Mesure de l'information et bibliométrie.

Application à l'information scientifique et technique pour l'étude des processus de construction d'information élaborée

Hervé Rostaing
Université Paul Cézanne Aix-Marseille III
herve.rostaing@univ-cezanne.fr

IRSIC - EA 4262
Institut de Recherche en Sciences de l'Information et de la Communicatiton

Sommaire

- **Positionnement de la bibliométrie dans les SIC**
- **Positionnement de mes travaux en bibliométrie**
- **Application de la bibliométrie à la veille technologique**
- **Application de la bibliométrie comme instrument d'aide au pilotage de la recherche**

Récent regain d'intérêt pour l'infométrie et la bibliométrie

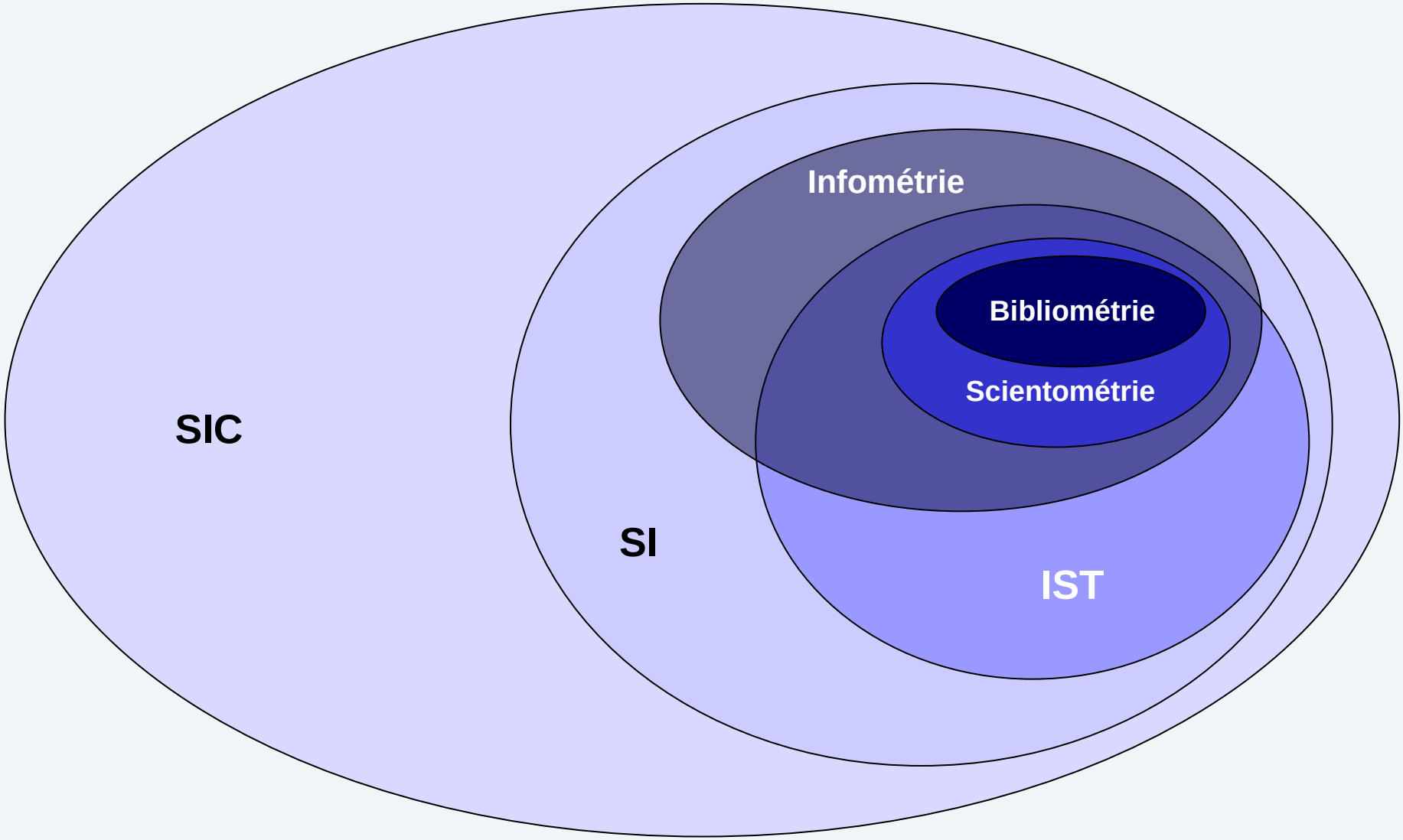
- **Les palmarès**
  - Shanghai et autres…
- **Les indicateurs d'évaluation de la recherche**
  - h-index, facteur d'impact et autres…
- **La réforme de la LOLF**
  - logique de financement par objectif
  - logique de mesure pour évaluer les objectifs atteints

➢ **Un besoin de mesures -> se réduit souvent à de simples indicateurs quantitatifs**

Infométrie et Sciences de l'information et de la communication

- **Infométrie :**
  - décrire quantitativement des phénomènes informationnels
    - Étudier tout type de données d'un processus informationnel
- **Scientométrie :**
  - étudier les activités scientifiques par des mesures quantitatives
    - Étudier les intrants, les extrants et les mécanismes de transformation mis en jeu dans les activités scientifiques
- **Bibliométrie**
  - statistique bibliographique : mesure quantitative de l'activité scientifique réduite à l'étude de la publication de travaux scientifiques
    - Étudier uniquement un type d'extrants des activités scientifiques

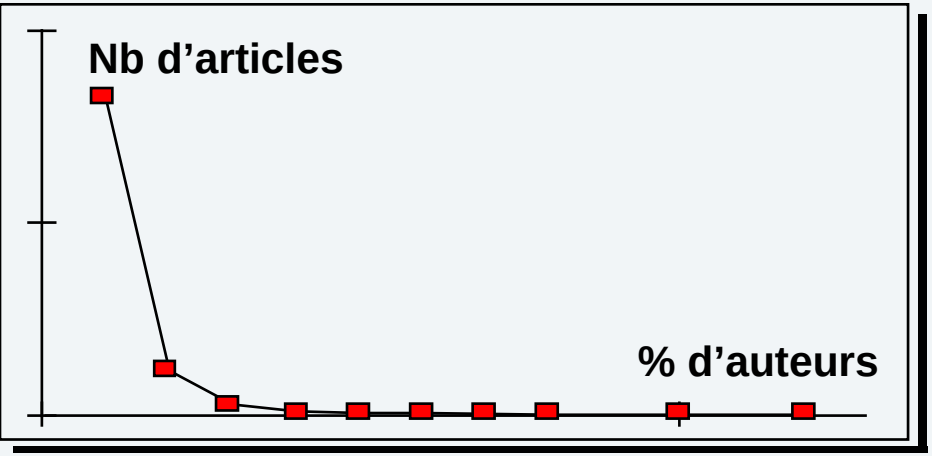
Positionnement de la bibliométrie dans les SIC

