

Les nouvelles formes d'évaluation scientifique : quelles évolutions en sciences, technique et médecine ?

Manuel Durand-Barthez,
Université Paul Sabatier (Toulouse 3)

Exiger d'un chercheur qu'il soit cité dans des revues à fort facteur d'impact est un acte institutionnel courant, supposé garantir d'une part un avancement individuel plus rapide au chercheur lui-même, et d'autre part la consolidation des arguments de demandes de subventions pour son équipe.

Les instances d'évaluation auraient tendance à privilégier la méthode bibliométrique fondée sur le recensement statistique rapide des citations. L'aisance et la célérité avec laquelle ce procédé s'exécute le font sans doute préférer à l'examen approfondi des contenus, pénalisant sur la durée et nécessitant la constitution de comités de lecture ad hoc. S'il est vrai que rapide ne signifie pas hâtif ou expéditif (le premier qualificatif recouvre objectivement la notion de durée ; les deux autres peuvent être interprétés comme des procès d'intention) il n'en reste pas moins que cette méthode revêt en apparence un caractère plutôt quantitatif que qualitatif à proprement parler.

Dès 1960, l'Institute for Scientific Information (I.S.I.) de Philadelphie, sous l'impulsion d'Eugene Garfield, a mis en œuvre le *Science Citation Index* (SCI) pour l'évaluation des auteurs, suivi en 1975 par le *Journal Citation Reports* (JCR) pour celle des revues¹.

On connaît la très forte influence exercée par l'ISI sur la communauté scientifique internationale. Le modèle statistique qu'il a mis en œuvre conditionne l'avancement des chercheurs dans un nombre croissant de pays. Les administrations qui en ont la tutelle utilisent fréquemment la combinaison Citations des articles + Facteur d'Impact des revues dans lesquelles ils paraissent (*Impact Factor*) pour « noter » ces chercheurs.

Si nous sommes amenés à développer ci-après la déclinaison de quelques critiques à l'encontre de ce modèle, c'est pour ensuite explorer la voie de nouvelles approches :

- l'algorithme mis au point sur le site *Citebase*²
- *Scholar Google*, avatar du moteur généraliste standard lancé sur le Net en novembre 2004
- deux alternatives récentes à la définition d'un facteur d'impact basée sur l'évaluation sur les Auteurs d'une part et les Sources de l'autre:
 - o le facteur *h* : proposé par J.E. Hirsch [Hirsch, 2005], ce facteur est lié à la production individuelle d'un chercheur
 - o le *Journal Status* : proposé par l'équipe de J. Bollen [Bollen et al., 2006] , cet indicateur est basé sur l'évaluation des titres
- nous mettrons enfin l'accent sur le modèle d'une évaluation effectuée par un « collectif », adopté par la *Faculty of 1000* dans les domaines biomédicaux. Son originalité par rapport aux précédents réside dans le primat de la qualité du contenu sur le principe statistique de la citation.

Mais d'abord, tentons de synthétiser une approche critique du modèle « historique », le modèle de l'ISI.

¹En 2004, le *SCI Sciences exactes et appliquées version On line* comprend environ 6000 titres (qui se reportent sur le corpus du JCR), tandis que le *SSCI* relatif aux Sciences humaines en compte approximativement 1700.

²Outil développé par l'équipe de Stevan Harnad et Timothy D. Brody à l'université de Southampton (Royaume-Uni) sur le miroir britannique du site d'ArXiv.org.

1. Le modèle de l'ISI.

Trois types d'observations critiques sembleraient émerger : le premier repose sur le mode de calcul, le second sur des critères purement éditoriaux liés au support de l'information, le troisième sur des problèmes de relations humaines.

1) Le mode de calcul³ :

- a) de 1961 à 1996, le SCI ne tenait compte que du premier auteur cité dans la séquence des noms d'auteurs figurant en tête d'un article. Ce problème est résolu depuis 1997 mais laisse encore des traces particulièrement sensibles dans la communauté scientifique. En effet les commissions d'évaluation ne cherchaient pas toujours à dépasser la première lecture des tableaux du SCI., et ce vice de forme a pu inciter les équipes à débattre sur le nom qui devait figurer en tête de liste au moment de la publication d'un article. Depuis 1997, lorsqu'un auteur figure en position $n+1$, son nom est mentionné en caractères minuscules sur les tableaux de référence, précédé d'un point de suspension. De plus, actuellement, les équipes de recherche constituées d'un très grand nombre d'auteurs s'identifient souvent par le biais exclusif de sigles qui se substituent à l'énumération des noms et qui semblent fréquemment omis par le SCI [Dickersin et al., 2002].
- b) le calcul du Facteur d'impact (Impact Factor) d'une revue table sur les deux années de publication précédant l'année de référence du *Journal Citation Reports*. On totalise le nombre d'articles publiés par cette revue dans ce délai de deux ans. On considère ensuite le nombre total de fois où ces articles sont cités dans la même période par l'ensemble des quelque 6000 revues analysées par le JCR. Et l'on divise ce second chiffre par le premier pour obtenir le facteur d'impact⁴. Le délai de deux ans est en soi critiquable car certaines disciplines pourraient mieux supporter une période de cinq ans.

