

Réseaux de co-citations et Open Access : pour un renouveau des méthodes d'évaluation

Manuel Durand-Barthez

► **To cite this version:**

Manuel Durand-Barthez. Réseaux de co-citations et Open Access : pour un renouveau des méthodes d'évaluation. Chérifa Boukacem-Zeghmouri. L'information scientifique dans l'univers numérique : mesures et usages = Academic Online Resources: Usage and Assessment. , ADBS, pp.55-70, 2011, Sciences et techniques de l'information, 978-2-84365-123-6. <<http://www.adbs.fr/l-information-scientifique-et-technique-dans-l-univers-numerique-mesures-et-usages-academic-online-resources-usage-and-assessment-101163.htm?RH=1216305435750>>. <sic_00589630>

HAL Id: sic_00589630

https://archivesic.ccsd.cnrs.fr/sic_00589630

Submitted on 29 Apr 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Réseaux de co-citations et Open Access : pour un renouveau des méthodes d'évaluation

Manuel Durand-Barthez / S.C.D. Université Paul Sabatier (Toulouse III)
manuel.durand-barthez@enc.sorbonne.fr

Bionotes :

Conservateur à la Bibliothèque nationale (Bibliographie de la France / ISSN) puis au Conservatoire national des arts et métiers (bases de données, mémoires d'ingénieurs), Manuel Durand-Barthez a pris en charge la formation des usagers au service commun de documentation de l'Université Toulouse-3 (sciences, techniques, médecine) avant d'intégrer l'URFIST (Unité régionale de formation à l'information scientifique et technique) de Paris en 2009. Il s'intéresse tout particulièrement à la bibliométrie, à l'*open access* et à la propriété industrielle.

Résumé :

Depuis quelques années, la méthodologie en matière d'évaluation des publications scientifiques fait l'objet d'une réflexion accrue, tant de la part des chercheurs que des professionnels de la documentation appelés à l'appliquer. Le modèle classique, en usage dans la majorité des instances administratives, émanant de l'Institute for Scientific Information de Philadelphie (ISI) révèle un certain nombre de limites notamment dues à l'absence de nuances et à des modes de calcul bruts qui semblent privilégier l'aspect comptable et quantitatif au détriment de la qualité. Les modèles alternatifs proposent l'application fine et nuancée de la méthode mathématique de la « marche aléatoire » (*random walk*) avec l'exécution itérative de l'algorithme de Lawrence Page (utilisé par *Google* pour le tri et l'affichage des pages Web) algorithme pondéré connu sous l'appellation *Weighted PageRank*. Associée à la détection des co-citations, cette méthode permet de visualiser des collèges invisibles, des affinités entre titres de périodiques et surtout la traçabilité des citations. Celle-ci est déterminante dans la définition qualitative de l'évaluation qui semble minorée dans le modèle classique.

Mots-clés :

Bibliométrie ; Évaluation des publications ; Facteurs d'impact ; Indicateurs scientométriques

Le modèle d'évaluation dominant relatif aux publications scientifiques reste à ce jour le *Web of Science*¹ de l'ISI (Institute for Scientific Information, Philadelphie). Il s'appuie sur un système clos, un corpus essentiellement constitué de publications en série (environ 8000 titres) homologuées en fonction de différents critères de qualité tant scientifiques qu'éditoriaux². Le comptage des citations est brut, englobant de façon indifférenciée des

¹ http://isiwebofknowledge.com/products_tools/multidisciplinary/webofscience/

² Périodicité régulière, si possible rapprochée, minimisant les délais de publication, favorisant l'innovation et les axes émergents, comité éditorial de haut niveau et plurinational, standards de présentation de qualité, citations significativement appréciables. Cf. : <http://wokinfo.com/benefits/essays/journalselection/>

publications citantes de tout bord, indépendamment de leur origine disciplinaire et de leur qualité propre, en se limitant à une première « couronne ».

On voit surgir en réaction des algorithmes fondés sur l'exécution itérative du *weighted PageRank* [XING, 2004], i.e. le mode de classement de Lawrence Page appliqué notamment à *Google*, pondéré. C'est le cas de l'*Eigenfactor*³ de Bergstrom [WEST, 2008] appliqué au *Web of Science*, et du *SJR*⁴ de *Scimago* [GONZALEZ-PEREIRA, 2009] appliqué à *Scopus*⁵. Dans cette perspective, c'est le procédé de la marche aléatoire (*random walk*) qui est en usage.