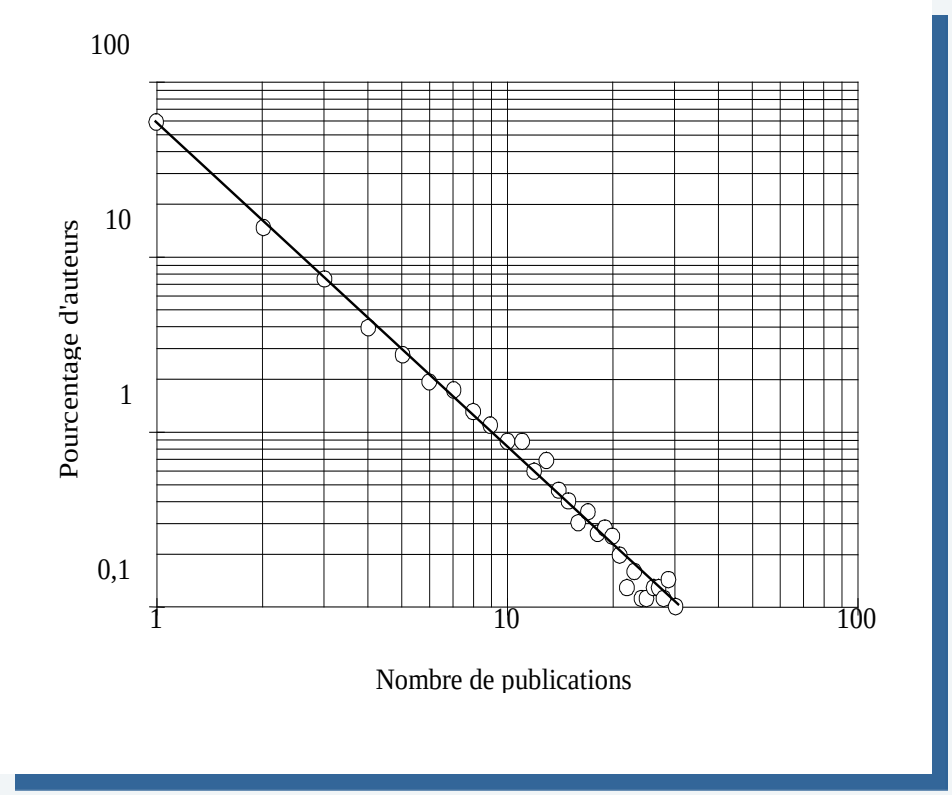


Les précurseurs de l'étude quantitative de l'information

- **Les premières études quantitatives ont cherché à établir des « lois »**
  - Loi = relation quantitative constante entre l'apparition d'un phénomène et des facteurs déterminants
  - Notion de régularité, de relation reproductible donc prévisible
- **Lotka en 1926**
  - Établit une relation entre le nombre d'auteurs et le nombre d'articles qu'ils publient
- **Bradford en 1934**
  - Établit une relation entre le nombre d'articles traitant d'un sujet et le nombre de revues qui les ont publié
- **Zipf en 1935**
  - Établit une relation entre la production d'un texte et la fréquence des mots qui le compose
- **Price en 1970**

Loi de lotka





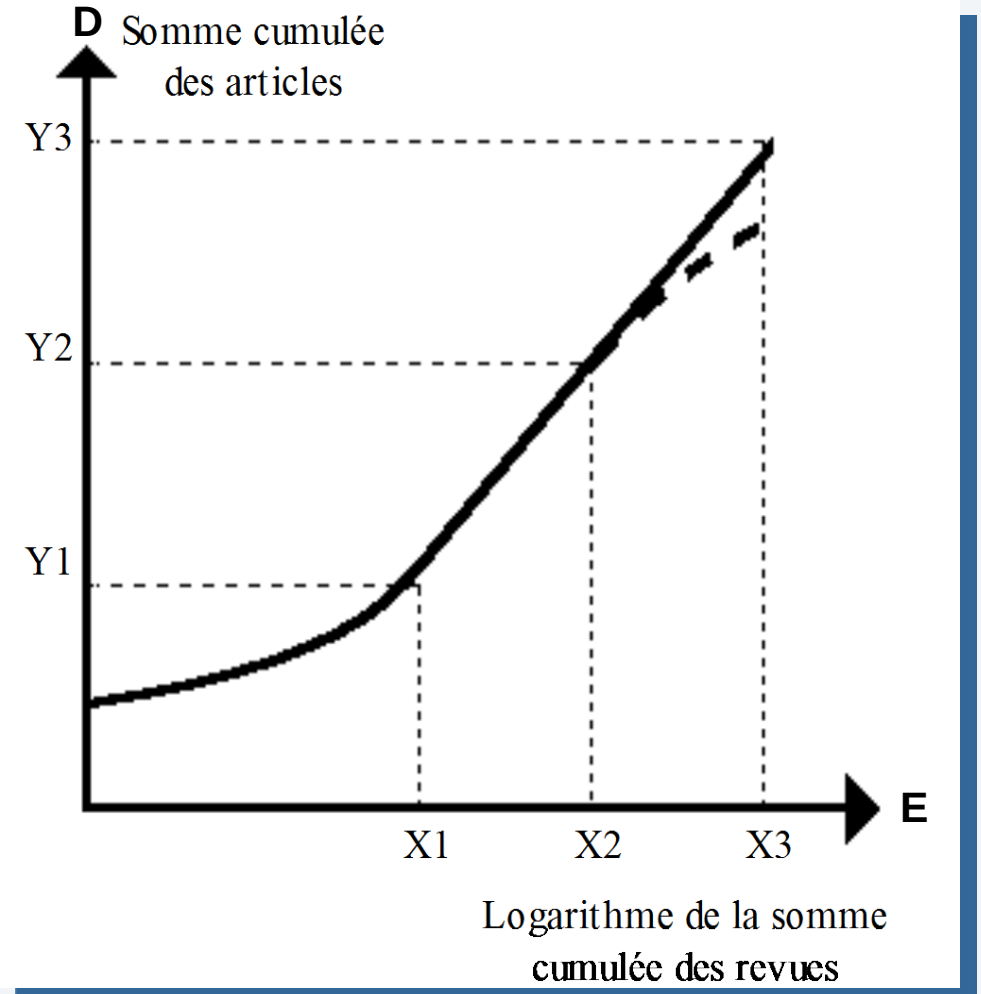
- **Objectif :**
  - Mesurer la part de contribution de chaque chercheur au progrès scientifique
- **Travaux :**
  - index du CAS entre 1907-16 : auteurs commençant par A-B
  - classement de ces auteurs par productivité (rang de fréquence)
  - Histogramme et graphe logarithmique
- **Loi :**
  - $Y = X^n \times C$ sur ses données
    $Y = C / X^2$
  - Par rapport au nb d 'auteurs ayant 1 pub, il y en a 4 fois moins ayant 2 pub, 9 fois moins ayant 3 pub...
- **Signification**
  - le nombre d'articles publiés n'est pas distribué de manière homogène entre les chercheurs. La production scientifique a tendance à se concentrer sur un nombre limité de chercheurs

