



HAL
open science

La bibliométrie au service des Sciences de l'Information et de la Communication

Hervé Rostaing

► **To cite this version:**

Hervé Rostaing. La bibliométrie au service des Sciences de l'Information et de la Communication. Captas, documentation et numérique, Institut Méditerranéen des Sciences de l'Information et de la Communication (Aix-Marseille Université), Jun 2021, Marseille, France. hal-03586140

HAL Id: hal-03586140

<https://amu.hal.science/hal-03586140v1>

Submitted on 23 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

La bibliométrie au service des SIC

Hervé Rostaing

Captas, documentation et numérique
Séminaire IMSIC - axe 1 – 18 juin 2021

Origine du terme Bibliométrie

- Etymologie : mesure (métrie) du livre (biblio) et de l'écrit
- Première apparition du terme : 1934, Paul Otlet dans son *Traité de documentation*
 - La bibliométrie définit comme le volet métrique de la bibliologie (science du livre)
 - Etude quantitative pour aboutir à une analyse objective des phénomènes
- Résurgence du terme en 1969
 - En France, R Estivals reprend le terme selon la définition de Otlet dans l'article *La statistique bibliographique*, *BBF*, 14(12)
 - En Angleterre, A Pritchard propose l'emploi du terme au lieu de l'expression *Statistical bibliography* jugé trop ambiguë dans l'article *Statistical bibliography or bibliometrics*, *Journal of Documentation*, 25(4)
- Origine commune : la statistique bibliographique

Bibliographie, bibliologie et bibliométrie

Propositions de P Otlet et R Estivals

- La bibliologie = science de la communication écrite
 - Communication écrite = processus d'inscription de la pensée intellectuelle et artistique
 - Bibliologie = permet l'étude de la création intellectuelle et artistique
- L'étude des phénomènes de la communication écrite implique 2 phases
 - L'observation → inventorier, décrire, classifier → connaissance descriptive (graphie) → la bibliographie
 - La compréhension → expliquer les phénomènes → connaissance théorique (logie) → la bibliologie
- La bibliométrie est le pont entre la bibliographie et la bibliologie
 - fournit les résultats statistiques répondant aux besoins d'explication de la bibliologie

Les bibliographies spécialisées

Propositions de R Estivals

- Les bibliographies nationales, régionales
 - Étude des créations intellectuelles internationales, nationale ou régionales
 - *Application bibliométrique = observatoires nationaux des sciences et techniques*
- Les bibliographies rétrospectives
 - Etude statistique synchronique
 - *Application bibliométrique : étude de l'activité scientifique d'une université, d'un laboratoire, étude d'un domaine scientifique...*
- Les bibliographies courantes
 - Etude des fluctuations de la création intellectuelle
 - *Application bibliométrique : veille scientifique et technique*

Les bibliographies signalétiques

Propositions de R Estivals

- Etudier les acteurs de la communication écrite
 - Les auteurs
 - Étude des contributeurs à la création intellectuelle
 - *Application bibliométrique = la loi de Lotka*
 - Les éditeurs
 - Étude des diffuseurs à la création intellectuelle
 - *Application bibliométrique = la loi de Bradford*
- Autres données bibliographiques
 - Etudier les supports de diffusion :
 - type de support, nombre de pages
 - Etudier le contenu des écrits
 - Titres, résumés, descripteurs documentaires, textes complets, sous-partie des textes complets
 - Unité de découpage : caractères, mots, phrases, expressions, paragraphes, chapitres, tables des matières, index, citations, bibliographie...

Les bibliographies particulières

Propositions de R Estivals

- Etudier les catalogues de bibliothèque
 - Étude de la politique d'acquisition
- Etudier les registres de ventes des éditeurs et librairies
 - Etude de l'orientation des achats et des goûts des lecteurs
- Etudier les registres des prêts des bibliothèques
 - Etude de la circulation des connaissances

Acception internationale de la bibliométrie

- Définition de A Pritchard
“the application of mathematics and statistical methods to books and other media of communication”
- Très proche de celle de Otlet/Estivals
- Mais en pratique la bibliométrie est
 - surtout connue comme l'étude quantitative des écrits scientifiques (ou techniques)
 - parfois confondue avec la scientométrie (l'étude quantitative des sciences ou techniques)

Périmètres des différentes études quantitatives portant sur des activités de recherche scientifique (ou technique)

Proposition de Hausteine Stefanie (2016)

Hausteine, S. (2016). Grand challenges in altmetrics : Heterogeneity, data quality and dependencies. *Scientometrics*, 108(1), 413-423. <https://doi.org/10.1007/s11192-016-1910-9>

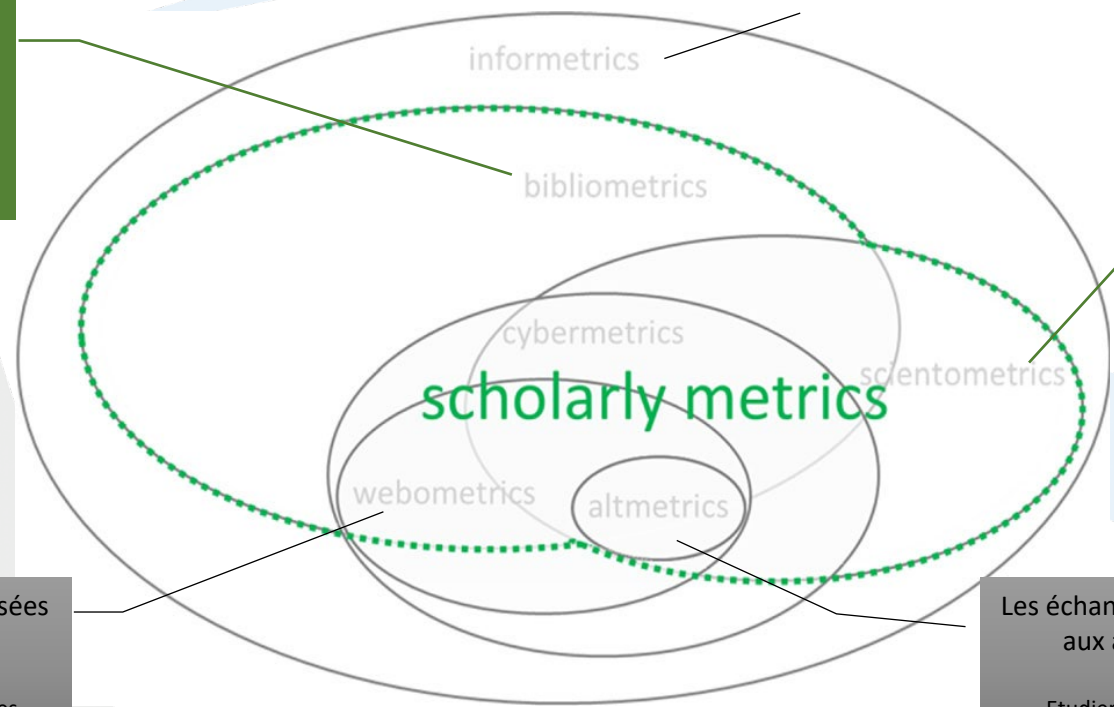
Les écrits scientifiques
Etudier la production académique des sciences