Ce modèle induit notamment l'occurrence des co-citations comme facteur d'appréciation des publications au sein de réseaux ou de collèges invisibles.

On étudiera successivement l'incidence de la marche aléatoire sur l'analyse évaluatrice, puis son interaction avec le principe des co-citations.

La marche aléatoire

Il s'agit d'un modèle mathématique inspiré des analyses de Markov, qui met en évidence la progression constante d'une séquence à la fois dynamique et discrète. Dynamique en ce qu'elle est caractérisée par une progression permanente, discrète en ce que cette séquence est composée d'une multitude de « pas » distincts et concaténés à la fois. Le paradoxe de cette atomisation et de cette concaténation caractérise la séquence d'étapes, tournée naturellement vers le futur, mais sans « mémoire » du passé. La séquence file « au hasard », d'où le fait que ce protocole soit apparenté à la théorie des probabilités. À chaque terme d'étape, l'amorce de la suivante est « hasardeuse » ou « aléatoire », quelle que soit la régularité du réseau ou la complexité du graphe sur lequel ce cheminement s'exécute. Dans ces conditions, l'état du présent est perpétuellement renouvelé, sans dépendance directe vis-à-vis de ce qui précède. Pour l'anecdote, ce modèle mathématique trouve écho en un phénomène observable dans la nature : celui que le physicien R. Brown a constaté vers 1855 dans le mouvement de grains de pollen immergés dans l'eau, dû à l'agitation thermique et dont le tracé semble aléatoire (mouvement brownien).

Ce tracé se retrouve dans un processus que l'on nomme « stochastique », issu d'un phénomène dont le déterminisme n'est pas absolu mais qui peut, par le biais de variables aléatoires, faire l'objet d'une analyse statistique.

Dans le domaine qui nous intéresse, cette marche s'effectue au long des bibliographies terminales d'articles à partir desquelles se génère une chaîne de traçabilité. Cette chaîne est-elle même figée par endroits en grumeaux ou nodosités définissant des collèges plus ou moins visibles, avec une intensité de liens plus ou moins grande [ROSVALL, 2008].

Considérant un article de référence, cette chaîne de « cité par.../cité par.../cité par.../ » se complète par une multiplicité d'autres chaînes du type « cité en même temps que.../ etc.

³ www.eigenfactor.org

⁴ <http://www.scimagojr.com/>

⁵ <http://www.info.scopus.com/scopus-in-detail/facts/> Concurrent Elsevier de l'ISI, revendique environ 18000 titres au lieu des 8000 du WOS avec un nombre de références supérieur d'environ 40%.

etc. », c'est-à-dire les co-citations déjà mises en œuvre par S. Harnad et T. Brody dans *Citebase*⁶, également présentes dans *Citeseer*⁷.

Par ailleurs, le concept de *Shared references* (partagées)⁸ existe depuis longtemps dans le *Science Citation Index*, sans pour autant qu'il soit exploité en tant que tel par l'ISI. Il l'est en revanche de façon systématique, étendu à des bases hétérogènes multiples, par des outils de veille technologique.

Citebase et *Citeseer* sont fondés sur la galaxie OAI. *Scopus* et *Scholar Google*⁹ incluent en partie la littérature de ce type. Le *WOS*, sans l'ignorer totalement, ne l'admet que très parcimonieusement (exemples : quelques revues de *BioMed Central*¹⁰ et *PLoS*¹¹).

Dans le cadre d'une réforme possible des critères d'évaluation actuels, nous envisageons de poser le problème de la façon suivante :

- Tant le *JCR* du *WOS* que l'*Eigenfactor* appliqué par Bergstrom à ce corpus, que le *SJR* appliqué à *Scopus*, - classifient, à partir d'un titre de publication en série, un support qui se traduit en catégories de type A, B, C... dans l'évaluation personnalisée ou locale des unités de recherche (cf. par exemple le formulaire de fiche individuelle AERES) ou en macroanalyse pour un pays donné par rapport à une région (comme l'Europe) ou au monde, ainsi que le fait l'OST (Observatoire des Sciences et des Techniques)¹².
- D'un autre côté, un outil comme *Citebase* exploite le moissonnage OAI pour qualifier des publications (i.e. articles, communications...) indépendamment de leur support. Il suggère des co-citations (onglet spécifique) en relation avec la publication considérée.
- Si l'on veut aller au-delà du simple affichage de ces co-citations, on peut, par le biais des marches aléatoires, cartographier des réseaux. Certes, de façon promotionnelle, l'ISI propose ce type de carte, souvent dans un esprit de compétition où la renommée s'affiche ouvertement : c'est le site *ScienceWatch* avec ses *Fast moving Fronts*¹³ assortis de *Research Front Maps*, révélant les « points chauds » de la recherche internationale, fréquemment de type appliqué. Du fait qu'elle n'a rien de systématique et qu'elle part d'une initiative épisodique et unilatérale de l'ISI, cette cartographie ne peut avoir qu'un caractère anecdotique.