Les précurseurs de l'étude quantitative de l'information

- **Les premières études quantitatives ont cherché à établir des « lois »**
  - Loi = relation quantitative constante entre l'apparition d'un phénomène et des facteurs déterminants
  - Notion de régularité, de relation reproductible donc prévisible
- **Lotka en 1926**
  - Établit une relation entre le nombre d'auteurs et le nombre d'articles qu'ils publient
- **Bradford en 1934**
  - Établit une relation entre le nombre d'articles traitant d'un sujet et le nombre de revues qui les ont publié
- **Zipf en 1935**
  - Établit une relation entre la production d'un texte et la fréquence des mots qui le compose
- **Price en 1970**

La loi de Bradford

A	B	C	D	E
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 périodiques d'un domaine
- **Travaux :**
  - Deux séries de données exp.
  - Tableau classant les revues par rang de «productivité»
  - Graphe de D = f(E)
- **Loi :**
  - noyau = (X1,Y1)
  - si Y2 = 2xY1 alors $X2 = X1^2$
  - affaissement de Groos
- **Signification**
  - les articles publiés sur un sujet ne sont pas distribués de manière homogène dans les revues scientifiques. La production scientifique sur un sujet a tendance à se concentrer sur un nombre limité de revues

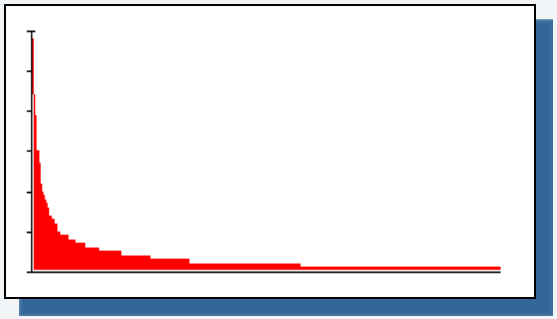
Les précurseurs de l'étude quantitative de l'information

- **Les premières études quantitatives ont cherché à établir des « lois »**
  - Loi = relation quantitative constante entre l'apparition d'un phénomène et des facteurs déterminants
  - Notion de régularité, de relation reproductible donc prévisible
- **Lotka en 1926**
  - Établit une relation entre le nombre d'auteurs et le nombre d'articles qu'ils publient
- **Bradford en 1934**
  - Établit une relation entre le nombre d'articles traitant d'un sujet et le nombre de revues qui les ont publié
- **Zipf en 1935**
  - Établit une relation entre la production d'un texte et la fréquence des mots qui le compose
- **Price en 1970**

La loi de Zipf

- **Objectif : étude linguistique des écrits littéraires par le biais de la fréquence d'utilisation des mots**
- **Travaux :**
  - comptage des mots présents dans *Ulysses* de Joyce :
    29 899 occurrences de mots
  - classé les mots par fréquence décroissante et numéroté par rang
  - le produit du rang par la fréquence donne un résultat proche d 'une constante
  - courbe de la forme 1/x
- **Signification :**
  - Les fréquences des mots d'un texte ne sont pas distribués de manière homogène. L'auteur utilise très fréquemment les mêmes mots.
  - « Principe du moindre effort »

Rang (I) (r)	Fréquence (II) (f)	Produit de I et II (r * f = C)
10	2.653	26.530
30	926	27.780
50	556	27.800
100	265	26.500
300	84	25.200
500	50	25.000
1.000	26	26.000
3.000	8	24.000
5.000	5	25.000
10.000	2	20.000
20.000	1	20.000
29.899	1	29.899



Les précurseurs de l'étude quantitative de l'information

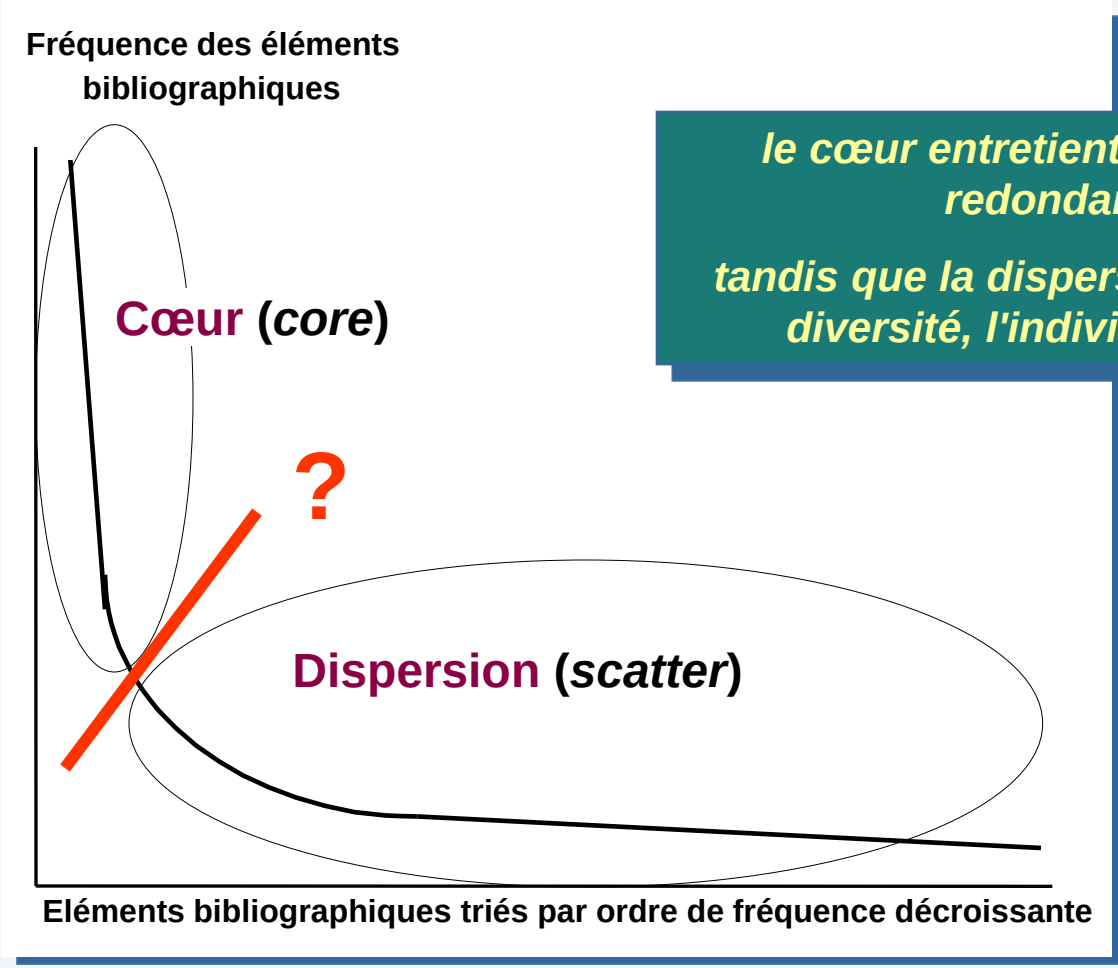
- **Les premières études quantitatives ont cherché à établir des « lois »**
  - Loi = relation quantitative constante entre l'apparition d'un phénomène et des facteurs déterminants
  - Notion de régularité, de relation reproductible donc prévisible
- **Lotka en 1926**
  - Établit une relation entre le nombre d'auteurs et le nombre d'articles qu'ils publient
- **Bradford en 1934**
  - Établit une relation entre le nombre d'articles traitant d'un sujet et le nombre de revues qui les ont publié
- **Zipf en 1935**
  - Établit une relation entre la production d'un texte et la fréquence des mots qui le compose
- **Les travaux de Price dans les années 70**

