

# Cartographie d'un champ de recherche à partir d'un corpus de thèses

Les Sciences de l'Information et de la  
Communication en France 1974-94

*Collaboration*

**Equipe RI3 - IUT II Grenoble**

**CRRM - Aix-Marseille**

# Plan de la présentation

- Le champ de recherche
- Constitution du corpus des thèses
- Etablissement de cartographies
- Principaux résultats
- Quelles perspectives ?

# Le champ de recherche

- Science(s) de l'information
- Sciences de l'information et de la communication
- Autres disciplines

# Constitution du corpus des thèses

- Répertoire des thèses
  - Docthèses
  - Parinfo, 1991
- Problèmes de fiabilité
- Problèmes d'exhaustivité

# Caractéristiques du corpus retenu

- Critères
  - champ Discipline
  - appartenance 71e section
  - directeurs de thèses
  - collecte thématique
- Résultat
  - 1009 thèses
    - 659 thèses : SIC
    - 170 thèses : SID
    - 180 thèses : Autres disciplines

# Indexation du corpus

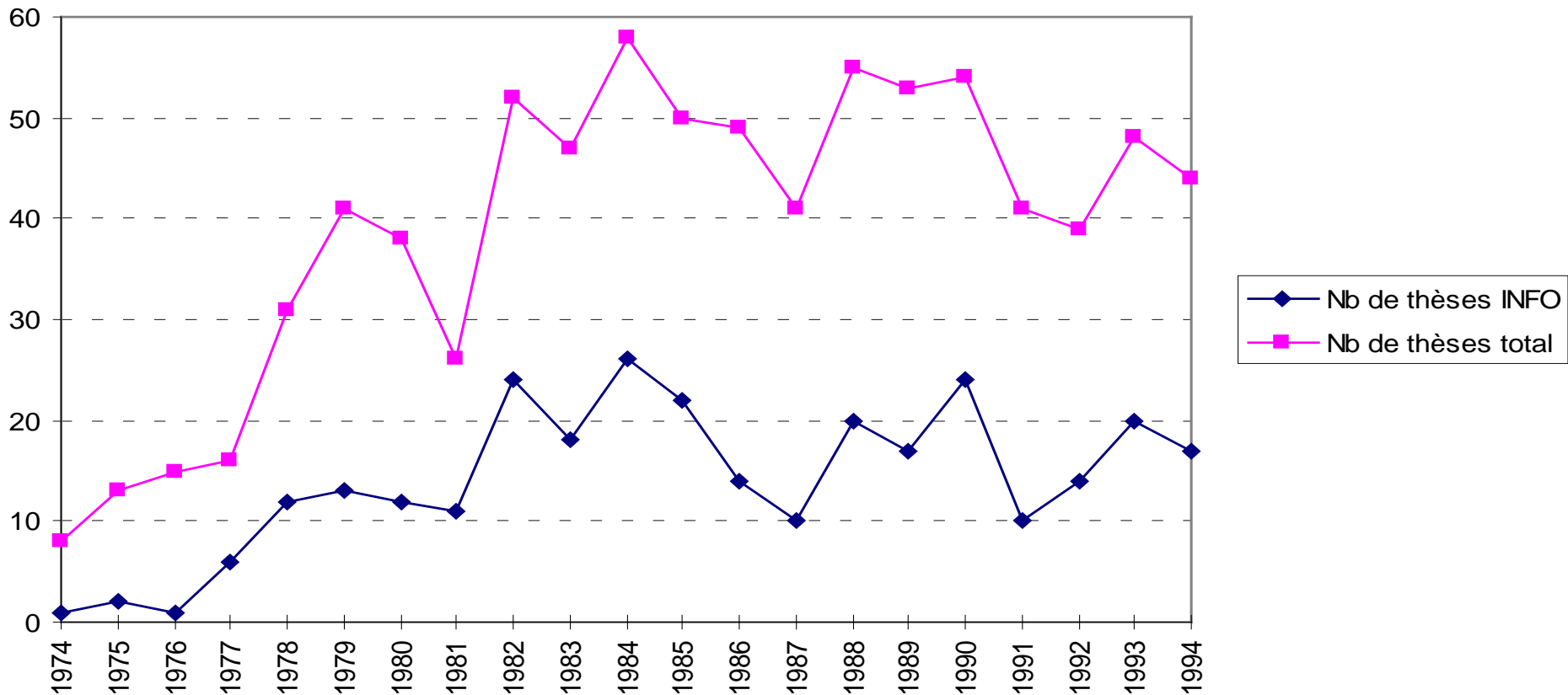
- 3 catégories
  - Sciences de la communication : 550 thèses (54%)
  - Sciences de l'information : 301 thèses (30%)
  - Hors champ et /ou En Marge : 167 thèses (16%)

# Etablissement de cartographies

- sur les éléments signalétiques
  - Date, université de soutenance, directeur de thèse, type de doctorat
- sur les éléments thématiques
  - Mots-clés, mots du titre, mots du résumé

