

[European Summer School for Bibliometrics](#) Vienne 11-13/09/11

Relevé de notes saisies lors du colloque ESSS11 incluant les liens vers les documents de référence.

par Manuel DURAND-BARTHEZ, URFIST de Paris

Des renvois vers les présentations PPT seront faits lors de leur insertion sur le site de l'ESSS

Introduction présentée collectivement par les intervenants

„Bibliometrics in a Nutshell“ - Crash Course for Newbies

Wolfgang Glänzel, Centre for R&D Monitoring (ECOOM), Katholieke Universiteit Leuven, Belgium / Juan Gorraiz, Bibliometrics Department, University of Vienna, Austria / Christian Gumpenberger, Bibliometrics Department, University of Vienna, Austria / Stefan Hornbostel, Institute for Research Information and Quality Assurance (iFQ), Germany / Sybille Hinze, Institute for Research Information and Quality Assurance (iFQ), Germany

Ce colloque est le fruit d'une coopération entre les universités de Vienne, la Humboldt de Berlin, l'Université catholique de Louvain et l'IFQ (Institut für Forschungsinformation und Qualitätssicherung, équivalent allemand de l'INIST) de Berlin

Les exposés seront marqués par un grand sens de la nuance, la nécessité de ne pas se limiter à des schémas réducteurs sur la bibliométrie. Par exemple, l'accent sera mis sur les spécificités disciplinaires et les différences d'usage en matière de publication.

Métaphore sur Vienne, lieu du colloque assez symbolique: évocation de la guerre froide, ville frontière, zone d'échange bilatéral, carrefour d'influences antagonistes ou convergentes pendant des décennies. Sybille Hinze voit dans cette image une incitation à réfléchir et critiquer sainement, analyser de façon lucide, singulièrement en regard du sujet qui nous concerne.

L'Univ. de Vienne dispense un enseignement spécialisé sur les Bibliothèques et la scientométrie.

Prendre en compte toutes les langues et tous les supports.

History and Institutionalization of Scientometrics

Wolfgang Glänzel, Centre for R&D Monitoring (ECOOM), Katholieke Universiteit Leuven, Belgium / Stefan Hornbostel, Institute for Research Information and Quality Assurance (iFQ), Germany

Bibliometric vs. Scientometrics : cf. Glossaire proposé par les intervenants. La seconde recouvre partiellement la première. Elle est plus orientée vers une analyse de la science et de son évolution en tant que telle.

Un graphique de Price ([Little Science, big science](#), 1963) montre, dès 1830, une progression constante des Review Journals, proportionnellement parallèle à celle des Scientific Journals.

Evocation de Lotka « The Frequency Distribution of Scientific Productivity » *J. Washington Acad. Sci*, Vol. 16, No. 12., pp. 317-323, 1926 : sur le principe de la distribution [qu'on retrouve dans l'indice de notoriété de l'INRA / M.-H. Magri]

La loi de Bradford « On the scattering of papers on scientific subjects in scientific periodicals » (*Engineering*, 1934) montre l'existence de revues cœur autour desquelles prolifèrent des articles citants suivant une progression mathématiquement constante.

Avant Garfield : Gross & Gross « [College libraries and chemical education](#) » (*Science*, 1927), première analyse des citations à partir du *JACS* pour orienter les acquisitions.

Zipf ([Human Behavior and the principle of least effort](#), 1949), recherche des biais permettant de concilier solution immédiate d'un problème avec celle de problèmes subséquents associés. Autrement dit : en quoi le quantitatif induit-il le qualitatif et dans quelle mesure le second est-il toujours directement associé au premier ? En quoi le moindre effort peut-il « piéger » la bibliométrie ?

Solla Price avait déjà pressenti en 1963 l'utilité des réseaux de (co)citations

Garfield, 2009 [From information retrieval to scientometrics : is the dog still wagging his tail ?](#) : souligne que son SCI n'était destiné ni à l'évaluation quantitative, ni aux facteurs d'impact, ni à l'étude de l'histoire des sciences de l'information.

Garfield & Sher : « [New Factors in the Evaluation of Scientific Literature Through Citation Indexing](#) », *American documentation*, 1963, à propos de l'I.F. : nécessité de comparer les journaux indépendamment de leur format, i.e. de leur volumétrie et de leur périodicité.

Objectifs Bibliométrie : Information science, sociologie des sciences et, plus récemment : science policy (financement notamment) ; cette troisième appli a été développée aux US dans les 70s

Aspect hybride entre outil d'évaluation et outil de recherche bibliographique.