Apport de ces premières études

- **Les données bibliométriques suivent des régularités respectant des lois hyperboliques**
  - Autant les lois en sciences naturelles suivent des répartitions gaussiennes, celles en sciences humaines et sociales sont de nature hyperboliques (Zipfienne)
  - Interprétation anglo-saxonne, productiviste et élitiste : Cœur / Dispersion (Pareto)
  - Interprétation lexicale et plus descriptive : Trivial / Information / Bruit

- **Conséquences de ces travaux**
  - 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

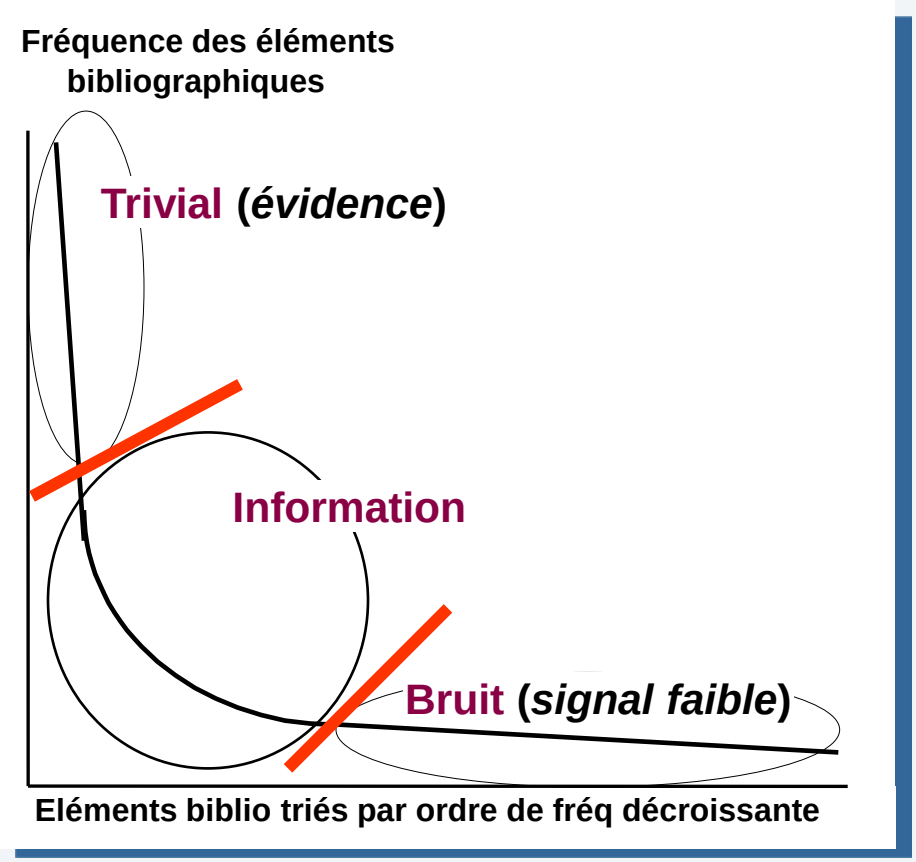
Interprétation anglo-saxonne des lois hyperboliques



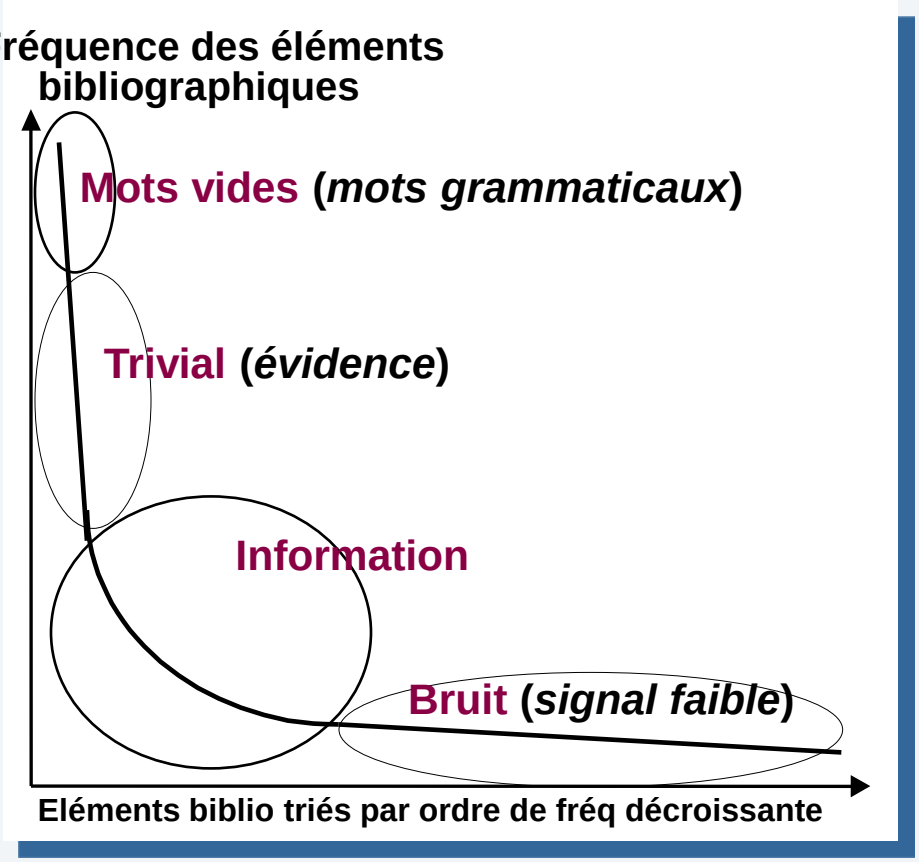
le cœur entretient l'identité, la redondance