Toutes données représentant une information
Etudier tout phénomène informationnel

Les activités scientifiques
Etudier les processus de construction et d'évolution des sciences

Les données diffusées sur le web
Etudier l'activité des contributions au web

Les échanges sur internet consacrés aux activités scientifiques
Etudier l'impact des résultats de la recherche dans la sphère internet



Bibliométrie vs scientométrie



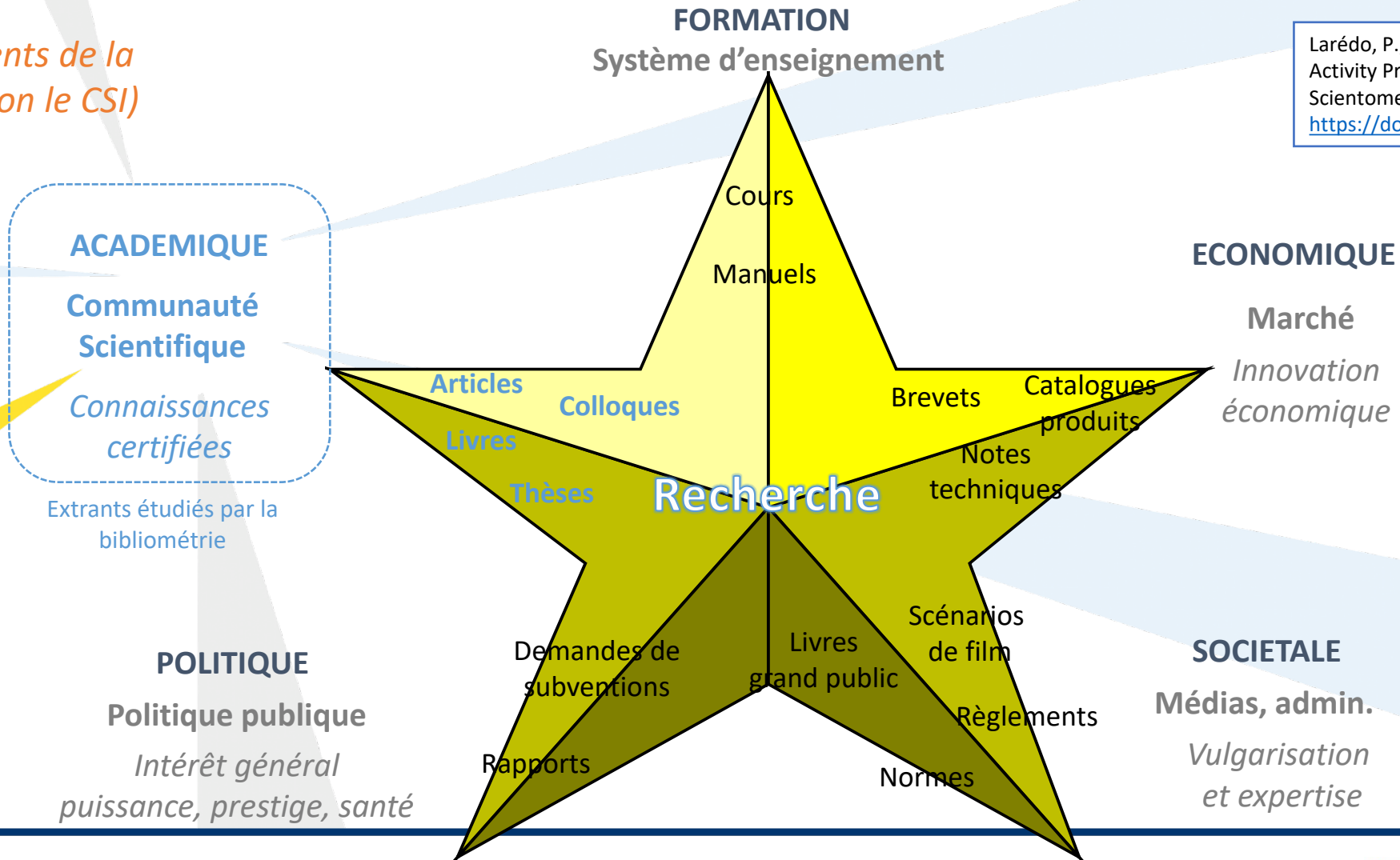
Modèle de référence des études quantitatives des activités de recherche

- La scientométrie a pour ambition d'étudier la science
 - donc se doit d'étudier toutes les activités de recherche
 - en prenant en considération les intrants et les extrants
- La bibliométrie n'étudie qu'un type d'extrant de la recherche
 - les écrits scientifiques

Diversité d'extrants produits par la recherche

La rose des vents de la recherche (selon le CSI)

Larédo, P., & Mustar, P. (2000). Laboratory Activity Profiles : An Exploratory Approach. *Scientometrics*, 47(3), 515-539.
<https://doi.org/10.1023/A:1005671901143>



Variété des études bibliométriques

- La bibliométrie comme instrument statistique : *les études des précurseurs*
- La bibliométrie comme instrument de sociologie des sciences : *la scientométrie*
- La bibliométrie comme instrument d'évaluation de la recherche : *le règne des indicateurs*
- La bibliométrie comme instrument de cartographie des sciences : *l'émergence de la cartographie de l'information*
- La bibliométrie comme instrument d'exploration des informations scientifiques : *la cartographie pour l'aide à l'analyse*

