

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).

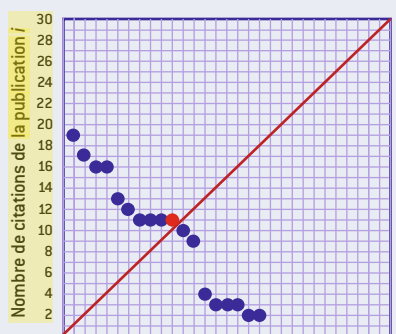
Avoir un indice h de 10 peut résulter de profils très différents (c'est

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

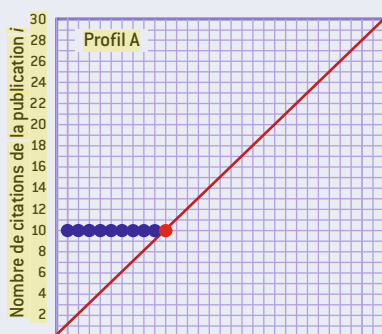
- Profil A. C'est le « chercheur optimal », il obtient 10 avec 10 articles cités chacun 10 fois. Nul ne peut avoir un indice h de 10 en étant moins cité.
- Profil B. C'est le « chercheur linéaire équilibré » : en moyenne, chacune de ses 19 publications est citée 10 fois et il a autant de citations inutiles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

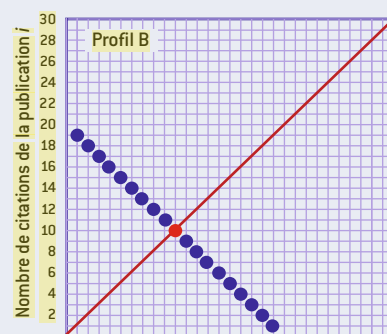
- Profil C. Encore un « chercheur linéaire ». Cette fois, 90 des citations de ses neuf premières publications sont inutiles (à son indice h de 10) et 30 autres après le point d'intersection. Conseil : il doit éviter les



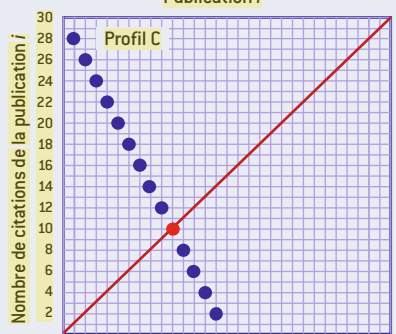
Publication i



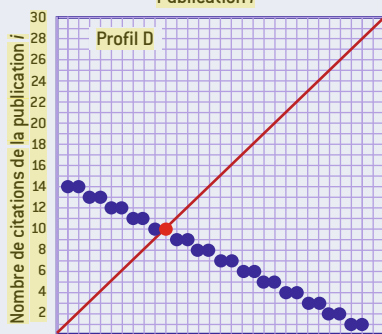
Publication i



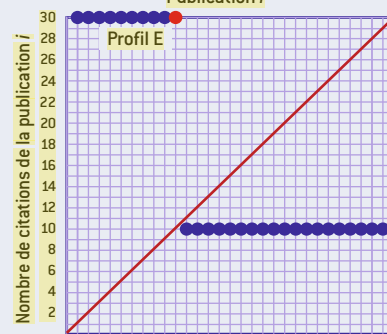
Publication i



Publication i



Publication i



Publication i

Le nombre total de citations NTC d'un chercheur est informatif, mais son domaine de variation est assez large et il a plusieurs défauts qui le rendent inapte à opérer des comparaisons fines. Un chercheur ayant publié un seul travail dans un domaine central ou à la mode et qui, du coup, est très visible et largement cité sera considéré comme l'égal d'un autre qui aura publié une série de travaux difficiles et plus spécialisés, donc moins cités, alors que la contribution scientifique réelle du second peut être très supérieure. Un chercheur publiant des articles de synthèse sera favorisé, même s'il est peu

créatif. Les auteurs de livres seront en général défavorisés car, d'une part, les livres sont moins bien pris en compte que les articles dans les bases de données et, d'autre part, le travail pour écrire un livre est en général bien plus grand que pour un article.

L'indice h de J. Hirsch

C'est sans doute pour cela qu'en 2005, Jorge Hirsch, physicien de l'Université de Californie à San Diego, a introduit une nouvelle idée qui, depuis, est devenue le critère préféré pour évaluer les chercheurs. L'idée de l'indice

de Hirsch a donné lieu à tant d'études et de travaux de bibliométrie qu'aujourd'hui la publication de J. Hirsch de 2005 est citée 1 252 fois alors que sa publication la plus citée en physique ne l'est que 410 fois... ce qui est pourtant très bien !

L'indice h (ou h -index en anglais) de Hirsch prend en considération uniquement les meilleures publications d'un chercheur. L'indice h d'un chercheur est le plus grand entier k tel que les k articles les plus cités de ce chercheur ont chacun été cités au moins k fois. Pour avoir un indice h de 1, il faut avoir 1 publication citée 1 fois. Pour avoir un indice h de 2,

publications trop citées (par exemple en les fractionnant).

- Profil D. À nouveau un « chercheur linéaire », mais ses publications après l'intersection sont les plus inutiles et gâchent les citations (90 pertes après le point d'intersection, et 20 avant). Conseil : il doit se faire citer plus.
- Profil E. Ce chercheur est le plus « maladroit » de notre lot : 200 citations sont inutiles avant l'intersection,

et 200 après. On imagine sans mal des cas pires encore.

Examinons maintenant l'évolution de l'indice h de chercheurs réguliers. Imaginons un chercheur (modèle A) qui publie un article chaque année et supposons que dès l'année où l'article a été publié, il est cité une fois par an exactement (y compris l'année de sa publication). Au bout d'un an, l'indice h

de notre chercheur imaginaire est 1 (il a un article cité une fois). Il lui faut attendre la fin de sa troisième année pour avoir un indice h de 2 et la fin de l'année $2k - 1$ pour avoir un indice h égal à k . L'indice h d'un tel chercheur croît linéairement.

L'hypothèse de citations régulières et indéfiniment prolongées pour chaque article publié est assez optimiste. En effet, les phénomènes de

mode et l'oubli ont pour conséquence qu'après avoir atteint un maximum, un article verra en général l'intérêt qu'il suscite diminuer. Si l'on suppose qu'au bout de cinq ans, on cesse de mentionner les articles de notre chercheur imaginaire (modèle B), cela aura un grave impact sur son indice h qui, après avoir atteint 5, cessera de croître, car aucun de ses articles ne sera jamais cité plus de cinq fois.

