

« Pourquoi ne devrions-nous pas retourner les outils de la science sur la science elle-même ? Pourquoi ne pas mesurer et généraliser, faire des hypothèses, et tirer des conclusions ? »
D.J. de Solla Price (1963)¹

La recherche scientifique se structure aujourd'hui par la publication des résultats/découvertes sous forme d'articles dans des revues spécialisées [1]. L'évaluation des laboratoires de recherche et des carrières individuelles repose en grande partie sur le dénombrement de ces publications. Le fait qu'un article soit cité par un autre article publié est, logiquement, une forme de reconnaissance. La mise en place de bases de données répertoriant tous les articles publiés et leurs citations a conduit à l'explosion d'une discipline propre, la bibliométrie qui, en 50 ans, est devenue incontournable dans la pratique de la recherche [2] (voir *Glossaire*). Une des conséquences les plus visibles en est le classement des journaux scientifiques par leur facteur d'impact (IF), un indice mesurant la tendance des articles publiés à être souvent cités à leur tour. La notion même de « qualité » d'un travail scientifique, très difficile à appréhender, est souvent approchée par une métrique simple : le nombre de fois où le travail a été cité. Le nombre de citations des publications d'un chercheur est maintenant un élément de son *curriculum vitae* et est pris en compte dans son évaluation ou ses demandes de financement. La citation peut ainsi être utilisée comme un étalon-or, même si une utilisation trop simpliste des nombres de citations est de plus en plus critiquée. En réalité, chaque citation est une flèche pointant d'un article (dans un journal) vers d'autres articles (potentiellement dans d'autres journaux). L'ensemble de ces connexions définit un réseau, avec des relations de proximité entre journaux ou entre auteurs. L'étude de ce réseau permet ainsi de « cartographier » la science sans spécifier de disciplines *a priori*, en regroupant les journaux et domaines qui sont le plus connectés entre eux. De même, l'étude des coauteurs des articles, et des liens de citations entre eux, per-

¹ De Solla Price D.J. *Little science, big science*. New York : Columbia University Press, 1963.

La face cachée de la science

Ce que les flux de soumission d'articles peuvent nous apprendre

Vincent Calcagno¹, Émilie Demoinet²



¹Institut Sophia Agrobiotech, Inra PACA, 400, route des Chappes, BP 167, 0690 Sophia Antipolis, France ;

²Department of biology, McGill University, Montréal, Canada.

vincent.calcagno@sophia.inra.fr

met de définir des réseaux de collaboration au sein d'un institut ou entre instituts. La modélisation du flux des citations dans le réseau (par des méthodes voisines de celles utilisées par Google pour classer les pages web) permet de calculer l'importance d'un journal de façon plus précise que le simple facteur d'impact [3]. Ainsi, le recueil systématique des données associées aux publications (le « méta-savoir » au sens de Evans et Foster [4]) permet une meilleure compréhension du fonctionnement de la science.

Les citations, partie émergée de l'iceberg

Quel que soit leur niveau de raffinement, toutes ces approches partent de la même donnée de base : les publications finales. Mais une part importante de la recherche se situe en amont : conception d'un projet, réalisation, préparation d'un manuscrit, etc. La publication finale n'est que l'aboutissement d'un long processus. Se focaliser sur les produits finaux peut donner une image partielle de la science (voir par exemple [5]). Le processus de publication lui-même est complexe : il s'agit de choisir un journal cible, lui soumettre un manuscrit, le réviser le cas échéant. Le manuscrit peut être rejeté et il faut tenter un autre journal, parfois plusieurs, avant sa publication. Ces étapes prennent un temps non négligeable dans la vie du chercheur, et choisir un journal pour y soumettre ses résultats est une décision stratégique. Elles suscitent de nombreuses interrogations et frustrations dans la communauté scientifique [6, 7]. Or, contrairement à d'autres formes de méta-savoir, l'histoire des manuscrits avant leur publication est une boîte noire : aucune base de données ne les recense, et les journaux eux-mêmes ne connaissent pas l'histoire des manuscrits qu'ils reçoivent. Seuls les auteurs possèdent cette information.