tandis que la dispersion contient la diversité, l'individualisation

Interprétation lexicale des lois hyperboliques



Cas général d'éléments de champs bibliographiques contrôlés

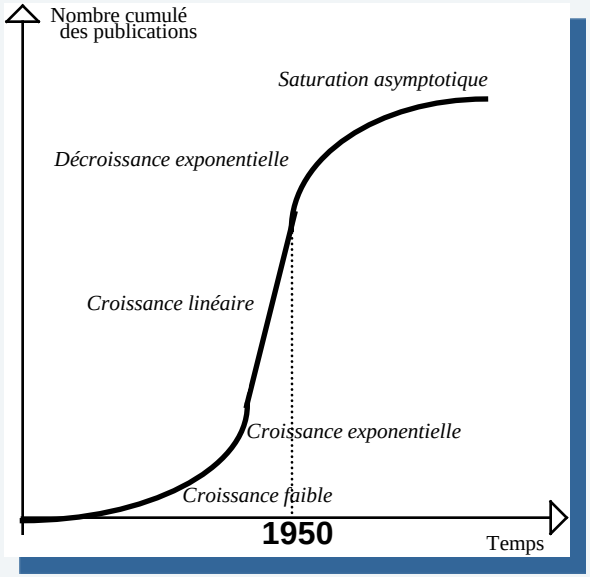


Cas d'éléments de champs bibliographiques non contrôlés

Les travaux de Price dans les années 70

- **Objectifs :**
  - définir les lois qui régissent l'évolution et le comportement de la science
    - Scientométrie : la *science de la science*
  - Les principales mesures employées : la production des publications et les citations entre publications
    - ne plus mesurer que la production mais aussi le degré d' « utilisation » des articles

- **Travaux :**
  - structure cumulative de la science
  - croissance exponentielle de la connaissance puis saturation logistique
  - généralisation de la loi hyperbolique à la productivité scientifique (analogie avec la loi de Pareto des 80/20)
  - théorie des *avantages cumulés* : le succès est récompensé alors que l'échec n'a aucun effet -> formule mathématique unificatrice des lois hyperboles..
  - les *collèges invisibles* : groupes d'élite qui se constitueraient au sommet de la communauté scientifique et autour d'un front de recherche



La vague des travaux sur la citation et l'apparition des premiers indicateurs

- **Collaboration de Price avec l'ISI**
  - creusets de la science et fronts de recherche
  - facteur d'immédiateté

    nombre de citations reçues l'année *y* pour les articles publiés la même année
    dans un domaine scientifique divisé par le nombre d'articles publiés cette année là

$$I_I = \frac{c_y}{p_y}$$

- **Application de ces travaux par l'ISI (Garfield)**
  - mesure de la citation comme indicateur de la notoriété des revues
  - pondération par l'application du facteur d'immédiateté aux revues
  - puis amélioration de la pondération : facteur d'impact

    nombre de citations reçues l'année *x* pour les articles publiés pendant les deux années précédentes,
    divisé par le nombre d'articles publiés par cette même revue pour ces deux années précédentes

$$I_F = \frac{c_x}{p_{x-1} + p_{x-2}}$$

- **Premières représentation de la structure de l'organisation de la science (Garfield & Small)**
  - les cartes de la science par l'analyse des co-citations

Les années 80 : déploiement des techniques d'analyse des relations

- **Besoins de représenter les relations qu'entretiennent les publications entre elles ou les autres éléments bibliographiques entre eux (acteurs de la science, concepts et connaissances**
  - démontrer les hypothèses de Price : la structure cumulative de l'organisation des sciences, les fronts de recherche, les collèges invisibles…
  - décrire la structure et l'évolution de la science et des disciplines

- **Emploi des techniques statistiques multidimensionnelles et cartographiques faisant appel aux méthodes statistiques descriptives (analyse des données)**
  - ne cherche plus à formuler mathématiquement une régularité statistique (ajustement des données à un modèle)
  - cherche à décrire les relations que les données entretiennent entre elles sans présumé de la reproductibilité de cette relation (pas de modèle sous-jacent établi)

Principaux travaux d'analyse relationnelle en bibliométrie

- **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
    - ...

Postulats et hypothèses en bibliométrie

- **Postulat 1 : l'essentiel de l'activité de la recherche scientifique est de produire des connaissances**
  - les produits de la science ne sont pas des objets mais des idées, des moyens de communication, des réactions aux idées des autres : des connaissances
- **Postulat 2 : l'écrit scientifique est le produit objectif cette production de connaissances**
  - une publication est une représentation objective de l'activité de recherche de son auteur
- **Postulat 3 : il existe une relation entre les écrits scientifiques (construction « cumulative » des connaissances)**
  - cette relation entre tous les travaux scientifiques publiés est mesurable
- **Hypothèse supplémentaire : réductionnisme bibliométrique**
  - Les travaux scientifiques importants sont essentiellement publiés dans des revues scientifiques internationales à forte notoriété
  - Cette hypothèse doit être vérifier à chaque étude bibliométrique menée

Bibliométrie : instrument d'étude des processus en IST

- **Décrire et mesurer les phénomènes et le processus mettant en jeu des informations scientifiques et techniques**
  - processus de construction de l'information scientifique et technique
    - Comment s'organise les acteurs de la science ? Comment émerge les nouvelles connaissance ? …
  - processus de communication de l'information scientifique et technique
    - Comment les acteurs de la science communiquent ? Influence des médias et médiateurs en IST ?...
  - processus d'usage de l'information scientifique et technique
    - Comment les acteurs de la science utilisent les travaux des autres ? Comment se diffusent les travaux scientifiques et leur influence ? …

Sommaire

- Positionnement de la bibliométrie dans les SIC
- Positionnement de mes travaux en bibliométrie
- Application de la bibliométrie à la veille technologique
- Application de la bibliométrie comme instrument d'aide au pilotage de la recherche