Les précurseurs : bibliométrie comme instrument statistique

- Les premières études quantitatives ont cherché à établir des « lois »
 - Etudier la régularité de distributions statistiques
 - Lotka en 1926
 - Etudie la régularité de la contribution des chercheurs à une domaine scientifique
 - Bradford en 1934
 - Etudie la régularité de la contribution des revues scientifiques à la diffusion d'articles traitant d'un domaine scientifique

Loi de Lotka (1926)

- Objectif :
 - Mesurer la part de contribution de chaque chercheur au progrès de la science
- Travaux :
 - 2 séries statistiques expérimentales
 - 2 sources : index de chimie (CAS entre 1907-16) et index de physique (Geschichtstafeln der Physik de Auerbach, 1910)
 - 2 bibliographies des auteurs (commençant par A et B pour CAS et tous pour l'index de physique)
 - 2 distributions des effectifs des auteurs ayant publié le même nombre d'articles (classe de production)
 - Histogramme et graphe logarithmique des 2 distributions
- Loi :
 - graphe logarithmique de la forme $Y = X^n \times C$ avec $n=-2$ pour ces données
 - donc, loi de l'inverse du carré : $Y = C / X^2$
- Signification
 - Ce qui signifie que si N auteurs ont 1 article, il y en a 4 fois moins avec 2 articles, 9 fois moins avec 3 articles...
 - le nombre d'articles publiés n'est pas distribué de manière homogène entre les auteurs. La production scientifique a tendance à se concentrer sur un nombre limité de chercheurs

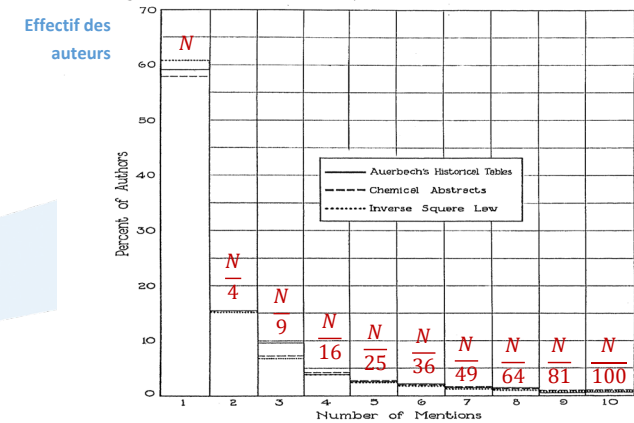


Fig. 1.—Frequency diagram showing per cent of authors mentioned once, twice, etc., in Auerbach's *Geschichtstafeln der Physik*, entire alphabet, and in the decennial index of *Chemical Abstracts* 1907-1916, letters A and B. The dotted line indicates frequencies computed according to the inverse square law.

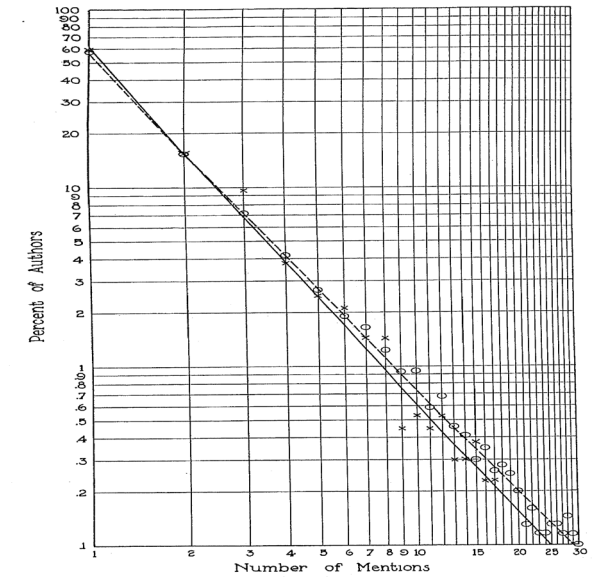


Fig. 2.—Logarithmic frequency diagram showing number of authors mentioned once, twice, etc., in Auerbach's tables (points indicated by crosses), and in *Chemical Abstracts*, letters A and B (points indicated by circles). The fully drawn line indicates points given by inverse square law, exponent = 2; the line of dashes corresponds to exponent 1.89.

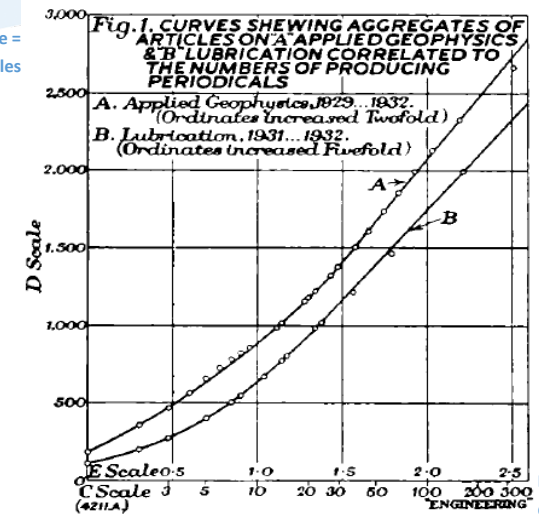
LOTKA, A. J. (1926). The Frequency Distribution of Scientific Productivity. *Journal of the Washington Academy of Sciences*, 16(12), 17-2.

La loi de Bradford (1934)

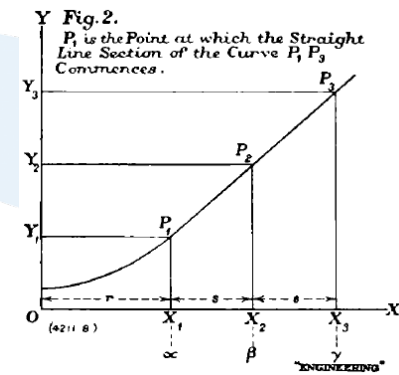
Nb revues (publiant le même nombre d'articles)	Nb articles correspondant	Nb cumulé des revues	[D] Nb cumulé d'articles	[E] Log(nb cumulé des revues)
n_1	f_1	$c_1=n_1$	$d_1=f_1*n_1$	$\log(c_1)$
n_2	f_2	$c_2=c_1+n_2$	$d_2=d_1+f_2*n_2$	$\log(c_2)$