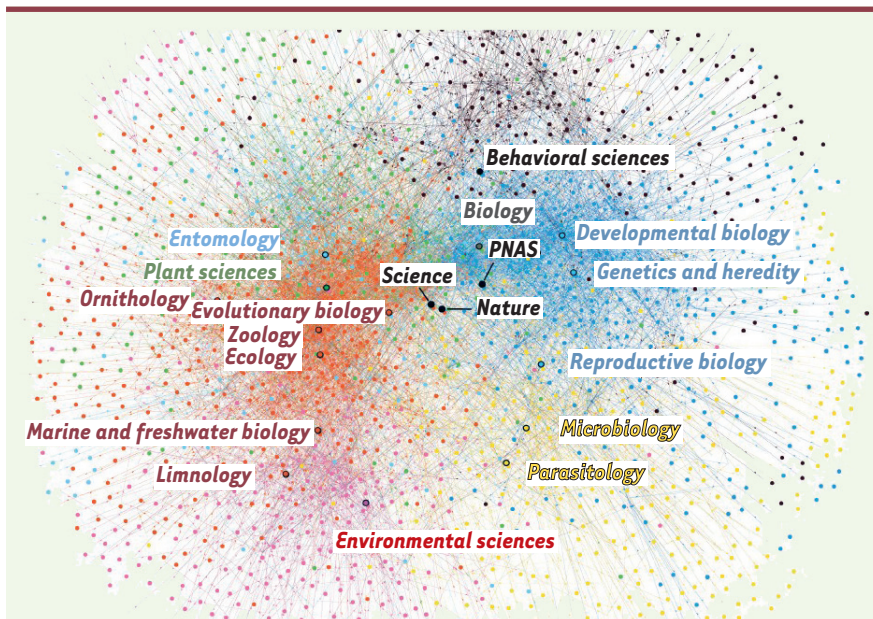


Figure 1. Le réseau des journaux de biologie basé sur les soumissions de manuscrits. Chaque point est un journal. Leur disposition résulte d'un algorithme tendant à éloigner les journaux les uns des autres comme des particules électriques chargées, mais en contraignant les mouvements par les liens entre eux. Chacune des sept couleurs identifie un groupe de journaux qui échangent beaucoup de manuscrits entre eux selon nos données. Les disciplines telles que déterminées par ISI sont aussi positionnées, pour comparaison.

Le premier réseau des (re)soumissions d'articles

Le seul moyen de récolter l'histoire de soumission des articles est donc de mener une enquête auprès des auteurs. Pour cette raison, très peu de données sont disponibles : essentiellement des enquêtes à petite échelle, souvent faites par des journaux eux-mêmes, ou bien des expériences personnelles de chercheurs. En 2008, frustré de voir un article issu de ma thèse de doctorat être rejeté sept fois d'affilée, je (V.C.) me suis intéressé à cette question. Si les bases de données de citations tenues par ISI² (*Institute for scientific information*) ne recensent pas l'histoire de soumission des articles, elles contiennent l'adresse e-mail de l'auteur référent. Avec un peu de programmation, il est donc possible d'écrire un automate qui, pour chaque article, envoie à l'auteur un e-mail personnalisé lui demandant s'il avait tenté une soumission du manuscrit à un autre journal avant cette publication et, si oui, lequel. C'est ce que nous avons fait pour les articles publiés entre 2006 et 2008 dans 16 domaines des sciences biologiques, ainsi que dans *Nature*, *Science* et *Proceedings of the National Academy of Sciences of USA* (PNAS). Ceci représente 923 journaux, et plus de 200 000 e-mails envoyés. À partir des quelques 100 000 réponses obtenues, c'est l'histoire de soumission d'environ 80 000 articles qui a ainsi été déterminée. Chaque réponse constituant un lien depuis un journal A vers un journal B, nous avons pu construire un « réseau de soumissions » connectant tous les journaux entre eux (*Figure 1*) [8].

² Maintenant Thomson Reuters Web of KnowledgeSM.