2) Les critères éditoriaux⁵ :

il s'agit de considérations d'ordre purement arithmétique dans la prise en compte des types de support. Un mensuel a plus de chances d'être cité qu'un semestriel ; un titre de revue qui change remet son compteur à zéro pour l'évaluation du facteur d'impact ; un nouveau titre (notamment « libre » de type OAI) doit effectuer un parcours minimal d'environ cinq ans pour être « connu » etc... Autant de variables indépendantes du contenu mais relevant du seul « contenant ».

3) Le facteur humain :

C'est sans doute le plus important. Quarante ans de fonctionnement policé par l'ISI ont fini par créer un espace cooptatif quasiment tautologique, et celui-ci est de plus en plus critiqué. En 2002 et 2003, d'éminents chercheurs ont initié un débat significatif sur ce thème dans la revue *Nature*, à travers une chronique intitulée [*Errors in citation statistics*, 2002] (le titre est suffisamment

³ concerne aussi bien le SCI que le JCR

⁴ Exemple : la revue *EMBO Journal* (biologie moléculaire et biochimie) a publié 710 articles en 2001 et 677 en 2002. Soit un total de 1387. Au sein du corpus des 5907 revues analysées par le JCR dans son édition 2003, ces articles 2001-2002 de l' *EMBO Journal* ont été cités : 7998 fois en 2001 et 6504 fois en 2002, soit un total de 14502. Si l'on divise ce second total par le premier, on obtient l'Impact Factor d' *EMBO Journal*, soit 10,456.

⁵ ne concerne que le JCR

éloquent). Peter Lawrence⁶ n'hésite pas à s'exprimer sur le népotisme croissant qui ronge les comités de lecture [Lawrence, 2003], et David Colquhoun⁷, évoque des problèmes identiques sous un titre parlant : *Challenging the tyranny of impact factors* [Colquhoun, 2003].

2. *L'alternative: pourquoi ?*

Il va de soi que l'analyse experte d'un papier soumis s'impose absolument. Qu'il s'agisse d'examiner la cohérence scientifique de son contenu à la lecture des formules ou du raisonnement, ou même de la réitération des expériences dans des circonstances similaires, cela peut être effectué dans des conditions à la fois courtoises et raisonnables par un groupe anonyme (vis-à-vis de l'auteur) de *referees* patentés, que cette fonction généralement bénévole peut honorer dans leur carrière. On peut citer l'exemple assez libéral des règles observées auprès du groupe *BioMed Central*⁸ : les auteurs proposent quatre noms de *referees* potentiels qu'ils jugent compétents dans le domaine, sans lien aucun avec eux, ni hiérarchique ni institutionnel. Le Comité peut toutefois se réserver le droit d'en nommer d'autres. Des témoignages convergent sur le bon fonctionnement du système.

Toutefois, ce protocole assez "fluide" n'est malheureusement pas généralisé. Ces deux dernières décennies ont vu s'accroître de façon significative le nombre des chercheurs et des applications industrielles. Peu à peu, l'examen serein et impartial devient perturbé par des considérations exorbitantes du seul domaine scientifique, même si les relations humaines ont toujours exercé un poids non négligeable. Le facteur d'impact de l'analyse quantitative, passablement expéditive et peu nuancée, a tendance à primer sur l'aspect scientifiquement qualitatif du processus et les alternatives offrent des moyens de contrer un certain nombre de dérives susceptibles de s'accroître au fil du temps⁹.

Tout d'abord, les *referees* ont un statut hybride, dans la mesure où leur fonction d'auteur et souvent de responsable d'équipe peut empiéter sur leur activité au sein d'un tel comité. Les revues à fort facteur d'impact ont aussi tendance à surcharger leurs comités de rédaction, ce qui ne facilite pas la sérénité de l'examen. Certains *referees* sont contraints d'émettre un avis tranché en moins de trois quarts d'heure de lecture [Lawrence, 2003], demandant à l'auteur candidat de faire des modifications notables, voire de renouveler telle ou telle expérience. Il arrive que l'éditeur demande à l'auteur de renvoyer son manuscrit corrigé après un premier examen, à l'ensemble des *referees* et d'obtenir leur accord global (*BioMed* va jusqu'à deux navettes).