Positionnement de mes travaux en bibliométrie

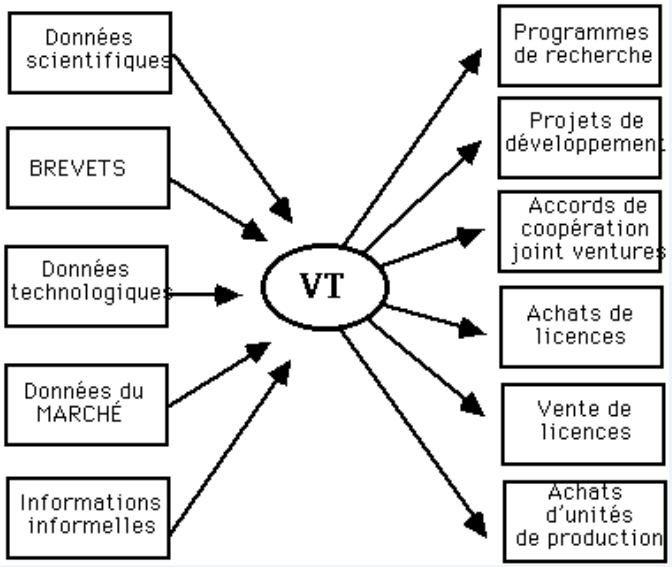
- **Bibliométrie comme instrument d'aide à la construction d'information élaborée**
  - Information élaborée : information à valeur ajoutée issue d'une processus de construction (analyse et interprétation) d'un ensemble de données variées et parcellaires
  - L'information élaborée s'inscrit dans un besoin informationnel préalablement défini (généralement pour d'accompagner une démarche de prise de décision)
  - Les processus informationnels étudiés :
    - définition du besoin en information élaborée
    - construction de l'information élaborée ⟸ **Bibliométrie**
    - exploitation de l'information élaborée
- **Domaine d'application de ces travaux**
  - La veille technologique
  - L'aide au pilotage de la recherche

Sommaire

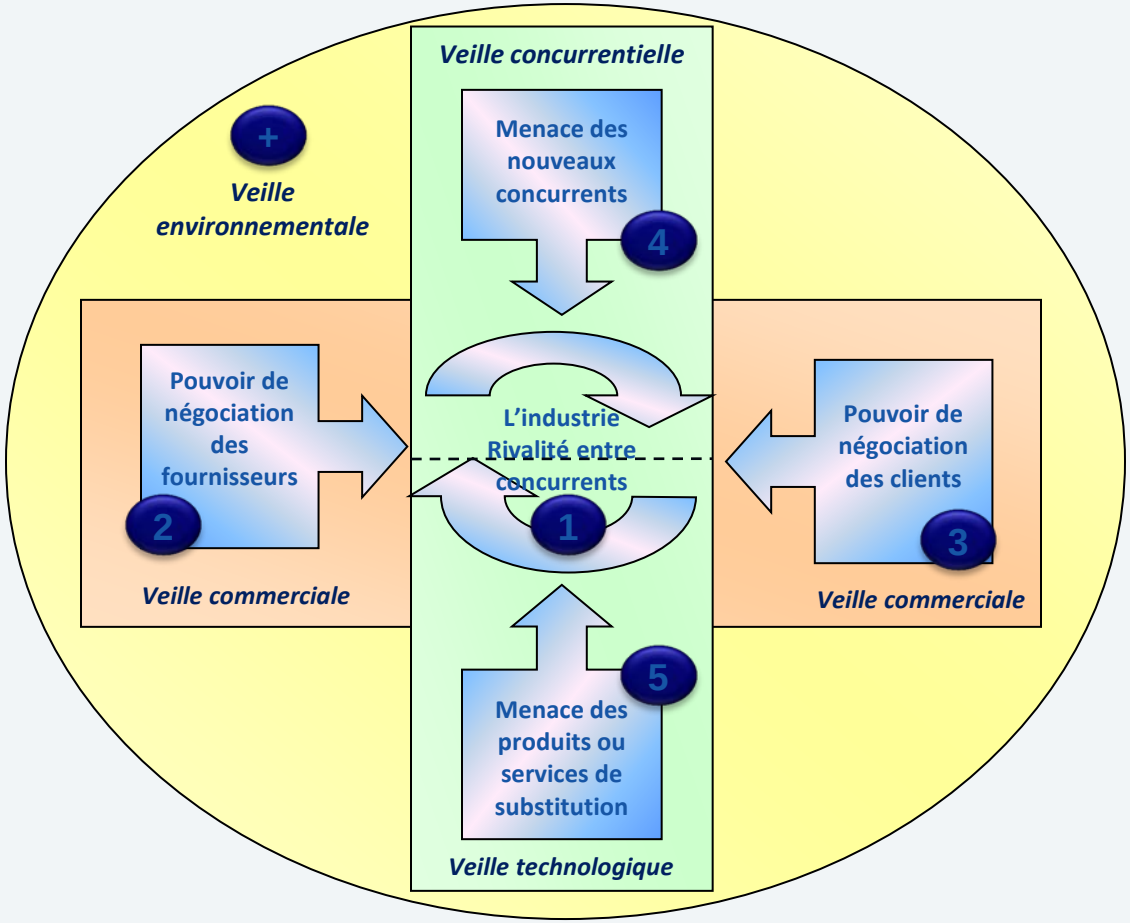
- Positionnement de la bibliométrie dans les SIC
- Positionnement de mes travaux en bibliométrie
- Application de la bibliométrie à la veille technologique
- Application de la bibliométrie comme instrument d'aide au pilotage de la recherche

Définition de la veille technologique

- **La Veille Technologique pour une entreprise peut être définie comme étant :**
  - l'observation et **l'analyse** de l'environnement scientifique, technique et technologique pour dégager les **menaces** et saisir les **opportunités** de développement



Les différentes veilles inspirées des 5 forces de M Porter



Intelligence compétitive ou veille stratégique

- **La veille concurrentielle (force 1 & 4 de M Porter)**
  - porte sur les concurrents actuels ou potentiels, les nouveaux entrants sur le marché
- **La veille commerciale (force 2 & 3 de M Porter)**
  - porte sur l'évolution des besoins des clients, mais aussi des fournisseurs, des sous-traitants, des partenaires de l'entreprise
- **La veille technologique (force 1 & 5 de M Porter)**
  - Porte sur l'évolution des sciences, techniques, technologies et innovations
- **La veille environnementale (forces supplémentaires)**
  - englobe le reste de l'environnement de l'entreprise intégrant les éléments de l'environnement politique, social, culturel et juridique… veille sociétale ; veille juridique et réglementaire ; veille politique, diplomatique et risque pays, veille financière, actionnariale et boursière; veille image ; veille tendance…

