



Rechercher

ok

Quelles sont les personnalités historiques les plus influentes sur Wikipédia ?

[précédent](#)

5 mars 2015

[LPT - UMR 5152](#)

Carl Von Linné, Jésus Christ, Aristote, Napoléon, Adolf Hitler, Jules César, Platon, William Shakespeare et Albert Einstein. Voici l'ordre d'importance des personnalités historiques donné par l'algorithme de PageRank appliqué à Wikipédia. C'est ce que viennent de proposer deux physiciens du Laboratoire de physique théorique (CNRS/Université de Toulouse III – Paul Sabatier), qui ont adapté des méthodes d'analyse des réseaux complexes à celui formé par les hyperliens entre les pages de Wikipédia, et ce, en 24 langues.



Ces travaux, qui permettent de mesurer l'influence culturelle à l'échelle locale et mondiale des personnalités de l'encyclopédie, sont publiés le 4 mars 2015 dans la revue *Plos One*.

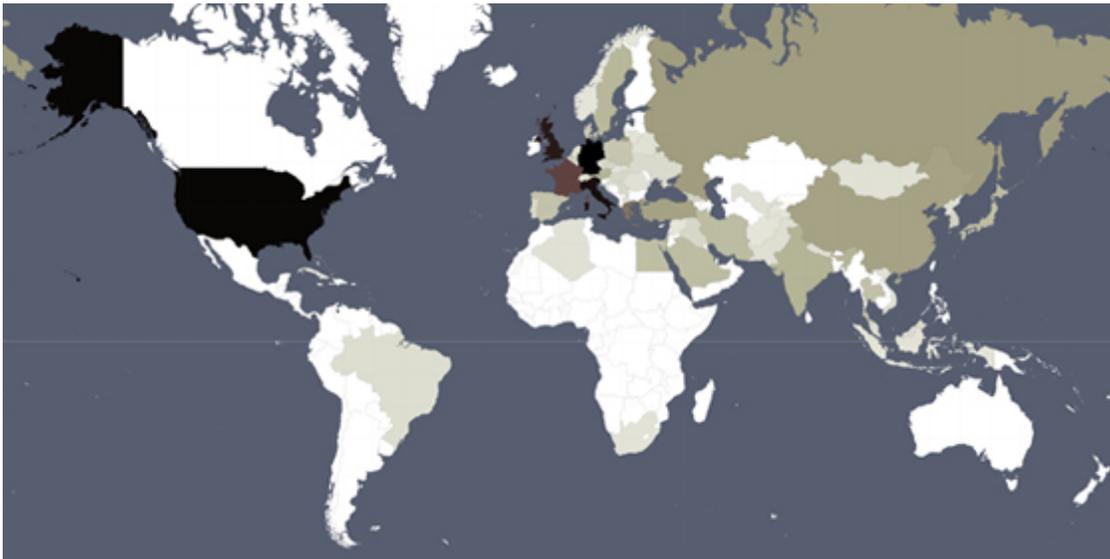
Le réseau complexe constitué par les pages de l'encyclopédie en ligne *Wikipédia*, et les nombreux (hyper)liens internes entre elles (liens entrants et sortants), représentent une photographie extraordinaire de la connaissance humaine, qui peut être exploitée afin d'identifier l'influence de personnalités, et l'évolution de l'impact et des interactions entre diverses cultures au cours de l'histoire. L'équipe interdisciplinaire et internationale de chercheurs impliqués dans cette étude, incluant deux physiciens du Laboratoire de physique théorique de Toulouse, a ainsi classé les articles biographiques de *Wikipédia* par leur « importance/influence » au moyen des algorithmes *PageRank* de *Google*, *CheiRank*, et *2DRank* (ces deux derniers, largement développés au LPT), bien adaptés à l'étude des réseaux complexes (voir l'encadré).

La liste globale obtenue des 100 personnes les plus influentes reproduit 43 % de [la liste historique de Hart](#), et peut être consultée (ainsi que pour chacune des listes associées aux 24 éditions locales de *Wikipédia* étudiées) sur la [page web du LPT dédiée à cette étude](#), avec des cartes interactives illustrant la distribution de ces personnes dans le monde. Par exemple, pour *Wikipédia* en français, le trio de tête des personnalités françaises les plus influentes sont : Napoléon, Louis XIV, René Descartes (algorithme *PageRank*) ; Napoléon, Jules Verne, Louis XIV (*2DRank*). Pour *Wikipédia* en anglais, les trois personnalités de langues anglaises les plus influentes sont William Shakespeare, Elizabeth II, Isaac Newton

(*PageRank*), mais Napoléon domine le classement complet en langue anglaise ! Globalement, les huit personnalités les plus influentes selon le *PageRank* sont Carl von Linné, Jésus Christ, Aristote, Napoléon, Adolf Hitler, Jules César, Platon, William Shakespeare, et Albert Einstein. La 1^{ère} position *a priori* étonnante du naturaliste suédois Carl von Linné (respectivement classé 2^{ème} et 3^{ème} par le *PageRank* dans la version française et anglaise) tient à son rôle de fondateur des bases du système moderne de la nomenclature binominale (et du classement) des espèces vivantes, et au très grand nombre d'articles de *Wikipédia* consacrées à celles-ci.

