

Masse critique d'un groupe de recherche

Ralph Kenna⁽¹⁾ et Bertrand Berche⁽²⁾ (Bertrand.Berche@ijl.nancy-universite.fr)

(1) Applied Mathematics Research Center, Coventry University, Coventry CV1 5FB, UK.

(2) Institut Jean Lamour, Université de Nancy, BP 70239, 54706 Vandœuvre-les-Nancy Cedex

Nous basant sur les résultats de l'exercice britannique 2008 d'évaluation des établissements d'enseignement supérieur et de recherche, nous proposons un modèle empirique simplifié permettant de corrélérer la qualité d'un groupe de recherche à sa taille.

L'interprétation du modèle met en évidence le rôle essentiel joué par les interactions entre chercheurs dans la valeur supposée du groupe. Elle permet d'estimer une taille critique, dépendant de la discipline académique, au-dessus de laquelle un groupe est plus performant.

S'agissant d'une application de nature « sociologique », les auteurs attribuent au modèle empirique présenté une certaine valeur dans la description de propriétés moyennes, sans pour autant prétendre à une application à des situations extrêmes.

Existe-t-il une notion intuitive de masse critique d'un groupe de recherche ?

La notion de masse critique en recherche est assez mal définie, mais peut être comprise comme la valeur seuil du nombre de chercheurs dans un groupe au-dessus de laquelle certains standards de qualité et de productivité sont atteints. L'idée que le « bénéfice » augmente par augmentation du potentiel de chercheurs (la taille du groupe) est sous-jacente et pousse à concentrer les moyens vers un nombre limité d'institutions de tailles significatives (voir en France les Equipex, Labex, Idex). Pourtant, malgré un certain nombre d'analyses bibliométriques, on n'a pas pu mettre en évidence un tel seuil de manière rigoureuse à ce jour.

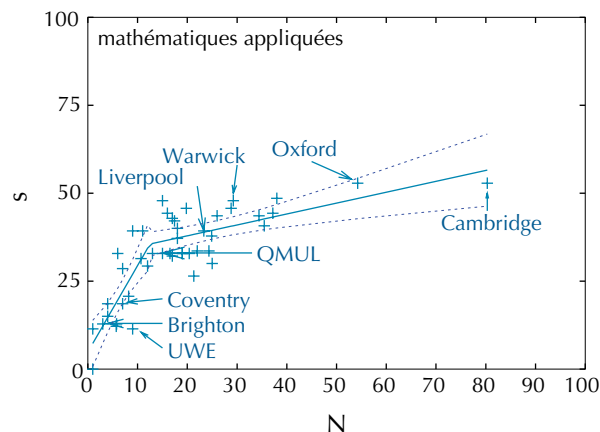
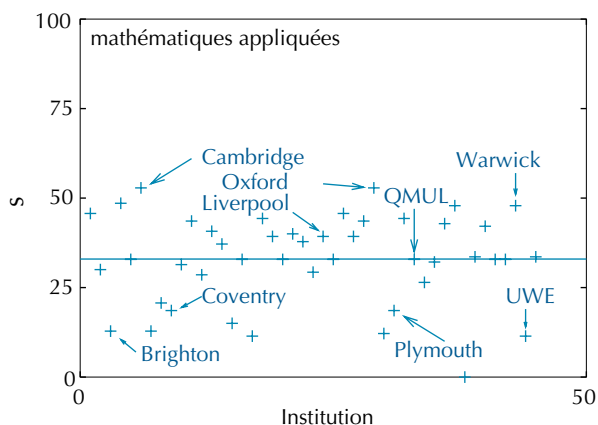
Nous avons récemment développé un modèle très simple [1] pour expliquer comment la qualité de la recherche produite dépend de la taille des groupes de recherche. Le modèle montre que l'effet dominant qui détermine la qualité de la recherche est l'interaction entre individus au sein du groupe et prédit en fait qu'il y a plutôt

deux masses critiques en recherche, dont les valeurs dépendent des disciplines [2,3,4]. La qualité (mesurée par chercheur) varie en moyenne linéairement avec le nombre de chercheurs dans le groupe jusqu'à une masse critique que l'on pourrait appeler *masse critique supérieure*, N_c , au-delà de laquelle la qualité sature pour devenir beaucoup moins dépendante du nombre d'individus. On peut interpréter la masse critique supérieure comme le nombre maximum moyen de collègues avec lesquels un chercheur peut interagir efficacement. Une fois atteinte cette valeur, le groupe se fragmente en sous-groupes qui interagissent également entre eux, mais avec une efficacité moindre.

