplômé de l'École normale supérieure de chimie de Rennes. Il a aussi obtenu un DEA en physique du solide, et a préparé, au Commissariat à l'énergie atomique, un doctorat d'ingénieur dans le laboratoire du professeur Néel, prix Nobel de physique. À la faveur d'un séjour au Canada, en tant qu'expert de coopération technique au ministère de la Santé, il s'intéresse à la politique de la science et à celle de l'information scientifique. De retour en France, il prépare un doctorat d'État sur « La structure et la dynamique des systèmes d'information scientifique ». Après un séjour de recherche aux États-Unis, il met en place, de 1982 à 1993, au sein du ministère de l'Éducation et du ministère de la Recherche, différents programmes de recherches en sciences de l'information. Il a enseigné au Cnam de 1987 à 2008. Yves-François Le Coadic a publié de nombreux articles et ouvrages, dont les derniers parus sont : Usages et usagers de l'information (coll. « 128 », Armand Colin, 2004), La science de l'information (coll. « Que sais-je ? », PUF, 2004), Éléments de statistique et de mathématique de l'information : infométrie, bibliométrie, médiométrie, scientométrie, muséométrie, webométrie (Presses de l'Enssib, 2002).

La bibliométrie¹ est accusée de tous les maux. On a pu lire ainsi sur le site Sauvons la recherche² un certain nombre d'assertions sur : « les mythes de la bibliométrie », « les comptes truqués du facteur d'impact », « les facéties de la bibliométrie », « des effets délétères de la bibliométrie sur la recherche en... », « contre la bibliométrie, nouvelles du front... », « l'absurdité de la bibliométrie individuelle », « faisons barrage aux usages absurdes de la bibliométrie, aux calculs des indices H et autres imbécillités de ce genre », etc. Et une analyse lapidaire : « Pour le meilleur et pour le pire, les grandes manœuvres sont lancées dans la recherche française sur le front de la bibliométrie : celle-ci cesse désormais d'être un outil réservé aux documentalistes et spécialistes de l'information. Les princes qui nous gouvernent s'en sont emparés à des fins d'évaluation des disciplines, des laboratoires, des individus. » C'est le classement des revues scientifiques qui a déclenché cette tempête, alors que les listes des revues avaient été établies pour chaque champ par des comités d'experts incluant des représentants des instances d'évaluation de la recherche et d'associations ou sociétés savantes, ainsi que des documentalistes³.

Mais la bibliométrie est-elle connue et comprise par ceux et celles qui la stigmatisent ? On peut en douter et mesurer l'ignorance qui semble exister dans les milieux académiques français concernant cette discipline à travers deux exemples. Ignorance ainsi d'une

enseignante de psychologie de l'université de Paris 5 qui, sur le site de Mediapart⁴, le « nouveau quotidien de référence (sic) » parle de biométrie en lieu et place de bibliométrie ! Ignorance patente aussi de l'auteur d'un article sur Paul Otlet, dans le dernier numéro de la revue Cités (revue qui n'a pas été classée dans les listes ci-dessus) au titre provocateur « L'idéologie de l'évaluation. La grande imposture⁵ ». Ignorance qui ne l'empêche guère de porter l'opprobre sur le « pauvre Otlet à qui l'on attribue la paternité de l'internet et de la bibliométrie, mais dont la croyance rationnelle contenait probablement déjà les erreurs et les servitudes volontaires de notre présent ».

Les développements qui suivent sur l'histoire de la bibliométrie, sur le nouveau paradigme métrique en science de l'information, sur les fondements mathématiques et leurs applications, gagneraient à être connus des enseignants et des chercheurs français⁶. On peut, sans trop s'avancer, en douter.

Une histoire de la bibliométrie

Les premières analyses bibliométriques ont vu le jour dans les années 1920. La première loi bibliométrique a été énoncée en 1926 par Alfred Lotka⁷.

4. www.mediapart.fr

5. Cités : Philosophie, politique, histoire, n° 37, 2009, Paris, PUF. Voir la critique par Yves Desrichard dans le BBF, 2009, n° 5, p. 90.