- Objectif :
 - Gestion des abonnements
 - Connaître le «noyau» des revues publiant le plus d'articles dans un domaine scientifique
- Travaux :
 - 2 séries statistiques expérimentales
 - 2 domaines : géophysique appliquée et lubrification
 - 2 bibliographies produites par la bibliothèque du Musée des sciences de Londres
 - 2 distributions des effectifs des revues ayant publié le même nombre d'articles (classe de production)
 - Calcul des effectifs cumulés des revues et du nombre d'articles publiés correspondant pour les 2 séries
 - Graphe du nombre d'articles selon logarithme des effectifs cumulés $D = f(E)$
- Loi :
 - noyau défini par le point (X_1, Y_1)
 - pour $Y_2 = 2 \times Y_1$ alors $X_2 = X_1^2$
 - affaissement de Groos
- Signification
 - les articles publiés dans un domaine ne sont pas distribués de manière homogène dans les revues scientifiques.
 - La production scientifique dans un domaine a tendance à se concentrer sur un nombre limité de revues

Mesure =
nombre d'articles

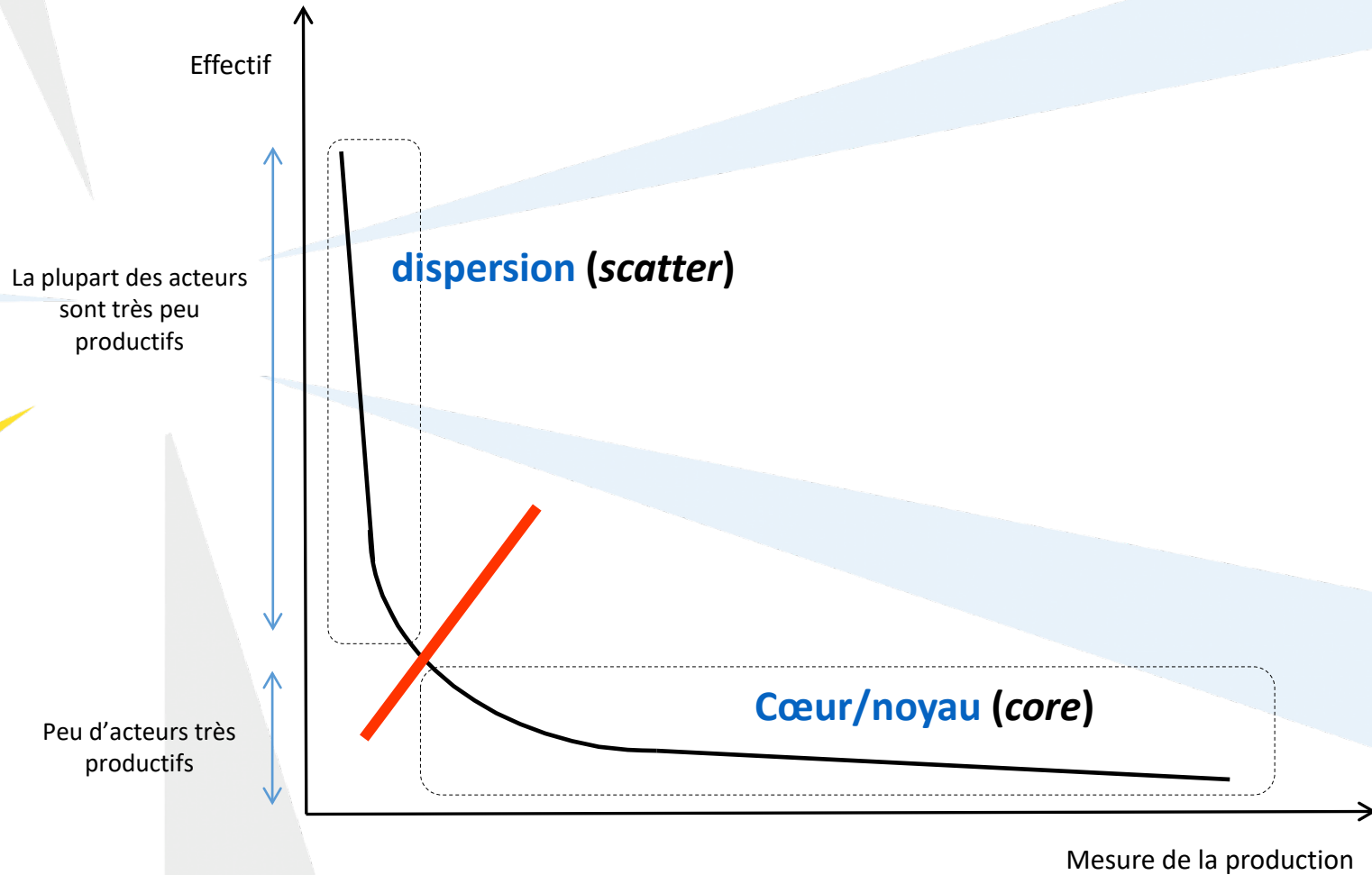


Log de l'effectif
des revues



Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering. An Illustrated Weekly Journal*, 137(3550), 176-180. <https://doi.org/10.1177/016555158501000407>

La production scientifique suit des lois hyperboliques



Apport de ces premières études

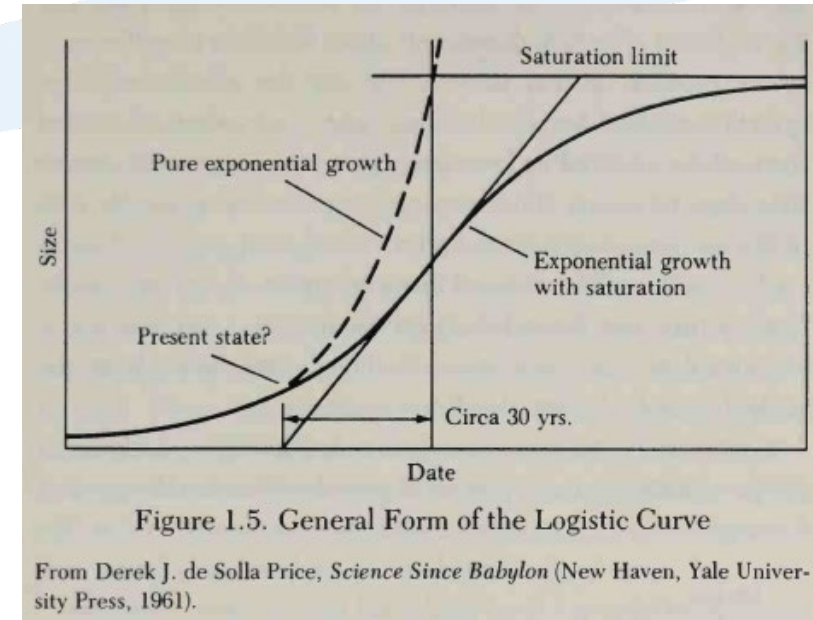
- Les données bibliométriques suivent des régularités respectant des lois hyperboliques
 - Répartition très peu homogène de la production
 - Production concentrée dans un nombre restreint d'acteurs (Pareto)
- Conséquences de ces travaux
 - Interprétation productiviste et élitiste de ces distributions : découpage en 2 parties **Cœur / Dispersion**
 - **Pas d'indicateurs synthétiques** de ces régularités qui faciliteraient leur interprétation et comparaison (la moyenne et l'écart type n'ont pas de sens)
 - ➔ Toutes les méthodes statistiques ne sont pas applicables
 - ➔ **Les quantiles** sont les seuls indicateurs pour représenter ces distributions