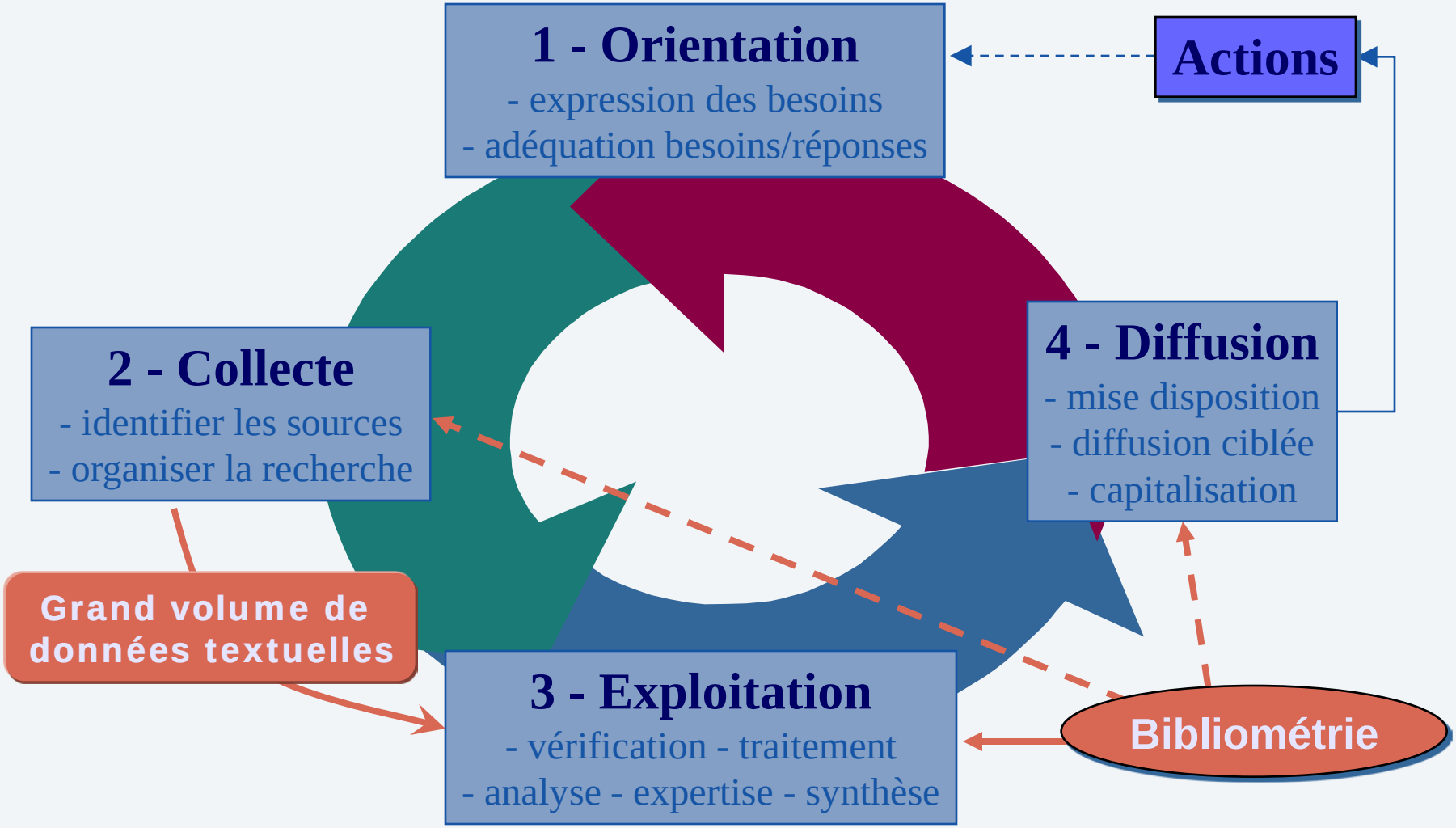
Besoins en information élaborée en veille technologique

- **Parfaitement connaître l'environnement scientifique et technique**

Par exemple :

- Positionnement concurrentiel en R&D
- Identification des acteurs d'un domaine et leurs spécialités
- Cartographie des recherches & des techniques et de leurs interactions
- Etudes dynamiques : émergence d'acteurs et de nouvelles technologies
- Analyse des portes feuilles brevets
- …

Application de la bibliométrie dans le processus de la Veille Technologique



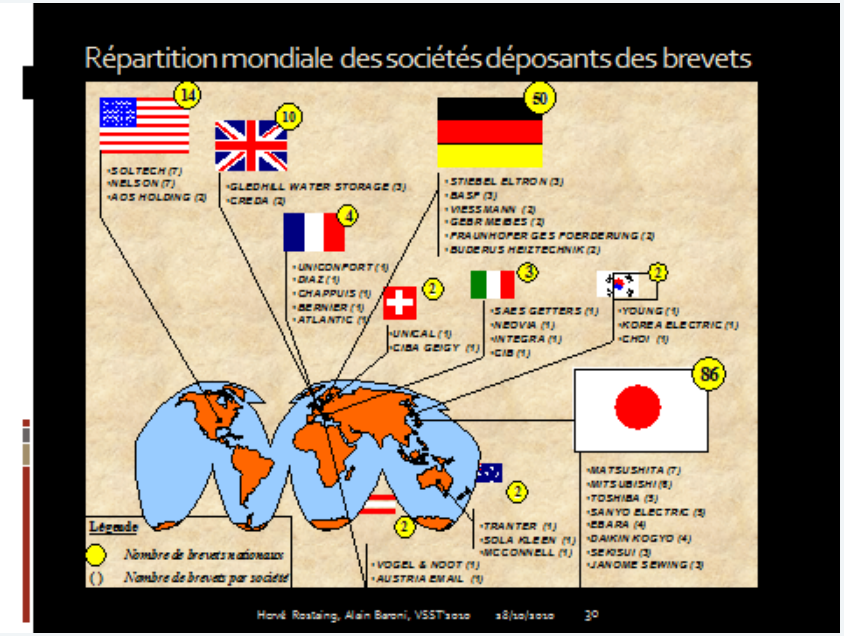
Apport de la Bibliométrie en Veille Technologique

- Exploitation du volume croissant des données à traiter
- Compréhension de la complexité croissante des connaissances due à la fragmentation des données
- Instrument d'aide au processus d'analyse par les experts en offrant des visions structurées des données