# Evolution du nombre de thèses 1974-1994

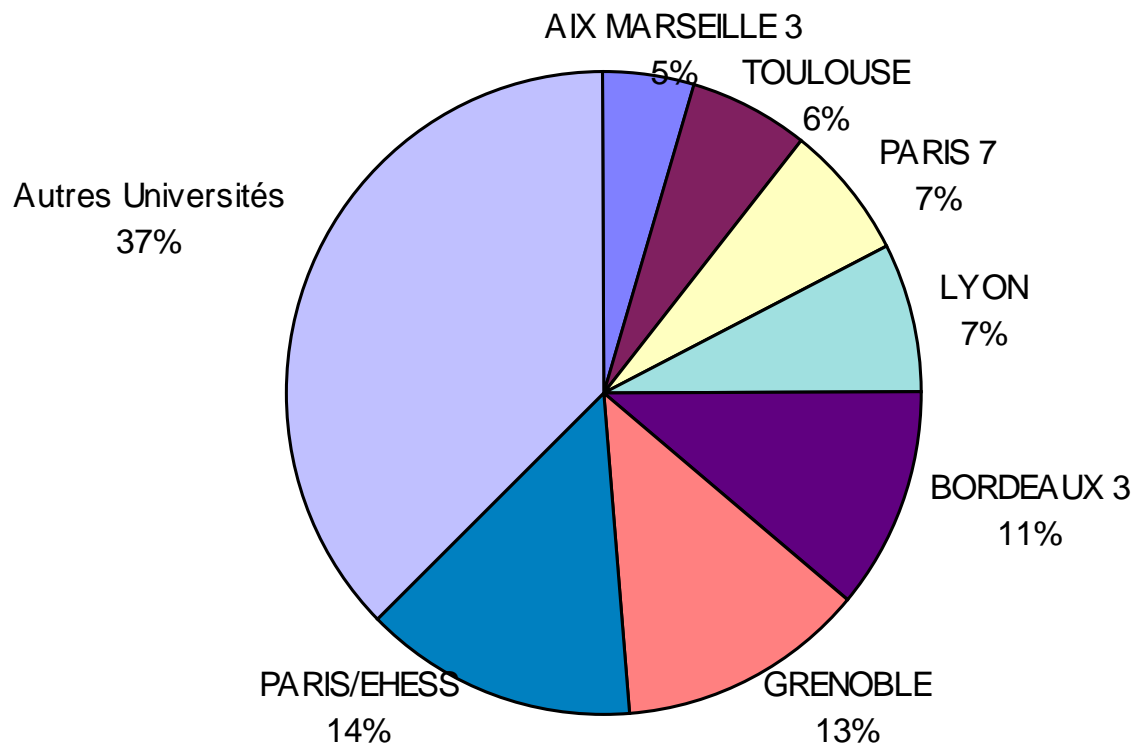
Evolution comparée du nombre de thèses INFO par rapport au nombre de thèses total