Il existe également une *masse critique inférieure*, qui est la taille au-dessous de laquelle un groupe est difficilement viable sur le long terme. Cette seconde masse critique vaut $N_k = N_c / 2$, elle est définie en examinant le bénéfice collectif que la société (l'université, l'institut, etc.) peut avoir à engager un nouveau chercheur dans une équipe. L'étude révèle que le bénéfice est maximum entre les deux masses critiques, et non pas au-delà de la masse critique supérieure.

>>>

“ À la concentration des moyens dans un nombre limité d'instituts ou de groupes de grande taille, il est préférable de développer une politique de soutien aux « poches d'excellence », là où elles se trouvent. ”



1. La qualité par chercheur $s = \Sigma / N$ (estimée par le RAE) des groupes de recherche en mathématiques appliquées (incluant le plus souvent la physique théorique), reportée alphabétiquement (à gauche) et en fonction du nombre d'individus composant le groupe (à droite), dans le cas des universités britanniques. QMUL : Queen Mary University of London ; UWE : University of the West of England Bristol.

>>>

Un modèle simple, basé sur les interactions entre chercheurs

L'entité de base est le groupe de recherche académique, composé de chercheurs permanents et non permanents. Les données utilisées pour estimer la qualité des groupes sont issues de l'exercice d'évaluation britannique 2008 (RAE⁽¹⁾ 2008 [5,6]), et l'indicateur choisi pour mesurer la qualité est celui défini par le HEFCE⁽²⁾. Les résultats du RAE prennent la forme d'un profil pour chaque groupe, avec une estimation du pourcentage de l'activité de recherche classifiée, selon la proportion de travaux considérés d'un niveau d'excellence international, national ou sous national, en 5 niveaux : 4* (le plus haut niveau), 3*, 2*, 1* ou non classée [5]. Selon la formule post-RAE du HEFCE, les financements attribués à des activités de rang 4* et 3* sont respectivement 7 fois et 3 fois plus élevés que des activités jugées 2*, les deux autres catégories ne contribuant pas au budget reçu.

Le groupe est défini par l'université qui soumet un dossier à l'évaluation. Le périmètre peut évoluer entre celui du département disciplinaire et celui du laboratoire de recherche.

La figure 1 montre les scores de qualité par chercheur (notés s) en fonction de la taille des groupes pour les universités britanniques ayant déposé un dossier d'évaluation en *mathématiques appliquées*. La qualité maximale correspondant à un groupe dont la totalité de la production de recherche serait jugée de rang 4* serait de 100 sur cette échelle, mais bien entendu aucun

groupe n'atteint ce niveau, les meilleurs obtenant une performance environ moitié. Sur cette figure, l'existence d'une corrélation entre qualité s et taille N ne fait aucun doute. De plus, comme évoqué plus haut, deux régimes distincts sont apparents : on observe tout d'abord une croissance assez rapide de la qualité avec la taille pour les groupes les plus petits, puis une saturation pour les groupes les plus grands. Invoquant le principe de simplicité (le fameux rasoir d'Ockham⁽³⁾), il est naturel d'affiner les données par une première droite de pente assez élevée traduisant pour les groupes les plus petits une qualité totale $\Sigma = Ns$ qui varie quadratiquement avec la taille, c'est-à-dire comme le nombre d'interactions entre individus. Ceci est vrai jusqu'à la taille critique supérieure N_c , de l'ordre de 12 ici. Au-delà, un affinement avec une droite de pente réduite traduit la fragmentation du groupe en « équipes » encore susceptibles d'interactions entre elles, mais l'augmentation de qualité est toutefois freinée.

On peut noter que plus la valeur de N_c est élevée, plus la saturation est marquée, un effet expliqué très simplement par le fait que la pente est alors déterminée par l'intensité de l'interaction entre équipes divisée par N_c^2 . Les courbes en tirets représentent l'intervalle de confiance à 95% de l'affinement.