Organisations ou institutions d'envergure nationale contribuant à l'élaboration d'indicateurs

- Tibor Braun : ISSRU Budapest.
- CWTS Leuden, 1982>
- Amsterdam : cf. nombreux articles de Leydesdorff, L.
- En France : 2 items :
 - o Turner Callon Courtial = mapping sémantique

- OST

Espagne : Consejo Superior de Investigaciones científicas (CSIC)

3 revues clés :

- *Scientometrics* (T. Braun, 1978>)
- *Research Evaluation* (Van Raan, 1991 >)
- *Journal of Informetrics* (L.Egghe [*promoteur du G-index*], 2007 >)

3 conférences importantes:

- Int Conf for Informetrics and scientometrics (bienn ,1987>)
- Int Conf for Sci & Tech Indicators (bienn 1988 >)
- [CollNet](#) (Global Interdisciplinary Research Network for the Study of all Aspects of Collaboration in Science and in Technology - annuel)

Exemple national (ou « régional » ?...): [Flemish Indicator Book on Science, Technology and Innovation](#). ECOOM (Expertisecentrum O & D Monitoring) Fait intervenir le Mean Observed Citation Rate: MOCR. et son complément, le Mean Citation Rate: MECR, avec leur ratio : $RCR = MOCR/MECR$. On retrouve là, nous semble-t-il, bien naturellement, de nombreux éléments présents dans les rapports de l'OST français. Mentionnons quand même une très abondante et intéressante bibliographie sur la bibliométrie, page 170 sq.

RAE britannique post-2008 : bibliométrie utilisée « entre autres » pour le financement. Idem en Australie et Nouvelle-Zélande. A noter une fois de plus les récentes mesures politiques australiennes en vue de valoriser la qualité des travaux vs. la mesure purement quantitative.

Importance croissante des analyses de co-citations. Co-citation analysis cf. H.Small, [« Co-citation in the scientific literature: a new measure of the relationship between two documents. »](#) *Journal of American Society of Information Science*, 24, 256–269. JASIS 1973 et I. Marshakova « System of documentation connections based on references (SCI). *Nauchno-Tekhnicheskaya Informatsiya* »2, 6, 3–8.

Clusters cartographiés par F.Janssens et al. : ["A hybrid mapping of information science"](#) dans *Scientometrics*, (75) 3, 2008. => 5 Nœuds pris dans un corpus expérimental, autour :

- des brevets
- du Net (webometrics)
- de la sociologie informationnelle
- des techniques informatiques associées (Information Retrieval)
- de la bibliométrie

Pour chacun de ces noeuds du corpus, deux articles ont été mis en avant qui présentent des caractéristiques communes au niveau de la co-citation, reliant ces clusters entre eux.

Cf. en particulier la « Figure 8. Term networks with for each of five clusters the best 20 stemmed terms or phrases from titles or abstracts according to mean TF-IDF scores » page 624 et les commentaires pp. 623 et 625.

La méthode TF-IDF correspond au recensement statistique des occurrences de chaînes de caractères dans les full-texts : *Term-Frequency times Inverse DocumentFrequency*, i.e. il ne s'agit pas de la comptabilisation brute des termes, mais de l'évaluation de leur nombre en proportion pondérée dans le corpus. Autrement dit, en fonction du corpus considéré dans sa globalité, la mention maintes fois réitérée de ce mot est-elle conceptuellement significative ou juste marginale ? Sur cette définition, voir aussi : Pazzani, M.-J. et al.: [« Content-Based Recommendation Systems »](#), *The adaptive web*, 2007 – Springer, pp. 327-328.

A plusieurs reprises, on met l'accent sur la nécessité de lutter contre le « quick and dirty » = Évaluation quantitative brute, hâtive et délétère.

New developments in bibliometric methods for evaluation and mapping of scientific research

Anthony van Raan, CWTS - Centre for Science and Technology Studies, Leiden University, The Netherlands

(Van Raan est le promoteur du [classement de Leyde](#))

Lutter contre l'inflation des publis.

3 clés :

- citation primaire
- co-citation
- réf. partagées

1^{ère} clé = impact associé à la *différence*

2^{ème} et 3^{ème} clés = réseaux secondaires rattachés par la *similarité*.

Liens Patents / Articles (cités par les brevets) de plus en plus exploités.

Pb définition des domaines. Exemple : Radiologie. On détecte des articles relevant spécifiquement de cette discipline dans *Nature*, ainsi que dans les biblios terminales d'articles de cette revue prestigieuse et, à certains égards, pluridisciplinaire. Cette détection sera rendue possible par un repérage via le réseau secondaire défini précédemment. C'est notamment par le biais de ce repérage qu'un institut hospitalo-universitaire pourra mieux démontrer sa spécificité vis-à-vis de son administration.