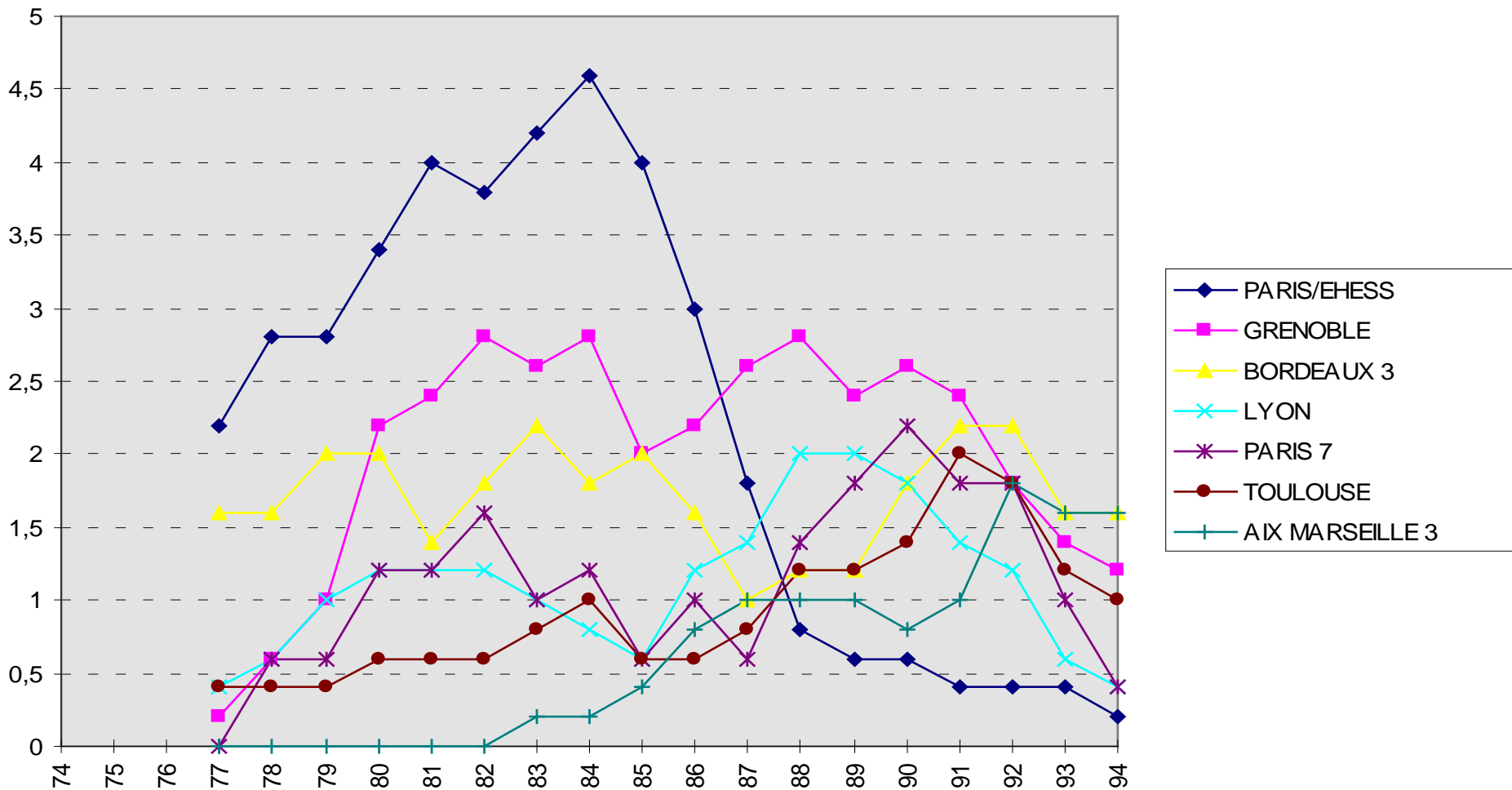


# Répartition par université de soutenance

Répartition des thèses INFO par Université de soutenance



# Evolution des pôles de formation



# Répartition par directeurs de thèses

- 117 profs pour 264 thèses
- 107 profs (92% des profs) n'ont dirigé que de 1 à 5 thèses en 20 ans (moyenne 1,46 thèse) --- >158 thèses
- 9 profs ont dirigé de 6 à 20 thèses en 20 ans (moyenne 11,78 thèses)---->106 thèses

