

## DANS L'ESPRIT DU PAGERANK

### Une enquête sur l'algorithme de Google

**Dominique Cardon**

**La Découverte** | *Réseaux*

2013/1 - n° 177  
pages 63 à 95

ISSN 0751-7971

Article disponible en ligne à l'adresse:

-----  
<http://www.cairn.info/revue-reseaux-2013-1-page-63.htm>  
-----

Pour citer cet article :

-----  
Cardon Dominique, « Dans l'esprit du PageRank » Une enquête sur l'algorithme de Google,  
*Réseaux*, 2013/1 n° 177, p. 63-95. DOI : 10.3917/res.177.0063  
-----

Distribution électronique Cairn.info pour La Découverte.

© La Découverte. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

# DANS L'ESPRIT DU PAGERANK

Une enquête sur l'algorithme de Google

Dominique CARDON

**A** quoi rêve Google<sup>1</sup>? Comment l'entreprise qui symbolise le mieux la forme la plus avancée du capitalisme cognitif se représente-t-elle le web et à quoi voudrait-elle qu'il ressemble ? Quelle portée faut-il donner aux propos de jeunes entrepreneurs intrépides qui se présentent devant les marchés financiers en déclarant vouloir « *faire du monde un meilleur endroit* » ? L'hypothèse que l'on soutient dans cet article est que pour répondre à cette question, il est nécessaire d'entrer intimement dans l'architecture calculatoire du PageRank, l'algorithme qui a fait la fortune de Google et constitue son joyau le plus précieux. Le PageRank est une machine morale. Il enferme un système de valeurs, donnant la prééminence à ceux qui ont été jugés méritants par les autres, et déployant une volonté : faire du web un espace où l'échange des mérites n'est ni freiné ni déformé. L'esprit qui anime le PageRank se trouve cependant si profondément mélangé à tant d'autres considérations, d'autres intérêts et d'autres sujets d'inquiétude que son écho est presque devenu inaudible. Les débats entourant l'entreprise Google ont étouffé la voix du PageRank. L'objectif de ce travail est de comprendre ce que Google a fait au web et du web en réanimant les principes qui nourrissent son algorithme et commandent beaucoup des comportements, des décisions et des choix stratégiques de la firme de Mountain View. C'est en explorant les procédures de l'algorithme, son mode de fonctionnement, ses décisions, ses interdits, et tout l'appareil de codification du comportement des internautes qu'il a fabriqué, que l'on peut saisir la manière dont Google a imposé au web un ordre qui lui est propre. C'est aussi comprendre comment, sous l'effet des transformations récentes du web, d'autres principes de mise en ordre de l'information sont venus fragiliser l'esprit du PageRank.

## L'INVENTION DU PAGERANK

Avant Google, le web était une vaste loterie. Les réponses aux requêtes des internautes étaient hasardeuses, souvent fantaisistes, massivement truquées et

---

1. Ce travail a été réalisé dans le cadre du projet ANR « Politique des algorithmes » (ALGO-POL - NR 2012 CORD 01804).

occasionnellement pertinentes. Les premiers moteurs de recherche fonctionnaient à partir de mots clés et mesuraient la densité de la présence du terme recherché sur les différentes pages du web. En 1990, les pionniers, Archie et Veronica, n'indexaient que le titre du document, avant qu'en 1994, le Web-Crawler de Brain Pinkerton ne prenne en compte l'ensemble du texte de la page. L'innovant moteur d'AltaVista conçu par Louis Monnier pour DEC fut le premier à se donner pour ambition d'indexer l'ensemble du web. Lors de son ouverture au public, en décembre 1995, il piochait moins maladroitement que les autres dans un index de 16 millions de documents (Batelle, 2005, p. 40). Jusqu'à ce qu'un jeune chercheur de Cornell University, Jon Kleinberg, ne le propose à IBM en 1996, personne n'avait vraiment pensé à prêter attention à la structure des liens hypertextes plutôt qu'à l'analyse sémantique des pages. En 1998, ils étaient pourtant nombreux qui travaillaient à une solution efficace pour améliorer les moteurs de recherche dont la qualité se détériorait toujours plus avec la croissance du nombre de pages. Il était alors très facile de tromper les algorithmes lexicaux. Il suffisait que les webmestres copient de multiples fois les termes les plus recherchés en blanc sur fond blanc pour améliorer la visibilité de leur site. Pour lever le verrou que constituait l'impasse de la recherche lexicale, une révolution dans la conception de l'algorithme s'avérait nécessaire. Celle-ci sera portée sur le marché par deux étudiants de Stanford, Sergey Brin et Larry Page (1998). Mais l'intuition qui aura présidé à cette rupture venait de loin. Elle trouve sa source dans l'esprit qui a nourri la conception du réseau des réseaux : tirer parti de la structure relationnelle des pages, qui se tiennent entre elles par des liens hypertextes, pour en extraire un ordre qui capitalise le sens des interactions qu'ont entre eux les internautes qui publient sur le web.

### Sociométrie et scientométrie

Fournissant tous les codes possibles à la fabrication du mythe de la Silicon Valley, l'histoire de la création de Google n'est plus à faire (Levy, 2011). Aussi ne s'intéresse-t-on ici qu'à la manière dont la conception du PageRank a été durablement attachée à une représentation particulière d'Internet qui exerce un effet structurant sur l'écosystème que forment désormais le web et son moteur dominant. L'intuition fondatrice dont Google est l'enfant est de considérer le lien *entrant* (et non *sortant*, comme le faisait alors Lycos) comme support de toute opération visant à classer les sites de l'Internet. Mais le principe qui sous-tend cette démarche n'est pas nouveau et s'enracine dans

deux traditions différentes : celle de la sociométrie, qui va réunir psychologie et mathématique des graphes autour des propriétés de la forme réseau et celle de la scientométrie qui va porter les savoirs de la bibliothéconomie vers l'évaluation de l'activité scientifique. Même si les échanges entre ces deux traditions n'ont pas été très nombreux, celles-ci convergent au moins sur une question qui sera décisive dans l'élaboration du PageRank : définir des métriques destinées à décrire les formes *relationnelles* du social. Que ce soit par le truchement de l'influence en sociométrie ou de la citation en scientométrie, un déplacement s'opère pour ne pas faire porter l'analyse sur des objets fixes et autosuffisants, qu'il s'agisse d'acteurs sociaux ou de documents, mais sur les relations qu'ils entretiennent les uns avec les autres.

L'idée d'utiliser le lien de citation pour définir le classement de l'information remonte d'abord à la révolution sociométrique de Moreno qui, dans les années 1930, voulait décrire la structure de la société à partir des liens entre individus, plutôt qu'à partir des catégories servant à identifier et différencier les personnes (Mayer, 2009). En demandant aux individus de désigner autour d'eux les personnes qu'ils aimaient (admiraient, avec lesquelles ils avaient le plus de contacts, etc.) le plus et celles qu'ils aimaient (admiraient, etc.) le moins, la psychologie de groupe de Moreno introduisait à la fois l'idée de représenter le réseau social des individus sous la forme d'un *sociogramme* et celle de l'organiser en fonction d'un principe d'attractivité et de répulsion qui témoignerait de l'influence réciproque que les uns exerçaient sur les autres. Comme l'indique Bernhard Rieder (2012) dans son enquête sur la généalogie computationnelle du PageRank, si la sociométrie de Moreno n'a pas connu le passage par la mathématisation, d'autres s'en sont emparé, notamment Elaine Forsyth et Leo Katz (1946), pour développer un ensemble de matrices et de calculs qui contribueront à la naissance d'une mathématique sociale dans la théorie des graphes naissante.

Mais le PageRank s'inscrit surtout dans une autre tradition de recherche promise à un très large succès, celle de la scientométrie, qui allait prendre forme avec le *Science Citation Index* (SCI), fondé en 1964 par Eugène Garfield au sein de l'Institut de l'information scientifique<sup>2</sup>. Son projet était de faciliter la circulation à l'intérieur des contenus scientifiques en favorisant le déplacement de citation en citation entre les articles scientifiques. Consultant indépendant dans le domaine de la documentation, Eugène Garfield devrait lancer

2. Voir l'article de D. Pontille et D. Torny dans ce numéro.

plusieurs initiatives destinées à produire un index des citations scientifiques dont l'objectif n'était pas alors de mesurer la réputation des chercheurs, mais de « fournir au scientifique consciencieux un moyen de connaître les critiques dont les articles qu'ils citent ont fait l'objet » en révélant des « associations d'idées » (Garfield, 1955, cité dans Wouters, 2006, p. 14). Au terme d'un long travail de persuasion, il parviendra à convaincre la National Science Foundation et le National Institute of Health de soutenir la mise en place d'un instrument permettant à la fois la centralisation de la production scientifique sous forme d'une base de données de citations des publications et son objectivation par une série de mesures. Pour sa première édition en 1964, le Science Citation Index (SCI) collecte à la main et à grand coût 1,4 million de citations des articles de 613 revues publiées en 1961. Les résultats seront enregistrés sur une bande magnétique pour constituer l'une des premières grandes bases de données de l'histoire de l'informatique naissante.