L'auteur peut à bon droit s'interroger sur l'utilité réelle des modifications requises et, plus encore, des nouvelles expériences sollicitées. Est-ce du temps perdu, surtout vis-à-vis de *referees* qui ne sont peut-être pas totalement au fait de sa spécialité ? Investir du temps dans l'envoi à d'autres revues, c'est, pour l'impétrant, perdre l'opportunité de satisfaire éventuellement une revue à fort facteur d'impact, dont le titre dans sa bibliographie personnelle peut optimiser son CV. Dans certains cas, ce sont plusieurs mois de travaux qui sont aussi perdus et l'auteur encourt le risque de se voir dépasser entre-temps dans le même domaine scientifique par des laboratoires concurrents. D'une façon générale, présenter son papier à plusieurs revues à la fois, dont les thématiques peuvent se chevaucher sans être identiques, fait courir un certain risque aux auteurs dont la recherche originale sort des sentiers battus¹⁰. C'est souvent le fait de chercheurs peu aguerris aux politiques éditoriales qui

⁶ du Laboratoire de Biologie moléculaire de Cambridge

⁷ du Département de Pharmacologie de l'University College de Londres

⁸ Société d'édition électronique de droit britannique, créée en 1999, pionnière en matière de revues libres d'accès dans le domaine biomédical : <http://www.biomedcentral.com/home/>

⁹ www.arxiv.org et www.citebase.org

¹⁰ C'est aussi une pratique interdite par certaines revues.

s'exposent inutilement et dangereusement en éparpillant leur envoi. Plus la teneur des recherches est diffusée sans certitude de publication, plus elle devient vulnérable.

Autre question délicate : un *referee* peut-il raisonnablement soutenir le papier d'un chercheur "exogène" susceptible de nuire aux travaux en cours de son propre laboratoire ou à ceux de ses doctorants ? Une telle occurrence peut expliquer la différence entre deux avis émis sur le même papier. L'examen par trois *referees* permet l'approbation ou le rejet par deux sur trois. S'il est assorti d'observations pertinentes et détaillées, cela permet en outre d'éviter à l'impétrant – ce qui arrive parfois – de se retourner vers le comité de lecture, de le "démarcher" abusivement, de flatter les *referees a priori* favorables, de dénigrer les instigateurs de rejet, et de demander la désignation de tel ou tel autre *referee*. Certains auteurs vont même jusqu'à harceler des éditeurs pour identifier le nom de *referees* rétifs restés naturellement anonymes. Car l'anonymat reste « théoriquement » de mise. L'objectivité est aussi émoussée par la cooptation entre *referees* : se fréquenter régulièrement, s'inviter les uns les autres à des colloques, attribuer un prix ou "nommer" tel ou tel, peut opacifier le suivi des examens.

Des litiges peuvent aussi toucher l'appropriation indue de la teneur d'un papier par le *referee*, qui profitera des attermoissements dans le processus d'examen pour publier sur le sujet en s'arrogeant la primeur. C'est l'éternel problème de la confidentialité qui, dans certains cas joue un rôle majeur en termes de propriété industrielle : satisfaire l'exigence d'un avancement de carrière rapide en soumettant prématurément un projet de publication ? Courir le risque d'un dérapage en matière de dépôt de brevet ?

Enfin, outre l'évaluation liée à l'aspect strictement technique du papier, on demande de plus en plus au *referee* de juger si l'article pourrait être valablement appréhendé par un lecteur d'une revue du type de *Science*, i.e. « générale d'excellent niveau », autrement dit par un éventail de destinataires plus large. C'est là une (autre) source d'arguments de rejet plus ou moins spécieux, car ils sont exorbitants de la scientificité stricte du document.

Sortons cependant de ces configurations passablement négatives, bien réelles quoiqu'un peu caricaturales, et voyons maintenant quelles sont les alternatives à un système fortement influencé, qu'on le veuille ou non, par les retombées de l'Impact Factor.

2.1 Citebase

La première alternative à laquelle nous faisons initialement allusion est celle que proposent Stevan Harnad et Tim Brody sur le site *Citebase*¹¹, à partir du miroir britannique du réservoir *ArXiv.org*¹². C'est en 1991 que le Los Alamos National Laboratory promeut une base « libre » en Physique théorique des Hautes énergies : *hep-th*¹³. Elle constituera le noyau d'*ArXiv*, autour duquel se développèrent d'autres bases en mathématiques, en astrophysique et en biologie quantitative notamment. Elle représente aussi un réservoir « pionnier » dans le domaine du libre accès.

Dans le droit fil de cette philosophie du droit à l'information outrepassant les barrières commerciales, Harnad et Brody conçurent le projet d'expérimenter un algorithme de citations qui, sans s'opposer frontalement à celui de l'ISI, propose dès 1999 une voie qui se veut plus objective et se concrétise dans l'interrogation de *Citebase*¹⁴.

¹¹ www.citebase.org

¹² www.arxiv.org

¹³ Elle est actuellement hébergée à la Cornell University.

¹⁴ On notera par ailleurs, au titre des initiatives « libres » de commentaires des articles, l'opportunité offerte par la base *Citeseer*. Mise au point par le NEC Research Institute, spécialisée en informatique et domaines connexes

Le principe de ce modèle repose sur le rapport existant entre les déchargements et les citations.