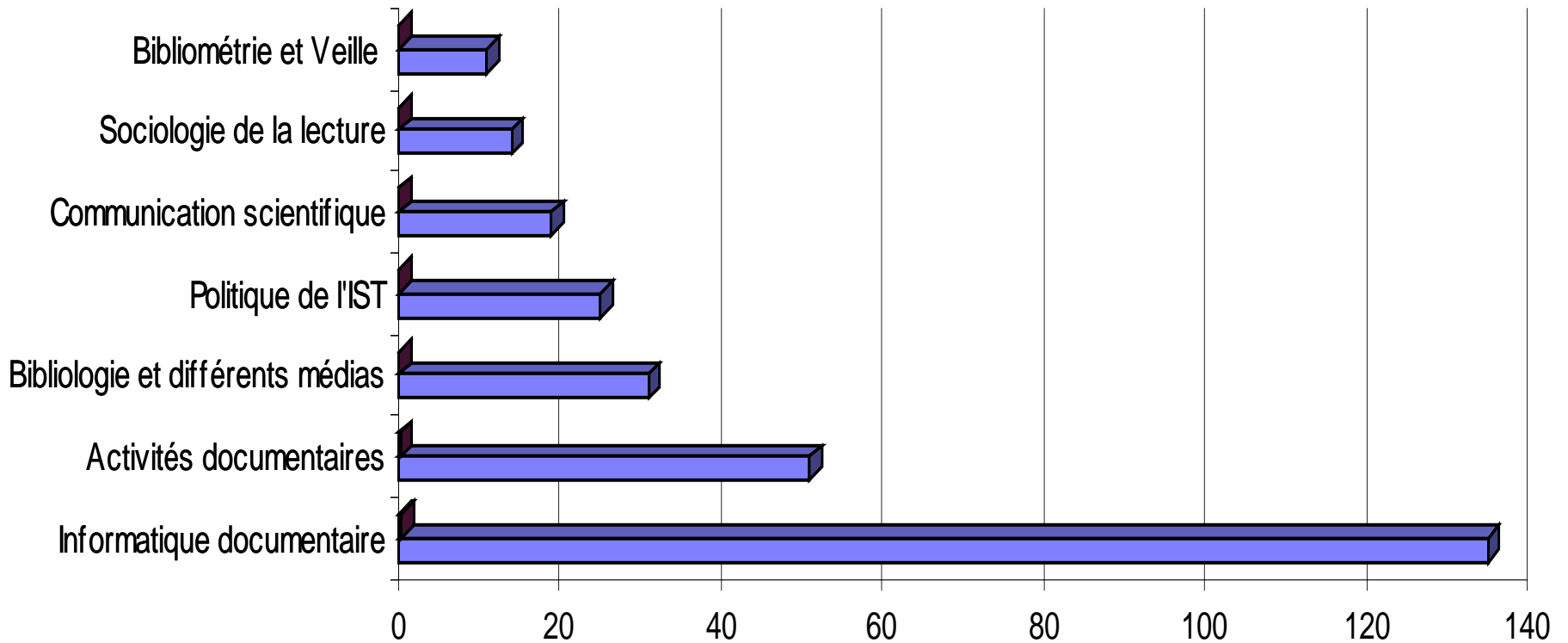
# Les cartographies thématiques

- Par enquête : Parinfo 1991
- Par classement d'«expert»+ tableur
- Leximappe
- Analyse réseau

# Parinfo 1991

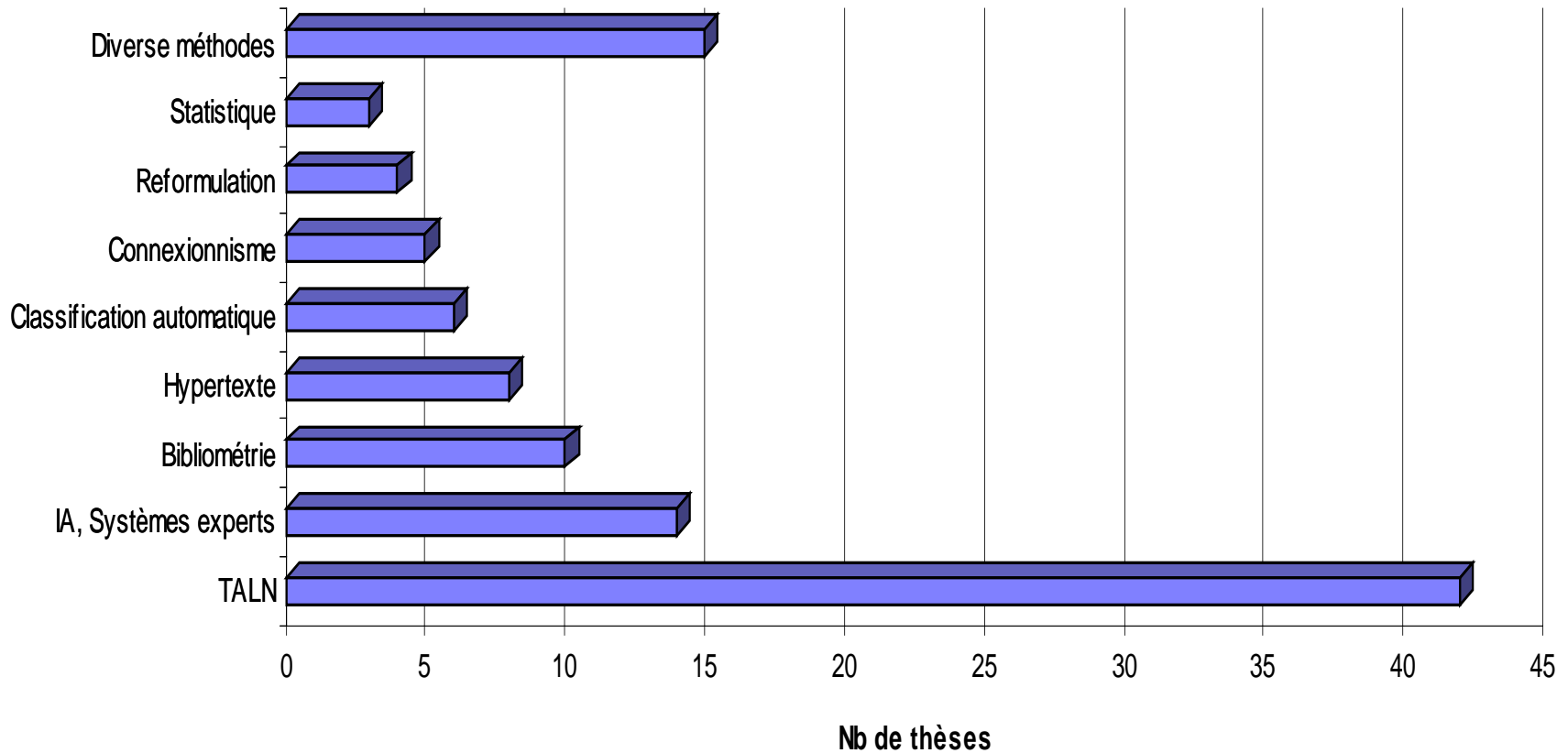
- Informatique, IA, systèmes experts
- Informatique linguistique
- Informatique documentaire
- Ergonomie, psychologie, philosophie
- Outils et méthodes d'analyse
- Droit
- Economie, Gestion
- Produits documentaires