### **Extériorité, abstraction, procéduralisme, neutralité, honnêteté**

Mais avant qu'elle ne devienne une métrique d'évaluation pour la bureaucratie scientifique, il faut s'attarder sur ce que présuppose l'ambition du *Science Citation Index* de *représenter la science par le tissu de ses citations*. Car tout en préservant un lien référentiel avec le monde qu'elle enregistre, cette représentation invente aussi un cadre cognitif bien particulier dont on voudrait dégager cinq propriétés épistémiques qui appareilleront le PageRank. La première de ces présuppositions est la revendication d'une position d'*extériorité*. Le SCI place l'instrument d'objectivation de la science en dehors de la science, afin d'en mesurer la qualité sans prendre de point d'appui normatif à l'intérieur même du champ scientifique, ce qu'est amenée à faire toute évaluation par des pairs. Cette extériorité est aussi ce qui lui permet d'acquérir une vue aérienne et complète de l'activité scientifique à laquelle ne peuvent accéder les chercheurs prisonniers de leurs disciplines. Cette vue d'ensemble est la condition de possibilité de la deuxième présupposition : celle d'*abstraire* la citation du contexte dans lequel elle a été émise. En effet, la principale opération que produit le SCI consiste à transformer une liste de *références* qui partent d'un article, données simples et immédiatement accessibles, en une liste de *citations* que reçoivent les articles des autres publications (information invisible depuis les articles eux-mêmes et qui ne peut être calculée qu'en se donnant un accès à la totalité des textes citeurs). L'opération cognitive à laquelle procède le SCI est donc une simple inversion de la référence

(une mention de l'article B dans l'article A) en citation (le fait pour l'article B d'avoir été cité par A). Cette minuscule opération qui réclame un prodigieux travail de collecte des données dans le monde analogique est justement celle que va bouleverser, en la matérialisant, le lien hypertexte qui rend activable la relation entre texte citeur et texte cité. Alors que la *référence* est investie de significations multiples par celui qui la place dans son texte (honorer, critiquer, marquer l'antériorité, préciser, faire acte d'ostentation, etc.), en revanche, la *citation* efface en une abstraction univoque la diversité des significations contextuelles de la référence qui l'a fait naître. Comme le souligne Paul Wouters, si toutes les références ne sont pas égales, l'opération d'abstraction de leur contexte d'émission que réalise l'inversion des références en citations invite à considérer toutes les citations comme égales. Cette transformation contribue donc à unifier le sens de la citation pour en faire une sorte de « monnaie de l'activité scientifique », à la fois standardisée, décontextualisée, univoque et égale (Wouters, 1999, pp. 108-109). Alors que la référence réfère au contexte de son émission, la citation ne réfère plus qu'à elle-même. La valeur de la citation devenue simple signe, prend un caractère autoréférentiel et se donne à la possibilité d'un calcul.

La troisième présupposition de ce dispositif d'objectivation de la science par ses citations est d'assumer son caractère résolument *procédural*. Ce que mesure le SCI n'est pas le contenu substantiel des appréciations scientifiques que se sont échangés les chercheurs dans leurs articles, mais la seule forme autoréférente de la citation avec ses propriétés indexables (nom de l'auteur, institution d'appartenance, titre, type de publication, etc.). « Alors que la littérature scientifique représente la science en se concentrant sur ses prétentions cognitives (le contenu des articles et des livres publiés) », écrit Paul Wouters (1999, p. 7), « le SCI représente la littérature scientifique en oblitérant le contenu au profit de ses propriétés formelles ». Le SCI opère une réduction délibérée de la variété substantielle du discours scientifique pour en faire un instrument d'exploration susceptible de circuler dans toutes les communautés disciplinaires sans avoir à se préoccuper de leurs idiosyncrasies. Ce formalisme se doit d'être absolument agnostique à l'égard des arguments scientifiques dans lesquels il ne fonde pas sa légitimité. Il n'est pas besoin de connaître la multiplicité des significations investies par les chercheurs dans l'acte de citer pour que le calcul global de leur nombre, bénéficiant des effets de la statistique des grands nombres, constitue une approximation pertinente de ce que l'on cherche à lui faire mesurer. « La grande utilité de l'approche d'un index des citations », écrit Joshua Lederberg, un généticien qui a soutenu activement Eugène Garfield dans son entreprise,

« est de passer outre le problème de la signification par le recours à une procédure automatisée » (cité dans Wouters, 1999, p. 20).

En quatrième lieu, la représentation de la science en un tissu de citations n'est pas pensée comme produite par ceux qui indexent, mais par ceux qui se citent entre eux : les chercheurs publiants. Incarnant, avec tous ses défauts simplificateurs, l'idéal transparent d'une objectivité non interventionniste, l'index des citations est le résultat cumulé des actions effectuées par les chercheurs se citant mutuellement sans qu'il n'ait en principe, la question fera évidemment retour en scientométrie comme en webométrie, agi en fonction de cette représentation qui est extérieure à leur activité. Le SCI importe ainsi dans l'univers de la connaissance de la science le présupposé naturaliste de l'« objectivité mécanique » (Daston et Galison, 2012). Cependant, à observer un monde social qui, à la différence du monde naturel, est réflexif, la présupposition de transparence de cette technologie cognitive ne peut écarter le risque que ceux qu'elle enregistre agissent en fonction de ceux qui les mesurent. Aussi la présupposition de *neutralité* implique-t-elle une revendication complémentaire d'*invisibilité*, afin de ne pas troubler le monde dont elle capture les agissements. Se faire oublier, on le verra, constitue le rêve de moins en moins satisfait du PageRank.

En dernier lieu, le SCI s'appuie sur une présupposition principielle sur laquelle s'adosent toutes les autres : une confiance raisonnée dans l'*honnêteté* de la citation. Le procéduralisme de la nouvelle technologie de mesure de la science nécessite un fond de justification substantielle qui confère à la citation scientifique les valeurs de l'ethos académique. Il n'est possible de compter les citations sans chercher à les comprendre d'une part qu'en présupposant qu'elles sont globalement considérées par les chercheurs, et ceci en dépit de la diversité des usages que ces derniers en font, comme un des signes les plus évident de leur appartenance à la communauté, du respect d'une de ses règles les plus essentielles et d'autre part sans tenir compte du fait que, soumises à la critique vigilante de la communauté, ces citations sont émises selon des principes justifiables devant la communauté. Sous cette condition, les chercheurs peuvent investir de toutes les significations possibles l'acte de citer sans que les bibliomètres ne s'en préoccupent. Ces derniers peuvent se contenter d'accorder une confiance raisonnée à la conception de l'activité scientifique comme compétition pour la reconnaissance entre pairs développée par Robert Merton (1957). C'est parce que la science est un savoir public, et non privé, que les chercheurs doivent verser leurs travaux dans le domaine public et les faire reconnaître comme les leurs. Mais comme la production scientifique est un tissu d'interdépendances entre différents travaux, il est indispensable de



citer les sources dont chaque publication hérite, au risque de voir la communauté signaler un manquement. « Citations et références se réalisent donc à l'intérieur d'un cadre qui est à la fois moral et cognitif », insiste Merton (1977). La citation procure ainsi aux scientomètres la trace d'une pratique régulière et objectivable qui incorpore suffisamment les normes de l'activité scientifique pour être soumise à une procédure de calcul.

### **Le lien est un vote**

À l'origine, le sujet de recherche présenté par Sergei Brin et Larry Page à leur enseignant, Terry Winograd, était la conception d'un système permettant de porter des annotations sur les sites du web (Levy, 2011, pp. 16-17). Très vite cependant, les jeunes étudiants de Stanford réalisent que le lien hypertexte équivaut à une citation et que, à sa manière, il peut être considéré comme un vote. Du *Science Citation Index* au PageRank, la filiation est explicite. Tous les deux fils d'universitaires, les fondateurs de Google ne cesseront de souligner qu'« un grand nombre de citations dans des publications scientifiques signifie que vos travaux sont importants, dans la mesure où d'autres personnes ont pensé que cela valait la peine de vous citer » (Vise et Malseed, 2006, pp. 34-35). Dans l'article de Jon Kleinberg, « Authoritative Sources in a Hyperlinked Environment » (1998), qui influencera Larry Page dans la conception du PageRank, la réputation des chercheurs mesurée en bibliométrie est très clairement affichée comme principale source d'inspiration. Il insiste sur le fait que, comme la citation scientifique, le lien hypertexte est à la fois un acte de reconnaissance et un signe d'autorité. « Les liens hypertextes », écrit-il, « encodent une somme considérable de jugements humains latents et nous prétendons que c'est exactement ce type de jugement qui est requis pour formuler la notion d'autorité. » Le lien hypertexte circonscrit le *domaine de pertinence* du texte du citeur, reconnaît la *valeur du contenu* cité et, lorsque ce dernier reçoit des approbations multiples et diversifiées, affirme son *importance* sur une échelle méritocratique qui distribue les honneurs à ceux qui ont été désignés par leurs pairs. Il constitue exactement le genre de traces qui, devenues métriques, permettent de hiérarchiser les objets informationnels selon la rationalité qui prévaut dans le monde de la recherche, en orientant l'attention vers les contenus présentant une sorte de prééminence. Mesurée conventionnellement par le nombre de citations, elles-mêmes autorisées, cette prééminence constitue la meilleure approximation de la certitude épistémique. Jon Kleinberg (1998) précise que le seul dénombrement des

liens suffit à capturer l'autorité de tout document : « La création d'un lien sur le *www* représente une indication concrète d'un jugement de ce type : le créateur de la page *p*, en incluant un lien vers la page *q*, a dans une certaine mesure conféré de l'autorité à *q*. Qui plus est, les liens nous donnent l'opportunité de découvrir des autorités potentielles par le truchement unique des pages qui pointent vers elles. » En tout point héritée des propriétés d'*abstraction* et de *procéduralisme* du *Science Citation Index*, cette intuition crée une rupture décisive en faisant de la qualité des informations présentes sur les sites non pas une propriété interne qui doit être recherchée à l'intérieur du document par une analyse toujours plus fine de son contenu lexical, mais une propriété externe qui est façonnée par les attributions respectives que les sites se font les uns envers les autres en se reconnaissant. La qualité est une construction sociale que projettent les interactions sur les documents. Larry Page ne le dit jamais aussi clairement que dans le brevet qui livre, sans le détailler, le fonctionnement du PageRank : « De façon intuitive, un document doit être important (sans considération pour son contenu) s'il est beaucoup cité par d'autres documents » (Page, 1998). Le lien hypertexte est une simple enveloppe, une « concrétion d'intelligence » (Pasquinelli, 2009, p. 155), qui ne doit pas être ouverte, afin de conserver un caractère calculable. Ses balises sont aisément identifiables par les robots qui aspirent le web. Il n'est pas nécessaire de savoir pourquoi il a été créé, ni quelle somme d'intentions, de sous-entendus, de calculs et d'appréciations diverses et variés ont été investis dans sa création. Il suffit, comme dans l'urne, de les dénombrer. De façon audacieuse, les fondateurs de Google vont étendre cette conception de l'autorité scientifique en prolongeant la métaphore du lien comme citation à celle du lien comme vote. À la section « Pourquoi Google », l'entreprise de Mountain View n'hésite pas à présenter son algorithme comme le code source de la démocratie : « PageRank est un champion de la démocratie [...] : tout lien pointant de la page A à la page B est considéré comme un vote de la page A en faveur de la page B. Toutefois, Google ne limite pas son évaluation au nombre de "votes" (liens) reçus par la page ; il procède également à une analyse de la page qui contient le lien. Les liens présents dans des pages jugées importantes par Google ont plus de "poids", et contribuent ainsi à "élire" d'autres pages » (cité dans Cassin, 2007, pp. 102-103).