Sur la dynamique de la connaissance associée au principe de « variété » et aux « flux croisés », voir notamment : [« Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience »](#), I.Rafols et M.Meyer, *Scientometrics* 2010.

Importance des *Sleeping beauties* : articles dont on parle 10 ans après leur parution.

De même : les *Hidden Jewels (or Pearls)* [concept d'ailleurs repris par *Faculty of 1000*].

Van Raan condamne IF et H.

Comparer : Equipe / Journal / Champ de recherche : c'est [l'objectif du FCS](#) (Field Citation Score) du classement de Leyde où le taux de citations, applicable à une équipe, est mesuré à l'intérieur d'un champ d'investigation (Field ou Subject Category) en regard des revues dans lesquelles cette équipe publie.

Montre que la fenêtre étroite de 2 ans bride la citation et fausse les calculs.

Combiner *concept-based maps* et *local citation-density measurement*. Evaluer la performance d'une équipe à partir de la restitution de ses citations à l'intérieur d'un domaine et non d'un cluster de journaux inclus dans un ensemble disciplinaire plus vaste non défini spécifiquement sur le plan disciplinaire.

Bibliometrics in the History and Philosophy of Science

Werner Marx, Central Information Service, Max Planck Institute for Solid State Research, Stuttgart, Germany (M.Planck)

Associe nécessairement l'utilisation de bases bibliographiques spécialisées (pour son domaine, la physique) telles qu'Inspec ou Phys Rev Online Archive (PROLA) à celles de bases spécifiquement bibliométriques (WoS + Scopus)

Inspec couvre ce que WoS ne couvre pas (ex. anecdotique : journaux de physique allemands pendant la 2^e guerre mondiale)

Taux d'erreur de citat^o standard dans le WoS : 3,5 %

Pour l'auteur G.Wassermann (1933) 81,94 % d'erreurs dans la manière de citer cet auteur et ses publis

Einstein cité dans un texte d'article, juste son nom, genre « les travaux d'Einstein sur... », sans réf. bibliogr. : c'est bien sûr encore mieux que d'être cité en référence, (célébrité ou réputation suffisante) mais ce n'est pas compté.

Montre un graphique éloquent sur l'évolution du pt de vue scientifique s/ tectonique des plaques, de 1883 à 1971 : la nature et le contexte du référencement bibliographique sont modifiés tant au niveau disciplinaire qu'à celui du générique/spécifique dans la définition des concepts.

New developments in the use of bibliometric tools in research assessment

Henk Moed, Senior Staff member at the Centre for Science and Technology Studies, in the Department (Faculty) of Social Sciences at Leiden University,

Moed présente deux slides importantes se rapportant à une matrice de typemulti-dimensionnel. Elles sont extraites de l'article intitulé [multidimensional research assessment matrix](#) (Henk Moed and Andrew Plume *Research trends* (23) May 2011) dont nous présentons les schémas ci-dessous, correspondant précisément à ces deux slides.

Unit of assessment	Purpose	Output dimensions	Bibliometric indicators	Other indicators
Individual	Allocate resources	Research productivity	Publications	Peer review
Research group	Improve performance	Quality, scholarly impact	Journal citation impact	Patents, licences, spin offs
Department	Increase regional engagement	Innovation and social benefit	Actual citation impact	Invitations for conferences
Institution	Stimulate international collaboration	Sustainability & Scale	International co-authorship	External research income
Research field	Promotion, hiring	Research infrastructure	Citation 'prestige'	PhD completion rates

Table 1 — The multi-dimensional research assessment matrix. This table presents a core part of the matrix, not the entire matrix. It aims to illustrate what the matrix looks like. It should be read column-wise: each column represents a different dimension.

Type (generation)	Description	Typical examples
First	Basic indicators; relatively easy to obtain from sources that have been available for decades	Number of publications; number of citations; journal impact metrics
Second	Relative or normalized indicators, correcting for particular biases (e.g., differences in citation practices between subject fields)	Relative or field-normalized citation rates
Third	Based on advanced network analysis using parameters such as network centrality	Influence weights; SCImago Journal Rank; 'prestige' indicators

Table 2 — Types of bibliometric indicators.

Il met l'accent s/

- Competitive research income
- ratio active staff in research /Total staff
- % Patent cit° + collab. avec l'industrie + research reports

Pb des publis multidisciplinaires : quels fonds alloués, qui subventionne, quel type de co-autorat ?

2 types d'analyse :

- top down (classique = partir de l'institution et recenser ses publis de façon brute)
- bottom-top (partir de noms d'auteurs en les agrégeant en groupes, champs, institutions, utilisat° des co-citat°, clusters => *diagramme stratégique*) un I.F. normalisé, i.e. en référence à un champ, moyenne mondiale à comparer à celle d'un groupe d'auteurs.