6. Thierry Lafouge, Yves-François Le Coadic, Christine Michel, *Éléments de statistique et de mathématique de l'information : infométrie, bibliométrie, médiométrie, scientométrie, muséométrie, webométrie*, Presses de l'Enssib, Villeurbanne, 2002.

7. Alfred James Lotka, « The frequency distribution of scientific production », *Journal of the Washington Academy of Sciences*, 1926.

1. Littéralement : mesure des livres. Le terme de bibliométrie continue à être utilisé alors que celui d'infométrie gagnerait à l'être.

2. www.sauvonsluniversite.com/spip.php?rubrique7

3. www.aeres-evaluation.fr/index.php/Publications/Methodologie-de-l-evaluation/Documents-generaux-de-referance

Étudiant l'index décennal 1907-1916 des *Chemical Abstracts*, ce dernier constata l'existence d'une relation simple entre le nombre des auteurs et le nombre d'articles qu'ils avaient publiés. Quelques années plus tard, en 1934, Samuel C. Bradford⁸, un bibliothécaire britannique, va élaborer, à l'aide d'un modèle mathématique simple, une méthode d'organisation de la documentation permettant de déterminer comment les articles portant sur un sujet donné sont distribués dans les revues scientifiques et techniques. En 1934, Paul Otlet⁹, dans son *Traité de la documentation*, consacre un chapitre au livre et à la mesure.

Dépassant ces analyses bibliométriques, pour répondre aux nécessités de l'analyse du développement des activités scientifiques, sont apparues, en Union Soviétique, entre les deux guerres, la « science de la science » (*Nauchno y nauchno*) et la scientométrie, mesure des activités de recherche scientifique (à travers les *inputs* – main-d'œuvre, budgets – et les *outputs* – dispositifs, produits, publications, livres). Si les principales mesures des « entrées », des ressources, consistaient à évaluer les dépenses et la main-d'œuvre, les mesures des sorties ne pouvaient être que les produits principaux de ces activités, à savoir les publications.

On passa ensuite, dans les années 1980, de l'article, du livre ou du brevet, à l'information qu'ils contenaient : débutait alors l'infométrie, c'est-à-dire la mesure des activités d'information, en particulier les activités d'information scientifique et technique. L'infométrie a d'abord été une quantification des informations bibliographiques à des fins d'analyse (d'où l'usage continué du terme bibliométrie). La littérature scientifique et technique se prête en effet à une quantification, en ce sens que chaque article de revue, chaque rapport, chaque livre, note, etc., contient des éléments bibliographiques paratextuels (comme les titres

des articles et des brevets, les noms des auteurs, les mots-clés, les références bibliographiques) et des éléments textuels. Ces éléments correctement agencés constituent un ensemble de variables textuelles, personnelles et temporelles, sur lesquelles on peut effectuer des mesures et des analyses quantitatives puis qualitatives des thématiques de recherche et des relations entre les disciplines.

Le paradigme métrique

Aujourd'hui, dans une société marquée par les développements scientifiques et techniques, la mesure semble dicter la moindre décision. Elle renvoie de multiples manières aux conditions de l'existence humaine. Mais on ne parvient à mesurer qu'en s'éloignant des perceptions singulières au profit de procédures réglées. C'est dire s'il est important de comprendre – ce que sont loin de faire les farouches opposants à l'évaluation – les fondements des techniques bibliométriques aujourd'hui en vigueur, les pratiques de mesure et leurs limites dans l'analyse des comportements sociaux.

En science de l'information, que ce soit pour la construction des connaissances, pour la production des informations et la communication de ces informations ou pour leurs usages, de nouvelles armatures intellectuelles ont été mobilisées, de nouveaux principes scientifiques sont apparus, caractéristiques de nouveaux paradigmes scientifiques et technologiques. La mesure est l'un de ces nouveaux paradigmes.