### **La pondération des pages**

Si le PageRank fait voter les hyperliens, son régime politique n'est cependant pas une démocratie dans laquelle chaque votant disposerait du même poids,

mais une méritocratie censitaire, qui n'accorde pas le même pouvoir à chaque vote. Alors que le *Journal Impact Factor* (JIF), l'algorithme développé par Eugène Garfield pour mesurer la réputation des revues, considère que chaque citation est équivalente, le PageRank va, lui, reprendre à son compte un mécanisme récursif permettant d'affecter des poids différents aux pages citeuses. Celui-ci s'inspire de l'algorithme de l'*influence weight* développé par Gabriel Pinski et Francis Narin (1976)<sup>3</sup>. Ce dernier propose d'abord d'établir un ratio du rapport entre citations entrantes et sortantes afin d'éviter que certaines revues reçoivent beaucoup de prestige pour la seule raison qu'elles publient beaucoup d'articles. Le poids d'une revue est mesuré par le nombre des citations entrantes divisé par le nombre des citations sortantes ; ce ratio a pour effet de faire de l'autorité un bien circulant, à la fois reçu et distribué, donnant un solde positif à ceux qui en reçoivent plus qu'ils n'en distribuent. Devenu ratio, l'indice d'autorité réfère les citations à elles-mêmes pour en faire une véritable monnaie. Car Francis Narin propose aussi et surtout de considérer que toutes les citations n'ont pas le même poids et qu'il faut leur appliquer un attribut récursif permettant de calculer l'autorité du citeur dans le réseau en fonction du nombre de citations qu'il a lui-même reçues des autres.

L'égalité des citations dénombrées par le *Journal Impact Factor* a du sens dans la communauté étroite de la base de données du *Science Citation Index* de l'ISI qui ne réunit que des revues académiques. La citation est égalitaire et son comptage démocratique si, censitaire, le corps électoral est restreint aux universitaires. Il est, en effet, vertueux de dénombrer des votes égaux lorsque le filtre de l'autorité s'est préalablement exercé avec la sélection du statut universitaire des citeurs qui, seuls, ont accès aux revues académiques. Dans un contexte de marché restreint, le fait de différencier l'autorité des votes apparaît même comme une contravention aux principes égalitaires de la communauté scientifique. Cependant, si le nombre des revues collectées par le *Science Citation Index* a pu, en son temps, apparaître très important, il est absolument infime au regard du volume gigantesque des pages qui se lient entre elles sur le web. Inclusif, le web considère que tous les internautes qui publient sont des citeurs et ne leur réclame aucun diplôme. Aussi, l'autorité ne se mesure-t-elle plus à l'entrée, mais à l'intérieur de la base de données. La pondération de l'autorité des pages citeuses apparaît comme un effet de la démocratisation des citeurs. Dans un monde ouvert à tous et à quiconque, une égalité des votes viendrait installer un principe de *popularité* et non d'*autorité*

---

3. Voir dans ce numéro l'article de D. Pontille et D. Torny (Pontille et Torny, 2013).

(Cardon, 2011). Comment, dès lors, reconnaître la prééminence de certains documents lorsque ceux qui les citent ne sont pas des pairs ? Revenant en 2004 sur cette question, alors que le web est devenu si étendu et diversifié qu'il est de plus en plus difficile de lui donner comme modèle le monde scientifique, Sergey Brin utilise la métaphore de la recommandation sociale pour décrire le PageRank. L'autorité que confère le jugement des chercheurs sur leurs travaux respectifs est étendue à la confiance accordée dans la vie quotidienne à des experts. Ce que doit désormais faire l'algorithme n'est pas de faire voter une communauté d'égaux préalablement sélectionnés, mais de dégager parmi tous, ceux qui sont reconnus par les autres comme des experts pour que leur vote ait davantage de poids. Tout le monde n'est pas également digne de confiance, mais tout le monde est capable de désigner ceux qui le sont :

« Si je suis à la recherche d'un bon docteur dans la région », explique Sergei Brin, « je vais regarder autour de moi et demander à mes amis de me recommander les bons docteurs. Ils pourront me désigner des personnes qui s'y connaissent mieux qu'eux – "ce gars connaît tous les docteurs de la Baie". J'irais ensuite voir cette personne pour la questionner. La même chose s'applique aux sites du web. Ils réfèrent à un autre avec des liens, c'est un système qui encourage la recommandation » (Sheff, 2004).

L'*influence weight* offre une solution algorithmique permettant de résoudre la tension entre la démocratisation des citeurs et les traits méritocratiques du principe d'autorité. Dans le monde ouvert du web, il est impossible de donner à tous la même autorité comme dans le monde confiné de la science. La solution qu'apporte le PageRank est très élégante : l'ouverture à la diversité des citeurs doit aider à désigner l'autorité des *pages* du web, et non des *internautes*, qui les ont produites. « Nous nous sommes rendu compte que toutes les pages n'étaient pas égales », soutient Sergei Brin. « Les gens le sont, mais pas les pages du web. De façon inhérente, certaines pages, sont, non pas plus mauvaises que d'autres, mais moins importantes et c'est autour de cette idée que nous avons développé une analyse de la structure du graphe qui utilise l'importance des pages<sup>4</sup>. » Le PageRank considère que les internautes publiants sont égaux, mais que leurs pages ne le sont pas et il fait de cette *séparation entre la personne et la page* une manière de préserver le

---

4. Vidéo d'une conférence de Sergei Brin à la School Information de l'Université de Berkeley le 3 octobre 2005 : UC Berkeley, « SIMS 141 – Search, Google, and Life : Sergey Brin – Google », *YouTube*, 20 août 2007.

principe d'autorité lorsque le droit de publier est ouvert à tous. Le PageRank s'appuie pour cela sur une idéalisation qui est au cœur même de la conception du web par les pionniers. Structure élémentaire d'un réseau de documents sans couture, le lien hypertexte constitue la réalisation la plus aboutie d'une utopie dans laquelle les textes parviendraient à entretenir des relations entre eux, en se soustrayant à l'autorité de leur producteur. Cette rêverie animait, dès 1945, le texte visionnaire de Vannevar Bush, *As we may think* si influent chez les pionniers de l'Internet. Elle a ensuite nourri le projet Xanadu de Ted Nelson (1965), le système HyperCard de Bill Atkinson (1986), et l'invention fondatrice du World Wide Web par Tim Berners-Lee en 1990. Pure intertextualité, le graphe du web ne serait constitué que d'associations entre énoncés, sans qu'il soit nécessaire de qualifier les personnes qui les ont produits. La disparition de l'énonciateur est au cœur de cette vision idéalisée d'un monde d'idées dialoguant entre elles dans un rapport d'argumentation et de raison débarrassé du poids des intérêts, de la personnalité ou de la psychologie de ceux qui les ont émises (Lévy, 1991, p. 62). Le graphe du web sur lequel le PageRank fait circuler son indicateur récursif est un graphe de documents et non de personnes. L'autorité qu'il mesure procède d'une opération qui ne s'appuie ni sur le seul contenu des textes se citant entre eux, comme le revendique une approche sémantique du web, ni sur le statut des personnes qui ont rédigé les textes, comme le fait le *Science Citation Index* en réservant l'accès aux revues scientifiques aux chercheurs. Si l'on reprend la distinction faite en pragmatique de l'énonciation, la force sociale dont le PageRank mesure l'autorité n'est pas celle de l'Énoncé, ni celle de la Personne de l'énonciation, mais celle du Sujet de l'énonciation<sup>5</sup>. Le PageRank présuppose qu'une distance s'est ouverte entre la Personne de l'énonciation et le Sujet de l'énonciation, afin que le tissage intertextuel des références ouvre à l'Auteur (*i. e.* le Sujet de l'énonciation) un « espace où le sujet écrivant ne cesse de disparaître » (Foucault, 2001, p. 821). Le PageRank ne tient pas le lien hypertexte pour une association sémantique entre énoncés ou pour un échange de gratifications entre personnes, mais comme un moyen d'évaluer l'autorité d'une page du web. Le lien hypertexte part d'un élément du texte citeur pour pointer l'URL d'une page citée dans son ensemble. Il attribue ainsi sa force à la page et propose ainsi une solution réaliste, et terriblement efficace, pour classer les documents.

---

5. Pour un développement plus complet sur cette question, voir Cardon (2013).

## Comment capturer la force du lien entrant ?

En 1998, dès que les premiers tests comparatifs sont réalisés, il apparaît d'emblée que les résultats proposés par le PageRank pour des requêtes variées sont d'une qualité infiniment supérieure que ceux de leurs concurrents. « C'était comme faire la différence entre juger un étranger sur son look ou recueillir les opinions sur lui de la part de tous ceux qui le connaissent » (Edwards, 2011, p. xii). En août 1999, les serveurs de Google reçoivent 3 millions de requêtes par jour. En août 2000, après un accord avec Yahoo!, 60 millions de requêtes s'adressent tous les jours aux serveurs de la jeune start-up (Batelle, 2005, p. 126). Pour faire fonctionner son algorithme, Google constitue une table de l'ensemble des sites du web aspiré et répertorié dans un gigantesque index auquel il attribue un ensemble de signaux. On en dénombre aujourd'hui plus de 200 pour chaque page enregistrée (Singhal, 2008). Ces signaux nourrissent deux dimensions différentes de la qualité d'une recherche : la *pertinence* de la page au regard de la demande effectuée dans la requête et l'*importance* de la page au regard des autres pages qui présentent un même niveau de pertinence. La première dimension cherche à spécifier le plus finement possible le sens de la requête, afin que les réponses proposées par le moteur correspondent à la question de l'internaute. Dans ce domaine, Google a développé un large panel d'indicateurs sémantiques qui vont contribuer à la complexification de l'algorithme et à l'augmentation du nombre des signaux. La seconde dimension cherche à mesurer, parmi les propositions pertinentes, l'autorité de la réponse en filtrant le mieux possible, parmi les liens entrants, ceux qui sont porteurs d'autorité et ceux qui ne le sont pas. C'est sur cette deuxième dimension que se déploie le calcul du PageRank. Celui-ci n'est désormais qu'un signal parmi les autres. Même si la question est de plus en plus discutée, ce calcul joue toujours un rôle prépondérant dans le fonctionnement global de l'algorithme et son esprit exerce une influence indirecte sur de très nombreux autres signaux qui viennent le préciser et le renforcer. Le PageRank déploie une note de 1 à 10 sur une échelle logarithmique qui mesure le nombre de liens reçus par la page venant d'autres pages en considérant que les sites s'envoient les uns aux autres une force qui, dans le jargon du référencement, va très vite être appelée le « Google Juice » ou « jus de liens ».