# Indexation des thèses : les thèmes



# Indexation des thèses : les méthodes

Méthodes et techniques utilisées



# Indexation des thèses : les secteurs d'application

- 19 : Médecine, Santé
- 8 : Secteur juridique
- 6 : Education
- 5 : Chimie
- 4 : Bureautique
- 3 : Industrie
- 3 : Agriculture
- 3 : Documentation technique



# Déficiência du champ Mot-clé

- Forte absence du champ mots-clés avant 1986

- 379 champs absents sur l'ensemble du corpus

Années	Nb Notices	Nb Champs MC absents	Pourcentage d'absence
1983	51	49	96%
1984	64	55	86%
1985	56	39	70%

- Trop grande diversité du vocabulaire des mots-clés

- 2 193 mots-clés différents
- 1686 mots-clés à fréquence 1 = 77%
- nombreuses variantes graphiques

BASE DE DONNEES

BASE DOCUMENTAIRE

BASE DONNEE BIBLIOGRAPHIQUE

BASE DONNEE INTERNE

BASE DONNEE REPARTIE

BASES DE DONNEES BIBLIOGRAPHIQUES

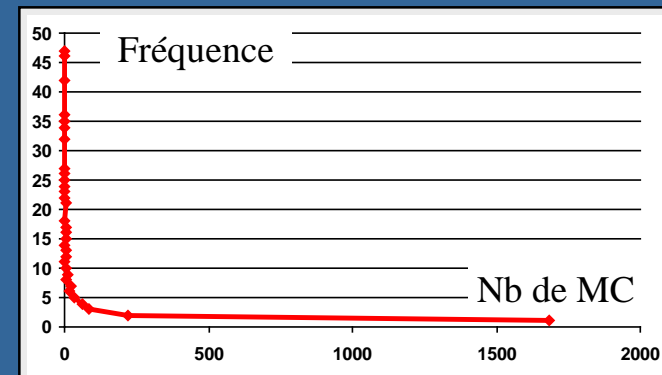
BASE DE DONNEES DOCUMENTAIRE

BASE DONNEE

BASE DONNEE DOCUMENTAIRE

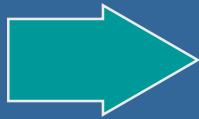
BASE DONNEE RELATIONNELLE

BASES DE DONNEES



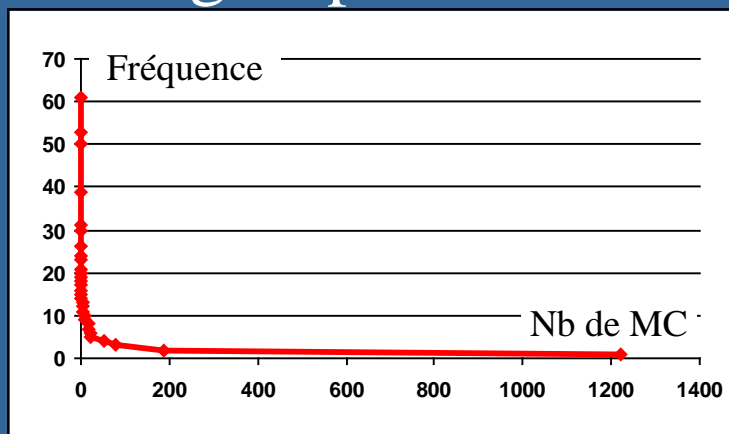
# Homogénéisation du corpus

- Restriction de l'étude aux notices de 1986 à 95



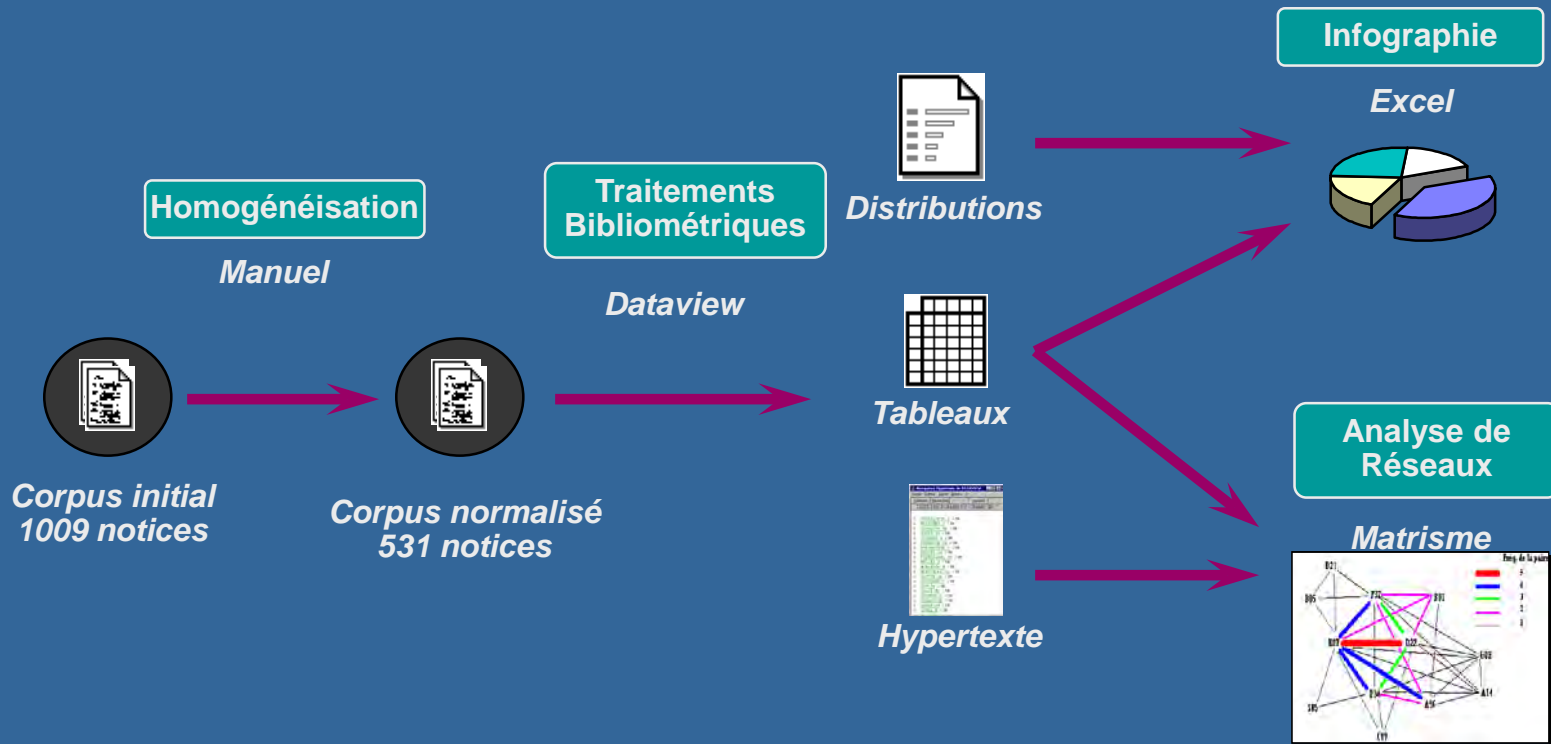
Corpus de 531 références de thèses soutenues sur 10 ans

- Normalisation du vocabulaire des mots-clés par regroupement manuel des formes graphiques



- 1662 mots-clés (- 25%)
- 1223 mots-clés à fréquence 1 (- 27%)

# Synoptique de l'étude bibliométrique



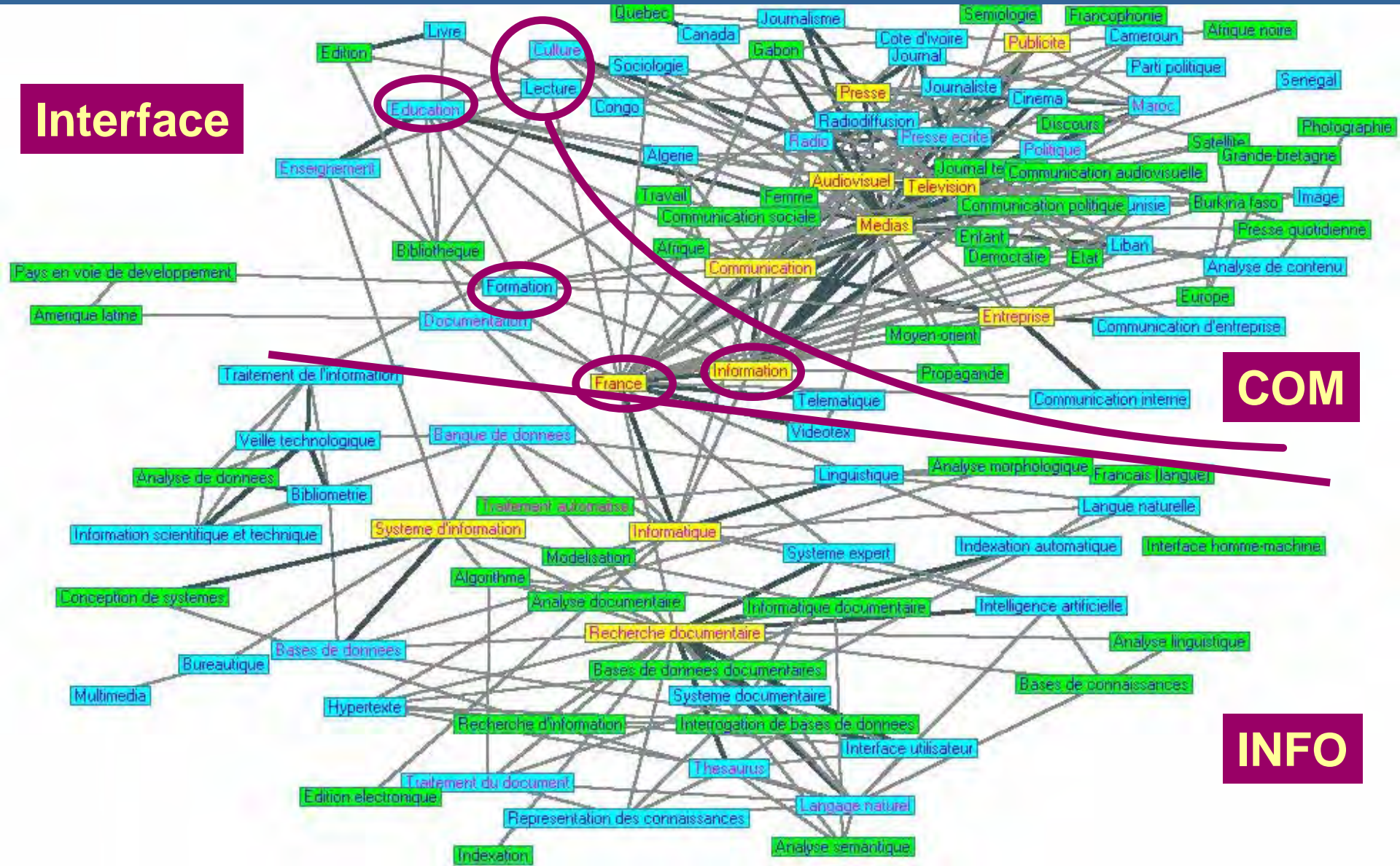
- Une étude globale du corpus -> étude des SIC (531 notices)
- Une étude sur un sous-corpus -> étude des SI (142 notices)

