



Comparer les productions scientifiques des universités : le projet UNI/Ref

Nicolas CARAYOL

Professeur d'économie à l'université Montesquieu – Bordeaux IV

L'OST, en particulier par l'intermédiaire du programme IPERU, a collecté une masse d'information homogène et de grande qualité sur la production scientifique des établissements d'enseignement supérieur français. Ne serait-il pas utile pour la communauté d'utiliser ces données à des fins de comparaisons ? Il s'agit là du point de départ de ce travail. Cependant, on a vu que les problèmes de comparaisons et de classements ne sont pas simples. En particulier ils butent sur de multiples problèmes d'agrégation : les critiques formulées dans l'intervention précédente à propos des classements d'universités donnent plusieurs exemples de ces problèmes. Les procédures d'agrégation et surtout leurs implications normatives ne sont jusqu'à présent pas suffisamment bien maîtrisées. Il s'agissait là d'un verrou scientifique qu'il nous fallait lever pour pouvoir établir des comparaisons de productions scientifiques des établissements dont les fondements seraient explicites et donc fassent sens pour les utilisateurs. Ce constat a stimulé un travail de recherche depuis presque deux ans au sein de l'OST. Nous entrons désormais dans une phase plus aval de la recherche nous permettant d'envisager une restitution de cet outil vers les acteurs et la mise au point d'un prototype/pilote. Le projet UNI/Ref a ainsi pour objectif de permettre la comparaison entre les productions scientifiques des établissements sur des bases explicites et discutées.

Les fondements de la réflexion vont être évoqués rapidement, puis seront illustrés par un exemple d'utilisation, enfin on parlera du « pilote » qui pourrait être mis en place.

Même si on ne se préoccupe que de la production scientifique, on a un premier problème d'agrégation qui est lié au fait que l'étude des articles publiés par un établissement peut être envisagée suivant deux dimensions simultanément, le nombre d'articles d'une part, pour évaluer le volume de production scientifique de l'établissement, et l'impact de chaque article d'autre part, pour en évaluer la qualité. La question est de savoir comment valoriser positivement ces deux aspects pour mener une analyse les prenant en compte simultanément. Pour simplifier, on peut par exemple se poser la question suivante : « *Dois-je préférer deux articles d'impact 3 ou trois articles d'impact 2 ?* ». Un responsable d'établissement peut être amené à se poser ce type de question et aura toujours dans sa réponse un principe d'agrégation implicite qui lui est propre. Ainsi toute comparaison bidimensionnelle repose sur un principe d'agrégation qui implique nécessairement un mode de valorisation, implicite ou explicite, de l'impact.

Pour illustrer cette question, nous avons traduit graphiquement (P.5 du diaporama associé) les préférences de différents individus en matière de valorisation de l'impact et défini six types de comportement. Ainsi, l'individu A ne va prendre en considération que le volume de production et va considérer que toutes les publications ont la même valeur quel que soit leur impact. Pour être exhaustif, on prend tous les cas de figure et ainsi l'individu B préférera les articles de faible impact aux articles de fort impact. L'individu C ne prendra en compte les publications qu'à partir d'un certain niveau d'impact et alors elles se vaudront toutes, tandis que l'individu D est celui qui préférera trois articles d'impact 2 à deux articles d'impact 3. L'individu E quant à lui, a une vision



linéaire des choses. Trois articles d'impact 2 ou deux d'impact 3 c'est exactement la même chose, en fait il additionnera les citations pour valoriser l'impact. Enfin, l'individu F préférera deux articles d'impact 3 à trois articles d'impact 2, valorisant ainsi l'excellence et le placement d'articles dans les niveaux de visibilité les plus élevés. Lorsque l'on discute avec les responsables d'établissement ou de laboratoire, leur position varie entre la vision des individus D, E et F, et la plupart d'entre eux ont une position proche de celle de l'individu F qui valorise l'excellence.

On a donc ici des personnalités dont les visions sont toutes différentes. Le problème consiste à s'accorder sur une manière de voir les choses au sein d'une assemblée. Dans quelle mesure ses membres peuvent-ils porter des jugements unanimes étant donné qu'ils s'entendent sur certaines propriétés de la fonction de valorisation implicite ? L'objectif est de dégager une unanimité des jugements au sein d'une communauté partageant certaines valeurs communes.

Par exemple (P.6 du diaporama associé), si tous les membres s'accordent uniquement sur le fait que la fonction de valorisation soit positive, alors ils partagent un focus sur le volume de publication. S'ils s'accordent sur le fait que la fonction soit non seulement positive mais aussi croissante, cela traduit une orientation partagée plus marquée vers la qualité des publications. Si en outre, ils exigent que cette fonction soit convexe alors cela traduit une orientation partagée vers l'excellence.

On a ainsi des propriétés de fonctions qui rendent compte de focus plus ou moins orientés vers l'excellence. Le travail de recherche fondamentale a permis de transcrire ces conditions d'unanimité assez générales en quelque chose de calculable sur la seule base des données disponibles. Ainsi, nous pouvons établir des comparaisons en rendant explicites les principes d'agrégation auparavant implicites.

Ce que l'on nomme un « système de valeur partagé », c'est bien un principe d'agrégation mais pas seulement, car c'est aussi de savoir répondre à une série d'autres questions, par exemple :

- Quel focus disciplinaire (une seule discipline, une spécialité, ou au niveau interdisciplinaire) ;
- Comment mesure-t-on l'impact des articles (citations directes, facteur d'impact – l'aptitude des auteurs à placer leur article dans les bonnes revues, facteur d'impact relatif – l'aptitude des auteurs à placer leur article dans les bonnes revues de la spécialité) ;
- Quel degré de sélectivité vis-à-vis des articles pris en compte en fonction de l'impact (100 % ou les 10 % meilleurs, voire le 1 %, etc.).