## L'ORGANIQUE ET LE STRATÉGIQUE

Si le PageRank a, d'emblée, constitué une rupture dans la qualité des recherches sur le web, c'est parce qu'il a le mieux su épouser l'esprit d'Internet en offrant, comme un miroir, un artefact cognitif faisant de la distribution des interactions entre documents une métrique qui leur révélait leur autorité respective. Dans son principe, le PageRank ne fait que rendre aux internautes les jugements qu'ils se sont portés les uns les autres à travers leurs liens. « Le génie de Google », souligne James Grimmelmann (2009, p. 941), « est que ses créateurs n'ont pas cherché à imposer un grand schéma organisationnel pour le web. À la place, ils ont demandé à tous les autres de le faire pour eux. » La critique sera souvent faite, Google aspirerait un travail qui n'est pas le sien, augmenterait sa pertinence de l'énergie bénévole des autres et, plus encore, accumulerait une richesse qui devrait tout au travail vivant des internautes (Pasquinelli, 2009 ; Moullier-Boutang et Rebiscol, 2009 ; Kyrou, 2010 ; Vaidhyanathan, 2011). Il faut cependant interroger plus avant ce paradoxe, car en soutenant avec une inépuisable énergie, et parfois contre l'évidence, l'idée que le classement que produit son algorithme est *naturel*, ou « organique » selon la terminologie consacrée, Google s'efforce aussi de satisfaire une contrainte statistique qui est nécessaire à la pertinence du PageRank : être absent des intentions des internautes.

### Sagesse des foules

De quel principe l'approche algorithmique de Google peut-elle se recommander pour justifier la préférence qu'elle accorde à l'agrégation automatique des jugements incertains, dispersés et aléatoires de la foule des internautes ? Alors que le *Science Citation Index* s'appuyait, *in fine*, sur une présupposition d'*honnêteté* de la citation inscrite dans la structure normative du fonctionnement du champ scientifique, le PageRank ne dispose pas d'un tel fondement substantiel pour justifier son procéduralisme. Il lui substitue un autre type de justification, beaucoup plus formelle car d'inspiration principalement statistique, qui a pris le nom de « sagesse des foules » (Surowiecki, 2008 ; Orrigi, 2008). La thèse de l'intelligence collective des internautes, dont le PageRank est toujours cité comme l'illustration la plus exemplaire, s'appuie sur un ensemble de travaux mêlant mathématiques et philosophie politique pour prouver la supériorité épistémique des grands nombres. Parmi les diverses conceptualisations de l'« intelligence des foules » (Landemore et Elster, 2012), trois thèses offrent une justification à la revendication qu'affiche le PageRank de

mesurer l'autorité sur le web. La première y reconnaît d'abord le *miracle de l'agrégation* dont on fait remonter l'origine au théorème du jury de Condorcet qui pose que pour trouver la bonne solution à une question épistémique, il suffit de faire voter le plus grand nombre possible de personnes, à condition que la majorité des participants ait une probabilité positive de trouver la bonne solution et que ceux-ci ne s'influencent pas les uns les autres. Si ces conditions sont satisfaites, plus le nombre de votants est important, plus il est certain que le vote majoritaire sera le bon. Ce théorème est aussi au principe de la fameuse expérience de Galton en 1906 faisant voter le public d'un marché aux bestiaux pour estimer le poids d'une vache. Le public, pris comme un tout, est donc épistémologiquement plus fiable que chacun des membres qui le composent, aussi experts que soient certains d'entre eux (Landemore, 2010). À cette propriété, les travaux de Scott Page (2006) vont ajouter une dimension nouvelle en démontrant qu'il est plus important de valoriser la diversité cognitive des votants que leur intelligence. Ce soubassement statistique nourrira la thèse du « miracle de l'agrégation » qui considère qu'il est important d'éviter les effets de coordination et d'influence que peuvent exercer les votants les uns sur les autres et encourage le développement de dispositifs de jugements individualisant comme les marchés prédictifs (Sunstein, 2006).

À la différence de cette approche « agrégative », une seconde interprétation, « délibérative », de la thèse de la sagesse des foules insiste sur les effets auto-organisés de la coordination des jugements sur Internet. Sous l'inspiration du modèle de discussion habermassien, on en trouve une version très optimiste dans *La richesse des réseaux* de Yochai Benkler (2009, p. 309 et sq.). Prenant exemple sur de multiples analyses de la blogosphère<sup>6</sup>, il souligne les mécanismes d'auto-organisation permettant aux petits cercles épars de conversation sur Internet d'entrer en communication les uns avec les autres pour gagner, par sélection successive, de la visibilité dans le moteur de recherche de Google. Ces formes de coordination entre conversations décentralisées<sup>7</sup> permettent, par une série d'itérations, d'extraire des contenus de leur cercle de production initial pour les faire reconnaître par d'autres et faciliter leur circulation dans l'espace étagé de la visibilité du web. Ces parcours de mise

6. Notamment l'article de Farrell et Drezner (2008) dont les résultats ont été contestés par Hindman (2008).

7. Goodwin (2003) soutient ainsi un modèle de « disjointed deliberation » permettant à une pluralité de cercles de conversations de faire circuler entre eux les attendus de leurs débats grâce à la coappartenance de certains membres à différents groupes.



en visibilité ne doivent cependant pas être comparés à la mesure de l'opinion des internautes telle que la produisent des dispositifs d'interrogation individualisant comme le sondage d'opinion. Ils sont le produit émergent d'une coordination par le bas, spontanée et sans organisation centrale.

Un troisième aspect de la thèse de la sagesse des foules, sans doute le plus exigeant, est indispensable à son fonctionnement. Les jugements que s'échangent les internautes à travers leurs liens doivent faire l'objet d'une « coordination non coordonnée » (Benkler, 2009, p. 33). L'agrégation des échanges entre hyperliens qu'enregistre le PageRank est le résultat d'actions individuelles qui n'avaient pas cette coordination pour intention, insistent Jon Kleinberg et Steve Lawrence (2001, p. 1849). Car la condition essentielle de son fonctionnement est que les internautes n'aient pas agi *en fonction* du PageRank, mais que leur choix de liens distribue de façon « naturelle » les honneurs et les oublis. Si les jugements qu'ils s'échangent à travers les liens ont été produits en fonction du méta-coordonateur qui les agrège, la pertinence épistémique du résultat en sera profondément altérée. Les différents modèles mis en place sous la bannière de la sagesse des foules font tous une distinction entre d'une part le caractère local, non intentionnel, et immédiat de la formation des jugements individuels (dans le théorème du jury) ou dans des enclaves de discussion (dans le modèle des délibérations décentralisées) et, d'autre part, l'outil formel d'agrégation qui permet de représenter ces jugements, sans être en rien, à leur origine. La présupposition d'*extériorité* est une condition de possibilité de l'intelligence collective. « Un système de sagesse des foules », souligne Daniel Andler (2012), « peut être considéré comme intelligent si l'on est prêt à rompre le lien entre deux composantes de l'intelligence : la compréhension du monde est atteinte, de façon répartie, par les membres individuels du groupe (chacun possédant une compréhension partielle mais réelle), tandis que la recherche d'une solution est obtenue par l'architecture du système d'une manière purement formelle (*i.e.* sémantiquement aveugle). L'exemple qui vient à l'esprit est fourni par l'Internet, en particulier un moteur de recherche comme Google. » Les choix de référencement du PageRank sont d'autant plus pertinents que l'agrégateur des jugements des internautes est absolument extérieur à leurs décisions. À la présupposition d'*honnêteté* de la citation requise par les normes substantielles de l'espace scientifique, le web substitue une exigence, simplement procédurale, mais si difficilement vérifiable, de *sincérité* : que les internautes n'aient pas pensé à Google.

## La machine Google

Alors que les fondateurs posaient dans l'article scientifique du PageRank qu'il existe une incompatibilité de principe entre la recherche d'information et la publicité (Brin et Page, 1998), Google va inventer un modèle publicitaire d'une efficacité sans pareille, en vendant aux annonceurs les mots clés des recherches des internautes. Lui aussi s'appuie sur un appareillage algorithmique singulier dont Google empruntera l'idée au service Goto de Bill Gross. Mais Google ajoute trois raffinements particuliers à *Adwords* : les enchères sur les mots clés sont dites « de Vickrey » (le gagnant paye le coût par clic de la seconde enchère), une analyse lexicale de la page publicitaire vérifie sa pertinence et une analyse des clics des internautes sur les différents liens publicitaires peut, par apprentissage, modifier l'ordre des classements. Cependant, la principale différence qu'apporte Google à Goto est de refuser fermement de mélanger les résultats du classement « naturel » de l'algorithme et ceux des enchères vendues aux annonceurs. C'est en séparant la recherche « naturelle » et les liens publicitaires que Google a fait la différence avec ses concurrents. Dans le jargon des *googlers*, cette barrière, véritable ligne de « séparation de l'Église et de l'État » (Cassin, 2007, p. 139), a pris le nom de « grande muraille de Chine ». Google ne propose pas seulement aux utilisateurs une interface isolant plus proprement que ne le faisaient ses concurrents la publicité et l'éditorial, mais érige une frontière sur le lieu du conflit entre les mathématiques et le marché qui traverse la culture de l'entreprise et la personnalité de ses fondateurs. La science des algorithmes doit poursuivre sa quête de perfection pour *refléter* au mieux les agissements des internautes, mais ceci sans jamais que les internautes n'agissent en fonction de Google, ni que les ingénieurs de Google ne mettent la main dans les classements. Google veut regarder ce monde comme *naturel*. Et parallèlement, un autre monde est ouvert aux annonceurs qui veulent se disputer les mots clés des enchères publicitaires. Et ce monde-ci, lui, ne se cache pas d'être pleinement *stratégique et instrumental*. Vu depuis la muraille de Chine, il existe deux manières d'acquérir de la visibilité sur la page de Google : soit par la réputation acquise auprès des autres, et ceci sans Google, soit en payant sa visibilité... à Google. La partition de la page de résultats en deux mondes, *organique et stratégique*, enferme une vision du web et des internautes que Google va imposer par tous les moyens possibles à l'ensemble de l'écosystème du web.