- a) il comptabilise et met en balance le nombre d'ouvertures de fichiers correspondant au texte intégral d'un article, et celui des citations y faisant référence dans d'autres articles également répertoriés dans la base
- b) il met en lumière le temps de latence existant entre la période d'ouverture et la période de citation ; plus l'écart est faible et plus les citations sont nombreuses, et plus l'on a de chances de tabler sur un article intéressant le domaine
- c) les éléments qui précèdent sont matérialisés par des courbes. Sur le long terme, celles-ci peuvent révéler le « cycle » d'un article : ouvert, il peut être cité, puis consulté par les lecteurs de l'article de second niveau qui, à leur tour, pourront le juger suffisamment « fondamental » pour le citer à nouveau et susciter d'autres ouvertures etc.

Dans les Foires Aux Questions d'ArXiv, les concepteurs répondent clairement « Non » à deux questions¹⁵ :

- 1) *la possibilité pour un auteur d'obtenir des statistiques sur le nombre de téléchargements de sa production (voire l'origine des téléchargements)*. Sans rentrer dans les détails de la réponse, la perspective d'une individualisation à outrance des résultats est catégoriquement refusée pour des raisons qui relèvent d'une déontologie particulière (et qui va justement à l'encontre d'une certaine philosophie du type de celle que peut engendrer l'ISI). Le contre-argument faisant allusion à *Big Brother* est aussi invoqué.
- 2) *la possibilité pour un auteur de recevoir directement les demandes de documents par messagerie, permettant ainsi d'identifier les personnes avec des intérêts similaires*. Là aussi, les concepteurs préfèrent le canal ascendant lecteur/auteur, via une invitation explicite à envoyer un e-mail de prise de contact, plutôt que l'inverse ; dans cette seconde hypothèse, l'auteur aurait accès aux sources d'interrogation de son texte et contacterait ses lecteurs sans demande de leur part.

2.2 Scholar Google

Apparu fin 2004 sur le Net, ce satellite de la galaxie Google constitue un réservoir de liens vers des articles de périodiques scientifiques, appartenant pour une écrasante majorité d'entre eux aux domaines des sciences exactes et appliquées.

L'algorithme de recherche est protégé et les contours précis de la couverture sont inconnus. Il s'agit en gros des grands éditeurs commerciaux classiques auxquels s'ajoutent les réservoirs « libres ».

Ainsi, à l'inverse du modèle de l'ISI mais au contraire à l'image du concept Brody/Harnad, *Scholar Google* est « ouvert ». Toutefois, alors que les communautés d'ArXiv sont relativement bien cernées, les réponses affichées par *Scholar Google*, quoique satisfaisantes en apparence, semblent échoir de façon quelque peu aléatoire. Ignorer les modalités exactes de la couverture expose l'outil à être considéré comme une aide certes très utile à la définition d'un paysage documentaire, mais sans la rigueur d'autres bases de citations.

Le module d'aide est focalisé très sensiblement sur l'insertion des références et s'adresse au chercheur qui publie ; il veut essentiellement répondre à la question basique d'un utilisateur classique du SCI : « comment puis-je retrouver mes propres articles dans *Scholar Google* ? »

(760.000 références), elle est construite suivant la même architecture que *Citebase*. Elle propose, à l'affichage des références, un lien de commentaire laissé libre au consultant, qui peut aussi voter en attribuant un indice de 1 à 5.

¹⁵ <http://fr.arxiv.org/help/faq/statfaq>

(question 1) et, plus loin : « pourquoi ne suis-je pas (encore) dans *Scholar Google* ? » (question 6). Plus simplement et plus objectivement, cette optique positionne *Scholar Google* comme une alternative au SCI, outil d'évaluation des auteurs pouvant accessoirement aider à l'élaboration d'une bibliographie. Il s'ensuit deux types d'attitudes :

- a) un étudiant avancé pourra utiliser cette base comme un outil pluridisciplinaire comparable à *Scirus* ou *Scienceresearch*¹⁶ se substituant avantageusement pour lui aux bases de données bibliographiques classiques inaccessibles sans contrat payant préalable. Il ne verra pas forcément l'utilité des citations.
- b) le chercheur confirmé aura tendance à substituer l'interrogation de *Scholar Google* à celle de l'ISI, en raison (comme dans le cas précédent) du caractère gratuit de l'un et (particulièrement) onéreux de l'autre. Mais sans voir que le SCI répond malgré tout à une notion extrêmement rigoureuse (du moins depuis 1997) des critères d'interrogation. *Scholar Google* peut en effet surprendre sur l'interrogation des auteurs si les initiales de prénoms sont mal concaténées aux patronymes, et les homonymes exigent, dans ce réservoir pluridisciplinaire, d'être associés à des « mots-clefs » de domaines pris au hasard pour les départager.

2.3 Le facteur de Hirsch

En novembre 2005, Jorge E. Hirsch¹⁷ a commis dans les PNAS un article intitulé *An index to quantify an individual's scientific research output* [Hirsch, 2005] dans lequel il propose le concept d' « indice *h* ».