Dans le Bottom-Top, apparemment clair entre définition des groupes et peer reviewing. Inverse si l'on considère le Top-Bottom où le peer reviewing s'incline devant la bibliométrie purement quantitative.

Cette symétrie « inversée » est expliquée par ailleurs dans la communication suivante, de Glänzel (triangles symétriques inversés).

5 critères majeurs d'analyse :

- ❖ Field Citation Score
- ❖ Traçabilité des citat°_ (prestige) . Il ne suffit pas d'être cité, encore faut-il l'être par des entités elles-mêmes fréquemment citées[cf. *Eigenfactor*, *Scimago Journal Rank* et *Journal Status* de Bollen]
- ❖ Impact hors du domaine usuel des publis du groupe
- ❖ Importance du contexte de la citation ; des outils « full text » l'indiquent et précisent son « poids ».
- ❖ Nb de downloads en full text[mais voir plus bas les réserves sur cette procédure en consultation d'O.A.]

Pb des art. cités mais non lus

Practical Aspects of Scientometrics

Koenraad Debackere, Centre for R&D Monitoring (ECOOM), Katholieke Universiteit Leuven, Belgium

La bibliométrie est aussi considérée comme outil d'investigation sur l'innovation pour

mesurer le progrès technique (cf. travaux économétriques de R.Solow)

Plusieurs modèles, manuels et programmes sous-tendent cette notion.

Actualité du modèle « Triple Helix » voir notamment les travaux de L.Leydersdorff et M.Meyer, ex. : « [Triple Helix indicators of knowledge-based innovation systems](#) » *Research Policy*, (35), 10, December 2006.

Ce modèle repose sur l'interaction de : « (1) wealth generation (industry), (2) novelty production (academia), and (3) public control (government). »

Importance toujours très influente du [Manuel de Frascati](#) (OCDE s/ R & D ; 2002) et du [Manuel d'Oslo](#) (OCDE 2005): *La mesure des activités scientifiques et technologiques : principes directeurs proposés pour le recueil et l'interprétation des données sur l'innovation technologique*. Que l'INSEE définit comme suit : « Le Manuel d'Oslo est la principale source internationale de principes directeurs en matière de collecte et d'utilisation d'informations sur les activités d'innovation. »

Le [Manuel de Canberra](#) (1995) également produit par l'OCDE : *La mesure des activités scientifiques et technologiques : manuel sur la mesure des ressources humaines consacrées à la science et à la technologie ; mesure les « talents »*.

[PISA : Programme for International Student Assessment](#). Son but : « Are students well prepared for future challenges? Can they analyse, reason and communicate effectively? Do they have the capacity to continue learning throughout life? » également élaboré par l'OCDE.

L'OCDE a aussi mis au point une [Carte de la recherche scientifique à haut niveau d'activité](#), (2008) et [bien d'autres produits](#) de visualisation en termes d'innovation.

L'indice de Balassa est notamment appliqué pour l'étude des brevets. Il est issu d'un modèle économétrique élaboré par Béla Balassa ([The Manchester school](#), 1965). Il est spécifiquement utilisé « for calculating the relative advantage or disadvantage of a certain country in a certain class of goods or services as evidenced by trade flows. It is based on the Ricardian comparative advantage concept. » (défin. [Wikipedia](#)). Voir les caractéristiques de son application à l'analyse des brevets, entre autres dans l'article du communicant, K.Debackere [Patent Data for Monitoring S&T Portfolios](#), chap.. 26 du *Handbook of Quantitative Science and Technology Research : The Use of Publication and Patent* edited by Henk F. Moed *Centre for Science and Technology Studies, University of Leiden, The Netherlands ; Wolfgang Glänzel Steunpunt O&O Statistieken, K.U. Leuven, Belgium and Ulrich Schmoch Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany* *Statistics in Studies of S&T Systems*, Kluwer, 2005, page 560 [Glänzel reprend cet élément dans sa communication ci-après].

En Allemagne, l'Institut Fraunhofer élabore des modèles bibliométriques relatifs à la classification par domaines des brevets. Voir notamment la fiche de présentation de [Peter Neuhäusler](#) sur le site de cet organisme.

B.Van Looy et al. combinent Bibliométrie, Brevets et R&D pour expliquer la performance du transfert de technologie en provenance des universités. « [Exploring the feasibility and accuracy of Latent Semantic Analysis based text mining techniques to detect similarity between patent documents and scientific publications](#) » UCL Louvain, Faculty of Business and Economics (2010)

A noter que les brevets, en fonction de leur institut d'émission, présentent des variations de

nature en fonction des usages (exemple : business methods et logiciels entre EPO et USPTO...). Cette nuance est très importante pour dresser des tableaux ou cartes.