# Le réseau des MC des SIC

Interface

COM

INFO

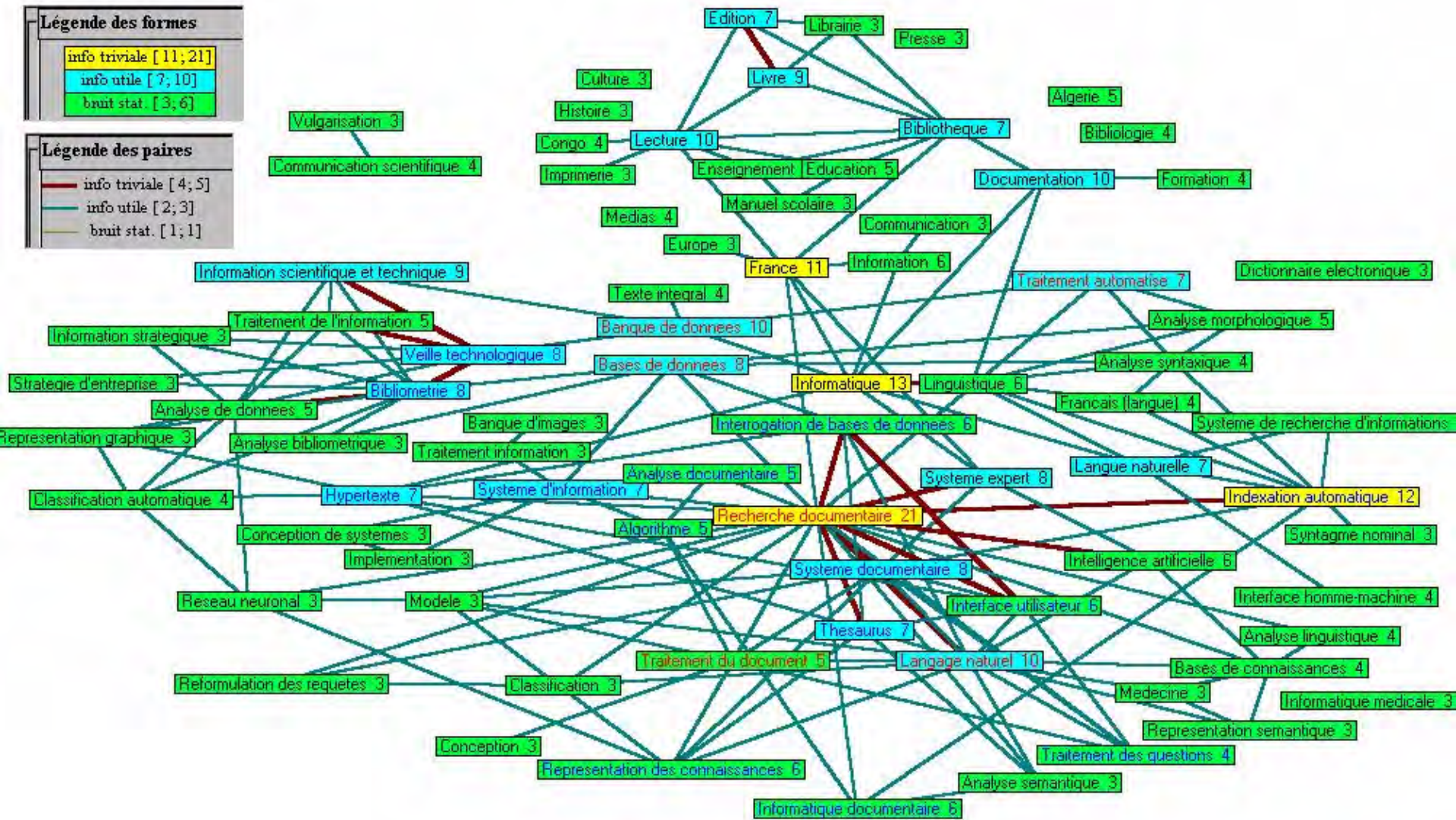


# Le réseau des MC des Sciences de