Si maintenant nous proposons de contourner le principe qui attache une publication à son support. Si nous souhaitons éviter que l'orientation d'un texte scientifique soit en partie régie par l'orientation du support auquel on le destine. Si l'on souhaite amoindrir considérablement le poids de lobbies éditoriaux et considérer de façon objective (disons : aussi objective que

⁶ <http://www.citebase.org/>

⁷ <http://citeseerx.ist.psu.edu/>

⁸ Un article est doté d'une bibliographie terminale ; il existe x articles dont la bibliographie terminale cite au moins l'une des références de cet article parent. D'où la notion de références « partagées ».

⁹ <http://scholar.google.fr>

¹⁰ <http://www.biomedcentral.com/>

¹¹ <http://www.plos.org/> Public Library of Science

¹² <http://www.obs-ost.fr/>

¹³ <http://sciencewatch.com/dr/fmf/>

possible, pour rester leibnizien...) la qualité intrinsèque d'un texte : prendre en compte l'appartenance à un réseau « fort » de co-citations peut être un biais appréciable¹⁴.

La transversalité disciplinaire de ces réseaux, pouvant relier SHS et STM au-delà d'une dichotomie (voire trichotomie...) de type *SCI / SSCI / AHCI*, va dans le sens de cette volonté souvent exprimée aujourd'hui de décompartmenter la science, par le biais de l'OAI notamment.

Les travaux de [RODRIGUEZ, 2006] ont mis en évidence la différence entre revues « populaires » et revues « prestigieuses ». Où l'on voit qu'une revue « modeste » issue d'une société savante dont la qualité est difficilement mise en doute, n'a pas forcément un bon Impact Factor au regard de l'ISI. [KLEINBERG, 1999] a mis en évidence deux concepts reliés à ces deux notions : les *hubs* et les *authorities*.

Les *hubs* sont des revues à fort facteur d'impact, quasi systématiquement citées par les autres revues. Il s'agit souvent de journaux compilant des *reviews* et ne rendant pas prioritairement compte d'innovations élaborées en laboratoire, mais servant de « référence » ou d'aiguillage (*hubs*) vers des sources. Exceptionnellement, ce seront des revues pluridisciplinaires d'excellence comme *Nature* ou *Science* qu'il faut effectivement bien séparer des autres. Du fait que cette seconde catégorie est très rare, la distinction est facile.

Les *authorities* sont des revues rendant effectivement compte d'innovations élaborées en laboratoire, auxquelles les auteurs spécialisés et réputés d'un domaine auront fréquemment recours. Ces revues sont « prestigieuses » au sens de [RODRIGUEZ, 2006], par opposition à celles qui sont « populaires », constituant des *hubs* pour la plupart. Le prestige des *authorities* se mesure notamment par le biais du *Weighted PageRank*.

De nombreuses analyses le prouvent, même des auteurs *Highly cited*¹⁵ fréquentent assidûment ce type de revue dites prestigieuses et peuvent par ce biais constituer des segments essentiels de ces réseaux de co-cités. Cet argument plaide en défaveur de la mise en avant du support comme critère fondamental d'évaluation.

Le *Weighted PageRank* est constitutif des deux formules respectivement appliquées par l'*Eigenfactor* de Bergstrom sur 5 ans et le *Scimago Journal Rank* sur 3 ans.

D'où la proposition allant dans le sens d'une meilleure exploitation des algorithmes de marche aléatoire pour dessiner des réseaux de co-cités à partir de l'OAI, à partir des textes eux-mêmes, indépendamment de leur support.

Problématique des auto-citations

La réplique consistant à mettre à l'index les citations de complaisance n'est évidemment pas à exclure. Celle de l'auto-citation aussi.

Ce phénomène d'auto-citation appelle plusieurs observations. *Stricto sensu*, l'auto-citation se définit aux deux niveaux du périodique et de l'auteur.

La référence en fin d'article à un article publié dans le même périodique correspond au premier niveau.