Il fait glisser deux curseurs en sens inverse : l'un sur une échelle du nombre de publications, l'autre sur une échelle du nombre de citations, et il les fait coïncider. Il prend l'exemple d'un physicien renommé, particulièrement fécond en citations : E. Witten. Les deux curseurs se rencontrent sur le nombre 110, i.e. il existe 110 publications citées chacune au moins 110 fois. C'est l'indice *h*. Il s'agit en fait d'une moyenne calculée sur un ensemble de données, non pas dans le cadre délimité de l'ISI, mais sur l'espace ouvert du Libre Accès : nous avons réexécuté ses calculs sur l'archive SPIRES (High-Energy Physics Literature Database, segment d'ArXiv faisant l'objet d'un site dédié¹⁸), réalisée par le Stanford Linear Accelerator Center et avons retrouvé ses chiffres. Les résultats obtenus à partir du réseau du Libre Accès sont très supérieurs à ceux du réservoir restreint de l'ISI¹⁹. Un biologiste allemand, Werner Müller²⁰, a fait le même constat en comparant le calcul de Hirsch avec un comptage sur la plate-forme *Scopus* élaborée par Elsevier [Müller, W., 2006]

Hirsch associe à son indice *h* une valeur *m* calculée en fonction du nombre d'années de carrière *n* du chercheur au moment de l'évaluation. La valeur *m* est égale au facteur *h* divisé par le nombre d'années d'ancienneté. Cela donne une répartition en trois groupes. :

- a) $m \approx 1$ ($h = 20$ après 20 ans d'activité) : chercheur de « bon niveau »
- b) $m \approx 2$ ($h = 40$ après 20 ans d'activité) : chercheur responsable d'une équipe de haut niveau dans une université très cotée
- c) $m \approx 3$ et + ($h = 60$ après 20 ans d'activité, voire 90 après 30 ans) : chercheur hors pair, Nobel etc...

¹⁶ www.scirus.com et www.scienceresearch.com

¹⁷ Physicien à l'Université de Californie à San Diego (La Jolla)

¹⁸ <http://www.slac.stanford.edu/spires/hep/>

¹⁹ Michael Schwartzbach, informaticien à l'université d'Aarhus (Danemark) a réalisé un moteur expérimental calculant l'indice de Hirsch sur des noms d'auteurs à partir de *Scholar Google*, réservoir typiquement « ouvert ». [Schwartzbach, 2006]

²⁰ Immunologiste au Helmholtz-Zentrum für Infektionsforschung de Braunschweig

À cette hiérarchie correspondent aussi des degrés moyens d'avancement : $h=12$ pour un « assistant », 18 pour un professeur, 15 à 20 pour un membre de l'American Physical Society, 45 pour un académicien...

Ce système suppose donc trois éléments : une ancienneté de carrière assez importante, un grand nombre de publications, générant en troisième lieu un grand nombre de citations. Cette combinaison est censée être moins partielle (et partielle) que l'association entre le nombre de citations et le facteur d'impact des revues correspondantes, au sens de l'ISI.

La méthode de Hirsch a fait l'objet de plusieurs réflexions sur des blogs de chercheurs. Ils déplorent la relation entre la nécessité d'avoir commis un grand nombre de publications amplement citées et le nombre d'années de carrière²¹. La corrélation entre ces deux facteurs n'a en effet rien d'obligatoire. Un chercheur objectivement « méritant » dans telle ou telle discipline, n'aura peut-être écrit qu'un petit nombre d'articles dont la portée aura sensiblement modifié le paysage scientifique, et ce – pourquoi pas ? - dès le début de sa carrière. Les critiques regrettent également la non prise en compte des différences de pratiques entre les disciplines [Iglesias & Pecharromán, 2006]. Certains champs disciplinaires, notamment fondamentaux, peuvent ne générer qu'un nombre modéré de publications. Ce qui n'empêche pas de nombreuses manipulations en laboratoire ou de multiples communications dans le cadre de sociétés savantes, ne donnant pas forcément lieu à publication systématique, mais pouvant déboucher sur l'obtention d'un prix ou d'une médaille. Un contributeur du Wiki de l'université de British Columbia (Canada)²² souligne le côté positif de l'impact personnel du chercheur, tel que Hirsch l'envisage, par opposition à l'impact de la revue, au sens de l'ISI.

Précisons enfin que Hirsch propose, comme l'ISI, un algorithme correctif pour les auto-citations.

2.4 Le *Journal Status* de Bollen et al.

Parallèlement, Johan Bollen, Marko A. Rodriguez et Herbert Van de Sompel proposent un *Journal Status* [Bollen, 2006] qui tient compte de la distinction entre les deux notions de « popularité » et de « prestige ». Elles sont inversement corrélées. Les revues « populaires » sont fréquemment citées par celles qui sont peu prestigieuses ; elles sont dotées d'un Impact Factor ISI très élevé. Les revues « prestigieuses » sont définies de façon strictement inverse par rapport aux « populaires ».