On peut associer Bibliometrics et Technometrics, cette seconde discipline touchant les brevets et l'analyse du progrès technologique. Les « indicateurs technométriques » sont fournis par les principaux offices de Propriété industrielle : EPO, WIPO, USPTO...[voir aussi la partie Brevets des indicateurs OST]

Importance croissante de l'étude des citations de brevets dans les publis académiques et vice-versa.

Nécessité de se référer à des catégories bien définies : voir notamment « [A new classification scheme of science fields and subfields designed for scientometric evaluation purposes](#) » de W.Glänzel et A.Schubert, *Scientometrics* Vol. 56, No. 3 (2003) 357–367

Cette catégorisation permet entre autres l'édition de schémas de type radar (inter)nationaux par subject categories. [Cf.par exemple spécialités OST]

Tous ces calculs sont naturellement impossibles s'ils ne sont pas précédés d'un nettoyage des données (data cleaning) ; au niveau des auteurs, des affiliations, des sources (intitulés et numérotation) etc...

Parmi les indicateurs susceptibles d'établir des comparaisons fondées sur les publis ou brevets dans telle ou telle discipline, on distingue la *real performance* de l'*expected performance* par champ disciplinaire : cette dernière nuance est extrêmement importante (cf. le FCS de Leyde). En effet, dans le premier cas (performance brute), les écoles d'ingénieurs, de petite taille, sont « enfoncées », disqualifiées. Dans le second cas, elles sont réhabilitées à leur juste valeur.

Metrics for Research Evaluation:

Indicators, Methods and Mathematical Foundations (1)

Wolfgang Glänzel, Centre for R&D Monitoring (ECOOM), Katholieke Universiteit Leuven, Belgium

La bibliométrie n'a pas pour unique objectif l'évaluation de la recherche. Elle peut aussi servir d'auxiliaire au peer reviewing

3 axes :

- mesure de la productivité (pub output)
- co-authorship (collaborations)
- taux de citations (mesure d'impact).

Les 2 derniers induisent la production d'indicateurs associés à la notion de réseau.

Schéma de Glänzel (2011, *slide originale à venir*) en deux trianglesymétriques contigus et inversés : l'un représente le Peer reviewing (qualit) et l'autre la Bibliométrie (quant), du macro vers le micro, du pays aux individus, en proportion inverse : cela donne une échelle Top/Bottom :

Pays, champs, universités, disciplines, journaux, départements, équipes, individus.

La base du triangle Peer rev (en bas de la figure) est large et part de l'individu (country étant sa pointe supérieure) , la base du triangle Bibliométrie (en haut de la figure) est symétriquement inverse de l'autre et va (du haut en bas) de *macro*(base supérieure) vers *micro*(pointe inférieure) en passant par *méso* dont le niveau (médian) se situe entre « disciplines » et « journaux » sur l'échelle déclinée ci-dessus (Top/Bottom – Pays/Individu)

Rappel sur les modèles stochastiques (random walk). Applicables au réseaux de citations. [Simulations déjà visibles dans l'Eigenfactor et le SJR de Scimago (voir aussi Durand-Barthez [Réseaux de co-citations et Open Access : pour un renouveau des méthodes d'évaluation](#))]

Ces indicateurs donnent de l'info sur la fiabilité statistique, les erreurs stochastiques ([random errors](#)) et les [confidence intervals](#)(intervalle de confiance) des indicateurs. Ils permettent la prédictibilité

Types de documents : exclusion de : recensions de livres, éditoriaux, errata, résumés de communications, reprints. Une étude bibliométrique peut également se concentrer sur les types de document en tant que tels. Exemple de visualisation internationale (12 pays) des types de document [Document-type country profiles](#) L.Zhang et al., JASIST (62) 7, 2011

Fondements théoriques des indicateurs : cf. Glänzel [Bibliometrics as a Research Field, 2003](#) : insiste sur le co-autorat, les noyaux durs dont peuvent être issus des ensembles d'articles présentant une homogénéité du point de vue des axes de recherche. Cf. *ibidem* Fig. 2.8, p. 22 montrée en slide.

Revient sur les comptes de présence et fractionnaire [cf. [définition OST](#) *Note méthodologique B-5*, §3.3, page 525] Insiste sur l'obsolescence des comptes fondés sur la première adresse. Montre un modèle de type Normadresse

Autres facteurs déterminants :

- l'âge du chercheur
- son environnement ou son statut social
- le sujet
- le type de document
- la période sur laquelle s'étend l'analyse.
-

Classifications : *Mathematical reviews* = Mat Subj Classif System (MSC)

JCR subject categories...