## Ne pas toucher l'algorithme à la main

La séparation du référencement naturel et du référencement publicitaire est d'abord présentée comme la réplique numérique du modèle économique des médias traditionnels. En 2004, les fondateurs de Google écrivent une lettre aux futurs actionnaires expliquant que ce que Google mesure colonne de gauche ne doit pas être contaminé par ce que Google vend dans la colonne de droite. C'est le modèle éditorial de la presse qui sert de justification : « Nos résultats de recherche sont ce que nous savons produire de meilleur », se vante Google. « Ils ne sont pas biaisés et sont objectifs. Nous n'acceptons pas de paiement pour eux [...]. Nous affichons aussi de la publicité, en travaillant beaucoup pour la rendre pertinente, et nous la labélisons clairement. Ceci est similaire à la presse où la publicité est claire et les articles ne sont pas influencés par le paiement des publicitaires<sup>8</sup>. » Google a beau jeu d'afficher devant des journalistes sa confiance dans un partage qui est au cœur de l'éthique professionnelle des médias. Mais la réalité de sa conception d'un monde éditorial « objectif » et « neutre » est différente de celle du journalisme professionnel. L'« objectivité » à laquelle se réfère Google est « mécanique » et plonge ses racines dans la disciplinarisation du travail scientifique qui s'est initiée à la fin du XIX<sup>e</sup> siècle dans le désir impérieux de réprimer toute intervention humaine au profit de méthodes et de machines capables d'imprimer directement la nature sur l'écran du chercheur (Daston et Galison, 2012). Le détachement impartial qui suppose du journaliste un haut niveau d'autocontrôle est, pour Google, une vertu inaccessible. Aussi objective que soit leur déontologie, les *gatekeepers* des médias traditionnels seront toujours soumis à des passions, des choix ou des intérêts qui introduisent des biais dans leur mise en ordre de l'information<sup>9</sup>. La visée d'objectivité, comme l'a montré Michel Porter (1995), a finalement moins à voir avec la vérité de la nature qu'avec le travail d'éviction du jugement humain que mènent les scientifiques contre leur propre subjectivité. Nourrie de science, de mathématique et de grand nombre, la conviction fondamentale de l'entreprise de Mountain View est que pour neutraliser les aléas du jugement humain, il est préférable de

8. Google Inc., *Letter from the Founders: "An Owner's Manual" for Google's Shareholders*, in Forms S-1 registration Statement Under the Securities Act of 1933.

9. De façon pour le moins surprenante pour un article d'informatique, les fondateurs de Google ont glissé dans la bibliographie du plus célèbre papier de la conférence WWW (Brin et Page, 1998) une citation de l'ouvrage de Ben H. Bagdikian, *The Media Monopoly* (1997), sévère dénonciation des biais journalistiques et des effets de la concentration économique sur le pluralisme de la presse.

faire confiance aux algorithmes et de s'en tenir là – toute tentative de corriger à la main un résultat insatisfaisant constituant le début d'une corruption du service. Dans un texte titré « Pourquoi nous vendons des publicités, pas des résultats », Google refuse d'être tenu pour responsable des résultats de la recherche « organique ». L'entreprise se place comme à l'extérieur des activités de son algorithme : « Nos résultats reflètent ce que la communauté considère comme important, pas ce que nous ou nos partenaires pensons que vous devez voir<sup>10</sup>. » Ce souci de déléguer à une règle de calcul la responsabilité des classements préserve l'entreprise d'avoir à se justifier contre les mille et une accusations de partialité dont elle fait l'objet. Lors des affaires *google bombing* (i.e. l'action coordonnée des internautes pour associer un site à une requête spécifique, par exemple la page officielle de George Bush à la requête « miserable failure »), lorsque les suggestions de son système d'aide à la recherche font remonter des termes antisémites, lorsqu'une entreprise juge que sa concurrente est injustement mieux classée qu'elle, etc., Google refusera de corriger à la main ou de poser un filtre sur son algorithme (Grimmelmann, 2009). Répondant à la question « Est-ce que Google édite à la main ses résultats ? », Amit Singhal, l'architecte du PageRank se montre ironique :

« Laissez-moi répondre à cela avec notre troisième philosophie : pas d'intervention manuelle. De notre point de vue, le web est construit par les gens. Vous êtes ceux qui créent des pages et des liens vers d'autres pages. Nous utilisons toute cette contribution humaine à travers nos algorithmes. L'ordonnement final des résultats est décidé par nos algorithmes utilisant les contributions de la grande communauté de l'Internet. Nous pensons que le jugement subjectif de n'importe quel individu est, disons... subjectif et l'information extraite par nos algorithmes de l'énorme somme de savoir encodé par les pages du web est meilleure que la subjectivité individuelle<sup>11</sup>. »

Confronté à des accros du résultat de l'algorithme, lorsque des erreurs locales sont détectées dans le classement de tel ou tel site, Google se refuse à porter « à la main » un patch local sur l'algorithme pour rétablir un classement correct. Les ingénieurs de l'équipe Search cherchent toujours à trouver une règle automatisable permettant de traiter en général les imperfections détectées. Les machines ont des qualités que n'ont pas les humains, mais surtout des vertus

10. Google Inc., “Why We Sell Advertising, Not Results”, *Google.com*, 2004. [<http://www.google.com/honestresults.html>].

11. Singhal (A.), “Introduction to Google ranking”, *Official Google Blog*, 9 juillet 2008.

qu'elles tiennent de leur faiblesse. Comme l'a spectaculairement affirmé Eric Schmidt (2004), leur force tient à leur bêtise : « Je peux vous l'assurer. Il n'y a pas de biais. Ce sont des ordinateurs, ils sont à mourir d'ennui. Je suis désolé que vous ne le compreniez pas. » Beaucoup de critiques « anthropomorphisent » l'algorithme de Google pour lui demander de se comporter comme un humain. Les reproches qu'ils lui adressent sont les mêmes que ceux que la sociologie des médias adresse à la rédaction d'un journal : partialité, goût du commun, oubli des périphéries, conformisme. Mais Google ne considère pas son algorithme comme un être humain. Il le dote d'une autre ontologie. Les décisions du ranking machinique sont procédurales, alors que les jugements humains, tout informés par des règles ou des déontologies soient-ils, restent substantiels. Ce qui disqualifie le jugement humain, c'est cette embarrassante disposition à juger et à évaluer substantiellement, à toujours vouloir apprécier la validité, la rationalité ou le bon sens de l'ordre des réponses. Cela, les algorithmes ne peuvent le faire. Ils sont idiots et cette idiotie est le meilleur garant de leur « neutralité ».

### **Ne vous préoccupez pas de nous !**

Sans doute, la meilleure communication de Google à l'adresse des internautes serait-elle de se taire. Mais se faire oublier est vain et Google est contraint de communiquer son désir d'effacement. Avec obstination, l'entreprise ne cesse de demander aux internautes de « faire comme si les moteurs de recherche n'existaient pas »<sup>12</sup>. En 2011, c'est Amit Singhal, le responsable du département Search Quality, qui revient à la charge : « Le conseil que nous donnons aux éditeurs est toujours le même : faites le nécessaire pour satisfaire au mieux les internautes qui visitent votre site web et ne vous préoccupez pas inutilement des algorithmes ou des paramètres utilisés par Google pour le classement<sup>13</sup>. » Google réclame l'invisibilité et se contente de recommander un ensemble de pratiques de bon sens qui permet d'être mieux vu par le moteur de recherche : optimiser les mots clés en vérifiant que sa page contient bien les termes les plus fréquemment utilisés par les utilisateurs, travailler le design du site afin qu'il soit clair et lisible, non seulement pour l'utilisateur, mais aussi pour le robot de Google, augmenter la vitesse de chargement du site. Ce que demande Matt

12. Centre d'aide pour les webmasters, « Consignes aux webmasters ».

<http://www.google.com/support/webmasters/bin/answer.py?answer=35769>.

13. Google Webmaster Central, « Informations supplémentaires concernant la création de sites de qualité » : <https://sites.google.com/site/webmasterhelpforum/informations-supplementaires-concernant-la-creation-de-sites-de-qualite>.