¹⁴ *Scimago* propose des cartes de co-citations, mais elles ne se mettent en évidence que des domaines et pas des équipes.

¹⁵ <http://isihighlycited.com/> : il s'agit d'une sorte de *Who's who* ou *Gotha* de la recherche internationale élaboré par l'ISI. L'inclusion du nom d'un chercheur compte pour 20 % dans les critères d'évaluation d'un organisme au classement de Shanghai.

La référence en fin d'article à un article quelconque publié par l'un des co-auteurs de cet article correspond au second niveau.

Le WOS s'efforce d'éliminer, dans le décompte des citations, les articles cités dont seul le premier auteur fait partie des co-auteurs de l'article citant. Cela altère évidemment quelque peu le résultat final.

Scholar Google, quant à lui, ne permet pas d'identifier clairement les auto-citations. La difficulté que suscite ce moteur à distinguer les homonymies ajoute à cette imprécision.

Par ailleurs, sans exagérer le trait, on peut imaginer que des stratégies soient mises en œuvre dans le cadre d'une communauté d'organismes pour mettre au point une politique de citations mutuelles.

Du côté des comités de rédaction des périodiques, il peut également arriver que l'on favorise la publication d'un article dans la mesure où celui-ci fait préférentiellement référence à des travaux parus dans la même revue. Des cas de figure extrêmes, pas forcément si rares, peuvent concerner des articles dont l'objet n'a qu'un rapport plutôt éloigné avec celui de l'article citant. Péter Jacsó, expert à l'agence d'information américaine Gale, relate ce phénomène [JACSO, 2005] qui, au-delà de l'anecdote, mérite peut-être que l'on s'y penche de plus près.

Il existe également ce que l'on pourrait appeler des « auto-citations d'excellence », lorsqu'une équipe est contrainte de se citer elle-même du fait de sa capacité à se détacher nettement au niveau international sur l'axe de recherche considéré. Dans le cas de certains auteurs faisant partie du répertoire *Highly Cited* de l'ISI, la proportion d'auto-citations avoisine le quart, voire plus. Rappelons que le *Highly Cited* est une base concentrant les auteurs « les plus cités » du *Web of Science* avec une répartition en 21 disciplines. Cette base est assortie d'une biographie des auteurs et d'un lien vers la bibliographie analytique du *Web of Science*.

Le *Journal Citation Reports*, dans les fiches descriptives de chaque périodique, mentionne son taux d'auto-citation. Il est clair qu'en fonction du titre de la revue, de la discipline et de ses usages, ce taux n'aura pas la même signification et qu'il serait absurde de partir du principe brut qu'une auto-citation est en général à proscrire absolument.

L'*Eigenfactor* de Bergstrom, s'il comporte l'avantage notable de reposer sur le principe de la marche aléatoire, observe une règle quelque peu sujette à caution en matière d'auto-citation. En effet Bergstrom utilise un algorithme de type PageRank attribuant un certain poids à chaque citation, générant le calcul d'un type de facteur d'impact qui est fonction de moyennes pondérées de citations. En soi, ce principe est plutôt appréciable. Mais par ailleurs, ces modalités de calcul sont complexes et peuvent altérer une appréhension rapide et claire des résultats. Ils excluent les auto-citations, définies dans cette application comme les références à des articles parus dans le même journal que celui où figure le papier analysé dans les cinq ans qui précèdent. C'est assurément là un abus qui peut porter atteinte à l'évaluation des travaux dans certains domaines. [ADLER, 2008] note que la proportion d'auto-citations légitimes qui disparaissent de ce fait peut être importante. Si l'on applique ces modalités à une base telle que la Math Reviews Database, environ un tiers des citations devient invisible.

Certes, une redondance peut être générée par les phénomènes d'auto-citation. De cette redondance liée à l'exécution itérative du PageRank sont issus des cycles avec des « retours sur soi » quasiment en boucle. Mais, comme le fait remarquer [SIDIROPOULOS, 2005, p. 55], ces cycles peuvent contribuer à la mise en évidence de noyaux définissant un axe de recherche et des équipes le constituant. Si ces « auto-citations légitimes » sont mises à part, les graphes résultants peuvent être faussés. C'est pourquoi, le concepteur de la base SCEAS (Scientific Collection Evaluator with Advanced Scoring)¹⁶ juge qu'en dépit de ses avantages,

¹⁶ <http://sceas.csd.auth.gr/php/index.php4> . Mis au point par Antonis Sidiropoulos de l'université de Thessalonique, analyse les citations à partir de la base DBLP (Digital

l'utilisation du PageRank n'est pas forcément la panacée.