Il existe en effet, dans le domaine de l'information, des régularités, des distributions et des rapports mesurables, universels. Mais ce n'est que très récemment que le corps de ces connaissances mathématiques a été rassemblé et a commencé à avoir des débuts d'applications, des secteurs culturels aux secteurs marchands de l'information, donnant naissance à l'infométrie et à ses dépendances, la bibliométrie, la muséométrie, la médiamétrie, la scientométrie et la webométrie. À ceux ou celles que cette mathématisation pourrait étonner, signalons que l'application de la mathématique à l'analyse des phénomènes

sociaux et humains ne date pas d'hier : Buffon sur l'arithmétique morale, Condorcet sur le problème de l'intérêt général, le dépouillement des scrutins, etc., ont ouvert la voie. Mais c'est l'économie (à la fin du XIX^e siècle), la démographie, la psychologie (1910) et plus tardivement la sociologie (1950) qui ont, les premières, donné lieu à des recherches mathématiques. D'où l'économétrie, la psychométrie, la sociométrie. Cela ne s'est pas fait facilement. Ainsi, en biologie, les travaux quantitatifs ont mis quelque temps à avoir droit de cité. Pour pouvoir publier et donc légitimer des travaux quantitatifs en biologie, au début de ce siècle, Pearson, un statisticien (bien connu par son coefficient), a été conduit à créer une nouvelle revue scientifique, la revue *Biometrika*.

En science de l'information, il est donc possible aujourd'hui de dénombrer, de classer, de distribuer et de mesurer en utilisant des outils et des objets statistiques et mathématiques ; les premiers d'entre eux, mais non les moindres, étant les nombres.

Les fondements statistiques et mathématiques

La statistique, une branche de la mathématique, s'applique à l'analyse des valeurs numériques ; en particulier, celles pour lesquelles une étude exhaustive est impossible, à cause de leur grand nombre et de leur complexité. L'analyse peut être simplement descriptive, donnant par exemple un état des usages faits de l'information ou du système d'information par les usagers. Elle peut être aussi interprétative, permettant de dire ce que signifient ces valeurs. Ce sont alors les statistiques bidimensionnelle (qui décrit et mesure la liaison entre deux variables informationnelles) et multidimensionnelle (qui décrit les relations existant entre trois ou plus de trois variables informationnelles).

Le dimensionnement de ces analyses sera différent selon que l'on a en vue un travail consistant, c'est-à-dire de recherche approfondie, ou une évaluation rapide. La démarche traditionnelle

8. Samuel C. Bradford, « Sources of information on specific subjects », *Engineering*, vol. 26, 1934.

9. Paul Otlet, *Traité de documentation, le livre sur le livre, théorie et pratique*, éditions Mundanéum – Palais mondial, 1934.

statistique, qui consiste à confirmer les hypothèses formulées, a considérablement évolué avec la généralisation d'outils d'analyse statistique multidimensionnelle (encore appelé analyse de données) qui, en particulier grâce aux outils infographiques, permettent de formuler des hypothèses que l'on vérifiera ensuite en utilisant d'autres méthodes, comme les statistiques exploratrices ou « fouilles de données » (*text mining, data mining, web mining*).

Une statistique efficace fournit donc des méthodes descriptives, interprétatives et exploratrices permettant d'évaluer la validité des modélisations des phénomènes informationnels qu'elle propose. Mais ce peut être un moyen de mentir ! Stade suprême de l'impérialisme mathématique, la statistique prétend formaliser la démarche scientifique en proposant des règles pour évaluer la validité d'un modèle. Il est bien entendu que l'on peut développer toutes sortes de modèles statistiques autour des phénomènes sociaux et en particulier des phénomènes informationnels.

Quelles sont alors les statistiques efficaces pour décrire, expliquer et maîtriser les phénomènes informationnels ?

