

# L'évaluation bibliométrique des chercheurs :

## même pas juste... même pas fausse !

Franck Laloë <sup>(1)</sup> (lalo@lkb.ens.fr) et Rémy Mosseri <sup>(2)</sup>

(1) Laboratoire Kastler Brossel, École normale supérieure, CNRS et UPMC, 24 rue Lhomond, 75005 Paris

(2) Laboratoire de Physique Théorique de la Matière Condensée, UMR CNRS 7600, Université Pierre et Marie Curie, 4 place Jussieu, 75252 Paris Cedex 05

Utiliser les indices bibliométriques pour l'évaluation des chercheurs résulte d'une extrapolation injustifiée entre des cas triviaux où ces indices ont un sens, vers les cas réels où ils sont mis en œuvre. Les vérifications les plus élémentaires n'ont pas été faites pour comprendre l'influence des méthodes de calcul arbitrairement choisies sur les résultats obtenus. Actuellement, leur utilisation relève plus de la pseudo-science, comme l'astrologie, que d'une démarche scientifique.

On dit que Wolfgang Pauli, un des géniaux fondateurs de la mécanique quantique, furieux contre un article de physique sans aucun intérêt, s'était écrié : « Ce n'est pas juste et, pire, ce n'est même pas faux ! ». Il est vrai que, par essence, ce qui n'est ni juste ni faux ne peut être scientifique : une assertion doit être réfutable pour mériter ce qualificatif. La remarque de Pauli s'appliquerait fort bien à de nombreuses applications de la bibliométrie, promues par ceux dont la croyance semble être que, du moment qu'on manipule des chiffres, on raisonne scientifiquement.

L'évaluation bibliométrique des chercheurs n'est effectivement « même pas fausse » : oui, si l'on compare un chercheur reconnu internationalement à un doux farfêlu qui n'a jamais été cité que par lui-même, les indices du premier sont bien supérieurs à ceux du second ; personne ne le conteste. Si le but de l'exercice était de reconnaître des chercheurs exceptionnels des chercheurs très médiocres, nul doute qu'on pourrait recourir à la bibliométrie pour retrouver... ce que chacun sait déjà. Mais supposons que l'on veuille réellement obtenir une information utile, comme par exemple classer les chercheurs au sein d'un groupe homogène, disons les chercheurs d'un bon laboratoire. On constate alors immédiatement de nettes fluctuations de leurs indices (H, G...), et avec une certaine surprise : des valeurs très différentes peuvent être attribuées à des chercheurs dont la qualité de production scientifique est perçue comme très similaire par la communauté scientifique. Pourquoi ?

Plusieurs raisons expliquent pourquoi les méthodes bibliométriques donnent une vue simpliste des contributions scientifiques individuelles. Elles sont sensibles à la qualité, certes, mais ce « signal » est noyé dans le « bruit » créé par une forte dépendance en fonction d'autres variables. Prenons par exemple un indice bibliométrique H, fonction d'une variable X dont nous supposons qu'elle soit assimilable à la qualité du travail scientifique, de la variable Y qui est le style du chercheur (travaille-t-il plutôt seul ou en équipe

constituée, est-ce plutôt un pionnier ou quelqu'un qui préfère des domaines déjà relativement à la mode, proche des applications ou non, etc.), de la variable Z qui est son style de publication (est-il plutôt tourné vers les courtes lettres ou les articles de fond, voire les ouvrages ? est-il attiré par les revues dites de prestige, genre *Nature* ou *Science*, même si elles sont moins utilisées dans son domaine ?), et enfin W (appartient-il à une école de recherche très reconnue depuis des années, ou a-t-il choisi un petit domaine émergent, etc.). Cette liste de variables n'est pas limitative ; on pourrait par exemple y ajouter le sens de la communication du chercheur et son goût pour les congrès, qui ne sont pas toujours liés à sa créativité. N'importe quel scientifique sait que, si l'on recherche la mise en évidence de la variable X, les nombreuses autres variables vont se comporter comme un « bruit statistique ». La seule façon d'avoir accès à la variable recherchée est d'éliminer ce bruit par des moyennes. Si elles sont effectuées sur de très gros échantillons d'individus, statistiquement les variables Y, Z, W prendront un peu toutes les valeurs possibles, et leur influence disparaîtra, laissant apparaître celle de X. C'est ce qui permet à la bibliométrie d'obtenir des chiffres pertinents pour, par exemple, une évaluation comparative de la production nationale dans un grand domaine de recherche. En revanche, utiliser H pour connaître la variable X au niveau individuel est tout simplement une erreur de raisonnement qu'on ne pardonnerait dans aucun laboratoire de recherche sérieux.