Sur les hiérarchies catégorielles :F. Narin, *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity* (monograph), NSF

C-637,(National Science Foundation, 1976). {Contract NSF C-627. NTIS Accession No. PB252339/AS}On en extrait :

-index d'activité

- indice de spécialisation à l'intérieur des disciplines (*Relative specialization Index RSI*)

Le modèle de Balassa s'applique – on l'a vu avec K.Debackere - à la bibliométrie : l'*Activity Index* de J.-D.Frame appliqué à la « Mainstream research in Latin America and the Caribbean » *Interciencia* 1977 et A.Schubert & T.Braun « [Relative indicators and relational charts for comparative assessment of publication output and citation impact](#) » *Scientometrics*(9) 5-6, 1986. Comparable aux *indices de spécialisation scientifique (publis) et technologique (brevets)* de l'OST.

Exemple de RSI dans Glänzel et al. [STI Leiden](#) 2004. Permet de comparer en schéma de type radar des secteurs disciplinaires rattachés à des pays. Ces schémas indiquent également de type d'affiliations (public/privé etc...) : cf. W.Glänzel & B.Schlemmer « [Science in a changing Europe: East vs west. National scientific profiles by subject fields](#) », *ISSI Newsletter*, 2009

Influence

- des variations en pratiques informationnelles
- de la mobilité des chercheurs

sur les sphères de coopération.

Importance également des *Acknowledgements*.

Il faut analyser le co-autorship avec circonspection.

- 18 motifs pour lesquels des chercheurs recherchent la coopération :
- Table 1: *The purposes for which people collaborate*
- [1] Access to expertise.
- [2] Access to equipment, resources, or "stuff" one doesn't have.
- [3] Improve access to funds.
- [4] To obtain prestige or visibility; for professional advancement.
- [5] Efficiency: multiplies hands and minds; easier to learn the tacit knowledge that goes with a technique.
- [6] To make progress more rapidly.
- [7] To tackle "bigger" problems [more important, more comprehensive, more difficult, global].
- [8] To enhance productivity.
- [9] To get to know people, to create a network, like an "invisible college".
- [10] To retool, learn new skills or techniques, usually to break into a new field, subfield, or problem.
- [11] To satisfy curiosity, intellectual interest.
- [12] To share the excitement of an area with other people.
- [13] To find flaws more efficiently, reduce errors and mistakes.
- [14] To keep one more focussed on research, because others are counting on one to do so.
- [15] To reduce isolation, and to recharge one's energy and excitement.
- [16] To educate [a student, graduate student, or, oneself].

- [17] To advance knowledge and learning.
- [18] For fun, amusement, and pleasure.

In : D.Beaver « [Reflections on Scientific Collaboration \(and its Study\): Past, Present, and Future](#) », *Collaboration in Science and in Technology. Proceedings of the Second Berlin Workshop on Scientometrics and Informetrics*, September 1 – 3, 2000, page 37

- Aussi Glänzel & Schubert « [Analysing Scientific Networks Through Co-Authorship](#) » *Handbook of Quantitative Science and Technology Research*

2005, Part 2, 257-276, DOI: 10.1007/1-4020-2755-9_12 : Coopération entre individus.

- *extramural* vs. *domestic* cooperation : la première est plus valorisante (sortir de son labo) [Glänzel & Schubert *ibidem*, p.258]
- entre secteurs disciplinaires
- internationale

Glänzel revient sur le compte fractionnaire et le calcul de ce qui peut de fait revenir à *chaque co-auteur*, en vue notamment d'une redistribution des moyens, i.e. pour affiner les critères de décision visant à financer l'équipe dont fait partie tel ou tel auteur (même si les travaux de cet auteur ont été réalisés dans une autre institution lors d'une période antérieure à ladite décision).

Pb de l'*hyperauthorship*, i.e. co-autorat très important (ex. : High Energy Physics) ou anonyme (ex. : projets internationaux astrophysique) reste insoluble actuellement. De toute façon, on ne peut jamais assigner systématiquement à un auteur une affiliation de façon pérenne. Beaucoup de travail manuel reste à faire en aval.

Développement de la cartographie internationale des collaborations.

Motivations pour citer les autres : Garfield, *Current Contents*, 1975 en donne 15. Il est vraisemblable qu'elles ont muté depuis en raison du rôle de la bibliométrie aujourd'hui.

Sur la manipulation des autocitations de revues pour accroître leur IF : Weingart, [Impact of bibliometrics upon the science system: Inadvertent consequences?](#) *Scientometrics*, 2005.

La coopération internationale renforce le taux de citations.

