

teurs d'impact des revues pour juger la qualité des publications d'un chercheur. En 2007, la *European Association of Science Editors* (EASE) a d'ailleurs émis l'avis suivant : « Le facteur d'impact n'est pas toujours un instrument fiable. En conséquence, on ne doit l'utiliser – avec précaution – que pour comparer l'influence des revues, mais pas pour évaluer un article et encore moins un chercheur ou un programme de recherche. »

Une idée plus simple et sans doute bien plus saine est de prendre en compte le nombre total de citations (NTC) faites aux travaux du chercheur à évaluer. Si un chercheur est cité, c'est que ses travaux sont reconnus et appréciés. Plus un chercheur est cité, plus son NTC augmente, plus il devient clair qu'il est efficace, a de l'influence et produit des idées et résultats pertinents. Au moins en première approximation, il semble raisonnable de jauger un chercheur à partir de son NTC.

Calculer le NTC d'un chercheur exige de disposer de bases de données contenant les listes d'articles publiés mais aussi, pour chaque publication, la liste des publications qu'elle mentionne. De l'ensemble

de ces listes on déduit, grâce à des programmes informatiques, le nombre de citations renvoyant à une publication donnée. Les trois principales bases de données sont la propriété d'acteurs privés : *Thompson Reuters* dispose de la base *SCI (Science Citation Index)*, via le *Web of Science* ; *Elsevier* a lancé en 2004 la base *Scopus* ; *Google* propose *Google Scholar* dont l'accès, contrairement aux deux autres, est gratuit.

Compter les citations

Les deux premières incluent surtout des journaux de recherche rigoureusement sélectionnés et mesurent donc l'activité de recherche dans un sens étroit. La base *Google Scholar*, plus large, donnera une idée de l'activité du chercheur en prenant mieux en compte les actes de conférences et congrès (très importants en informatique), ainsi qu'une part du travail de publication didactique. Selon les domaines et selon ce qu'on souhaite mesurer, on choisira l'une ou l'autre, mais toutes les trois sont assez bonnes et donnent des évaluations rarement contradictoires.

Considérons quelques exemples calculés en utilisant la base *Google Scholar* et la page Internet qui en synthétise les résultats (<http://quadsearch.csd.auth.gr/index.php?lan=1&s=2>).

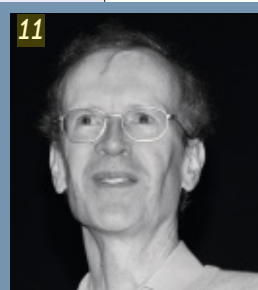
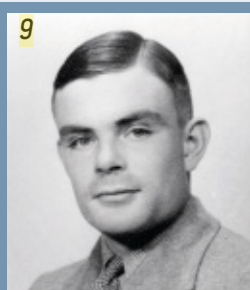
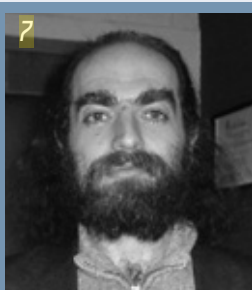
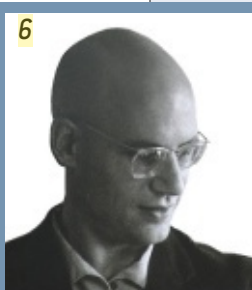
Nous trouvons que les travaux de Paul Erdős ont été cités 15 040 fois, chiffre considérable, mais peu étonnant puisqu'il est sans doute le plus prolifique de tous les mathématiciens du XX^e siècle avec plus de 1 500 articles à son actif (dont 975 sont identifiés par le système). Jean Dieudonné, un des piliers du groupe Bourbaki, a lui été cité 12 580 fois. L'influent Nicolas Bourbaki obtient 16 127. Alexandre Grothendieck, mathématicien français vénéré, a un NTC de 9 110 et Cédric Villani, l'un des médaillés Fields 2010, un NTC de 3 426. Notons que seuls les mathématiciens de bonne renommée atteignent un NTC de 1 000.

Quelques tests et l'examen de ces nombres montrent que le nombre total des citations NTC est une note imparfaite attribuée à un chercheur. D'abord, les problèmes typographiques sont difficiles à éliminer. Une requête pour « N Bourbaki » donne 25 642,

1. Le classement de quelques mathématiciens

Pour chacun de ces 11 mathématiciens renommés, on a noté le nombre *N* des publications repérées, le nombre total de citations (NTC) qu'elles ont reçu, l'indice *h*, l'indice *g* et le taux moyen de citations (TMC). Ces données brutes ont été calculées le 29 décembre 2010 à l'aide du site *Uad Search* (<http://quadsearch.csd.auth.gr/index.php?lan=1&s=2>) qui utilise la base bibliographique *Google Scholar*. Paul Erdős est le chercheur de notre liste qui a le plus publié, mais c'est lui qui a le taux moyen de citations par article le plus faible. Les deux plus célèbres mathématiciens de notre sélection sont sans doute Andrew Wiles et Grigori Perelman à cause de leur démonstration, pour l'un du grand théorème de Fermat, pour l'autre de la conjecture de Poincaré. Or ce sont les moins cités, ceux qui ont le moins de publications identifiées et les plus petits indices *h* et *g* !