Les incidences du facteur H

Qu'en est-il maintenant des indices individuels du type facteur H¹⁷ ? Dans la mesure où les instances administratives l'associent à un taux de publications plus ou moins citées en fonction notamment du facteur d'impact de leur support, conçu au sens de l'ISI, il n'est pas adapté à la configuration que nous souhaitons mettre ici en évidence. Il serait de loin beaucoup plus intéressant de s'en référer aux multiples nuances exprimées par Harzing dans son outil *Pop (Publish or Perish)*¹⁸.

Par exemple :

- le *Contemporary H-Index* qui se réfère à l'ancienneté de l'article vs. le nombre de citations. Il est clair qu'un article de 2007 recueillant en 2009 un total cumulé de 32 citations sera peut-être plus estimable qu'un autre de dix ans d'âge cité autant de fois. A cet égard, *Citebase* propose des graphiques dont une exploitation poussée permettrait de faire le rapport entre ces données: le temps s'écoulant entre la mise sur le Web et les premières citations (faisant bien sûr elles-mêmes suite aux déchargements), la proportion de citations mesurée par rapport à l'évolution dans le temps de l'exploitation du texte.
- *L'indice individuel* issu de la division de l'indice de Hirsch par le nombre moyen d'auteurs ayant contribué à cet indice afin de réduire les effets induits par le groupe d'auteurs.
- *L'Age-weighted citation rate (AWCR)* qui prend en compte la totalité du corpus d'une base de référence et non pas seulement les items qui sont au-dessus de la barre du H. Cela autorise en particulier des articles plus récents et donc moins cités de contribuer à cet AWCR, même s'ils ne participent pas de l'indice de Hirsch.

On distingue dans ce qui précède deux éléments positifs :

- la prise en compte de l'âge de la publication
- la pondération exercée en fonction du nombre d'auteurs

Restent deux zones d'ombre :

- le taux de citations d'une part et le nombre moyen annuel de publications d'autre part varient amplement suivant les disciplines
- la participation effective de chaque auteur reste méconnue, sauf dans le cadre spécifique de certaines revues du type *Nature* dont le comité de rédaction exige parfois une indication sur le taux effectif de contribution des auteurs

On peut répondre à la première objection que l'un des principes fondamentaux de l'*Eigenfactor* (facteur propre appliqué à la bibliométrie) est la prise en compte des variations d'usage de la citation en fonction des disciplines.

Bibliography & Library Project), base bibliographique internationale en accès libre dévolue à l'informatique, implantée à l'université de Trêves (Trier) <http://www.informatik.uni-trier.de/~ley/db/index.html>

¹⁷ Au-delà des individus, le H peut qualifier une revue, comme le fait *Scimago*

¹⁸ <http://www.harzing.com/pop.htm>

- Si l'on applique la transversalité des disciplines à un réseau de co-citations, ce nouvel H pourra être pondéré.
- Si le réseau se limite dans une très forte proportion à une discipline où l'usage de la citation est soit très large, soit très faible, le H devra être ramené à une indication pondérée permettant des comparaisons dans un environnement équilibré.

Il est très difficile d'annihiler la seconde objection. On ne peut pas dire que la proportion de la participation effective de tel ou tel membre d'une équipe peut être ramenée, dans un réseau de co-citations, à la fréquence de sa présence dans le maillage du réseau. Dans certains cas extrêmes qui restent anecdotiques, un auteur peut être très fréquemment cité, au sein d'une équipe dont les autres membres varient, sans qu'il soit pour autant scientifiquement « méritant »¹⁹. La fréquence de la présence d'un auteur existe à la base dans le système d'évaluation standard de l'ISI, sans que la notion de co-citation intervienne.

Ce point relève semble-t-il plutôt de la déontologie : dans certains secteurs de la recherche, médicale par exemple, la déclinaison des auteurs se fait dans un ordre qui permet le décryptage de la proportion des contributions individuelles²⁰.

Le facteur H appliqué à l'individu a fait l'objet de nombreuses controverses [GINGRAS, 2008, p. 10] [BORNMANN, 2009]. On a vu qu'il s'appliquait aussi aux revues en tant que telles (ex. : *Scimago*).

Or il a été introduit dans le WOS pour pallier le caractère limité de l'I.F. (Impact Factor) à des titres de publications en série, au détriment précisément des individus, i.e. des auteurs.