Compromis à rechercher pour la largeur des fenêtres d'analyse. Bien sûr, plus elle est large, meilleurs sont la distanciation et le recul. Mais à l'inverse, c'est sur la période récente que l'on souhaite avoir le plus d'informations.

Comparaison 3 vs. 5 ans de fenêtre : Glänzel et al. « [A new generation of relational charts for comparative assessment of citation impact](#) » *Arch. Immunol. Ther. Exp.*(56) 6,2008

Affinage du RCR de l'ISI notamment par les FCSM de Leyde

Indicators, Methods and Mathematical Foundations (2)

Sybille Hinze, Institute for Research Information and Quality Assurance (iFQ), Germany

Rapports : Nat. Science Board (NSB) : [Science and engineering indicators 2010](#)

E.U. idem [2008/2009 European Report on Sci & Tech Indicators \(european research area\)](#)
cf. notamment chap.. 3 sur la R & D

Le rapport NSB éclaire sur la comparaison US /EU, et notamment le co-autorat (ccop entre les 2 entités géographiques)

Allemagne : [Research innovation and technological performance in Germany](#)
(Expertenkommission... EFI) 2010

Ulrich Schmoch et al. « Performance and structures of the German System in an International Comparison 2010 with a Special Analysis of Public Non-university Research Institutions : Analyses carried out for the annual report of the Expert Commission on Research and Innovation » [Studien zum deutsche innovationssystem](#), n°8, 2011, p. 13 = tableau des performances comparées entre pays et disciplines , en situant la position de l'Allemagne.
Rapport entre la croissance d'un champ et la croissance de tous les champs : leur ratio donne une idée de l'évolution globale de l'activité d'un pays. Ou même d'une région : dans quoi investir, en fonction des résultats acquis vs. telle ou telle discipline ?

Mapping Science

Bart Thijs, Centre for R&D Monitoring (ECOOM), Dept MSI, Katholieke Universiteit Leuven, Belgium

A développé un modèle de classif avec Glänzel en 2008 : [« A structural analysis of publication profiles for the classification of European research institutes »](#) Selected Papers Presented at the 9th International Conference on Science and Technology Indicators. *Scientometrics* (74) 2

Insiste sur la cocitation.

La similarité de contenu est repérée lexicalement (chercher des liens sur tous ces items):

- Co-Word (Callon, 1993) « [The use of patent titles for identifying the topics of invention and forecasting trends](#) » *Scientometrics*, (26) 2
- Keywords (Noyons, 1999) « [Integrating research performance analysis and science mapping](#) » *Scientometrics*, (46) 3
- TF x iDF= term frequency multiplied by the inverse document frequency (not the

pure number of occurrences) = on fait intervenir la proportion représentée par ce terme dans le document et par ce terme dans la totalité des documents du corpus

- Latent semantic analysis (indexing) with Singular Value decomposition : modèle permettant de réduire le vocabulaire à analyser à un corpus hautement significatif de termes.

La similarité de contenu est aussi repérée par le biais de la classification disciplinaire dont relève (en tenant compte également des cas d'appartenance à plusieurs Subject categories) la revue où l'article est publié.

Application du *Jaccard similarity coefficient* et du *Salton's cosine* aux auteurs : sur ces deux modèles, cf. l'article de L.Leydesdorf « [On the Normalization and Visualization of Author Co-Citation Data : Salton's Cosine versus the Jaccard Index](#) » JASIS (59) 1, 2008

Calcul des « angles » entre documents : notions de bords (edges) et d'emprise (cooccurrence). Dendrogrammes. La distance entre les clusters est fonction de la distance moyenne entre tous les membres.

Les clusters peuvent être mesurés par le [modèle de Ward](#). (WARD, J. H. (1963). [Hierarchical grouping to optimize an objective function](#). *Journal of the American Statistical Association*, 58 : 236–244.) Cf. Kamada, T., & Kawai, S. (1989). « [An algorithm for drawing graphs](#) ». *Information Processing Letters*, 31, (exemple sur Energy & Fuels)

Network Analyses

Edgar Schiebel, Head of Technology Management, AIT Austrian Institute of Technology GmbH, Austria

- bibliographic coupling = research frontiers
- co-authorship : networks of authors and co-authors
- co-citation : knowledge bases

Matrices de similarité permettant de croiser des co-auteurs pour mettre en évidence un carrefour de connexions

Modèle de type « spring embedder » sur les réseaux de co-auteurs. Voir le schéma explicatif sous « [Visualisation interactive de documents](#) » par Mountaz Hascoët, labo. LIRMM, Univ. Montpellier II, slide 31 page 16 et une application à la visualisation d'un réseau de co-authorship sous Sato.S, et al. « [Readable Representations for Large-Scale Bipartite Graphs](#) » *Knowledge-Based Intelligent Information and Engineering Systems ; Lecture Notes in Computer Science*, 2008, Volume 5178/2008, 831-838, DOI: 10.1007/978-3-540-85565-1_103

