

Les limites du classement de Shanghai

Le premier « **classement académique international des universités** » a été publié en 2003 sur le site de l'université Jiao Tong de Shanghai. Ce *classement de Shanghai* a connu un succès médiatique qui se confirme chaque année lors de son actualisation. Il est mis en avant par les universités pour justifier de leurs performances. Il est aussi utilisé par de nombreux décideurs politiques pour promouvoir des projets de réforme. Mais quel crédit faut-il accorder à ce classement ? Est-il un instrument pertinent et fiable permettant de répondre à la question de savoir quelle est la meilleure université au monde ? L'analyse proposée dans [Billaut et al, 2009] montre que ce classement présente de grandes faiblesses.

Comment fonctionne le classement de Shanghai ?

Chaque université est évaluée sur six critères.

- 1 ALU : nombre d'**anciens élèves** ayant reçu un prix Nobel (sauf Paix et Littérature) ou une médaille Fields,
- 2 STA : nombre de professeurs de l'institution ayant reçu un prix **Nobel** ou une médaille Fields,
- 3 HiCi : nombre de **chercheurs** les plus cités tel que calculé par *Thomson Scientific*,
- 4 N&S : nombre d'articles publiés dans les revues **Nature et Science**,
- 5 PUB : nombre d'**articles** indexés par *Thomson Scientific*,
- 6 PROD : score total des cinq premiers indicateurs divisé par la taille du **corps professoral (équivalent temps plein)**.

Pour chaque critère, on normalise les évaluations de sorte que la meilleure université sur le critère ait un score de 100. Pour chaque institution, on calcule alors une somme pondérée de ces six scores normalisés. Les poids attribués aux critères sont égaux à 20 %, sauf pour ALU et PROD qui ont un poids de 10 %. Ce score est à nouveau normalisé. Les universités sont classées sur la base du score global ainsi obtenu. Une description complète du classement est donnée dans [Liu et Cheng, 2005]. Les données sont collectées sur Internet.

En 2008, parmi les 20 meilleures universités mondiales, on trouve 17 universités américaines, 2 britanniques et une japonaise.

Pourquoi les résultats du classement ne sont pas pertinents

Un classement pourquoi et pour qui ?

Un défaut important du classement de Shanghai vient du fait que le problème est mal posé. Plus précisément, ce classement apporte une réponse à un problème qui n'a jamais été posé par qui que ce soit. Dès lors, tout est sujet à questionnement et pour commencer, que cherche-t-on à évaluer ? pour qui ? dans quel but ?

Quelle est la définition d'une université, c'est-à-dire, quels sont les objets évalués par le classement ? La réponse n'est pas toujours simple. Par exemple, en France, se côtoient des universités publiques, des grandes écoles et des instituts de recherche. Ces institutions ne fonctionnent pas de la même façon, elles n'ont pas les mêmes objectifs et ne sont pas soumises aux mêmes contraintes en termes de budget ou de gouvernance. Ce qui doit, dans un tel cadre, compter comme une « institution universitaire », n'est pas évident. Cela a, par exemple, conduit les auteurs du classement à admettre de 2003 à 2005 que le Collège de France était une institution universitaire. Mais **le Collège de France n'a aucun étudiant et ne délivre aucun diplôme...**

À quel public ce classement est-il destiné ? Est-ce aux étudiants pour les aider dans leur mobilité ? Est-ce aux recruteurs de cadres de haut niveau ? Est-ce aux présidents d'université, pour les aider à identifier leurs forces et leurs faiblesses ou encore pour les aider à attirer les meilleurs chercheurs ? Est-ce aux décideurs politiques pour juger de l'efficacité d'un système d'enseignement supérieur ? La manière de concevoir un classement doit dépendre du public et des usages visés. Un classement « généraliste » destinés à tous et pour tous types d'usages n'a pas de sens.