	<i>N</i>	NTC	Indice <i>h</i>	Indice <i>g</i>	TMC
1. Nicolas Bourbaki	355	16127	47	125	45,4
2. Alain Connes	329	19892	61	138	59,4
3. Jean Dieudonné	310	12580	39	111	40,6
4. Paul Erdős	975	15040	60	110	15,4
5. Kurt Gödel	215	7471	28	85	34,6
6. A. Grothendieck	136	9110	32	95	66,9
7. Grisha Perelman	20	1643	11	20	82,1
8. Terence Tao	445	13387	49	107	30,1
9. Alan Turing	100	14359	22	100	143,6
10. Cédric Villani	102	3426	28	58	33,6
11. Andrew Wiles	88	3277	16	57	37,2



ce qui est presque le double de ce qu'on obtient avec la requête « Nicolas Bourbaki ». L'écart, qui n'est pas aussi grand pour les autres chercheurs, est sans doute dû au fait que les ouvrages de Bourbaki ne font pas apparaître complètement le prénom Nicolas.

Autre difficulté : une partie des citations comptabilisées pour Bourbaki devrait être reportée sur Jean Dieudonné, puisqu'il est l'un des auteurs anonymes des ouvrages de Bourbaki. Bien sûr, personne ne sait comment opérer ce report ! Le même problème se pose, cette fois à une bien plus grande échelle, pour l'encyclopédie Wiki-

pédia dont l'impact est devenu considérable... et impossible à prendre en compte.

Cédric Villani, le plus jeune des chercheurs mentionnés ici, est défavorisé puisque sa carrière est loin d'être terminée.

Par ailleurs, d'une discipline à l'autre, les usages en matière de publication et de citation sont très différents. Il est donc absurde de comparer des chercheurs de disciplines différentes en considérant leur NTC (voir la figure 3). En mathématiques, par exemple, on publie moins et les articles sont plus souvent publiés avec un seul nom d'auteur, alors qu'en physique des articles ayant plusieurs

dizaines de signataires ne sont pas rares. Ce problème des articles signés à plusieurs se retrouvera plus loin pour l'indice h .

En examinant, publication par publication, le nombre des citations faites à chacune, on remarque que, même chez les plus renommés des scientifiques, nombreux sont les articles à ne jamais être cités. Certes, il est faux de croire que les articles jamais cités sont inutiles, mais je doute que beaucoup de chercheurs sachent qu'une part importante de leurs travaux n'est jamais mentionnée, et cela même s'ils ont été publiés dans des revues prestigieuses.

2. Articles, citations, indice h

Considérons un chercheur ayant N publications $P_1, P_2, \dots, P_N$ à son actif. Notons C_i le nombre de fois que la publication P_i a été citée dans une autre publication. On supposera que l'on a classé les publications en plaçant les plus citées en tête : $C_1 \geq C_2 \geq C_3 \geq \dots \geq C_N$. Le nombre total NTC de citations de ce chercheur est : $C_1 + C_2 + \dots + C_N$. L'indice h de ce chercheur est par

définition le plus grand i tel que $i \leq C_i$. Géométriquement, l'indice h est donc l'abscisse du point le plus à droite du graphe de la fonction $i \rightarrow C_i$, placé au-dessus de la droite $y = x$, c'est-à-dire en gros l'abscisse de l'intersection de la droite avec la courbe (voir le graphique de gauche ci-dessous).

Avoir un indice h de 10 peut résulter de profils très différents (c'est

un reproche fait à l'indice h). Examinons cinq profils.

- Profil A. C'est le « chercheur optimal », il obtient 10 avec 10 articles cités chacun 10 fois. Nul ne peut avoir un indice h de 10 en étant moins cité.
- Profil B. C'est le « chercheur linéaire équilibré » : en moyenne, chacune de ses 19 publications est citée 10 fois et il a autant de citations inutiles (45) avant le point d'intersection de sa

courbe avec la droite $y = x$ qu'après. Son NTC est $1 + 2 + 3 + \dots + 19 = 190$. Comme on le voit ci-contre (modèle A), il peut espérer gagner un point d'indice h tous les deux ans.

- Profil C. Encore un « chercheur linéaire ». Cette fois, 90 des citations de ses neuf premières publications sont inutiles (à son indice h de 10) et 30 autres après le point d'intersection. Conseil : il doit éviter les