Pris ensemble, les différents choix définissent un système de valeur de référence explicite au sein duquel peuvent prendre place des comparaisons aux fondements clairs.

Prenons l'exemple dans des universités américaines, on se concentre sur la biologie fondamentale, à partir de l'étude de la production scientifique WoS de 2003 à 2005 et des citations de 2003 à 2005 des cent principales universités de recherche aux Etats-Unis. Trois formes de restitution des comparaisons sont possibles. On peut faire :

- un positionnement relatif des établissements dans des réseaux orientés de comparaison ;
- un positionnement global de ces établissements en produisant des classements ;
- un positionnement associatif par les classes de référence (listes de paires).

Par exemple en se concentrant sur les douze premières universités avec un système de référence dont le focus est la « qualité », le choix d'un impact mesuré par les citations et une restriction au



10% des publications les plus citées, on voit distinctement (P.9 du diaporama associé), une hiérarchie dans laquelle la flèche veut dire « fait mieux que ». Harvard est au sommet de la hiérarchie et il y a deux établissements (UCSF et Stanford) qui font moins bien que Harvard mais ne peuvent pas être comparés entre eux etc... On obtient ainsi un réseau de relations entre les établissements qui leur permet de savoir comment ils se positionnent exactement.

On peut faire varier le système de valeur ce qui va permettre d'établir certaines relations qui n'avaient pu être mises en évidence dans le focus précédent. Par exemple (P.10 du diaporama associé), en prenant un focus « d'excellence » sur toutes les publications, on clarifie la situation de l'université de Yale par rapport à l'UCDS et à l'U Penn dans le focus précédent.

Si l'on prend le même focus mais en privilégiant la capacité à placer les articles dans les meilleurs journaux (facteur d'impact), (P.12 du diaporama associé), on voit que la hiérarchie bouge et par exemple que Berkeley qui était tout en bas dans le focus précédent remonte à la hauteur de Yale.

On peut procéder à une deuxième forme de restitution des résultats de ces comparaisons lorsqu'il n'y a pas de lien entre deux établissements. On peut en effet conclure qu'ils sont des « pairs », et par conséquent qu'ils sont mutuellement dans la classe de référence de l'autre. Par exemple ici on peut dire que Berkeley est dans la classe de référence de Yale (et vice-versa).

Le positionnement des universités varie selon le focus adopté et le mode de mesure de l'impact (qualité / excellence avec impact mesuré par les citations / excellence avec impact mesuré par le facteur d'impact). La hiérarchie et les relations entre établissements diffèrent selon l'approche retenue.

Il est également possible de mener une analyse globale du réseau d'établissements en comptant le nombre d'établissements qui sont en aval de l'établissement concerné dans le réseau de relations et sur chacun des focus volume, qualité et excellence, pour établir le rang de chacun (P.13 du diaporama associé). Ici, Harvard se classe en haut de la hiérarchie pour les trois critères, mais d'autres établissements occupent une place différente dans le classement en fonction du focus retenu. Ainsi UCLA qui est juste derrière Harvard en volume perd des places lorsque l'on va vers le focus qualité et Stanford a lui un mouvement inverse. Classé quinzième en volume, il se place au second rang dans un focus excellence.

Voilà donc un exemple neutre, mais il reste pas mal de chemin à parcourir sur ce projet. Comme tout projet de recherche fondamentale, il est encore assez loin de pouvoir être utilisé par les usagers. Si nous avons en effet développé un outil permettant de construire des comparaisons entre établissements sur des fondements explicites définis en fonction d'objectifs analytiques, de nombreux ajustements restent néanmoins à réaliser pour pouvoir offrir à la communauté un tel outil. Nous souhaitons construire un pilote pour affiner plusieurs points avec un groupe de travail d'experts (chercheurs, enseignants-chercheurs et équivalents et/ou intervenants en exécutif recherche dans des établissements à dominante recherche, mais aussi dans des établissements plus diversifiés et/ou de taille intermédiaire).



Q : Le nombre de citations par établissement dépendant du nombre de chercheurs, avez-vous prévu une normalisation de l'Etp recherche au niveau de l'établissement ?

R : Nous nous concentrons sur l'analyse de la production sans prendre en compte le nombre d'ETP. Cette analyse peut sembler alors guidée essentiellement par le volume de l'établissement, dans laquelle les grands établissements seraient forcément mieux classés. Mais comme nous travaillons sur la qualité, des établissements de petite taille mais excellents se positionnent très bien.

Q : Le critère du nombre de citations, pris isolément, ne reflète pas forcément la qualité de l'article. Ainsi, un article d'histoire du moyen-âge a été abondamment cité pour les multiples erreurs qu'il contenait et le scandale qu'il constituait. Prendre cette valeur est-il du coup représentatif ?

R : Les sociologues se posent cette question depuis longtemps. Les citations sont généralement considérées comme une approximation acceptable de l'impact d'un travail de recherche. Certes, les profils de citations varient en fonction du type d'article et de disciplines et le sujet est complexe. Des corrections existent et nous en utilisons. Pour ce qui est de la variation des profils de citations entre articles de la même spécialité, on peut s'appuyer sur le facteur d'impact, puisque ce dernier prend en compte la qualité de la revue dans laquelle l'article est publié et par conséquent laisse moins de place au bruit.