La bibliométrie statistique peut être unidimensionnelle. Elle s'appuie sur des classifications, des nomenclatures préétablies, et consiste à effectuer le décompte du nombre de publications (articles ou brevets) ayant telle ou telle caractéristique, appartenant à telle rubrique de la classification. Elle permet aussi de mesurer la productivité d'un auteur, d'un pays, d'une institution, ainsi que l'impact d'une revue, d'un thème, d'un auteur (en utilisant les comptages des références qu'a reçues un travail donné pendant une période). Elle est également bidimensionnelle ou relationnelle, reposant sur la détection d'une relation entre les éléments d'information et visant par exemple à identifier la structure d'un domaine d'activité scientifique ou technique qu'on représente le plus souvent dans un espace-plan (carte). Elle est enfin multidimensionnelle (analyse de données) et probabiliste (processus poissonniens et markoviens).

L'efficacité des mathématiques

Traditionnellement, pour beaucoup, la mathématique s'applique pour construire des ponts, des machines. Elle s'applique aussi en physique, discipline particulièrement mathématisée, en chimie, en biologie. Et de plus en plus, aux sciences sociales comme l'économie, la psychologie, la sociologie et... la science de l'information. Mais dans l'esprit des professionnels de ce secteur, cela ne va pas forcément de soi. Les succès de la physique classique, puis de la relativité et de la mécanique quantique ont mis en lumière sa pleine fécondité. Mais ce sont les travaux de sociologie mathématique (Raymond Boudon¹⁰, James S. Coleman¹¹) qui nous ont révélé son incroyable efficacité.

Qu'est-ce que cette efficacité ? Elle apparaît au travers de trois capacités : une capacité prédictive, une capacité rétrodictive et une capacité explicative.

Une capacité prédictive

La mathématique est efficace dans la mesure où elle suggère la réalisation d'observations ou d'expérimentations et fournit des résultats numériques qui, à une certaine marge d'erreur près, rejoignent les résultats empiriques issus de ces observations ou de ces expérimentations.

Une capacité rétrodictive

La mathématique est efficace parce qu'elle reproduit des résultats déjà connus en les organisant dans un formalisme concis. La mathématique fournit ici des outils servant seulement à « sauver les phénomènes ». Par exemple, grâce à la méthode des moindres carrés, on recherche des courbes passant au plus près des points expérimentaux.

10. Raymond Boudon, *L'analyse mathématique des faits sociaux*, Paris, Plon, 1970.

11. James S. Coleman, *Introduction to mathematical sociology*, New York, The Free Press, 1964.

Une capacité explicative

Pour qu'une théorie mathématique soit vraiment efficace en science, il faut qu'elle rende manifeste une explication des phénomènes, c'est-à-dire une suite d'inférences reliant leurs descriptions à des principes reconnus comme fondamentaux. Cette capacité explicative va de pair avec une capacité unificatrice (expliquer, c'est ramener la diversité des phénomènes à un très petit nombre de principes) et une capacité générative (suggérer des concepts nouveaux, des stratégies nouvelles).

Une mathématique efficace est donc un formalisme doué de capacités prédictives, rétrodictives et explicatives ; autrement dit un langage permettant de décrire, d'expliquer et de maîtriser les phénomènes. Mais attention, si nous avons l'espoir que cette incroyable efficacité, que nos qualités de logique, de clarté, devraient aider la science de l'information, il peut aussi y avoir une contamination en sens inverse ! Dans la mesure où la culture mathématique est imposée de façon artificielle, de l'extérieur, sans qu'il y ait – comme ce fut le cas en physique – de véritable exigence interne, les mathématiques perdent de leur caractère de sûreté puisqu'elles pourraient en définitive s'appliquer sur n'importe quoi et n'importe comment.

Quelles sont alors les mathématiques efficaces pour décrire, expliquer et maîtriser les phénomènes informationnels ?

- La bibliométrie mathématique peut être *analytique* (analyse numérique, analyse fonctionnelle, équation différentielle). Les comptages informationnels sont légion. Ils fournissent des ensembles de nombres appelés suites dont la somme constitue dans certains cas des séries mathématiques.

- Elle peut être *algébrique* (théorie des ensembles, matrices). Les ensembles (et leur théorie) sont dans le secteur de l'information des objets omniprésents que l'on utilise souvent sans s'en rendre compte tant ils ont marqué les professions du secteur.