De plus, quand un chercheur rédige un article et qu'il y inclut des références, ce n'est pas un acte destiné à la bibliométrie : le but premier des citations n'est pas de dresser une sorte de palmarès, mais de donner au lecteur des informations qui lui sont utiles pour lire l'article en question. C'est donc un processus relatif, fortement contextuel. Par commodité, on peut par exemple citer un article de revue plutôt que les sources originales, pour gagner de la place. Parfois, on cite un article qui permet de raccourcir sa propre rédaction, et on choisira alors le texte juste pour une question >>>

### En savoir plus...

- Y. Gingras (Univ. Montréal), « La fièvre de l'évaluation de la recherche ; du mauvais usage des faux indicateurs », [http://www.cirst.uqam.ca/Portals/0/docs/note\\_rech/2008\\_05.pdf](http://www.cirst.uqam.ca/Portals/0/docs/note_rech/2008_05.pdf)
- D. Jérôme, « Redouter l'évaluation ? », *Reflets de la physique* 10 (2008) 18-20.

>>>

de similarité. On peut même citer un article qu'on considère comme faux dans le but d'en corriger les erreurs ! Comme il s'agit de faciliter la répétition des expériences dans d'autres laboratoires, on privilégiera dans les citations les articles qui décrivent des méthodes ou des appareillages. Pour les idées scientifiques plus abstraites, en revanche, c'est généralement des articles dérivés qu'on cite, pas le grand article original et fondateur. C'est donc un emploi très dérivé des citations, pour ne pas dire un contresens sur leur fonction réelle, que de les prendre comme élément de base pour l'évaluation de la qualité scientifique. Pire encore, par un effet pervers maintes fois signalé, cela risque d'entraîner des changements dans la manière dont les citations se feront à l'avenir, aux dépens de la qualité de la rédaction scientifique et donc de la recherche elle-même.

La base qui est utilisée pour une évaluation bibliométrique en « sciences dures » est celle du *Web of Science* de l'ISI.

Première remarque : les ouvrages scientifiques ne sont pas pris en compte dans le calcul du facteur H. Premier paradoxe, chacun s'accordant à penser qu'une des meilleures façons pour un chercheur de « laisser une trace dans un domaine » est précisément de publier un ouvrage de référence.

Deuxième remarque : les indices G, H, etc., classiquement utilisés pour classer les individus, sont aussi fondamentalement biaisés que le classement de Shanghai<sup>1</sup>. Dans le calcul de ces indices, la contribution d'un auteur est la même s'il est seul signataire ou

s'il a dix coauteurs ! Il pourrait paraître élémentaire de calculer des indices G', H', établis sur la base du nombre de citations divisé par le nombre d'auteurs, ce qui découlerait de la logique la plus élémentaire, mais personne ne le fait sur les données de l'ISI<sup>2</sup>. Le biais est évident : si trois amis décident de tout publier ensemble pendant toute leur carrière, chacun de leurs indices H fera un bond vers le haut.

Troisième faiblesse : tout est centré sur le court terme. Dans beaucoup de domaines, de petites percées techniques entraînent une bouffée de publications, vite oubliées, mais nombreuses. Ainsi les indices sont très orientés vers les sujets à la mode, même s'ils disparaissent vite. Il n'y a aucune raison particulière à cette faiblesse, et on pourrait facilement imaginer d'effectuer des calculs plus adaptés où l'on prenne en compte surtout les publications qui ont une influence à long terme ; mais là encore personne ne le fait<sup>3</sup>.

Quatrième faiblesse, basement technique peut-être, mais très réelle dans certains cas : la base ISI elle-même est inhomogène, ayant fluctué au gré des habitudes de travail des opérateurs de saisie qui l'ont alimentée au cours des décennies. Cela induit toute une série de corrections qui sont nécessaires, qu'il serait trop long de discuter ici, mais qui demandent un travail de spécialiste. Seul un travail précis permettrait de reconstruire des vrais indices, mais bien évidemment dans la pratique personne ne s'en donne la peine : il est tellement plus simple de faire un classement avec des chiffres faux obtenus en trois clics d'ordinateur !