Les chercheurs ont aussi analysé la distribution géographique (voir la figure), temporelle, homme/femme, de ces personnalités, en lien avec leur culture d'origine. Cette étude établit non seulement l'existence d'un biais naturel favorisant les personnalités de la culture locale (associée à une langue donnée de *Wikipédia*), mais révèle aussi un groupe de personnalités ayant une influence *véritablement globale*. Cette étude considère des personnes balayant une période de 35 siècles de l'histoire humaine, d'origine et de culture extraordinairement variées, ce qui a permis de reconstruire les interactions et « l'intrication » de ces cultures au cours du temps, et de quantifier objectivement leur degré d'influence.

L'étude des réseaux complexes est un domaine de recherche extrêmement actif et intrinsèquement interdisciplinaire, impliquant des chercheurs en physique, mathématiques, informatique, biologie, sciences humaines et sociales... Les réseaux considérés peuvent être issus de sociétés humaines (Internet ; web – comme ici, avec *Wikipédia* ; transport ; échanges commerciaux...), animales (réseau des galeries d'une termitière ; réseau d'interaction entre individus...), ou émergent dans d'autres sciences (réseau d'interaction de protéines ; réseau des neurones du cerveau ; propagation d'épidémies via les réseaux de transport aérien et/ou de migration aviaire ; réseau d'interaction des routines informatiques de Linux...). Les méthodes développées par les physiciens du LPT peuvent en fait s'appliquer (et l'ont été dans certains exemples cités ci-dessus) à des réseaux apparaissant dans ces contextes très divers. Elles permettent ainsi d'identifier et caractériser des communautés en interaction (sans que les acteurs soient nécessairement conscients de l'existence de leur communauté !), et de mesurer et classer les degrés d'influence des différents nœuds/acteurs du réseau.



Carte de l'influence globale selon l'algorithme PageRank (la couleur du pays est d'autant plus sombre que son degré d'influence est important).
De nombreuses cartes interactives générées par les différents algorithmes sont consultables sur la [page web du groupe Cohérence Quantique du Laboratoire de physique théorique \(LPT\) dédiée à ce travail](#).



« Les algorithmes de classement (Rank) » : les trois algorithmes PageRank, CheiRank, et 2DRank mentionnés dans le texte permettent de quantifier l'influence d'un nœud d'un réseau complexe quelconque. Partant initialement d'un nœud arbitraire, on se déplace ensuite aléatoirement sur le réseau (d'où le lien formel avec la physique d'une particule quantique dans un milieu/réseau désordonné) en explorant ses liens au hasard, à chaque pas de temps, tout en s'autorisant aussi parfois à sauter sur n'importe quel autre nœud du réseau, afin de ne pas rester bloqué sur un nœud sans lien de sortie. Le degré d'influence d'un nœud est alors défini comme étant le pourcentage du temps passé sur ce nœud, qui converge après un très grand nombre de sauts. Dans le contexte de Wikipédia, cela correspondrait à un internaute curieux (et patient ! Il existe environ 1,5

million de pages, rien que sur l'édition française) qui suivrait un lien sortant au hasard par page (et ceci, plusieurs milliards de fois), et appuierait avec une faible probabilité sur le bouton « Article au hasard » offert par l'encyclopédie. Au final, le degré d'influence d'un nœud du réseau n'est pas simplement conditionné par son nombre de liens sortants ou entrants, mais aussi par l'influence des nœuds auxquels il est connecté, et plus généralement, par la structure globale du réseau. Les trois algorithmes se distinguent par le fait que les sauts autorisés entre nœuds ne peuvent se faire que dans la direction des liens sortants (CheiRank) du réseau étudié, ou au contraire, en remontant uniquement les liens entrants (PageRank), ou encore, en combinant les informations fournies par les deux algorithmes (2DRank).

En savoir plus

Interactions of cultures and top people of Wikipedia from ranking of 24 language editions

Y.-H. Eom¹, P. Aragon², D. Laniado², A. Kaltenbrunner², S. Vigna³, D. L. Shepelyansky¹
PLOS One (2015)

- Retrouvez l'article sur la [base d'archives ouvertes arXiv](#)

Contact chercheur

[Clément Sire](#), directeur de recherche au CNRS, directeur du LPT
[Dima Shepelyansky](#), directeur de recherche CNRS

Informations complémentaires

¹ [Laboratoire de physique théorique \(LPT\)](#)

² [Barcelona Media Foundation](#), Barcelona

³ [Dipartimento di Informatica](#), Università degli Studi di Milano, Italy

Contacts INP

Jean-Michel Courty,
Catherine Dematteis,
Simon Jumel,
inp.com@cnrs.fr