Des critères discutables

Nous avons analysé en détail chacun des critères utilisés dans le classement. Notre conclusion principale est que les critères ont été choisis principalement sur la base de la disponibilité sur Internet des informations permettant de les renseigner, que chacun d'entre eux est lié de façon très approximative avec ce qu'il est censé mesurer et que leur évaluation fait intervenir des paramètres arbitraires et des microdécisions non documentées. L'impact de ces éléments sur le résultat final n'est pas examiné. Les données initiales utilisées ne sont pas rendues publiques et donc ne peuvent pas être vérifiées.

Illustrons ceci sur la base du seul critère STA, nombre de professeurs de l'institution ayant reçu un prix Nobel ou une médaille Fields. Observons en premier lieu que ce critère ne prend pas en compte l'ensemble des récompenses scientifiques prestigieuses (on peut penser à la **médaille Bruce en astronomie** ou encore aux **prix Turing en informatique**). Remarquons ensuite que, pour les prix Nobel, une longue période s'écoule généralement entre le moment où la découverte a été faite et le moment où elle est récompensée. Il est fréquent que le récipiendaire ne travaille plus dans l'institution où il travaillait quand il a fait sa découverte. Or le prix est attribué à l'institution où travaille le récipiendaire au moment de l'annonce du prix. Par exemple, Albert **Einstein** a conduit sa recherche alors qu'il était employé par l'**office suisse des brevets à Zurich** mais a reçu le prix Nobel lorsqu'il était affilié à l'université de Berlin. Pire encore, suite à la partition de l'Allemagne, l'**Université de Berlin** n'existe plus. Il faut alors décider qui de ses deux continuatrices (l'Université Libre de Berlin et l'Université Humbolt) a droit à ce prix. Dans la plupart des pays d'Europe, les guerres et les changements politiques ont entraîné de très nombreux changements institutionnels. C'est particulièrement vrai en France où la plupart des universités ont été scindées en de plus petites unités après 1968. Ainsi, on trouve sur le site officiel des prix Nobel, des prix attribuées à des institutions françaises qui n'existent plus (l'Université de Toulouse, par exemple) ou même à des institutions n'ayant jamais existé (la Sorbonne University). Ainsi une affectation correcte de ces prix demande une connaissance très fine du paysage institutionnel de chaque pays. Notons enfin que les prix Nobel sont attribués depuis 1901. Il faut alors prendre position sur le poids relatif attribué au prix Nobel d'Albert Fert (2007) et à celui de Louis de Broglie (1929). Le critère STA utilise à cet effet une procédure d'actualisation visant à donner un poids plus grand aux prix récents qu'aux prix anciens. Mais on pourrait envisager pour ce faire bien d'autres techniques toutes aussi raisonnables...

Une méthode doublement fautive

La manière dont les auteurs du classement agrègent l'information contenue dans les six critères est fautive car la position dans le classement de deux institutions ne dépend pas seulement de leurs performances. Il s'agit ici d'un problème lié au fait que les poids adoptés dans une somme pondérée sont liés à la normalisation des échelles. Si la normalisation des échelles change, on doit changer les poids. Si on ne le fait pas, on obtient des résultats *absurdes*. Illustrons ceci par un simple exemple numérique. Considérons quatre universités évaluées ici sur deux critères supposés de poids égaux. Avec la technique utilisée dans le classement de Shanghai, on obtiendrait alors les résultats suivants :

| | N&S | PUB | N&S norm | PUB norm | Score | Rang |
|---|-----|-----|----------|----------|-------|------|
| A | 11 | 40 | 35,5 | 100 | 67,7 | 1 |
| B | 16 | 32 | 51,6 | 80 | 65,8 | 2 |
| C | 24 | 20 | 77,4 | 50 | 63,7 | 3 |
| D | 31 | 10 | 100 | 25 | 62,5 | 4 |