Pour finir, ces indices d'évaluation individuelle :

1. N'ont jamais été testés rationnellement, pour les corrélés avec d'autres évaluations ; ce sont des méthodes qui se présentent elles-mêmes comme des méthodes d'évaluation de la recherche, mais qui paradoxalement ne sont pas passées par l'évaluation scientifique de leur fiabilité et de leur pertinence. Même les vérifications les plus élémentaires comme celles que nous avons signalées plus haut n'ont pas été faites, afin de comprendre l'influence des méthodes de calcul choisies sur les résultats obtenus.
2. Personne ne semble avoir pris le temps d'essayer honnêtement de les améliorer pour obtenir des chiffres qui soient plus pertinents. Porter le débat vers la recherche intelligente de la qualité réelle de la recherche ne semble pas d'actualité.
3. Ne sont « même pas faux », car certes ils contiennent un peu d'information sur les individus, mais la plupart du temps cette information est triviale et déjà connue. Lorsqu'ils sont utilisés sur une population homogène de chercheurs (dossiers au Comité National par exemple), ils déterminent plus un style de travail du chercheur (porté au travail en équipe, ou plutôt éclaireur dans son domaine, etc...), alors que tous ces styles sont nécessaires pour une recherche équilibrée et efficace. Les utiliser pour l'évaluation des chercheurs est donc une faute de raisonnement, une espèce d'extrapolation injustifiée entre des cas triviaux où les indices ont un sens vers les cas réels où ils seront utilisés.
4. Leur succès vient de la fausse facilité qu'ils procurent : avoir un chiffre en quelques clics est évidemment bien commode.
5. La foi dans ces indices est devenue une espèce de croyance, qui échappe au rationnel. La comparaison qui vient à l'esprit est l'astrologie ou la numérologie, qui elles aussi se parent de vertus du scientifique, mais ne se sont jamais passées avec succès sous les fourches caudines d'une véritable évaluation scientifique.

1. Il ne tient pas compte de la taille des établissements.
2. Il ne s'agit pas de dire que cette façon de faire serait bonne, mais juste qu'elle serait moins fautive que la méthode habituelle.
3. Dans le cas du calcul du facteur d'impact des journaux, c'est même exactement le contraire qu'on fait, puisqu'on ne prend en compte les citations que sur deux ans ; pour quelques domaines de recherche techniques ceci peut éventuellement avoir un sens, mais certainement pas pour le cœur des grandes disciplines scientifiques

## Recette

*Si vous êtes un bon chercheur et si vous voulez un meilleur indice H, voici quelques conseils :*

1. Travaillez dans un groupe d'au moins 5 ou 6 personnes, si possible plus, dont toutes les publications sont systématiquement communes ; cela doit permettre au moins de doubler votre indice, voire mieux. De plus, ce regroupement d'efforts permet la mise en commun de « moyens » (matériels et humains, postdocs par exemple) qui peut même parfois augmenter encore votre productivité réelle, ce qui n'est pas à négliger. Inutile de préciser que, plus ces collègues seront brillants, plus vous en profiterez ; choisissez-les donc bien !
2. Favorisez les grands domaines ; on constate une corrélation entre le taux de citations et la taille du domaine scientifique, due au fait que les articles des petits domaines citent beaucoup d'articles plus généraux et pas l'inverse. Évitez les travaux en marge du courant général de votre domaine, même si vous êtes génial : il faudra 10 ans pour que vos travaux soient vraiment reconnus, et alors ce seront les travaux dérivés des vôtres qui seront cités. Bref, ne vous laissez pas trop obnubiler par l'intérêt scientifique de vos recherches : prendre des risques pour tenter une percée scientifique est rarement payant avant des décennies !
3. Ne perdez surtout pas de temps à publier des ouvrages, quel que soit leur impact international ; ce sont des pertes sèches pour les indicateurs.
4. N'accordez pas trop d'importance à la mission fondamentale des chercheurs, la production de connaissance, en particulier quand vous rédigez vos articles ; c'est la communication qui passe d'abord.