En revanche, dans la perspective qui nous intéresse, autrement dit les réseaux de co-citations à partir de l'OAI (Open Access Initiative), nous avons affaire d'une part à des collectivités (des ensembles de personnes agrégées par leur thématique de recherche, même transversale) et d'autre part à des agrégats de supports de publication inclus dans des systèmes ouverts (Open Access) et non des corpus clos (JCR).

Parallèlement, les instances administratives exigent la production d'un indice adapté à la description de l'activité de chaque chercheur. En l'occurrence le H.

La question se pose donc de savoir comment adapter un H, ou assimilé, à un « nœud » pris en tant qu'agrégat de personnes et de publications, et comment ou dans quelle proportion ce nouvel H pourra être réparti à l'échelle de chaque personne qui le compose.

Le Facteur Y et son contexte

Nous avons évoqué les nuances émises par Sidiropoulos au sujet du *Weighted PageRank* comme solution prétendument idéale. L'équipe de [RODRIGUEZ, 2006, p. 5] a conçu un moyen terme entre l'*Impact Factor* classique et le *Weighted PageRank*, compromis supposé relativement équilibré.

Il s'agit d'un facteur désigné par la lettre y.

¹⁹ Ce peut être le cas, certes exceptionnel, du technicien de laboratoire mentionné par courtoisie comme co-auteur chaque fois que l'équipe dont il est membre publie...

²⁰ Ainsi arrive-t-il souvent, dans le domaine médical en France, que le dernier nom d'auteur mentionné soit celui du chef de service. Néanmoins, cela ne signifie pas qu'il ait pris une part majoritaire dans l'élaboration de la publication...

Le facteur γ se définit comme le produit de l'I.F.-ISI par le *Weighted PageRank*.

À titre d'exemple, si l'on considère avec les auteurs la totalité du corpus analysé par le JCR 2003, le plus fort I.F. est celui de l'*Annual Review of Immunology* alors que *Nature* emporte la première place à la fois pour le *Weighted PageRank* et pour le facteur γ . *Nature* est classé en 9^{ème} position pour l'I.F.²¹

Les exemples sont plus significatifs dans le cadre disciplinaire. En Physique, dans la liste de référence des dix premiers I.F. du JCR 2003, 6 revues sont absentes des dix premières de la liste « facteur γ ».

Cette disparité paraît encore plus flagrante dans l'analyse des rangs associés respectivement :

- au Facteur d'impact classique du JCR, sur 2 ans (version 2008, 6620 titres)
- à la variante du précédent sur 5 ans
- à l'*Eigenfactor* de Bergstrom calculé sur 5 ans à partir du JCR (version 2007, 7937 titres)
- au *Scimago Journal Rank* appliqué sur 3 ans par le consortium universitaire espagnol SJR²² à la liste Scopus d'Elsevier (version 2008, 17124 titres)
- au facteur H des journaux (suivant un principe identique à celui qui s'applique aux auteurs : x articles du périodique cité au moins x fois) calculé par *Scimago*

Les trois tableaux qui suivent révèlent différents aspects de cette disparité.

Le Facteur d'impact du JCR est calculé suivant la formule classique bien connue de l'ISI ; L'*Eigenfactor* et le *SJR* sont calculés suivant des formules différentes et spécifiques [WEST, 2008] [GONZALEZ-PEREIRA, 2009]

Les durées d'analyse sont différentes (2,3,5 ans) et l'ampleur des corpus varie, ainsi que leurs champs disciplinaires.

Néanmoins, seules les STM (Sciences, Technologie, Médecine) sont visibles dans cette tranche supérieure.

Les noyaux (plus grands communs dénominateurs parmi les titres) sont quasiment identiques. Ce qui ressort de ces tableaux, ce sont des disparités de classement qui constituent en soi une information intéressante, *indépendamment de la disparité des modalités de calcul*.

Le tableau 1 montre le *Top Ten* des revues dans l'ordre décroissant de facteur d'impact JCR 2008 : « normal » sur 2 ans, sur 5 ans, leur *Eigenfactor* sur 5 ans et leur *Scimago Journal Rank (SJR)* sur 3 ans avec leur *facteur H* corrélé au *SJR*.