Cutts, le porte-parole de Google vers les webmestres, c'est une coopération entre les robots du moteur et le designer du site. Le PageRank a installé son ordre sur le web en domestiquant les techniques d'écriture des webmestres. Ceux-ci structurent leur site en fonction des subtilités du moteur, installant ainsi une intimité de plus en plus étroite entre leur expression et ce que sait en lire le moteur. Ils ont appris la manière dont le robot parcourait les pages du web : d'abord, l'URL, puis le titre et les sous-titres, l'importance des caractères en gras et en italique, le fait que le PageRank ne garde que la première ancre d'un lien lorsqu'une même page est citée plusieurs fois, l'importance des tags dans les fichiers images, l'impossibilité de crawler les fichiers pdf, etc. Cette intense familiarisation aux moindres procédures de l'algorithme est devenue un savoir-faire décliné en formations, en guides de conseil, et en outils de mesure de la Google-compatibilité des sites. La normalisation de l'écosystème formé par le web et son moteur dominant est devenue un enjeu industriel de grande importance. Mais Google donne aussi sa définition d'un site de qualité en mettant en avant des critères de rigueur, de rationalité et d'originalité directement empruntés à la déontologie de la documentation la plus traditionnelle. Véritable programme de rééducation destiné à faire oublier aux webmestres les mauvaises questions qu'ils se posent sur la manière d'être vu par l'algorithme, Google leur propose une liste de bonnes questions qui dessine, en creux, sa conception d'un site de qualité. Il faut d'abord que l'article soit écrit « par un expert ou une personne connaissant bien le sujet », une « autorité compétente » et que le site lui-même soit une « autorité reconnue dans le domaine abordé ». La meilleure manière de le savoir est de se demander : « Est-ce le genre de page que vous aimeriez ajouter à vos favoris, partager avec un ami ou recommander ? ». Il importe ensuite que les informations soient « fiables », que la « qualité du contenu [puisse être] contrôlée ». L'absence de fiabilité, souligne Google, laisse des traces que les robots ont appris à détecter. Il importe que le contenu ne soit pas copié « en double », ne contienne pas « des fautes d'orthographe, de style ou des faits inexacts » et qu'il livre « des contenus, des informations, des recherches, des analyses ou des rapports originaux ». Il est préférable d'offrir « quelque chose en plus par rapport aux autres », de donner « plusieurs points de vue sur ce qui s'est passé » en proposant « une description complète du sujet ». Austère et professoral, Google en vient même à réclamer « une analyse poussée ou des informations ayant demandé un certain travail de réflexion ». Car Google n'aime pas les contenus non « soignés », « rédigés à la va-vite », sans « beaucoup de soin et de rigueur », « inutiles, car trop courts, trop superficiels ou trop vagues », « produits en masse, récupérés auprès d'un grand nombre de sources externes », ou contenant « un nombre excessif d'an-

nonces qui distraient le lecteur ou l'empêchent d'accéder au contenu principal ». On ne saurait mieux exprimer cette vision documentaire de la qualité informationnelle qu'en se demandant *in fine* : « Pourriez-vous trouver cet article dans un magazine, une encyclopédie ou un livre en version papier ? »<sup>14</sup>. Vue depuis le Googleplex, la qualité de l'information numérique est encore et toujours à mesurer aux standards de l'édition papier.

### Des internautes calculateurs

Pour Google, la réputation se mérite ou s'achète et cette distinction préside à la séparation entre le référencement naturel et celui des *adwords* dont l'entreprise a fait un modèle économique très efficace. Épousant l'esprit des pionniers de l'Internet, ce clivage suppose aussi un partage clair entre le monde marchand des entreprises invitées à satisfaire leur désir de visibilité en achetant des mots clés et le monde non marchand des internautes qui ne calculent pas leur visibilité, mais s'échangent des liens sincères. Cependant, les internautes ne vont pas se montrer à la hauteur des vertus morales que leur prête le PageRank. Leur monde n'est pas si « naturel » et certains de leurs liens ne sont pas « sincères ». La distinction entre les mondes marchand et non marchand sépare moins deux populations différentes qu'elle ne traverse chaque internaute publiant en quête de réputation et de visibilité. Beaucoup d'entre eux ne cessent de calculer, cherchent à se faire voir et sont en compétition pour obtenir une place de choix dans les résultats « organiques » du moteur de recherche. Et pour cela, ils vont entreprendre de déformer à leur profit la structure des liens de la Toile, afin de capturer un surcroît de l'autorité dispensée par le PageRank. En agissant en fonction de l'algorithme, les internautes stratèges défont la position d'extériorité et d'invisibilité à laquelle prétend le PageRank, agissent réflexivement sur la structure du web (Espeland, 2007) et posent à Google un problème à la fois mathématique et moral. Dès lors que les jugements, *i.e.* les liens, ont été produits de façon stratégique, ils apportent une information biaisée qui érode la pertinence du résultat global de la recherche. Mais, à vouloir le corriger, Google est obligé de s'affranchir de son procéduralisme pour produire une définition substantielle de la qualité des liens et s'ériger en police du web.

---

14. Tous ces extraits de citations, sont repris dans Google Webmaster Central, « Informations supplémentaires concernant la création de sites de qualité » : <https://sites.google.com/site/webmasterhelpforum/informations-supplementaires-concernant-la-creation-de-sites-de-qualite>.

On peut retracer toute l'histoire de l'évolution de l'algorithme comme un jeu de chat et de souris entre les webmestres et la firme de Mountain View pour, d'une part, essayer d'agir stratégiquement sur le PageRank et, d'autre part, détecter et punir ces comportements en réformant l'algorithme avant que la détection d'une autre faille n'ouvre une nouvelle porte aux comportements stratégiques des internautes. Sans doute mesure-t-on mal les enjeux technologiques, marchands et moraux de cette guerre de basse intensité qui traverse toute l'histoire du web depuis que les moteurs sont devenus la principale voie d'accès à l'information numérique. Sans doute évalue-t-on aussi mal le rapport de forces entre le Goliath surpuissant et dominateur du web et les milliers de David bricoleurs et rusés qui l'assaillent de mille flèches. Car, de fait, sous bien des aspects, la position du PageRank face à ses assaillants est extraordinairement fragile et certains observateurs considèrent que Google a depuis longtemps perdu la bataille d'un classement optimal du web (Ippolita, 2011 ; Mowshowitz et Kawaguchi, 2002 ; Diaz, 2005 ; Granka, 2010). Le développement du marché du référencement (*Search Engine Optimization* – SEO) a transformé une partie du web en une gigantesque compétition des acteurs publiants pour se faire voir des algorithmes. Une partie de cette activité de conseil, appelée « white hat SEO », consiste à conformer les sites web aux robots de Google en produisant le code html le plus approprié (URL, ancre des liens, choix des mots clés, etc.). Mais un autre pan de l'activité du SEO (« black hat SEO ») consiste à vendre de la réputation. Se rendre visible, c'est obtenir des autres du « jus de lien ». Si celui-ci ne vient pas naturellement, il faut alors l'extorquer, l'acheter ou le produire artificiellement. Les techniques auxquelles se livrent les webmestres pour obtenir de la notoriété en produisant des liens factices n'ont cessé de se raffiner, au point de devenir une véritable industrie. En inscrivant d'abord leurs sites sur une galaxie d'annuaires, de catalogues ou d'index, les webmestres font venir à eux une série de liens. En cherchant, ensuite, à placer un lien vers leurs sites chez les autres, par exemple dans les commentaires de blogs réputés ou sur Wikipédia, les webmestres ont longtemps été aspirer du « jus de lien » chez les mieux dotés (pratique dite de *spamdexing*). Mais Google a rendu improductive cette pratique en créant la balise `<nofollow>` qui permet aux sites de démonétiser certains des liens qui partent de chez eux. L'encyclopédie Wikipédia, par exemple, est aujourd'hui entièrement en `<nofollow>` et ne distribue plus son autorité à ceux qu'elle cite. Un marché noir du lien s'est aussi ouvert permettant à deux sites de s'échanger des liens alors qu'ils n'ont aucune proximité, à un site de vendre des liens à un autre, de créer des « fermes de liens » organisant un véritable village Potemkine de faux sites liés entre eux pour se donner du Page-



Rank et redistribuant ensuite leur force capitalisée vers le site-client (*money site* dans le jargon du SEO). Au moment où l'amélioration de l'algorithme de Google a permis de détecter les fermes de liens, les truqueurs se sont lancés dans la constitution de « fermes de contenus » produisant à la chaîne, souvent à l'aide de robots linguistiques que corrigent des stagiaires sous-payés, des contenus ineptes à base de proximité synonymique destinés à envoyer un lien vers le site de l'entreprise cliente. Qualifiés de sites de CP (pour Communiqué de presse), ces plates-formes aux contenus informes ne sont en réalité écrites que pour les robots. Panda, la dernière évolution de l'algorithme, cherche précisément à déclasser ces sites à contenu faible et dupliqué.

Cette concurrence entre le marché des liens et l'algorithme met en tension deux principes contradictoires régissant la visibilité sur le web : l'*audience* et l'*autorité*. En défendant le lien naturel, le PageRank considère qu'il fait circuler des actes de reconnaissance sur la Toile et doit donc prendre source dans la qualité du texte citeur. En revanche, le marché des liens conçoit le lien hypertexte comme un pourvoyeur de trafic, un simple signalement de direction qui n'a pas besoin de s'ancrer dans un texte de qualité. Cette concurrence sur la définition de ce que met en circulation le lien hypertexte permet d'expliquer les évolutions successives de l'algorithme de Google. Révision après révision, l'algorithme exerce en effet un tri de plus en plus fin pour distinguer à l'intérieur des pages web les liens qui transportent de la reconnaissance (URL, titres, sous-titres, liens en gras, liens incorporés dans le contenu textuel de la page) et ceux qui en transportent moins ou pas (liens dans le paratexte de la page, liens commerciaux, liens en *<nofollow>*, etc.). Attaché à défendre une conception méritocratique de la force qui circule dans le lien hypertexte, Google a aussi entrepris de punir ceux qui créent des liens qui mettent en circulation de la « fausse autorité » en déclassant brutalement les sites qui trichent avec les règles fixées par Google, punition qui peut se révéler désastreuse pour les sites qui en sont victimes. Cette politique place cependant Google dans une position difficile qui met en tension deux éthiques contradictoires. Parce qu'elle est procédurale, la position du PageRank prétend s'exonérer de toute appréciation substantielle en imposant un formalisme abstrait sur le web (dénombrer les liens sans regarder leur contenu). Mais parce qu'elle défend une conception méritocratique du lien, elle est de plus en plus invitée à porter un jugement substantiel sur la nature des vraies et des fausses citations. En devenant le législateur et le policier des écritures du web, Google ne cesse de perdre sa position d'extériorité.

## LA CRISE DU PAGERANK

L'esprit du PageRank dont on vient d'esquisser les contours en explorant avec le plus de sympathie possible les justifications qui le nourrissent – principe de méthode nécessaire pour éviter de réduire trop rapidement, et avant analyse, les agissements de Google à ses intérêts économiques – est aujourd'hui en crise. La vertu d'*autorité* qu'il s'est attaché à promouvoir se trouve de plus en plus ébranlée par les tensions qu'exercent sur lui d'autres principes de classement de l'information : la *popularité* que poussent les logiques d'audience, l'*affinité* que lui oppose l'irrésistible ascension des réseaux sociaux et, enfin, l'*efficacité* mesurée par la satisfaction de l'internaute qui guide la personnalisation prédictive des algorithmes.