- Elle est aussi *géométrique* (espace vectoriel, théorie des graphes). La collection de livres, d'objets muséaux,

“Ce que montre la bibliométrie, dans les usages qu’il en est fait de plus en plus pour l’évaluation des activités scientifiques, c’est que la « cité savante » n’est pas aussi savante, en France tout du moins, qu’elle le fait croire”

le fonds documentaire, les archives audiovisuelles, maintenant les archives électroniques, sont autant de ces ensembles informationnels sur lesquels fonctionnent des dispositifs mettant en œuvre structures géométriques et algébriques comme les espaces vectoriels et les graphes.

Les applications

Ainsi, aujourd’hui, dans le secteur de la culture, de l’éducation, de la recherche, une bonne gestion des services publics nécessite de plus en plus l’utilisation d’une large gamme d’outils de gestion adaptés aux contextes culturels, éducatifs, scientifiques, à la taille et au caractère du service. Ce sont des outils d’analyse des besoins d’information de la communauté desservie, des outils de pilotage et d’évaluation et des outils de mesure des performances, permettant à l’établissement de disposer d’une batterie d’indicateurs de performance.

À l’autre extrême, la démarche marchande particulièrement envahissante sur internet et dans les médias, audiovisuels entre autres, entraîne la réalisation d’analyses statistiques élaborées de la « relation client » (Customer Relationship Management ou CRM) et des audiences grâce aux tableaux de bord de suivi de l’activité

des sites (audience par heure, date, nombre de sessions, de machines, de pages vues, de clics, etc. ; audiences des émissions de télévision, nombre d’entrées dans les cinémas, etc.).

Faute d’unités de mesure cohérentes, d’outils fiables et de méthodes éprouvées, il demeure encore difficile de connaître avec toute la précision souhaitée :

- les chiffres réels de consultation des sites, de suivi des émissions de radio ou de télévision, etc. ;

- les valeurs précises des changements sociodémographiques et socioprofessionnels du bassin de population desservi par l’organisme d’information, etc.

Mais, au moyen de la bibliométrie, nous pouvons explorer plus avant des univers inconnus de l’information : information infiniment croissante, information infiniment rapide, information infiniment complexe. Et comment la saisir pour mieux maîtriser sa production, sa communication et son usage.

Ce que montre la bibliométrie, dans les usages qu’il en est fait de plus en plus pour l’évaluation des activités scientifiques, c’est que la « cité savante » n’est pas aussi savante, en France tout du moins, qu’elle le fait croire. C’est ce qui expliquerait ce refus forcené de l’évaluation, que ce soit des recherches ou des enseignements, par les savants, et le rôle de

bouc émissaire qu’ils font jouer à la bibliométrie. À la petite échelle des usages des bibliothèques, des centres de documentation et des centres de ressources électroniques, nous savons déjà combien est faible l’activité documentaire des savants et combien est grand leur non-usage de ces ressources, d’où la faiblesse de leur production savante¹². ●

Avril 2010

Bibliographie complémentaire

- Christine L. Borgman, *Scholarly Communication and Bibliometrics*, Londres, Sage Publications, 1990.
- Burt R. Boyce, Charles T. Meadow, Donald H. Kraft, *Measurement in Information Science*, Academic Press, San Diego, 1994.
- Nicola De Bellis, *Bibliometrics and Citation Analysis: from the Science Citation Index to Cybermetrics*, New York, Scarecrow Press, 2009.
- Leo Egghe, Ronald Rousseau, *Introduction to Infometrics: Quantitative Methods in Library, Documentation and Information Science*, Amsterdam, Elsevier, 1990.
- Yehuda Elkana, *Towards a Metric of Science*, New York, John Wiley & sons, 1978.
- Henk F. Moed, *Citation Analysis in Research Evaluation*, Dordrecht, Springer, 2005.
- Michael Thelwall, *Introduction to Webometrics: Quantitative Web Research for the Social Sciences*, Morgan & Claypool Publishers, 2009.

12. Yves-François Le Coadic, « Pauvre recherche française ! », *Le Monde*, 13 juin 1997.