Deux concepts récurrents dans la littérature spécialisée :

- Base de connaissance (*knowledge base*) = cluster de papiers co-cités
- *Research fronts* = Cluster of bibliographically coupled articles

L'abondance des données de co-citations peut perturber la conception et l'exploitation des matrices. La densité des références est réduite par une « digital grid » qui sert de filtre. Traduction sous forme d'aiguilles pyramidales colorées à base plus ou moins large et donc plus ou moins hautes, donnant un aspect global « montagneux » où la visibilité des différentes altitudes permet de mesurer la densité des partages (High amount of concentrated locally similar publications). Permet de voir l'évolution et le déclin éventuel de tel ou tel thème en comparant les diagrammes d'aiguilles (ou d'éminences) sur plusieurs années.

Les signaux faibles sont détectés à la marge : petites éminences qu'il convient d'analyser sur une courte période pour voir si, en comparant par exemple sur deux échantillons de 4 et 2 ans, elles croissent ou décroissent. Dans le premier cas, le concept est abandonné. Voir le modèle de Ward.

Voir également le schéma de co-citations réalisé par deux communicants du colloque : E.Schiebel et J.Gorraiz, dans « [International publication output and research impact in social sciences: comparison of the Universities of Vienna, Zurich and Oslo](#) », *Research Evaluation*, 18(3), September 2009, pages 226, 229 et 231, avec le logiciel *BibTechMonTM*

What to do (and not to do) with corporate address data

Matthias Winterhager, Bielefeld University, Institute of Science and Technology Studies (IWT), Germany

L'abréviation des affiliations est souvent due au manque de place dans l'édition imprimée ancienne du SCI.

Aujourd'hui : C1... corporate address 1... ET séparément RP reprint address

Une adresse peut être périmée suivant l'espace de temps entre la rédaction et la publication. La nouvelle adresse peut être mentionnée, le cas échéant, mais uniquement sur la publication de l'article lui-même, pas dans le WoS => problème

L'article le plus cité du Dpt Economie de l'Univ de Bielefeld comporte comme auteur unique affilié à cette entité le nom d'un ancien étudiant qui n'a jamais fait partie du personnel. La mention des autres co-auteurs, venus d'ailleurs, sauve la mise.

Pb des attaches géographiques diverses : par ex., l'Inst. Max Planck n'est pas qu'en Allemagne...

Existents des systèmes modélisant des chaînes de caractères associées à des intitulés d'affiliations.

[Norme NISO sur l'adressage.](#)

La procédure consistant à instituer des Identifiants attribués aux affiliations peut être une solution. Mais sa mise en œuvre étendue au niveau international risque de prendre beaucoup de temps...

Introduction to Bibliometric Data Sources

Wolfgang Glänzel, Centre for R&D Monitoring (ECOOM), Katholieke Universiteit Leuven, Belgium / Juan Gorraiz, Bibliometrics Department, University of Vienna, Austria

Combiner, autant que possible, l'interro d'une BD bibliométrique avec une autre, spécifiquement bibliographique (Inspec, Medline, Math Reviews, CAS etc...)[cf. *supra* intervention de Marx]

Reprend les critiques classiques à l'encontre de G.Scholar (absence de référentiel, base « ouverte » etc...)

Image fréquemment évoquée du nombre d'Erdős qui mesure la distance entre des collaborateurs, conscients ou non de cette collaboration : « Le nombre d'Erdős d'un mathématicien M est le plus petit nombre d'Erdős de tous les mathématiciens avec qui M a cosigné un article mathématique, plus un (si M a un nombre de Erdős qui vaut 1, cela signifie qu'il a écrit un article avec Erdős, un nombre de Erdős égal à 2 signifie qu'il a cosigné un article avec un collaborateur direct de Erdős...) ». [Défin. Wikipédia](#).

On a dit plus haut que les BD spécifiquement bibliométriques ne suffisent pas. Font aussi partie des BD de référence :

- Les Thesauri
- Le brevets et les normes
- La doc. en OAI

OAI : on télécharge d'abord, parce que c'est gratuit : après, « on voit ». Quand c'est payant, c'est l'inverse : on analyse le résumé, les keywords, l'éventuelle réputation des auteurs à travers d'autres titres, puis on télécharge. Donc le nb de téléchargements sous OA ne correspond pas forcément à de la notoriété engendrant de la citation intellectuellement appréciable. Il est prudent de ne faut pas considérer le downloading OA comme un critère fiable en matière d'évaluation.