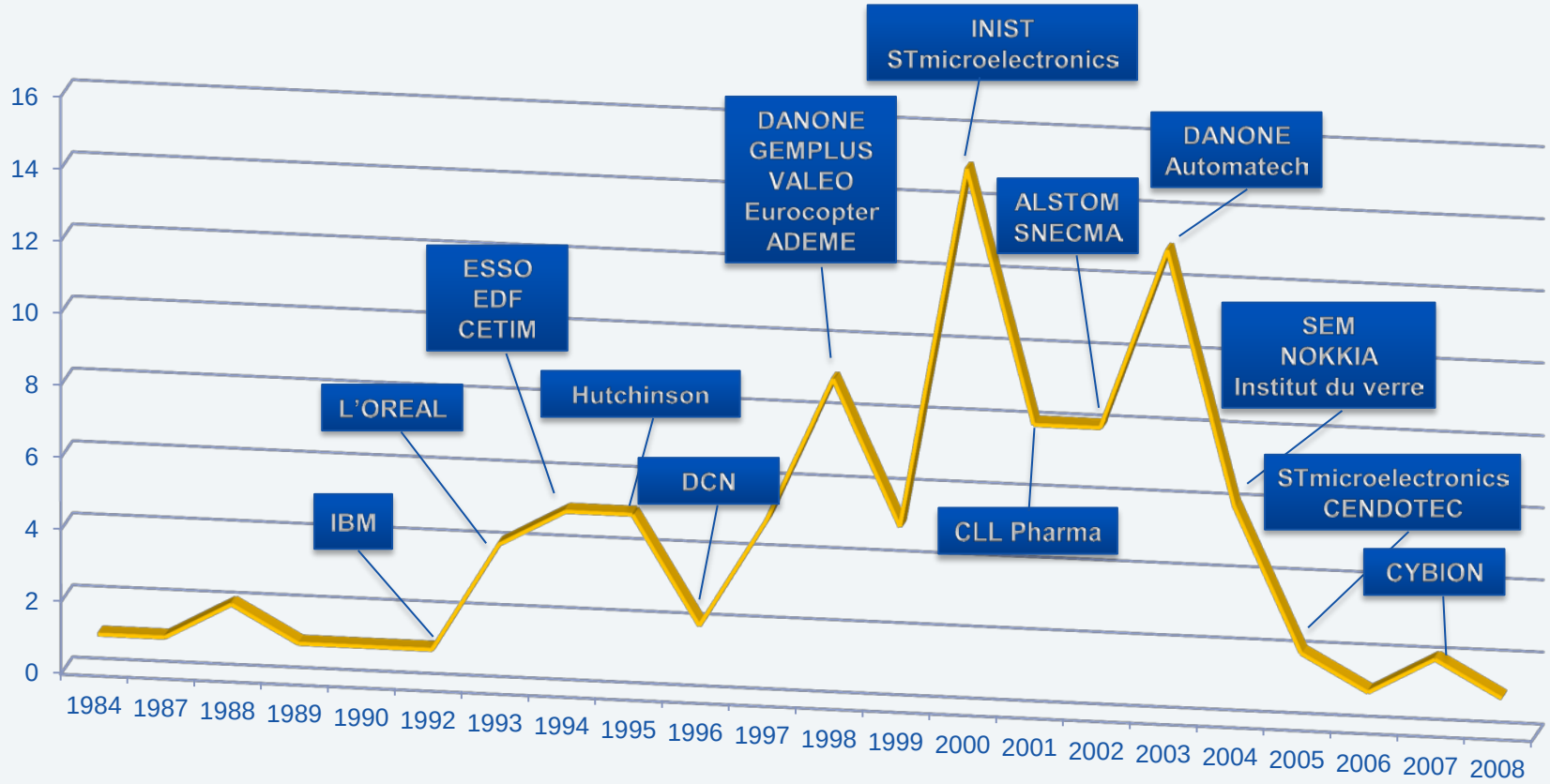
Offrir des "grilles de lecture" de grand volume de données représentants des connaissances scientifiques ou techniques

-> instruments d'aide à la création d'information élaborée

- **Processus de construction d'information élaborée assisté par une analyse bibliométrique**
  - Aide à la sélection et à l'expertise lors d'une étude de veille technologique



Thèses en entreprises – retour d'expériences veille technologique et bibliométrie



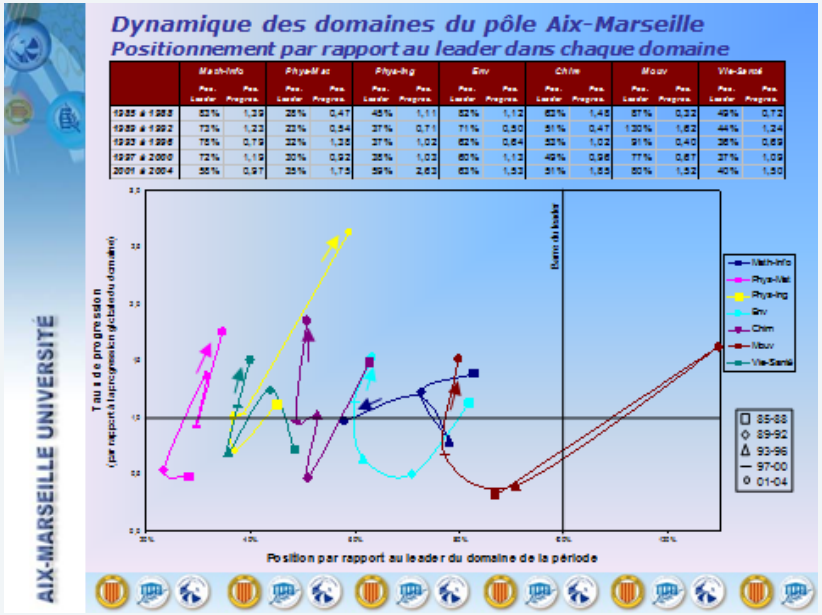
Sommaire

- **Positionnement de la bibliométrie dans les SIC**
- **Positionnement de mes travaux en bibliométrie**
- **Application de la bibliométrie à la veille technologique**
- **Application de la bibliométrie comme instrument d'aide au pilotage de la recherche**

Pourquoi cette nouvelle orientation ?

- **L'effet Palmarès de Shanghai**
  - 2003, contre étude commanditée par le Président de U3
- **Le projet d'un infocentre recherche**
  - 2004, la Présidence de U3 lance la création d'un infocentre recherche
- **Etude bibliométrique pour le COS**
  - 2006, étude bibliométrique du positionnement du pole scientifique d'Aix-Marseille demandé par le Comité de pilotage du Comité d'Orientation Scientifique
- **Indicateurs des publications pour le contrat quadriennal**
  - 2006, élaboration des indicateurs de publications de la Faculté des Sciences et Techniques pour le Contrat quadriennal
- **Participation au Programme150**
  - 2007, contribution à l'élaboration des indicateurs de publications de U3 demandé par le Ministère auprès de l'OST

- **Processus de construction d'information élaborée assisté par une analyse bibliométrique**
  - Positionnement scientifique du pôle universitaire d'Aix-Marseille



Idée de nouvel axe de recherche

- **L'influence de la bibliométrie et de l'évaluation quantitative de la recherche sur les pratiques des acteurs de la recherche ?**
  - Changements dans les instances dirigeantes des universités ?
  - Changements dans les directions des unités de recherche ?
  - Changements dans les pratiques de recherche des chercheurs ?