Un journal échange des manuscrits avec en moyenne 16 autres journaux. Une analyse de modularité de ce réseau révèle sept sous-groupes principaux (voir couleurs dans la *Figure 1*) qui correspondent bien, comme attendu, avec les catégories disciplinaires définies *a priori*.

Le changement de journal pour une resoumission n'est pas la règle

Première surprise : la majorité des articles (environ 75 %) ont été soumis en première intention au journal qui les a publiés. En d'autres termes, les chercheurs avaient bien « visé » dans la majorité des cas. Même en supposant que les auteurs étaient bien moins susceptibles de répondre dans le cas d'articles resoumis, la tendance ne s'inverse pas. D'expérience, nous aurions prédit le contraire : le rejet d'articles semble commun et les revues les plus prestigieuses affichent des taux de rejet très importants. Mais toute expérience est relative : plusieurs auteurs ont répondu, lors de l'enquête, « *Bien sûr que c'était le premier journal tenté : en xx années de recherche, je n'ai jamais eu un papier refusé.* ». En outre, les revues rejettent souvent un manuscrit en suggérant de le resoumettre, et l'article finit par être publié dans la revue. Ceci semblerait même être de plus en plus fréquent, peut être dans le but de réduire le temps affiché en soumission et publication.

Le facteur d'impact influence les soumissions...

Clairement, les journaux à haut facteur d'impact devraient attirer préférentiellement les soumissions d'articles. C'est en effet ce que nous avons mesuré dans nos données. Les journaux à fort impact sont les plus centraux dans le réseau (voir par exemple *Nature* ou *Science* sur la *Figure 1*), et les resoumissions sont fortement biaisées dans le sens descendant, en terme d'IF : en moyenne, le journal tenté en seconde intention a un IF inférieur de 42 %. Les chercheurs descendraient progressivement l'échelle des impacts, les bonds extrêmes étant observés moins fréquemment que ne le prédisait le seul hasard. Tout ceci est cohérent avec l'image d'une lutte pour l'impact parmi les chercheurs.