La caractéristique déterminante qui les différencie par rapport au nouveau facteur d'impact proposé par les auteurs, c'est leur « Weighted PageRank ». Cet algorithme a notamment été mis en évidence par [Xing & Ghorbani, 2004]. Il porte sur la connectivité non seulement de la page « racine » elle-même (ce que fait sur Google l'algorithme standard de Larry Page) mais au-delà sur la connectivité des pages auxquelles elle est liée, aussi bien dans le sens entrant que sortant. On sait que tous les liens d'une page ne sont ni forcément utiles ni forcément utilisés. Analyser la cohérence et la pertinence du suivi des liens aussi loin et profondément que possible dans leur cheminement d'une page à l'autre : tel est l'objectif du *Weighted PageRank*. Il a aussi été employé par [Liu et al., 2005] pour élaborer un *Author Rank* dans des corpus thématiques définis. Liu calcule par ce biais un Facteur d'impact fondé sur l'interactivité des co-auteurs dans des communautés plus ou moins éclatées. La notion de *Weighted* apparaît là en toute clarté : « peser » signifie alors mesurer le degré de cohérence et de pertinence de la séquence des liens (*degree centrality* : nombre d'articles parents et proportion de co-auteurs identiques), puis le degré de proximité absolue d'un article - et donc

²¹ *The Quantum Pontiff*, blog de David Bacon, informaticien à l'Université de Washington : 25 commentaires sur le facteur de Hirsch entre le 15 avril 2005 et le 13 décembre 2006 sur <http://dabacon.org/pontiff/?p=1066>

²² *Impact Factor*. http://hlwiki.slais.ubc.ca/index.php?title=Impact_factor (23-10-2006)

d'une équipe - à l'autre (*closeness centrality*) ou de proximité relative des mêmes éléments (*betweenness centrality* : deux « nœuds » de co-auteurs sont séparés par un troisième qui les relie dans la bibliographie d'un « tiers-article »). Dans le contexte du *Journal Status*, Bollen ne va pas utiliser ce *Weighted PageRank* isolément. Il est conscient des spécificités du modèle de l'ISI « Nous avons maintenant à notre disposition – explique-t-il – deux modèles statistiques, très différents mais très utilisés, de mesure du statut des revues scientifiques. Le Facteur d'impact de l'ISI repose sur la fréquence des citations et met de ce fait en valeur la “popularité” du statut d'une revue. Le *PageRank* pondéré (*weighted*) quant à lui, tel que défini plus haut, repose sur la transmission des valeurs de “prestige” d'une revue à l'autre et correspond mieux à cette notion que nous élaborons intuitivement, en vertu de laquelle le prestige ne ressort pas seulement du nombre de cautions intellectuelles (*endorsements*), mais aussi de l'identité de ceux qui les confèrent. » C'est précisément là qu'interviennent les réseaux d'auteurs auxquels [Liu et al., 2005] font référence. Bollen admet la coexistence des deux modèles et propose en conséquence un facteur, désigné par la lettre *y*, qui se définit comme le produit du facteur d'impact de l'ISI par le *Weighted PageRank*. Ce facteur *y* apparaîtrait donc comme un compromis assez équilibré dans l'évaluation des sources.

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

Les exemples sont plus significatifs dans le cadre disciplinaire. En Physique, dans la liste de référence des dix premiers facteurs d'impact du JCR, 6 revues sont absentes des dix premières de la liste « facteur *y* ».

2.5 Faculty of 1000 : le *F1000 Factor*

Il s'agit ici d'un modèle qui tranche totalement avec ceux qui précèdent. Il ne repose plus sur le nombre de citations mais sur une lecture coopérative effectuée simultanément par les membres d'un groupe de spécialistes constitué par cooptation dans un domaine biomédical précis. Ils se donnent pour tâche d'analyser des articles susceptibles d'engendrer une évolution scientifique significative, sur des questions jugées essentielles par la communauté concernée. Le service, créé par *BioMed Central*, se subdivise en *F-1000 Biology* et *F-1000 Medicine*. La section *Biology* a été mise en place en janvier 2002, *Medicine* en janvier 2006.

Prenons l'exemple de *Faculty of 1000 Biology*. Le domaine est subdivisé en 16 « facultés », i.e. Sujets, dont la supervision est confiée à 2, 3 ou 4 experts qualifiés *Heads of Faculty*. Chacune des Facultés se décline en Sections, dont le nombre oscille entre 3 et 12, dont la responsabilité est assumée par 2 ou 3 experts qui, à leur tour, contrôlent des équipes variant de 10 à 50 membres.

On observe donc une hiérarchie assez pointue entre Facultés et Sections. L'effectif des Facultés est proportionnel au nombre d'articles publiés en moyenne dans la discipline correspondante. 1400 membres se répartissent sur 73 sections, certaines d'entre elles se chevauchant sur le plan conceptuel : d'où la nécessité d'une indexation fine et transversale suivant le principe de l'arborescence.

Aux domaines classiques de la Biologie vont bientôt s'ajouter ceux de l'Ecologie, de la Pharmacologie et de la Physiologie.