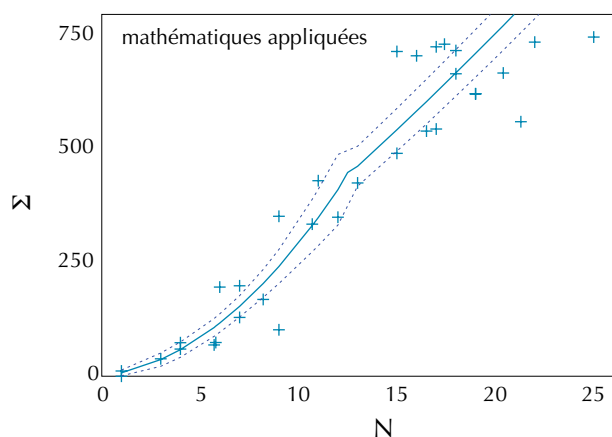
Une seconde masse critique peut être déterminée dès lors que l'on s'intéresse au bénéfice qu'il peut y avoir à augmenter la taille d'un groupe. En maximisant le gradient de la qualité totale (fig. 2), on constate que le bénéfice maximum est

atteint entre $N_k = N_c / 2$ et N_c , ce qui définit la taille critique inférieure au-dessous de laquelle les groupes sont vulnérables. Le modèle conduit donc à l'identification de trois types de groupes, « petits », « moyens » et « gros », et à privilégier le soutien en termes d'effectif aux groupes moyens pour les amener à la taille critique supérieure. Les groupes « petits » ont tout intérêt à atteindre la taille critique inférieure s'ils veulent être viables sur le long terme.

Quelques enseignements, quelques évidences

Le modèle est issu d'une observation empirique, mais plusieurs faits sont à souligner :

1. Les données semblent crédibles car l'évaluation par le RAE n'est pas contestée, les données sont bien indépendantes (elles passent les tests statistiques de normalité pour la plupart des disciplines académiques [2]) et les conclusions générales sont robustes en ce sens que même en modifiant la « formule » du HEFCE pour prendre celle appliquée en 2010⁽⁴⁾, le schéma proposé est inchangé.
2. L'affinement par deux droites a le double avantage de la simplicité et d'admettre une interprétation simple en termes d'interactions entre chercheurs.
3. De nombreuses autres disciplines académiques sont décrites de manière aussi satisfaisante par ce même modèle (cf. fig. 3 et références), et les valeurs de N_k qui en résultent ont du sens au regard des pratiques professionnelles (taille typique des collaborations selon les disciplines, nombre typique d'auteurs



2. La qualité Σ des groupes de recherche en mathématiques appliquées, en fonction du nombre d'individus composant le groupe.

sur les publications, etc). La masse critique N_k est inférieure à 2 pour les mathématiques pures [7] et elle est plus élevée pour des disciplines plus « collaboratives » (13 pour la physique expérimentale [8], 18 pour la chimie, et 7 pour la sociologie, par exemple).

4. La campagne 2008 d'évaluation des laboratoires français par l'AERES⁽⁵⁾ a également été utilisée en comparaison à l'évaluation britannique, et il en ressort que les deux systèmes produisent des résultats similaires, avec davantage de dispersion pour les résultats de l'AERES.

5. Enfin, le modèle s'impose avec une certaine évidence aux acteurs de la recherche – en tout cas en physique et dans les disciplines connexes – en ce sens qu'il ne fait que mettre en évidence des notions dont les chercheurs ont l'intuition.