### De l'autorité à la popularité

La première tension tient à l'effet de révélation qu'a exercé le PageRank en rendant visible et mesurable le fait que la distribution des liens sur la Toile était extraordinairement inégale. L'imaginaire dont le web se nourrit projette sur lui la vision pastorale d'un graphe de petits producteurs s'échangeant entre eux des liens afin de se désigner mutuellement leurs meilleurs produits. Mais les liens entre les sites ne se répartissent pas selon un ordre étal et égal, réservant à tous une position d'égalité, au moins initiale, devant la possibilité de recevoir des liens des autres. Depuis 2000, les travaux sur la structure du web répètent inlassablement qu'il n'a rien d'un rhizome : un très petit nombre de pages attire un nombre considérable de liens, pendant que la très grande majorité des sites sont liés à très peu de sites et ne sont souvent cités par aucun (Broder *et al.*, 2000 ; Adamic et Huberman, 2001). Le constat est sans appel : 90 % du PageRank du web est possédé par 10 % des sites (Pandurangan *et al.*, 2006). « Les hubs », souligne cruellement Albert-Laszlo Barabási (2002, p. 58), « sont l'argument le plus fort contre la vision utopique d'un cyberspace égalitaire. Oui, nous avons tous le droit de mettre tout ce que nous voulons sur le web. Mais qui va le remarquer ? [Les hubs] sont très faciles à trouver, quel que soit l'endroit où vous vous trouvez sur le web. Mais face à ces hubs, le reste du web est invisible. » La distribution de l'autorité en loi de puissance n'est pas seulement rendue visible par le PageRank, elle est aussi renforcée par lui en raison d'un ensemble d'effets de concentration, d'asymétrie et de hiérarchisation propres aux structures en réseau. Le plus connu est l'« effet Mathieu » (*Matthew effects*) dégagé par Robert Merton (1968) en scientométrie, qui montre que le système de reconnaissance de la communauté scientifique contribue à

ce que les « scientifiques renommés reçoivent une somme de reconnaissance disproportionnée pour leurs contributions, alors que la reconnaissance est ôtée aux scientifiques junior qui n'ont pas un nom établi ». Les « riches », gros acteurs du web et sites disposant d'un capital de popularité très fort à l'extérieur du web (entreprises, médias, institutions), deviennent encore plus riches (*rich get richer*), parce que la visibilité qu'ils acquièrent sur le réseau attire mécaniquement à eux de nouveaux liens surnuméraires. Les nœuds qui reçoivent le plus de liens exercent un effet de halo qui conduit les autres nœuds à rechercher activement leur proximité pour emprunter un peu de leur force ; phénomène qui conduit à donner une autorité imméritée à certains (phénomène souvent qualifié de *winners takes all*) tout en laissant dans l'ombre de nombreux méritants. Cet effet se renforce aussi des mécanismes d'« attachement préférentiel » qui poussent les sites à citer des sites ayant une autorité égale ou supérieure à la leur et à refuser de citer plus petits qu'eux (Cardon *et al.*, 2011). La conséquence de ces effets de renforcement est que lorsque l'on observe le haut du classement des sites effectué par les moteurs de recherche, l'autorité (mesurée par le nombre de liens) se confond avec la popularité (le nombre de clics d'internautes) (Hindman *et al.*, 2003)<sup>15</sup> : les sites des entreprises, des grands médias et des institutions, mais aussi des acteurs centraux du web comme Wikipédia, reçoivent de la reconnaissance (du jus de lien) autant que de l'audience (des clics), sans qu'il ne soit possible de déterminer la variable qui a agi sur l'autre. Le lien hypertexte n'enfermerait pas cette autorité-prééminence du jugement citationnel que lui conféraient les pionniers, mais la simple attention-réflexe que commandent les mécanismes mimétiques de la publicité. L'autorité méritocratique ne serait alors que le cache-sexe de l'autorité statutaire des puissants qu'ils doivent à leur centralité dans la vie sociale et à leur capital économique (Diaz, 2005). En dominant la hiérarchie des liens, ces puissants imposent aussi un classement qui donne une visibilité excessive aux sites centraux, dans la moyenne, conformistes, sans controverses ni originalité.

### Les réseaux sociaux et le *ranking* des personnes

La deuxième tension est une conséquence de la démocratisation de la participation des internautes, rendue possible par le développement de techniques de publication ne requérant qu'un très faible coût d'engagement (Cardon, 2010). Élitiste, le PageRank n'accorde qu'aux internautes publiants (*i.e.* produisant

---

15. La superposition des classements d'autorité et d'audience ne vaut cependant que pour la tête des listes de résultats. Ils se distinguent dès que l'on quitte le sommet des classements ou que les mots clés deviennent plus complexes ou moins centraux (Pennock *et al.*, 2002).

des liens hypertextes) le droit de participer au classement de l'information sur le web. Avec la massification des usages de l'Internet, ce monopole des publiants sur l'ordre de l'information se trouve de plus en plus contesté. Le développement des nouveaux formats d'écriture conversationnelle sur les réseaux sociaux numériques (phrases de statuts, commentaires, boutons « I like », « +1 », et outils de partage comme le RT de Twitter) a rapproché l'acte de publication d'un simple geste de retour de l'audience. Ces nouvelles formes expressives ont donné des droits nouveaux à des publics plus jeunes, davantage dispersés géographiquement et plus « populaires » que les producteurs méritants de liens hypertextes. Mais il organise aussi un tout autre principe de classement de l'information. Alors que le PageRank mesure les liens entre les documents, l'EdgeRank de Facebook classe les documents en fonction des jugements subjectifs que s'échangent des personnes liées par une relation d'affinité. Au lieu d'effacer la personne derrière son texte, l'énonciation conversationnelle des réseaux sociaux, assouplie, relâchée et immédiate, a conféré de la visibilité à la subjectivité des personnes pour faire de leur jugement un signal identitaire que les individus projettent vers leur sociabilité (Cardon, 2013). Alors que dans le web des documents, la force intrinsèque, illocutoire, du lien est déposée dans l'autorité de la page du texte citeur, dans le cas du web des personnes, c'est l'autorité numérique de l'énonciateur, sa e-réputation, qui appuie son énonciation. Les métriques d'affinité du web social distribuent vers les documents qu'elles classent une autorité qui s'enracine dans les personnes que voulait effacer le PageRank.

### Quand l'appareil devient machine

La dernière tension qui pèse sur l'esprit du PageRank a trait aux transformations que les ingénieurs de Google ne cessent d'apporter à l'algorithme dans leur combat contre les stratèges du référencement et le marché noir du lien. À force de révisions et de réglages, l'algorithme de Google apparaît de moins en moins comme un *appareil* posé sur le web pour l'enregistrer et de plus en plus comme une *machine* que pilote avec une précision toute stratégique l'équipe Quality Search d'Amit Singhal<sup>16</sup>. Sous l'effet des multiples pressions que connaît Google en raison de ses ambitions commerciales et de sa place dominante sur le marché des moteurs, l'entreprise est de plus en plus conduite, bien qu'elle s'en défende, à « agir à la main » sur les résultats de son algorithme et à faire le deuil de son souci de neutralité non interventionniste. Sous

---

16. Sur la distinction entre appareil et machine, voir Citton (2013).

la pression des juridictions nationales, elle a dû censurer certains sites racistes et antisémites en France et en Allemagne (Zittrain et Edelman, 2002). Sous la pression des industries culturelles, elle a été amenée à déclasser les résultats de recherche conduisant vers des sites proposant des contenus piratés (Menell, 2012). Sous la pression des entreprises, elle a accepté de refuser à leurs concurrents l'achat de certains *adwords*. Sous la pression de l'industrialisation des triches destinées à produire artificiellement du « jus de liens », Google s'est affranchi de sa position d'extériorité pour faire sa police en filtrant et en punissant les contrevenants. L'entreprise a même récemment dû mettre en place un dispositif de délation permettant aux internautes de dénoncer les sites fabricant de la réputation. Il ne fait guère de doute que l'idéal mécanique d'un appareillage de règles enregistrant le web pour le classer en aura été fortement ébranlé. Si Google a sacrifié sa position d'extériorité parce que les internautes se préoccupaient trop de lui, il est indéniable que la stratégie industrielle de Google, et notamment le fait que la firme de Mountain View ait développé de nombreux autres services dont la commercialisation, peut entrer en conflit avec la logique de neutralité du moteur de recherche contribue à rendre de plus en plus fragile, et de plus en plus rhétorique, la revendication de neutralité de la recherche naturelle. Par ailleurs, l'algorithme de Google incorpore de plus en plus des technologies dites d'apprentissage (*machine learning*) permettant de calculer les classements présentés à l'utilisateur. Ainsi, il n'est désormais plus nécessaire de fixer les multiples paramètres qui donnent du poids à tel ou tel des signaux, notamment le PageRank, extraits de chaque page du web pour être déposés dans l'Index de Google. Il suffit de laisser les techniques d'apprentissage ajuster au cas par cas ces paramètres en fonction des requêtes, de ce que Google sait des pratiques antérieures de l'utilisateur, de l'acquisition de connaissance permise par les liens cliqués par les autres internautes pour une même requête (Granka, 2010) et *in fine* des jugements humains sur la pertinence des sites recueillis par les *quality rater* embauchés par Google (Pot-PieGirl, 2011). Au principe d'autorité qui a fait la force du PageRank, Google substitue de plus en plus un principe d'*efficacité* qui renvoie de manière toujours plus appropriée vers l'internaute les choix que l'algorithme a appris de ses comportements<sup>17</sup>. De fait, la machine inventée par Google est devenue si complexe, si sensible aux tests statistiques qui ne cessent de la reparamétrer, si dévoreuse de variables et de traces, si auto-apprenante, que ses comportements ne peuvent plus désormais être compris et interprétés, pas même par ses géniteurs.

---

17. Sur la prétention des algorithmes d'apprentissage à gouverner le réel en l'épousant dans ses moindres plis, voir l'article de T. Berns et A. Rouvroy dans ce numéro.