Supposons à présent que l'année suivante, tout reste inchangé, à l'exception de la performance de l'université A sur le critère PUB qui passe de 40 à 50, toutes les autres évaluations étant inchangées. On obtiendrait alors les résultats suivants :

| | N&S | PUB | N&S norm | PUB norm | Score | Rang |
|---|-----|-----------|----------|----------|-------|------|
| A | 11 | 50 | 35,5 | 100 | 67,7 | 1 |
| B | 16 | 32 | 51,6 | 64 | 57,8 | 4 |
| C | 24 | 20 | 77,4 | 40 | 58,7 | 3 |
| D | 31 | 10 | 100 | 20 | 60 | 2 |

Ces résultats sont clairement absurdes puisque le classement des universités B, C et D a été complètement inversé alors que leurs performances sont restées inchangées.

Ce phénomène se produit dès lors que l'on change la normalisation d'un critère sans changer corrélativement les poids. Or les poids ne varient pas dans le classement de Shanghai tandis que la normalisation change chaque année car la performance de la meilleure université sur chaque critère varie. Ceci est le premier défaut de la méthode adoptée.

Le second problème est lié au fait que le classement mélange des critères visant à appréhender la *production scientifique* (ALU, STA, HiCi, N&S et PUB) avec un critère (PROD) visant à appréhender la *productivité scientifique*. Le bon sens, ainsi qu'une analyse économique élémentaire suggère qu'une telle opération conduit à un indicateur qui n'a pas de sens. On s'en rendra mieux compte en envisageant une

méthode classant les pays du monde selon leur richesse et qui consisterait à faire une moyenne pondérée du PNB du pays avec son PNB par habitant. Ceci est évidemment absurde.

Le CHERPA-Network

Le classement de Shanghai est un exercice de style qui n'a été commandé par personne et dont le résultat est sans aucune valeur sur le plan scientifique : sur ce plan notre analyse rejoint et complète celle de [Van Raan, 2005]. Qu'il tombe à point nommé et qu'il ait une valeur sur un plan politique est une autre question et chacun se fera son opinion.

Terminons sur une note positive. Le CHERPA-Network est un consortium qui a récemment emporté un appel d'offres de la commission européenne pour l'élaboration d'un classement européen des universités. Nous lui souhaitons une grande réussite dans cette entreprise, en espérant que le résultat sera plus convaincant que celui du classement de Shanghai. Compte tenu de ce qui précède, ce ne devrait pas être une tâche insurmontable.

J-C. Billaut, D. Bouyssou, Ph. Vincke

Références

[Billaut et al, 2009] J.-C. Billaut, D. Bouyssou et Ph. Vincke. Should you believe in the Shangaï ranking? An MCDM view. Cahier du LAMSADE # 283, LAMSADE, 2009. Disponible à <http://hal.archives-ouvertes.fr/hal-00388319/en/>.

[Liu et Cheng, 2005] N. C. Liu et Y. Cheng. The academic ranking of world universities. Higher Education in Europe, 30(2):127-136, 2005.

[Van Raan, 2005] A. F. J. van Raan. Fatal attraction: Ranking of universities by bibliometric methods. Scientometrics, 62:133-145, 2005.

Jean-Charles Billaut  jean-charles.billaut@univ-tours.fr Tél. 02 47 36 14 42 Fax 02 47 36 14 22
 Laboratoire d'informatique de l'Université François-Rabelais
64 avenue Jean-Portalis 37200 TOURS

Daniel Bouyssou  bouyssou@lamsade.dauphine.fr Tél. 01 44 05 48 98 Fax 01 44 05 40 91
 CNRS - LAMSADE UMR 7024 & Université Paris-Dauphine
Place Maréchal-de-Lattre-de-Tassigny 75775 PARIS CEDEX 16

Philippe Vincke  pvincke@ulb.ac.be Tél. +32 2 650 32 01 Fax +32 2 650 36 30
 Université Libre de Bruxelles
50 avenue Franklin-Roosevelt CP. 130 B 1050 BRUXELLES