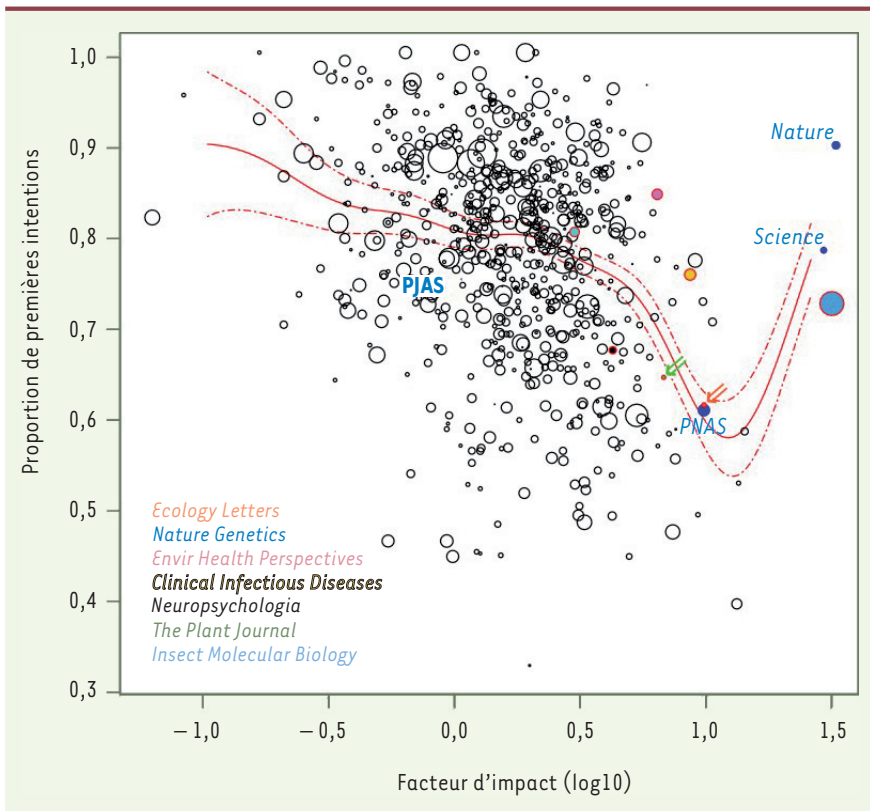


Figure 2. La proportion des articles publiés qui sont des soumissions en première intention au journal (axe vertical) en fonction du facteur d'impact du journal (axe horizontal). La surface des points augmente avec le nombre d'articles par journal. Un journal important de chacun des sept groupes principaux (Figure 1) est indiqué, ainsi que les journaux multidisciplinaires.

Dans ces conditions, les journaux à fort impact devraient publier majoritairement des articles qui leur ont été soumis en première intention, et ceux à faible impact plus de resoumissions, s'ils sont choisis plus souvent en second ou troisième choix. De façon inattendue, nous observons le contraire (Figure 2). Les journaux à faible impact publient presque exclusivement des manuscrits qui leur sont proposés en première intention, et ce taux diminue avec le facteur d'impact, tombant à près de 50 % autour des IF 10. Seuls *Nature* et *Science* font figure d'exceptions, mais sont malgré tout loin de publier uniquement des manuscrits soumis en première intention. Ce résultat suggère que les journaux à plus fort impact, bien que plus attractifs, sont aussi en plus grande compétition pour les articles. Leur proximité pour ces forts taux de rejet rend plus probable qu'ils reçoivent (et publient) un article resoumis. Cet effet de réseau semble suffisant pour inverser la tendance attendue en première approche.

...et l'histoire de soumission influence le nombre de citations

Jusqu'ici, nous avons montré que le facteur d'impact affecte les stratégies de soumission, que ce soit indirectement (car il reflète, au même titre que les soumissions, la valeur perçue des journaux) soit directement (car les chercheurs sont influencés par les valeurs publiées d'IF). Mais la réciproque est-elle vraie ? L'histoire de la soumission d'un article pourrait elle influencer son impact (nombre de citations) après publication ? Pour répondre à cette question, nous avons téléchargé depuis *ISI Web of Science* le nombre de citations attribuées aux articles étudiés en juillet 2011, soit entre trois et cinq

ans après leur publication. Il est alors possible de comparer, *a posteriori*, le nombre de citations – pour un journal et une année de publication donnés – des articles soumis en première intention et de ceux resoumis depuis un autre journal. Détecter un signal dans des nombres de citations est assez difficile, ce type de données étant notoirement variable et difficile à prédire. Néanmoins, à notre grande surprise, il y a une différence systématique entre les deux types d'articles : en moyenne, les articles resoumis depuis un autre journal ont

reçu significativement plus de citations que les premières intentions (Figure 3A). La tendance est particulièrement claire entre 0 et 50 citations reçues, ce qui représente la grande majorité des articles considérés. Les resoumissions étaient moins susceptibles de ne pas être citées du tout (- 30 %) ou citées juste une fois, mais plus susceptibles d'être citées entre 2 et 5 fois, 6 à 10 fois (+ 20 %), ou 11 à 50 fois.

Comment ceci peut-il s'expliquer ? Il y a plusieurs hypothèses. Celle qui vient généralement en premier à l'esprit est que les chercheurs reconnaissent la valeur potentielle de leur travail : les « bons » articles sont soumis à des journaux prestigieux à fort taux de rejet. Lorsqu'ils sont resoumis et publiés dans un autre journal, ils seraient plus cités car ils restent de qualité supérieure. Deux éléments vont néanmoins à l'encontre de cette interprétation. Premièrement, les resoumissions sont autant citées qu'elles se fassent dans le sens ascendant ou descendant en terme d'IF du journal. Deuxièmement, la différence est moins claire, voire inversée, pour les articles très cités (Figure 3A) alors qu'elle est encore plus prononcée lorsque l'on exclut les journaux les plus prestigieux (*Science*, *Nature*, *PNAS* ; Figure 3B). Une autre hypothèse, finalement plus simple, est que les articles resoumis ont fait l'objet d'un plus grand travail de préparation : plus d'avis de *reviewers* et d'éditeurs ont pu être pris en compte, et plus de