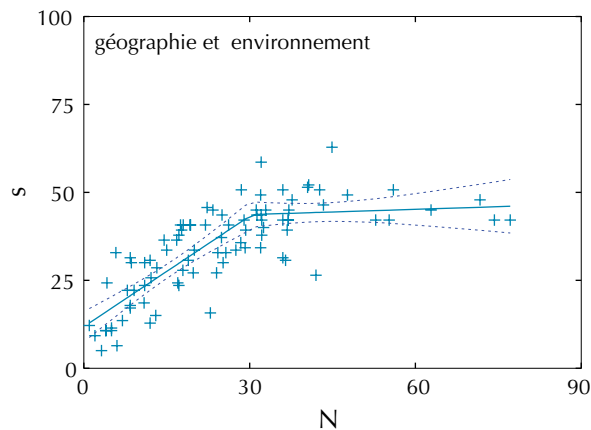
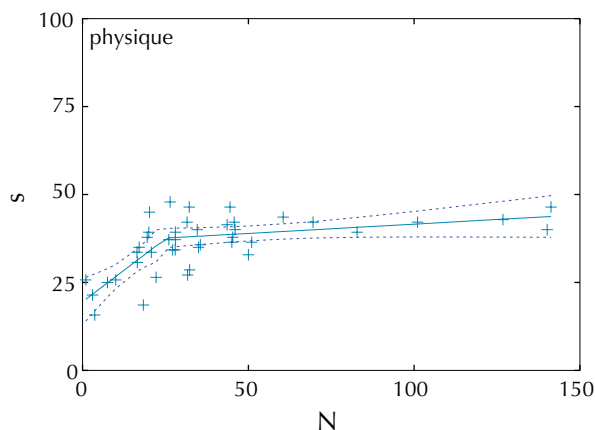
Le résultat principal issu de ce modèle est probablement que l'effet dominant qui gouverne la qualité des groupes de recherche académiques est l'interaction entre chercheurs. Les atouts tels que la stature individuelle, le prestige de l'université, ses équipements collectifs ou autres éléments similaires ne sont que des facteurs intervenant à un degré moindre. Que ce soit une évidence pour les chercheurs ne fait probablement pas de doute, la question demeure cependant d'en convaincre les « managers ».

Le modèle montre également qu'à la concentration des moyens dans un nombre limité d'instituts ou de groupes de grande taille (Iindex, Labex), il est préférable de développer une politique de soutien aux « poches d'excellence », là où elles se trouvent. ■

En savoir plus

- 1• R. Kenna et B. Berche, "The extensive nature of group quality", *EPL* **90** (2010) 58002.
- 2• R. Kenna et B. Berche, "Critical mass and the dependency of research quality on group size", *Scientometrics* **86** (2011) 527-540.
- 3• R. Kenna et B. Berche, "Normalization of research evaluation results across academic disciplines", *Research Evaluation* **20** (2011) 107-116.
- 4• R. Kenna et B. Berche, "Critical mass of education research groups", *Research Intelligence* **114** (2011) 14.
- 5• J. Enderby, « L'exercice d'évaluation de la recherche (le RAE) au Royaume-Uni », *Reflète de la physique* **9** (2008) 19-22.
- 6• Le site du RAE 2008 : www.rae.ac.uk/
- 7• R. Kenna et B. Berche, "Critical masses of research groups in the mathematical sciences", *Mathematics Today* **47** (2011) 114-116.
- 8• "Why 13 and 25 are magic numbers for physicists", <http://physicsworld.com/cws/article/news/46238>

- (1) Research Assessment Exercise.
- (2) Higher Education Funding Council for England.
- (3) Le rasoir d'Ockham est un principe de raisonnement que l'on attribue à Guillaume d'Ockham (XIV^e siècle), mais qui était connu et formulé avant lui : « Les multiples ne doivent pas être utilisés sans nécessité » (« *pluralitas non est ponenda sine necessitate* »). L'énoncé « *Entia non sunt multiplicanda praeter necessitatem* », littéralement « Les entités ne doivent pas être multipliées par-delà ce qui est nécessaire », est une variante souvent attribuée à Guillaume d'Ockham sans cependant qu'il y ait trace dans ses écrits. Une formulation plus moderne est que « les hypothèses les plus simples sont les plus vraisemblables ». C'est un des principes fondamentaux de la science (source : Wikipedia).
- (4) Dans la « formule 2010 », le poids des activités évaluées 4* vaut 9 fois plus – et non plus 7 fois plus – que celles évaluées 2*.
- (5) Agence d'évaluation de la recherche et de l'enseignement supérieur : www.aeres-evaluation.fr.



3. La qualité par chercheur reportée en fonction du nombre de membres du groupe pour la physique (essentiellement expérimentale) et la géographie, dans le cas des universités britanniques.