---

 RÉFÉRENCES
 

---

- ADAMIC L., HUBERMAN B., 2001, "The Web's Hidden Order", *Communications of the ACM*, vol. 44, n° 9, pp. 55-60.
- ANDLER D., 2012, "What has collective wisdom to do with wisdom?", in H. LANDEMORE et J. ELSTER, eds, *Collective Wisdom: Principles and Mechanisms*, Cambridge, Cambridge University Press.
- BARABÁSI A.-L., 2002, *Linked: The New Science of Networks*, Cambridge, Perseus Publication.
- BATELLE J., 2005, *The Search. How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*, New York, Portfolio, 2005.
- BENKLER Y., 2009, *La richesse des réseaux. Marchés et libertés à l'heure du partage social*, Lyon, Presses universitaires de Lyon.
- BRIN S., PAGE L., 1998, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", *Proceedings of the seventh international conference on World Wide Web*.
- BRODER A., KUMAR R., MAGHOUL F., RAGHAVAN P., RAJAGOPALAN S., STATA R., TOMPKINS A., WIENER J., 2000, "Graph Structure in the Web", *Computer Networks*, vol. 33, n° 16, pp. 309-320.
- CARDON D., 2010, *La démocratie Internet. Promesses et limites*, Paris, Seuil/République des idées.
- CARDON D., 2011, « L'ordre du Web », *Médium*, n° 29, octobre-décembre, pp. 191-202.
- CARDON D., 2013, « Du lien au like. Deux mesures de la réputation sur Internet », *Communication*, à paraître.
- CARDON D., FOUETILLOU G., ROTH C., 2011, "Two paths of glory. Structural position and trajectories of websites within their topical community", *ICWSM 2011*, Barcelone, 17-21 juillet.
- CASSIN B., 2007, *Google-moi. La deuxième mission de l'Amérique*, Paris, Albin Michel.
- CITTON Y., 2013, « Le retour de l'objectivité ? », *La Revue des livres*, n° 9, janvier-février, pp. 3-12.
- DASTON L., GALISON P., 2012, *Objectivité*, Paris, Les Presses du réel.
- DIAZ A. M., 2005, *Through the Google Goggles: Sociopolitical Bias in Search Engine Design*, Thesis, Stanford University, May.
- EDWARDS D., 2011, *I'm Feeling Lucky. The Confession of Google Employee Number 59*, London, Allen Lane/Penguin Books.

- ESPELAND W. N., 2007, "Rankings and Reactivity: How Public Measures Recreate Social Worlds", *American Journal of Sociology*, vol. 113, n° 1, juillet, pp. 1-40.
- FARRELL H., DREZNER D. W., 2008, "The Power and Politics of Blogs", *Public Choice*, 134, pp. 15-30.
- FORSYTH E., KATZ L., 1946, "A Matrix Approach to the Analysis of Sociometric Data: Preliminary Report", *Sociometry*, n° 9, 1946, pp. 340-347.
- FOUCAULT M., 2001, « Qu'est-ce qu'un auteur ? », *Dits et écrits I, 1954-1975*, Paris, Gallimard/Quarto, pp. 817-849.
- GARFIELD E., 1955, "Citation indexes for science. A new dimension in documentation through association of ideas", *Science*, n° 122.
- GERHART S., 2004, "Do Web Search Engines Suppress Controversy?", *First Monday*, vol. 9, n° 1.
- GOODIN R. E., 2003, "Democratic deliberation within", in J. FISHKIN, P. LASLETT, eds, *Debating Deliberative Democracy*, Malden, Blackwell Publishing.
- GRANKA L. A., 2010, "The Politics of Search: A Decade Retrospective", *The Information Society*, 26, 2010, pp. 364-374.
- GRIMMELMANN J., 2009, "The Google Dilemma", *New York Law School Law Review*, vol. 53.
- HINDMAN M., 2008, "What is the Online Public Sphere Good For ?", in J. TUROW, L. TSUI, eds, *The Hyperlinked Society*, Chicago, University of Michigan Press.
- HINDMAN M., 2009, *The Myth of Digital Democracy*, Princeton, Princeton University Press.
- HINDMAN M., TSIOUTSILOULIKLIS K., JOHNSON J. A., 2003, "« Googlearchy »: How a Few Heavily-Linked Sites Dominates Politics on the Web", Paper presented at the annual meeting of the Midwest Political Science Association.
- IPPOLITA (2011, *Le côté obscur de Google*, Paris, Rivages, 2011.
- KLEINBERG J., 1998, "Authorative Sources in a Hyperlinked Environment", *Proc. 9th ACM-SIAM Symposium on Discrete Algorithms*, 1998. Also appears as IBM Research Report RJ 10076, May 1997.
- KLEINBERG J., LAWRENCE S., 2001, "The Structure of the Web", *Science*, 294, novembre.
- KYROU A., 2010, *Google God. Big Brother n'existe pas, il est partout*, Paris, Inculte.
- LANDEMORE H., 2010, « La raison démocratique : Les mécanismes de l'intelligence collective en politique », *Raisons Publiques*, n° 12, pp. 9-55.
- LANDEMORE H., ELSTER J., 2012, eds, *Collective Wisdom: Principles and Mechanisms*, Cambridge, Cambridge University Press.

LANGVILLE A. N., MEYER C. D., 2006, *Google's PageRank and beyond: The science of search engine rankings*, Princeton, Princeton University Press.

LÉVY P., 1991, « L'hypertexte, instrument et métaphore de la communication », *Réseaux*, n° 46-47, pp. 59-68.

LEVY S., 2011, *In the Plex. How Google Thinks, Works and Shapes our Lives*, New York, Simon & Schuster.

MAYER K., 2009, "On the sociometry of Search Engine. A Historical Review of Methods", in K. BECKER, F. STALDER, eds, *Deep Search. The politics of search beyond Google*, Innsbruck, StudienVerlag, 2009, pp. 54-72.

MENELL P. S., 2012, "Google, PageRank and Symbiotic Technological Change", UC Berkeley Public Law Research Paper, n° 2136185, August 26.

MERTON R., 1977, "The sociology of science: An episodic memoir", in R. K. MERTON, J. GASTON, eds, *The sociology of science in Europe*, Southern Illinois University Press, Carbondale, pp. 3-141.

MERTON R. K., 1968, "The Matthew Effect in Science", *Science*, vol. 159, n° 3810, 1968, p. 56-63.

MERTON R. K., 1996, 1<sup>re</sup> ed. 1957, "The Reward System of Science (1957)", in R. MERTON, P. SZTOMPKA, eds, *On Social Structure and Science*, Chicago, Chicago University Press, pp. 286-304.

MOULLIER-BOUTANG Y., REBISCOUL A., 2009, « Peut-on faire l'économie de Google ? », *Multitudes*, n° 36, pp. 83-93.

MOWSHOWITZ A. KAWAGUCHI A., 2002, "Bias on the Web", *Communications of the ACM*, vol. 45, n° 9.

ORIGGI G., 2008, « Sagesse en réseaux : la passion d'évaluer », *La Vie des Idées*, 30 septembre.

PAGE L., 1998, "Method for node ranking in a linked database", Patent #6285999, 9 janvier.

PAGE S., 2006, *The Difference. How the Power of Diversity Creates Better groups, Firms, Schools and Societies*, Princeton, Princeton University Press, 2006.

PANDURANGAN G., RAGHAVAN P., UPFAL E., 2006, Pandurangan (Gopal, "Using PageRank to Characterize Web Structure", *Internet Mathematics*, vol. 3, n° 1, pp. 1-20.

PASQUINELLI M., 2009, "Google's PageRank. Diagram of the Cognitive Capitalism and Rentier of the Common Intellect", in K. BECKER, F. STALDER, eds, *Deep Search. The Politics of Search beyond Google*, Innsbruck, StudienVerlag.

PENNOCK D. M., FLAKES G. W., LAWRENCE S., GLOVER E. J., GILES C. L., 2002, "Winners don't take all: Characterizing the competition for links on the web", *Proceedings of the National Academy of Sciences*, vol. 99, n° 8, april, pp. 5207-5211.



- PINSKI G. NARIN F., 1976, "Citation Influence for Journal Aggregates of Scientific Publications", *Information Processing and Management*, 12, pp. 297-312.
- PONTILLE D., TORNAY D., 2013, « La manufacture de l'évaluation scientifique : algorithmes, jeux de données et outils bibliométriques », *Réseaux*, n° 177, pp. 25-62.
- PORTER T. M., 1995, *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*, Princeton, Princeton University Press.
- POTPIEGIRL, 2011, "Google Raters. Who are They?", *PotPieGirl.com*, 17 novembre.
- RIEDER B., 2012, "What is in PageRank? A historical and conceptual investigation of a recursive status index", *Computational Culture. A Journal of software studies*, n° 2, 28 septembre.
- SCHMIDT E., 2004, *Keynote Address at the 2004 Conference on Entrepreneurship*, Stanford University Graduate School of Business.
- SHEFF D., 2004, "Playboy Interview: Google Guys", *Playboy*, vol. 51, n° 9, September, pp. 55-60.
- SINGHAL A., 2008, "Introduction to Google ranking", *Official Google Blog*, 9 juillet.
- SUNSTEIN C. R., 2006, *Infotopia. How Many Minds Produce Knowledge*, New York, Oxford University Press.
- SUROWIECKI J., 2008, *La sagesse des foules*, Paris, Jean-Claude Lattès.
- VAIDHYANATHAN S., 2011, *The Googlization of Everything (and why we should worry)*, Berkeley, University of California Press.
- WISE D. A., MALSEED M., 2006, *Google Story. Enquête sur l'entreprise qui est en train de changer le monde*, Paris, Dunod.
- WOUTERS P., 1999, *The citation culture*, doctoral thesis, University of Amsterdam.
- WOUTERS P., 2006, « Aux origines de la scientométrie. La naissance du *Science Citation Index* », *Actes de la recherche en sciences sociales*, n° 164.
- ZITTRAIN J., EDELMAN B., 2002, "Localized Google Search Result Exclusions", Berkman Center for Internet & Society at Harvard Law School, 26 octobre.