Variété des études bibliométriques

- La bibliométrie comme instrument statistique : *les études des précurseurs*
- La bibliométrie comme instrument de sociologie des sciences : *la scientométrie*
- La bibliométrie comme instrument d'évaluation de la recherche : *le règne des indicateurs*
- La bibliométrie comme instrument de cartographie des sciences : *l'émergence de la cartographie de l'information*
- La bibliométrie comme instrument d'exploration des informations scientifiques : *la cartographie pour l'aide à l'analyse*

Les travaux de Price (« Père » de la scientométrie)

- Objectif :
 - Définir les modèles qui régissent l'évolution et la construction de la science -> la « science de la science »
 - Etudier l'impact de la production de connaissances sur la construction de la science → mesure de l'impact par la citation
- Travaux :
 - **Structure cumulative de la science** (1962)
 - **Croissance exponentielle de la connaissance** puis saturation logistique (1963)
 - **Les collègues invisibles** (1966) : « l'élite » des scientifiques qui contribueraient à la partie de la recherche la plus avancée (les fronts de recherche)
 - **Théorie de l'avantage cumulé** : le succès engendre le succès alors que l'échec n'a aucun effet -> formule mathématique unificatrice des lois hyperboliques (1976)



La bibliométrie comme instrument de sociologie des sciences



Besoin d'outils appropriés

Price, D. de S. (1963). *Little Science Big Science*. Columbia University Press.
<http://derekdesollaprice.org/little-science-big-science-full-text/>

La priorité est donnée à la métrique de la citation

- Travaux de Price :
 - Ne plus mesurer uniquement la production scientifique mais aussi son impact
 - étudier la citation entre les travaux scientifiques
 - disposer des données pour effectuer ces études
- Collaboration avec E Garfield fondateur de l'ISI (*Institute for Scientific Information*)
 - Le SCI (*Sciences Citation Index*) commercialisé en 1964
 - Restriction aux travaux publiés dans les revues scientifiques
 - Restriction au cœur de la science → uniquement les revues à plus forte notoriété dans chaque domaine
 - SCI(E) : en 1964, 613 revues rangés dans en 128 domaines
en 2019, 9 200 revues rangés dans en 245 domaines
 - Puis SSCI en 1973, A&HCI en 1978, et récemment CPCI et ESCI
 - Comment ces revues sont sélectionnées ? → le JCR
- Le JCR (*Journal Citation Report*) : **plusieurs indicateurs de mesure de l'impact des revues**
 - Indicateur d'immédiateté
 - Facteur d'impact
 - Facteur d'impact sans autocitation
 - Facteur d'impact à 5 ans
 - La demi-vie des citations : l'âge médian des publications d'une revue citées au cours d'une année
 - Répartition des revues selon les quartiles Q_1 , Q_2 , Q_3 de leur facteur d'impact
- L'ESI (*Essential Science Indicators*)
 - Calcul des premiers indicateurs de citation mondiaux → base de comparaison mondiale dans chaque domaine

Le JCR – étudier l'impact des revues

Les indicateurs de citation dépendent du domaine

Indicateurs	Valeurs du JCR 2019 pour <i>Scientometrics</i>	Valeurs du JCR 2019 pour <i>Lancet</i>	
II	0,548	15,011	Les publications 2019 de <i>Lancet</i> sont 27 fois plus citées en 2019
JIF	2,867	60,390	Les publications 2017-18 de <i>Lancet</i> sont 21 fois plus citées en 2019
JIF sans autocitation	2,279 (-20%)	59,207 (-2%)	L'autocitation de la revue <i>Scientometrics</i> est 10 fois plus importante en 2019
JIF à 5 ans	3,073 (+7%)	59,345 (-2%)	Pour <i>Scientometrics</i> , les publications 2014-16 sont plus souvent citées en 2019 que celles de 2017-18 (inverse pour <i>Lancet</i>)
½ vie des citations	6,8 ans	8,6 ans	Les publications de <i>Lancet</i> citées en 2019 sont plus anciennes

- **Les indicateurs de citation d'une revue dépendent du domaine scientifique** des travaux qu'elle publie :
 - dépendent de la taille de la communauté scientifique
 - dépendent de la dynamique de publication de cette communauté scientifique
 - dépendent de la pratique de citation de cette communauté scientifique
- L'indicateur d'impact d'une revue ne peut être comparé qu'avec ceux des revues du même domaine scientifique
 - SCIE classe les 9 200 revues dans 178 catégories de domaines scientifiques
 - SSCI classe les 3 400 revues dans 58 catégories de domaines scientifiques

Variété des études bibliométriques

- La bibliométrie comme instrument statistique : *les études des précurseurs*
- La bibliométrie comme instrument de sociologie des sciences : *la scientométrie*
- La bibliométrie comme instrument d'évaluation de la recherche : *le règne des indicateurs*
- La bibliométrie comme instrument de cartographie des sciences : *l'émergence de la cartographie de l'information*
- La bibliométrie comme instrument d'exploration des informations scientifiques : *la cartographie pour l'aide à l'analyse*

Comment la bibliométrie évalue la science ?