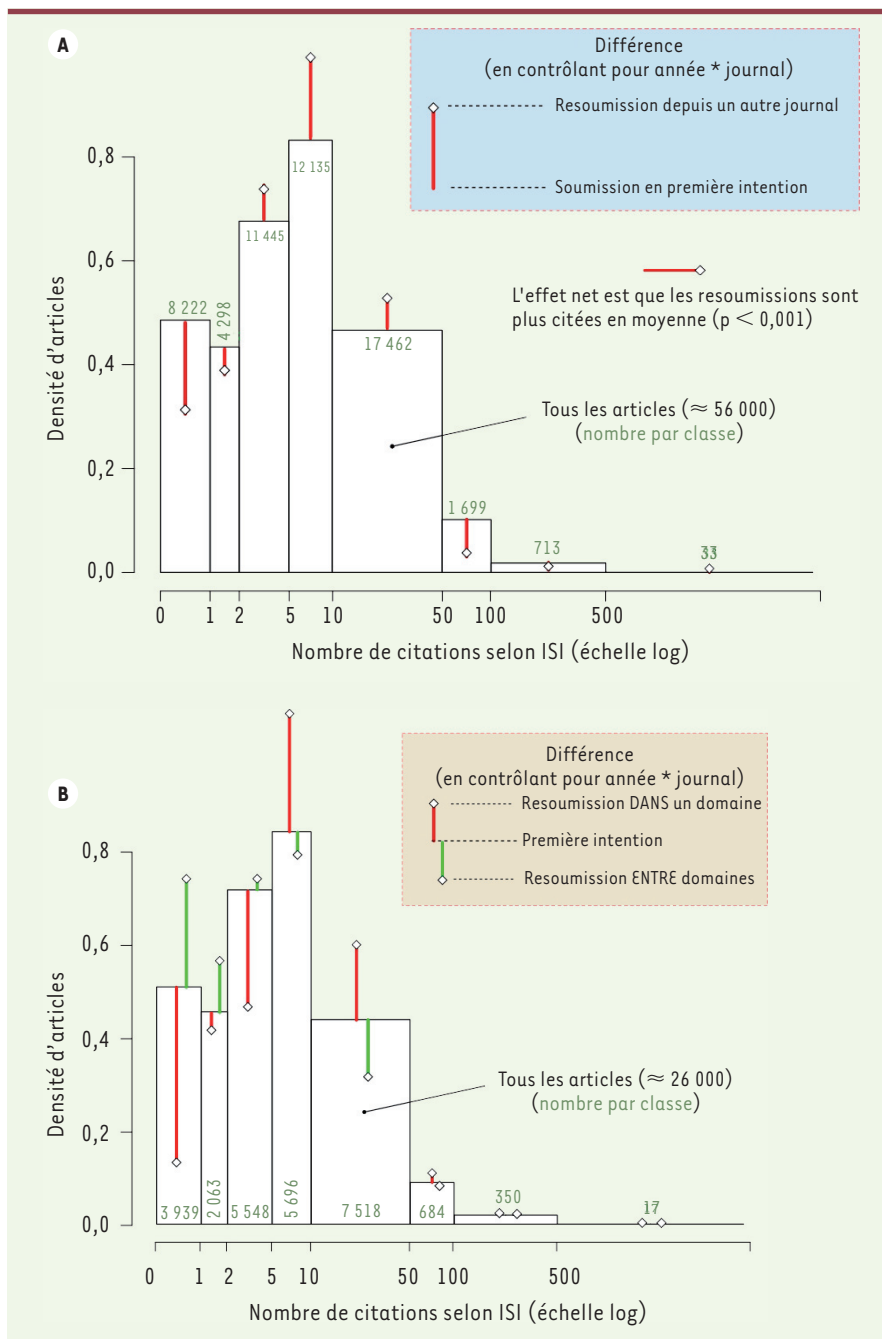


Figure 3. Effet de l'histoire de soumission sur le nombre de citations reçues après publication. **A.** Les resoumissions ont été plus citées que les premières intentions, dans un journal donné et pour une année de publication donnée. **B.** Les resoumissions entre des journaux appartenant à des domaines différents (vert) ont été moins citées que celles qui ont été resoumises entre des journaux du même groupe/domaine (rouge) (voir les couleurs Figure 1), et même que les premières intentions. Dans **B** seuls les journaux clairement assignés à l'un des sept domaines principaux sont inclus, ce qui exclut les journaux multidisciplinaires *Nature*, *Science* et *PNAS* (à la différence de **A**).

temps a été passé à peaufiner la rédaction, par nécessité. Ainsi, si le processus de *peer-review* améliore la qualité d'un article, comme il est supposé le faire, les articles resoumis en bénéficieraient plus, ce qui expliquerait qu'ils soient cités un plus grand nombre de fois.

Dans cette hypothèse, les articles rejetés sans évaluation détaillée ne devraient pas bénéficier de cet effet. Comme ce type de rejet est particulièrement fréquent pour les revues à très fort impact, la différence devrait être particulièrement marquée pour les articles « normaux » en termes de citations. De façon générale, comme nous n'avons pas pu distinguer les resoumissions ayant ou non bénéficié d'expertises détaillées, nos données tendraient à sous-estimer l'impact de l'histoire de la resoumission. Bien sûr, c'est la première fois qu'une telle différence peut être étudiée, et il sera très intéressant de comprendre précisément les mécanismes responsables du plus grand nombre de citations de resoumissions. À tout le moins, ce résultat suggère que tout n'est peut-être pas perdu lorsque un article est rejeté par une revue ! En outre, il légitime la stratégie de plus en plus de grands groupes de publication (par exemple *Wiley* ou *Nature Publishing Group*) qui cherchent à garder en leur sein les articles rejetés, en invitant les auteurs à resoumettre à un autre journal appartenant au groupe, parfois à un journal spécialisé dans ce type de « recyclage » de manuscrits (par exemple les journaux *Open Access* de *Wiley*, comme *Brain & Behavior*).