Chaque membre, à la base, est invité à commenter les 2, 3 ou 4 articles qu'ils ont jugés les plus intéressants sur une période d'un mois. Chaque commentaire est assorti de liens vers les séminaires ou conférences qui peuvent graviter autour de la rédaction de cet article, de même que vers d'autres articles plus ou moins directement liés à celui-ci. Le cas échéant, le lien vers le texte intégral est bien sûr proposé.

Puis, et c'est là que l'on touche la spécificité de ce système, chaque membre attribue le *F1000 Factor* en se conformant à une grille limitée à trois indices : 3 points = « Recommended » / 6 = « Must read » / 9 = « Exceptional »²³.

On fait ensuite la moyenne des deux indices les plus élevés attribués à cet article, à laquelle on ajoute un trentième de la somme de tous les indices attribués. Cela implique évidemment que l'article ait fait l'objet d'un minimum de trois commentaires, mais s'il est censé avoir un impact significatif, on peut naturellement supposer qu'il a été « détecté » et lu par un nombre appréciable d'examineurs.

Un commentaire est mis en ligne par le panel, qui *qualifie* l'article (découverte, avancée, hypothèse, confirmation, controverse) et le *catégorise* dans l'une des Sections.

Faculty of 1000 dresse par ailleurs périodiquement une liste de « perles » (*Hidden Jewels*), i.e. des articles dont l'intérêt semble déterminant dans un cadre thématique précis, parus dans des revues dont l'impact au sens classique du terme est jugé relativement faible. Il peut s'agir par exemple de revues publiées par des Sociétés savantes, dont le haut niveau qualitatif est incontournable dans la discipline, mais dont la diffusion peut rester plus ou moins confidentielle pour des raisons éditoriales matérielles. Si l'on se réfère aux systèmes évoqués plus haut, ces revues bénéficieraient d'un classement peu attractif selon l'ISI, mais très fort selon Bollen. Quotidiennement, le système génère le *Top 10* des *Hidden Jewels*, leur ordre séquentiel étant fonction du nombre de fois où l'évaluation de chaque article listé a été visualisé sur le site au long de la dernière période de 14 jours francs. On pourrait, à certains égards, comparer cette modélisation aux critères de calcul de *Citebase*.

Ceci posé, on retrouve là un système ouvert, i.e. non fondé sur une liste close de titres de revues.

En second lieu, il est conçu pour une utilisation « au fil de l'eau », beaucoup plus que comme un état synthétique reposant sur des données à la fois cumulées et stables, du type ISI. Peut-être est-ce là une philosophie nouvelle de l'évaluation scientifique qui, en dépit de son manque apparent de cadrage, comporte des facteurs humains d'appréciation tout à fait positifs ?

3. Approche de comportements typiques par disciplines

Lorsque le réservoir international *ArXiv* fut créé au début des années 90, la communauté qui le promut était essentiellement constituée par des physiciens des hautes énergies. La philosophie dominante était la mise sur le Web d'un compte rendu d'expérience dès que possible, afin qu'il soit universellement visible et puisse faire l'objet, de la part de la communauté, d'une contre-expérience ou d'une contre-expertise en vue d'une validation. On divulgue immédiatement pour être confirmé dans sa démarche.

La communauté biomédicale a depuis longtemps échangé assez librement sur des supports libres. *BioMed Central* et *PLOS (Public Library of Science)*²⁴ en sont des exemples. Sans doute une telle attitude est-elle appuyée par une visée plus ou moins humanitaire liée à la santé. Mais il faut parallèlement bien voir également que des intérêts associés à l'industrie pharmaceutique [Sotelo, 2006] peuvent sous-tendre le soutien à telle ou telle revue. Enfin, la gratuité de *PubMed* doit beaucoup à l'influence de l'ex vice-président américain Al Gore qui a favorisé le financement public de cette plate-forme documentaire remarquable dès 1997.

La chimie demeure majoritairement à l'écart du mouvement de l'accès libre. Ce domaine reste

²³ En fait, l'échelle est étalonnée de la façon suivante : de 3 à 4,5 puis de 4,6 à 7,5 enfin 7,6 et au-delà (e-mail de David Weedon, Managing Director, F1000 Biology, 22-01-07)

²⁴ www.plos.org

lié aux circuits de l'édition commerciale, particulièrement onéreuse. Peut-être cela s'explique-t-il par les nombreuses applications industrielles générées par cette discipline.

Les mathématiques ont tendance à représenter une famille relativement fermée. Elle comporte un très grand nombre de sociétés savantes. Elle est massivement représentée dans *ArXiv*. Sa communauté se réfère beaucoup au réseau américain *MathSciNet*²⁵ de l'American Mathematical Society, dont le noyau est la *Mathematical Review* créée en 1940. Son comité dispose d'une équipe de *referees* qui écrivent des recensions sur la littérature du domaine : plus de 60.000 *reviews* sont ainsi ajoutées chaque année à la base de données *MathSciNet*. En 2006, elle compte plus de 2 millions de références et 700.000 liens directs vers les textes intégraux. Beaucoup ont été scannés en remontant jusqu'à 1864. La base possède également un référencement par citations.