La bibliométrie mesure uniquement la production des écrits de la recherche

- En pratique ne mesure que la **production d'articles scientifiques publiés dans les revues à plus forte notoriété**
 - Articles référencés dans WoS ou Scopus (index multidisciplinaire, mondiale, contenant les données des citations)
- **Avec 2 mesures de production** = décompte de 2 extrants
 - Le comptage du nombre de publications scientifiques (essentiellement les articles),
 - Le comptage du nombre de citations reçues par ces publications scientifiques.
 - Aucune mesure de productivité
 - Productivité = décompte d'extrants par unité d'intrant = $\frac{\text{mesure d'une production}}{\text{mesure d'une ressource}}$
 - Absence de données sur les intrants qui soient fiables et comparables (grosses différences selon les systèmes nationaux de recherche)
- Ces 2 mesures peuvent s'appliquer à différentes entités
 - Les publications des travaux scientifiques (unité de base),
 - Un chercheur, une unité de recherche, une institution, un pays, une région géographique,
 - Une discipline scientifique, un domaine de recherche, un thème de recherche,
 - Un journal scientifique, une conférence scientifique, un ouvrage, ou tout autre type de modes de diffusion de la production scientifique

L'évaluation de la science tend à élaborer des indicateurs normalisés (comparables)

Exemple de la standardisation de l'indicateur d'impact

- Les 2 mesures de production restent insatisfaisantes
 - Car dépendantes de la taille des entités étudiés
- L'indicateur d'impact (de base)
 - Le ratio $\frac{\text{production de citations}}{\text{production de publications}}$ est indépendant de la taille des entités
 - Ce ratio (l'indicateur d'impact) reste dépendant du domaine scientifique et de type de publication
- Normalisation de l'indicateur d'impact

Impact relatif par discipline = $\frac{\text{impact de l'entité étudiée}}{\text{impact moyen mondial des publications de même type, de la même année et du même domaine}}$

$$\text{MNCS} = \frac{1}{P} \sum_{i=1}^P \frac{1}{D_i} \sum_{j=1}^{D_i} \frac{c_i}{IM_{ij}}$$

avec P = nombre de publication de l'entité
 c_i = nombre de citations reçues par la publication i
 D_i = nombre de domaines attribués à la publication i
 IM_{ij} = nombre moyen des citations des publications de même type, de la même année et du même domaine

- MNCS = 1,2 → les publications sont cités à 20% au dessus de la moyenne mondiale des publications de même type, de la même année et du même domaine
- Défaut = calcul de la moyenne pour des distributions très asymétriques

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator : An empirical analysis. *Scientometrics*, 87(3), 467-481.
<https://doi.org/10.1007/s11192-011-0354-5>

Autres indicateurs souvent employés pour l'évaluation de la science ?

- Indicateurs d' « d'excellence » : Top x%
 - Part des publications de l'entité qui appartiennent au premiers centiles (*pourcentiles*) des publications les plus citées
 - Top 10% =
$$\frac{\text{Nombre de publications de l'entité appartenant au 10\% des publications les plus citées dans le monde pour le même type, la même année et dans le même domaine}}{\text{Nombre total des publications de l'entité}}$$
 - Idem Top 5%, Top 1%
 - Autre mesure de performance mais dépendant d'effet de taille

Bornmann, L., Leydesdorff, L., & Mutz, R. (2013). The use of percentiles and percentile rank classes in the analysis of bibliometric data : Opportunities and limits. *Journal of Informetrics*, 7(1), 158-165. <https://doi.org/10.1016/j.joi.2012.10.001>

- Indicateurs de collaboration
 - Part des publications de l'entité qui sont le résultat d'une collaboration nationale, européenne ou internationale
 - $$\frac{\text{Nombre de publications résultant d'une collaboration nationale (ou internationale)}}{\text{Nombre total des publications de l'entité}}$$

Nagpaul, P. S., & Sharma, L. (1994). Research output and transnational cooperation in physics subfields : A multidimensional analysis. *Scientometrics*, 31(1), 97-122. <https://doi.org/10.1007/BF02018104>

- Indicateurs de spécialisation
 - Pour les entités ayant des activités multidisciplinaires (pays, organismes)
 - Comparer la contribution de l'entité à une domaine à une contribution de référence (nationale, internationale...)

$$RAI_{\text{Nat}} = \frac{\frac{\text{Nombre de publications de l'entité dans le domaine}}{\text{Nombre total de publications de l'entité}}}{\frac{\text{Nombre de publications nationales dans le même domaine}}{\text{Nombre total de publications nationales}}}$$

$$RSI = \frac{RAI - 1}{RAI + 1}$$

RSI ~ -1 → totalement inactif dans le domaine
 RSI ~ 1 → activité exclusivement dans le domaine
 RSI ~ 0 → activité dans le domaine identique à l'activité nationale/internationale

Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics*, 9(5), 281-291. <https://doi.org/10.1007/BF02017249>

Ces mesures sont fortement dépendantes des sources des données

- Bases de données bibliographiques commerciales et leurs plateformes bibliométriques

- ⇒ WoS ou Scopus (multidisciplinaire, mondiale, données des citations)
- ⇒ Incite ou SciVal (séries d'indicateurs bibliométriques)

- Bases de données bibliométriques

- ⇒ Construites par les centres nationaux or internationaux à des fins d'évaluation de la recherche

- Compléter les données
- Correction des données
- Normalisation de données
- Identification des acteurs nationaux et internationaux (selon plusieurs niveaux d'agrégation)
- Redécoupage des domaines scientifiques selon une nomenclature ad hoc
- Comptages des présences ou comptage fractionnaire
- Comptages des citations avec ou sans les autocitations
- ...

- ⇒ Ex.