Mieux vaut rester entre ami(e)s ?

Un autre effet de l'histoire de soumission d'un article sur son devenir en termes de nombres de citations est celui de la proximité (au sein du réseau de soumission) entre le journal auquel il a été préalablement soumis et celui qui le publie (Figures 1 et 3B). En comparant, pour un journal donné, les resoumissions effectuées depuis un journal du même groupe thématique (voire code couleur dans la Figure 1) et celles (plus rares par définition) provenant d'un journal d'un autre groupe, il apparaît que ces dernières sont nettement moins citées en moyenne. À



GLOSSAIRE

Bibliométrie : terme attribué à l'Américain A. Pritchards en 1969, mais la discipline apparaît dès la fin du XIX^e siècle, parfois sous le nom de « statistique bibliographique ». Elle consiste en l'utilisation de méthodes mathématiques et statistiques pour décrire les patrons observés dans la littérature, aussi bien quant aux écrits eux-mêmes qu'à leurs auteurs (les définitions exactes varient fortement). La bibliométrie est considérée comme une part importante de la scientométrie.

Infométrie : terme général apparu en 1979 pour désigner l'étude quantitative de l'information au sens large, ce qui inclut sa genèse, sa diffusion et son usage.

Libramétrie : terme plus ancien (introduit par le documentaliste indien S.R. Ranganathan en 1948), peu répandu en langue française, désignant l'analyse quantitative de différents aspects des activités et des documents de librairie. Le concept est largement chevauchant avec celui de bibliométrie mais souligne leur ancrage initial dans les sciences de la documentation, avant que dans les années 1970-1980 leur usage à des fins d'évaluation et de mesure de la performance ne prenne le dessus.

Métasavoir : selon l'usage réflexif du préfixe méta, « savoir sur le savoir ». Terme utilisé par J.A. Evans et J.G. Foster en 2011 (*metaknowledge*) pour désigner l'ensemble de la connaissance que l'on peut extraire d'une publication scientifique, au-delà de la connaissance « explicite » que la publication entend rapporter.