RG	TITRE ABREGE	ISSN	I.F. normal	I.F. 5 ans	Eigenf. 5 a.	SJR 3 a.	Fact. H sjr
1	CA - CANCER J CLIN	0007-9235	74,575	50,766	0,0365	11,331	69
2	NEW ENGL J MED	0028-4793	50,017	49,911	0,6803	3,349	497
3	ANNU REV IMMUNOL	0732-0582	41,059	46,200	0,0738	16,204	179
4	NAT REV MOL CELL BIO	1471-0072	35,423	34,221	0,1784	7,671	178
5	PHYSIOL REV	0031-9333	35,000	35,855	0,0561	7,090	183
6	REV MOD PHYS	0034-6861	33,985	40,395	0,0893	4,760	141

²¹ Pour information, l'*Annual Review of Immunology* est 3^{ème} et *Nature* 8^{ème} dans le JCR 2008. Cf. les tableaux dans le paragraphe suivant.

²² Consejo Superior de Investigaciones Científicas [CSIC], Universités de Grenade, d'Estrémadure, Carlos III de Madrid et d'Alcalá de Henares.

7	JAMA - J AM MED ASSOC	0098-7484	31,718	27,957	0,3810	2,103	351
8	NATURE	0028-0836	31,434	31,210	1,7635	6,727	599
9	CELL	0092-8674	31,253	30,149	0,6717	12,802	431
10	NAT REV CANCER	1474-175X	30,762	35,007	0,1353	4,405	162

Tableau 1

Le tableau 2 garde l'énoncé décroissant des revues au sens de l'I.F. JCR 2008, mais resitue leur rang en fonction des autres facteurs d'impact, ainsi que le rang du H. Les disparités sont criantes sur les 1^{er} et 10^{ème} rangs.

RG	TITRE ABREGE	ISSN	I.F. normal	Rang I.F. 5 ans	Rang Eigenf. 5 a.	Rang SJR 3 a.	Rang Fact. H sjr
1	CA - CANCER J CLIN	0007-9235	74,575	1	10	3	10
2	NEW ENGL J MED	0028-4793	50,017	2	2	9	2
3	ANNU REV IMMUNOL	0732-0582	41,059	3	8	1	6
4	NAT REV MOL CELL BIO	1471-0072	35,423	7	5	4	7
5	PHYSIOL REV	0031-9333	35,000	5	9	5	5
6	REV MOD PHYS	0034-6861	33,985	4	7	7	9
7	JAMA - J AM MED ASSOC	0098-7484	31,718	10	4	10	4
8	NATURE	0028-0836	31,434	8	1	6	1
9	CELL	0092-8674	31,253	9	3	2	3
10	NAT REV CANCER	1474-175X	30,762	6	6	8	8

Tableau 2

Le tableau 3 reprend les données précédentes en mentionnant le rang des revues au sein des 3 corpus : JCR, idem revu par *Eigenfactor* et le corpus de *Scopus* revu par *Scimago*, avec le nombre total de revues de ces trois réservoirs. Bien qu'il faille interpréter ces résultats avec discernement, cela met néanmoins en évidence les points forts du *Weighted PageRank*. La 789^{ème} position du numéro 1 du *JCR* dans le facteur H de *Scimago* peut s'expliquer du fait que de très nombreux articles différents sont cités régulièrement, au fil de l'eau, par les autres revues, avec un nombre modéré de citations, mais qu'à l'inverse, cette revue n'est pas dotée de nombreux « articles phares » cités de très nombreuses fois sur le long terme.

RG	TITRE ABREGE	ISSN	I.F. normal	Rang I.F. 5 ans s/6620 titres	Rang Eigenf. 5 a. s/7937 titres	Rang SJR 3 a. s/17124 titres	Rang Fact. H sjr s/17124 titres
1	CA - CANCER J CLIN	0007-9235	74,575	1	13	4	789
2	NEW ENGL J MED	0028-	50,017	2	6	44	3

		4793					
3	ANNU REV IMMUNOL	0732-0582	41,059	3	1	1	56
4	NAT REV MOL CELL BIO	1471-0072	35,423	8	4	11	58
5	PHYSIOL REV	0031-9333	35,000	5	15	12	51
6	REV MOD PHYS	0034-6861	33,985	4	2	28	122
7	JAMA - J AM MED ASSOC	0098-7484	31,718	16	27	84	7
8	NATURE	0028-0836	31,434	10	7	13	1
9	CELL	0092-8674	31,253	13	5	2	4
10	NAT REV CANCER	1474-175X	30,762	6	11	29	77

Tableau 3

Conclusion

Le concept de réseau de co-citations n'est certes pas dépourvu de failles. Toutes choses égales par ailleurs, il semble qu'il soit franchement meilleur qu'une évaluation brute, globale, en « première couronne », liée au support, telle qu'on la conçoit communément aujourd'hui.