- HCERES-OST (France) : <https://www.hceres.fr/fr/sources-et-donnees>
- CWTS (Leiden, Pays-bas) : <https://www.cwtsbv.nl/evidence-bibliometric/>

Variété des études bibliométriques

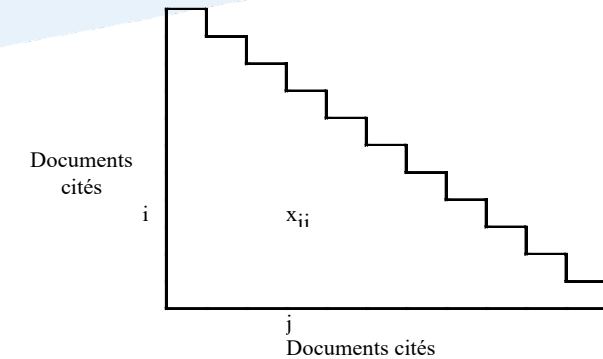
- La bibliométrie comme instrument statistique : *les études des précurseurs*
- La bibliométrie comme instrument de sociologie des sciences : *la scientométrie*
- La bibliométrie comme instrument d'évaluation de la recherche : *le règne des indicateurs*
- La bibliométrie comme instrument de cartographie des sciences : *l'émergence de la cartographie de l'information*
- La bibliométrie comme instrument d'exploration des informations scientifiques : *la cartographie pour l'aide à l'analyse*

Les travaux précurseurs de la cartographie des sciences

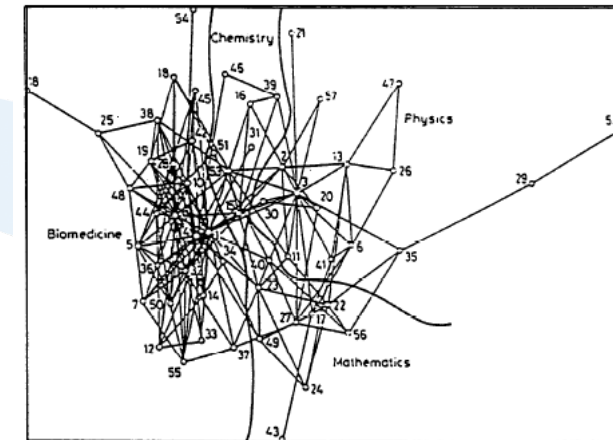
- Travaux fondés sur la mesure des citations et co-citations
 - l'association bibliographique (*bibliographic coupling*)
 - l'analyse des co-citations (*co-citation analysis*)
 - l'analyse des co-citations d'auteurs (*author co-citation analysis*)
 - l'analyse des citations croisées de revues (*cross-citation analysis*)
- Travaux fondés sur la mesure des relations d'éléments descriptifs
 - l'analyse des mots-associés (*co-word analysis*)
 - l'analyse des co-classifications documentaires (*co-heading analysis*)
 - l'analyse des co-auteurs (*co-author analysis*)
 - l'analyse des coopérations internationales
 - analyse d'un tableau de contingence
 - de mots-clés
 - de collaborations internationales
 - de villes x codes
 - de pays x mot-clés
 - ...

L'analyse des co-citations (*Co-citation analysis*)

- Objectif
identifier la « littérature-cœur » de la science et ses fronts de recherche
- Méthode
 - Inspirée de Kessler mais pour pallier le problème du surplus de données à traiter, ne conserve pas l'information concernant les documents « citants »
 - mesure les ressemblances entre les documents cités
 - tableau carrée symétrique doc. cités
 - normalisation du tableau : $x'_{ij} = \frac{x_{ij}}{\sqrt{c_i \cdot c_j}}$
 - Seuil de filtrage : $x_{ij} > 2$ et $x'_{ij} > 0,3$
 - classification automatique par CAH à lien simple -> agrégats de documents cités = littérature-coeur
 - identifier les documents citant ces agrégats = fronts de recherche (et mesure de recouvrement)
- L'Atlas de la science (85)
 - amélioration des agrégats + itération d'agrégation
 - 3 932 agrégats de 49 documents maximum
 - réduits à 502 agrégats puis 57 agrégats
 - Positionnement par MDS de ces agrégats



x_{ij} = nombre de documents qui ont cité le document i et le document j en même temps

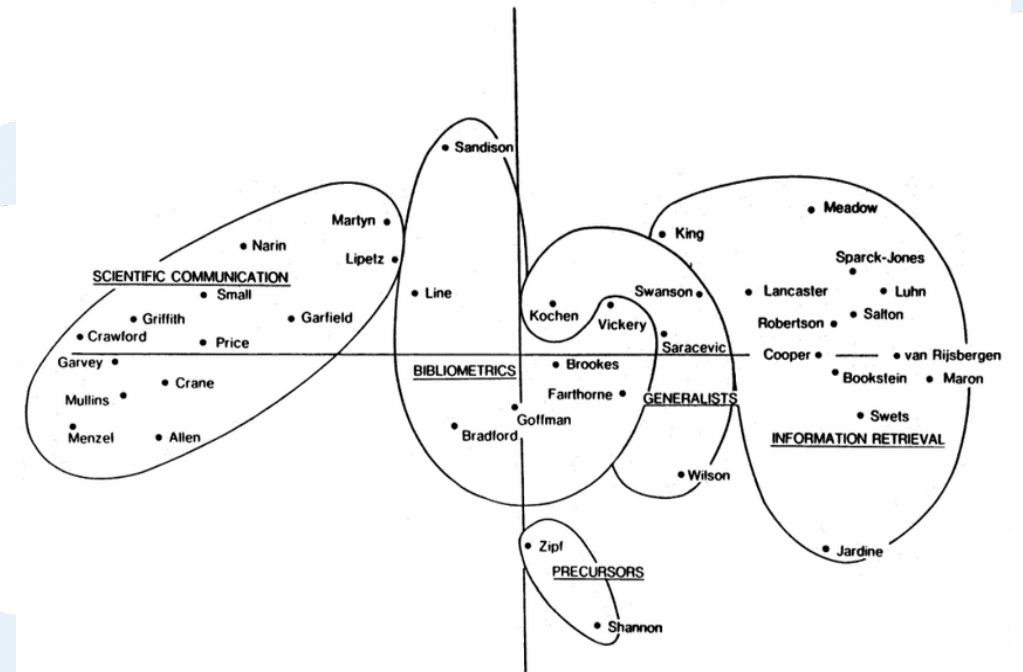


1979 C4 map of science

Small H G, "Co-citation in the scientific literature : a new measure of the relationship between two documents", Journal of the american society for information science, Vol 24, N° 4, p. 265-269, 1973

L'analyse des co-citations d'auteurs (*Author co-citation analysis*)

- Objectif
déplace l'unité d'analyse du document individualisé au groupe de documents identifiables comme l'œuvre d'un auteur
- Méthode
 - mesure les ressemblances entre les auteurs cités
 - tableau symétrique d'auteurs cités + normalisation
 - classification automatique -> agrégats d'auteurs cités = Auteurs-cœur
 - identifier les documents citants ces agrégats = fronts de recherche
- Etudes
 - Le résultat des "constellations" de points sur ces cartes représente des structures non plus construites par des consensus autour de travaux précis mais sur l'apport global d'un auteur à sa discipline. Les cœurs de littérature obtenus symbolisent en général les grandes écoles de pensée