Scientométrie : attribué au philosophe russe V.V. Valimov en 1969, le terme désigne aujourd'hui largement l'« étude des aspects quantitatifs et des caractéristiques de la science et de la recherche scientifique », selon la description du journal *Scientometrics* (fondé en 1978).

information sur le fonctionnement de la science, sans doute complémentaire de celle qui est obtenue en s'en tenant aux patrons de citations. Il est concevable que les comportements de soumission, plus coûteux en temps et plus contraints pour les auteurs que le simple fait de citer d'autres articles, soient plus proches de la perception par les chercheurs de la valeur des journaux. Il serait aussi intéressant de comparer plusieurs disciplines du point de vue de leurs réseaux de soumission. Quoiqu'il en soit, récolter ce type de « métasavoir » reste très difficile en l'absence de méthode systématique et de traçage des soumissions de manuscrits entre journaux et éditeurs. Peut-on imaginer une modernisation du processus de publication, avec la création d'un identifiant numérique unique pour un manuscrit, qui persiste au cours des resoumissions, et puisse être rendu public une fois l'article publié, gardant la trace de son histoire ? Outre le gain en métasavoir, ceci pourrait aider à améliorer les processus d'évaluation par les journaux, en optimisant le recours à des relecteurs de plus en plus sollicités. ♦

The dark side of science

LIENS D'INTÉRÊT

Les auteurs déclarent n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

1. Guédon JC. Repenser le sens de la communication scientifique : l'accès libre. *Med Sci (Paris)* 2008 ; 24 : 641-8.
2. Garfield E. Commentary: Fifty years of citation indexing. *Int J Epidemiol* 2006 ; 35 : 1127-8.
3. Bergstrom C. Eigenfactor: Measuring the value and prestige of scholarly journals. *Coll Res Libr News* 2007 ; 68 : 314-6. Voir aussi le très beau site associé www.eigenfactor.org
4. Evans JA, Foster JG. Metaknowledge. *Science* 2011 ; 331 : 721-5.
5. Bollen J, Van de Sompel H, Hagberg A, et al. Clickstream data yields high-resolution maps of science. *PLoS ONE* 2009 ; 4 : e4803.
6. Weber EJ, Katz PP, Waeckerle JF, Callahan ML. Author perception of peer review: Impact of review quality and acceptance on satisfaction. *JAMA* 2002 ; 287 : 2790-3.
7. Clapham P. Publish or perish. *Bioscience* 2005 ; 55 : 390-1.
8. Calcagno V, Demoinet E, Gollner K, et al. Flows of research manuscripts among scientific journals reveal hidden submission patterns. *Science* 2012 ; 338 : 1065-9.
9. Larivière V, Gingras Y. On the relationship between interdisciplinarity and scientific impact. *J Am Soc Inform Science Technol* 2010 ; 61 : 126.

tel point que ces resoumissions atypiques sont moins citées que les articles soumis en première intention, renversant l'effet précédent (Figure 3B). Ceci suggère que les frontières entre groupes de journaux qui émergent des pratiques de soumission sont raisonnables, dans la mesure où les franchir entraîne un coût en termes de nombres de citations. Ce résultat est intéressant car il pourrait indiquer que lorsque l'on change de contexte, de communauté, l'article est moins bien accueilli, peut-être en raison de son écriture, des références citées, ou tout simplement du fait que les auteurs sont moins connus. Cela pourrait être aussi lié au caractère interdisciplinaire de ces travaux, car il est très vraisemblable que les articles interdisciplinaires soient plus souvent resoumis de la sorte entre groupes disciplinaires différents [9]. De façon intuitive, mieux vaut ne pas « changer de crèmerie », et ceci pourrait en être une manifestation au niveau des disciplines scientifiques.

Faut-il casser la boîte noire ?

Ces premiers résultats indiquent que l'histoire des soumissions des articles scientifiques, une donnée cachée, peut apporter une autre

TIRÉS À PART

V. Calcagno