Conclusion

Compte tenu de ce qui précède, il apparaît hautement souhaitable d'explorer rapidement la voie d'un nouvel outil d'évaluation libre d'accès, dont les règles seraient clairement définies, tant au niveau de la couverture qu'à celui des critères d'analyse statistique.

La neutralité apparente de l'analyse statistique brute (*a priori* quantitative) effectuée sur des masses relativement importantes à partir du critère de la citation, risque, on l'a vu, de générer des dérives difficilement contrôlables, des « effets délétères », pour reprendre les termes de Jean-Marc Monteil, Directeur de l'Enseignement supérieur²⁶ qui juge indispensable une réflexion approfondie sur la bibliométrie. Celle-ci ne doit pas, avance-t-il, altérer l'analyse qualitative en tant que telle. Une recherche de type cooptatif, i.e. sur des thèmes déjà porteurs, n'impliquant aucun risque réel, en circuit plus ou moins fermé sur des valeurs peut-être trop sûres, s'éloigne des principes authentiquement novateurs de la recherche. Dans la conjoncture présente, une bibliométrie gérée sans nuance peut contribuer à freiner la recherche «*Blue Sky*» au sens où l'entend Donald Braben, chercheur au University College de Londres [Braben, 2004], à savoir : sortir des sentiers battus pour déboucher sans contraintes (mais aussi sans garantie) sur une réelle innovation, susceptible d'engendrer des répercussions durables sur la science et la technologie.

Jean-Marc Monteil estime²⁷ que les méthodes d'évaluation pourraient faire utilement l'objet d'un exposé auprès des doctorants. Il suggère enfin que les indicateurs de production scientifique des établissements devraient, dans l'état actuel de leurs méthodes d'élaboration, être considérées comme des « aides à la décision » plutôt que comme des « normes décisionnelles ».

Par ailleurs, s'il est exact que l'analyse qualitative des articles effectuée par les « collectifs » de *Faculty of 1000* semble plus « humaine », elle reste soumise aux aléas de la constitution des équipes lectrices et marquée par une incertitude relative quant à la couverture des revues analysées.

Un nouveau modèle pourrait combiner des éléments d'investigation envisagés dans les expériences relatées ci-dessus, associant avec une impartialité optimale les objectifs d'évaluation des sources et des auteurs. Sa mise au point paraît, dans la conjoncture actuelle, assez urgente.

²⁵ <http://www.ams.org/mathscinet>

²⁶ *Séminaire de l'AMUE [Agence de Mutualisation des Universités] : Les indicateurs de production scientifique des établissements : pour quoi faire ? comment faire ?* Paris, Université René Descartes, 22 novembre 2006, <http://canalc2.u-strasbg.fr/video.asp?idvideo=5503>

²⁷ *Séminaire de l'AMUE*, ibidem

Références

- Bollen, J., Rodriguez, M.A., & Van de Sompel, H. (2006) Journal Status. *Scientometrics*, 69(3) 669-687
- Braben, D.W. (2004) *Pioneering research : a risk worth taking*. Chichester: Wiley
- Colquhoun, D. (2003) Challenging the tyranny of impact factors. *Nature*, 423, 479
- Dickersin, K., Scherer, R., Suci, E.S.T. & Gil-Montero, M. (2002) Problems with Indexing and Citation of Articles with Group Authorship. *Journal of the American Medical Association*, 287, 2772-2774
- Errors in citation statistics [Opinion]. (2002) *Nature*, 415, 101
- Hirsch, J.E. (2005) An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569-16572
- Iglesias, J.E. & Pecharrómán, C. (2006) Scaling the h-index for different scientific ISI fields. www.arxiv.org/ftp/physics/papers/0607/0607224.pdf
- Lawrence, P. (2003) The politics of publication. *Nature*, 422, 259-261
- Liu, X., Bollen, J., Nelson, M.L., & Van de Sompel, H. (2005) Co-authorship networks in the digital library research community. *Information Processing & Management*, 41(6), 1462-1480.
- Müller, W. (2006) Hirsch Faktor : the new way to measure the scientific performance. *Eu-Genes*. www.eugene.de/page4/page4.html
- Schwartzbach, M. (2006) H-index Calculator of Scientist Impact and Influence. *Epidemiologic Inquiry*. www.epidemiologic.org/2006/12/h-index-calculator-of-scientist-impact.html
- Sotelo, J. (2006) Regulation of Clinical Research Sponsored by Pharmaceutical Companies: A Proposal. *PLoS Med* 3(7), e306
- Xing, W. & Ghorbani, A. (2004) Weighted PageRank algorithm. In *Second Annual Conference on Communication Networks and Services Research CNSR'04*, Fredericton, N.B., Canada, 305-314