Ce concept de réseau de co-citations repose, dans la perspective que l'on envisage ici, sur l'Open Access et promeut la transversalité disciplinaire.

Afin qu'il puisse évoluer et s'imposer plus largement, il est indispensable que le caractère trop souvent expérimental (que l'on ne peut, à ce stade, leur reprocher) de ces investigations puisse se traduire dans la mise au point de systèmes plus élaborés et surtout assumés : par la communauté scientifique d'abord et, dès que possible ensuite par les instances administratives.

Le réseau de co-citations est d'abord une méthode de prospective, une composante principale d'un système de veille technologique, utile par exemple à un service de R & D pour positionner un laboratoire, orienter des investissements, tant humains que financiers, évaluer la validité du concept qui se trouve au cœur d'un axe de recherche.

Mais peut-on assigner à cette notion de réseau un autre rôle que celui-ci ? Un rôle étroitement associé à l'évaluation, non plus seulement des orientations ou des concepts, mais à celle des entités et de leurs membres ?

Peut-on imaginer qu'au lieu d'être évalué dans la circonscription d'une unité ou d'une institution, on le soit à l'intérieur d'un nœud de co-citations, qui aurait un caractère communautaire, transversal et international ?

Peut-on imaginer une révolution copernicienne à la fois de l'I.F. et du H, s'inscrivant dans les réseaux de co-citations sous-tendus par l'Open Access ?

Bibliographie

ADLER R., EWING, J., TAYLOR, P., Joint Committee on Quantitative Assessment of Research (2008) *Citation Statistics : A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics.* <<http://arxiv.org/pdf/0910.3529>>

BORNMANN, L., MUTZ, R., DANIEL, H.D. (2009). Do we need the h Index and its variants in addition to standard bibliometric measures? *Journal of the American Society for Information Science and Technology*, vol. 60, n° 6, p. 1286-1289. <<http://www.lutz-bornmann.de/icons/BornmannRevised3.pdf>>

DURAND-BARTHEZ, M. (2009). L'évaluation des publications scientifiques : nouvelles approches, nouveaux enjeux. *Les Cahiers du Numérique*, vol. 5, n° 2, , p. 123-141.

GINGRAS, Y. (2008). *La fièvre de l'évaluation de la recherche. Du mauvais usage de faux indicateurs : note de recherche.* Montréal : Centre interuniversitaire de recherche sur la science et la technologie. <http://www.cirst.uqam.ca/Portals/0/docs/note_rech/2008_05.pdf>

GONZALEZ-PEREIRA, B., GUERRERO-BOTE, V., MOYA-ANEGON, F. (2009). The SJR indicator: A new indicator of journals' scientific prestige. <<http://arxiv.org/pdf/0912.4141>>

JACSO, P. (2005). Journal Citation Reports *Peter's Digital Reference Shelf.* <<http://www.jacso.info/gale/jcr/jcr.htm>>

KLEINBERG, J.M. (1999). Hubs, authorities, and communities. *ACM Computing Surveys (CSUR)*, vol. 31 , n° 4es, p. 1-3. <http://www.csee.umbc.edu/~kolari1/Mining/papers/ft_gateway.cfm.pdf>

RODRIGUEZ, A., BOLLEN, J., VAN DE SOMPEL, H. (2006). Journal Status. *Scientometrics*, vol. 69, n° 3, p. 669-687. <<http://arxiv.org/abs/cs/0601030>>

ROSVALL, M., BERGSTROM, C. (2008) Maps of random walks on complex networks reveal community structure. *PNAS*, vol. 105, n° 31, p. 1118-1123. <<http://arxiv.org/pdf/0707.0609v3>>

SIDIROPOULOS, A., MANOLOPOULOS, Y., (2005). A Citation-Based System to Assist Prize Awarding. *SIGMOD Record*, Vol. 34, n°. 4, p. 54-60 <<http://www.sigmod.org/sigmod/record/issues/0512/p54-article-sidiropoulos.pdf>>

WEST, J., ALTHOUSE, B., ROSVALL, M., BERGSTROM, C., (2008). Eigenfactor TM Score and Article Influence TM Score: Detailed methods. <<http://www.eigenfactor.org/methods.pdf>>

XING, W., GHORBANI, A. (2004). Weighted PageRank Algorithm. *Proceedings of the Second Annual Conference on Communication Networks and Services Research (CNSR'04).* <http://ias.cs.unb.ca/papers/xingw_weighted.pdf>