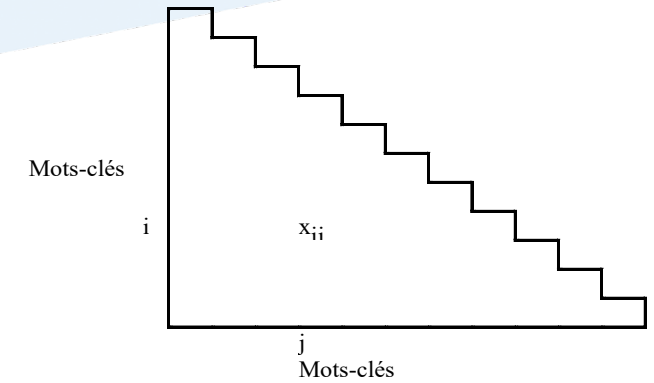


Carte des co-citations d'auteurs dans le domaine des sciences de l'information de 1972 à 1979

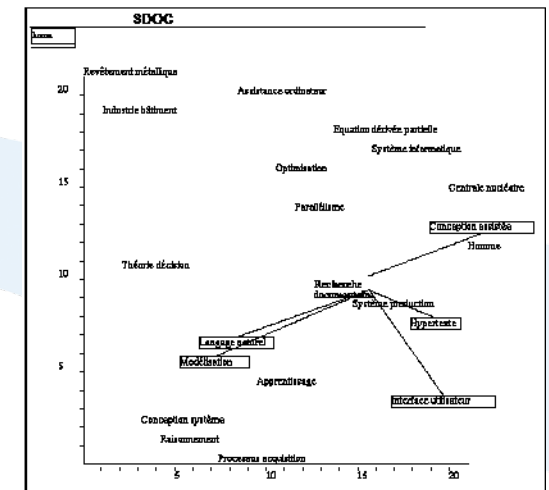
White H D, Griffith B C, "Author cocitation : a literature measure of intellectual structure", Journal of the american society for information science, Vol 32, N° 3, p. 163-172, 1981

L'analyse des mots associés (*co-word analysis*)

- **Objectif**
Pour décrire un domaine de recherche en représentant la structuration des concepts grâce aux termes des articles.
- **Hypothèse**
Une paire de mots-clés présente dans un grand nombre d'articles représente une forte probabilité d'association entre les concepts. Théorie des réseaux socio-techniques du CSI de l'Ecole des Mines
- **Méthode**
 - mesure de la distance d'équivalence entre les mots-clés
 - agrégation des mots-clés en groupe de taille identique
 - mesure des relations entre groupes (la centralité)
 - mesure d'un indicateur d'homogénéité des groupes (la densité)
 - carte stratégique : position des groupes en fonctions de leurs valeurs de centralité et de densité
- **Inconvénient**
 - technique d'agrégation très peu stable



X_{ij} = distance d'équivalence entre les mots-clés i et j



Carte stratégique de l'analyse des mots associés

Michelet B, L'analyse des mots associées, thèse Univ. Paris VII, 1988

Variété des études bibliométriques

- La bibliométrie comme instrument statistique : *les études des précurseurs*
- La bibliométrie comme instrument de sociologie des sciences : *la scientométrie*
- La bibliométrie comme instrument d'évaluation de la recherche : *le règne des indicateurs*
- La bibliométrie comme instrument de cartographie des sciences : *l'émergence de la cartographie de l'information*
- La bibliométrie comme instrument d'exploration des informations scientifiques : *la cartographie pour l'aide à l'analyse*

La bibliométrie comme instrument d'exploration des sciences et techniques : cartographie

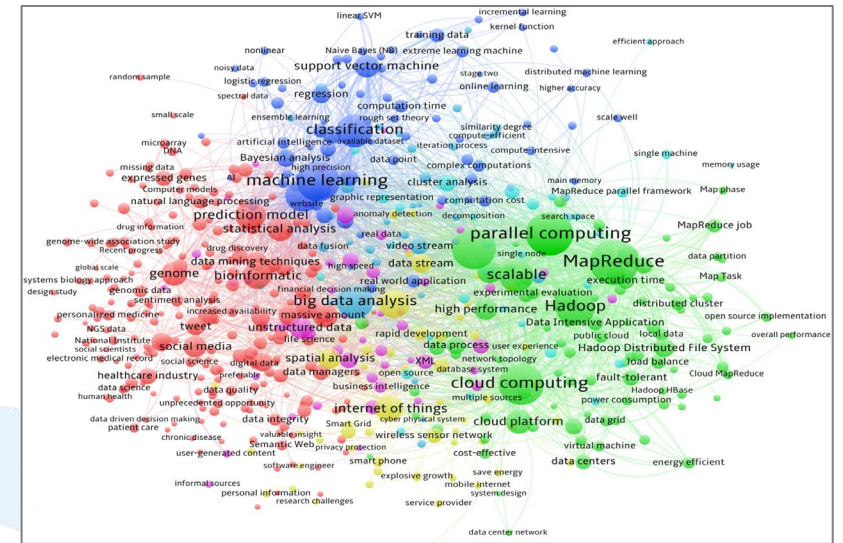
- La bibliométrie pour explorer les sciences & techniques

Découvrir les relations, les tendances et les modèles qui expriment la construction et l'évolution des domaines scientifiques et des émergences technologiques.

Inspiré des premières études bibliométrique de la cartographie des sciences : analyse des co-citations, analyse des mots associés, analyse des co-classifications... employant des **méthodes de clustering** couplées à des **méthodes de visualisation** (analyses factorielles ou cartographie de réseaux)

Les méthodes bibliométriques sont parfois combinées à **des techniques de text mining** ou **d'autres technologies de l'information** (méthodes probabilistes, *machine learning*, analyse de réseaux, modèles prédictifs...) (Zhang, Porter Chiavetta, 2017)

Combinaisons de méthodes employées pour la veille scientifique, la veille technologique et le management de l'innovation.



Term correlation map for big data research (Zhanga,et al, 2018)

Exploration de l'évolution technologique de la recherche sur le Big Data pour déterminer l'évolution et la disparition des technologies employées

Réseau de collaborations d'auteurs japonais (travaux sur les réseaux neuronaux)