l'Information

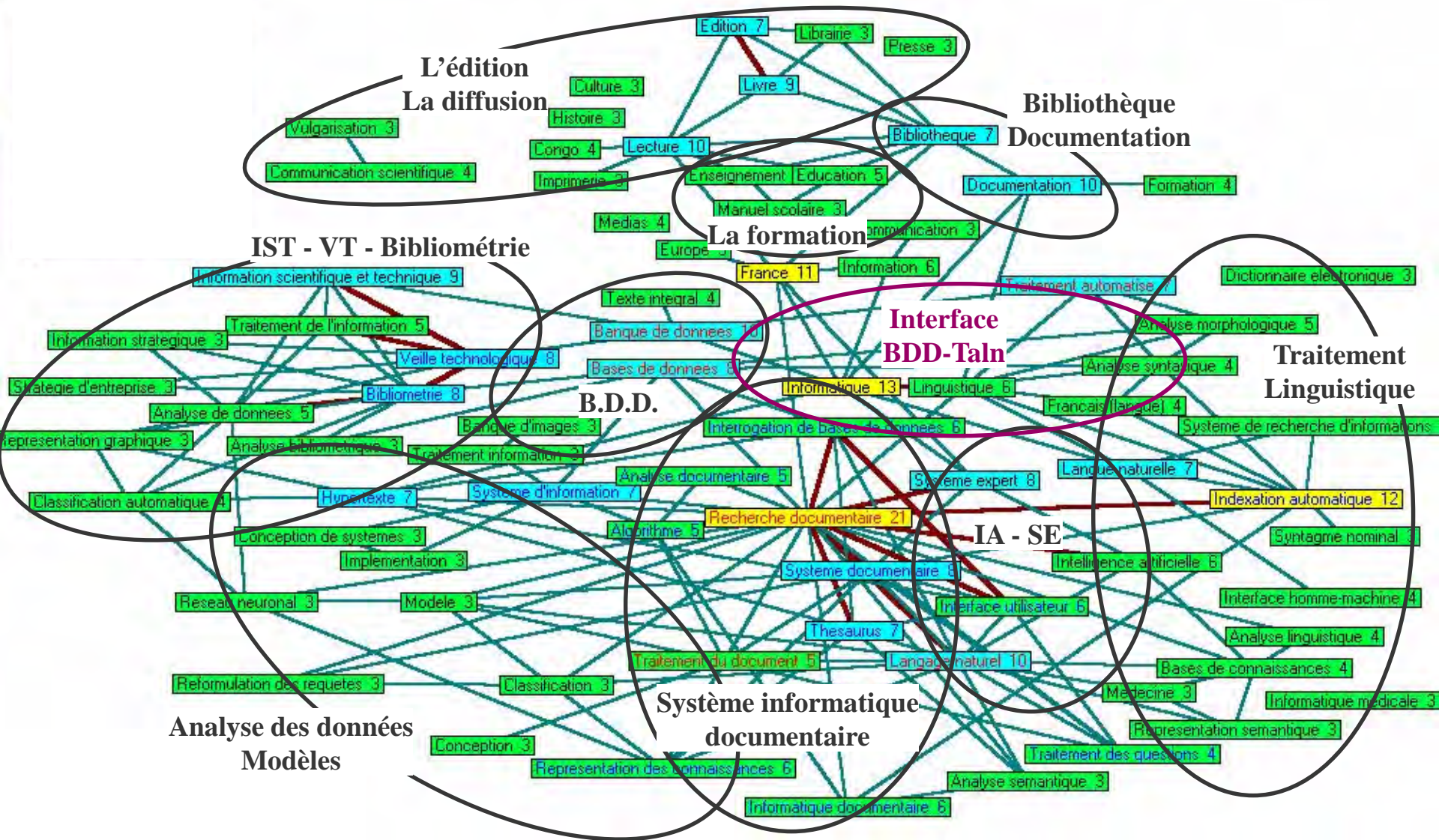
Légende des formes	
info triviale	[ 11; 21]
info utile	[ 7; 10]
bruit stat.	[ 3; 6]

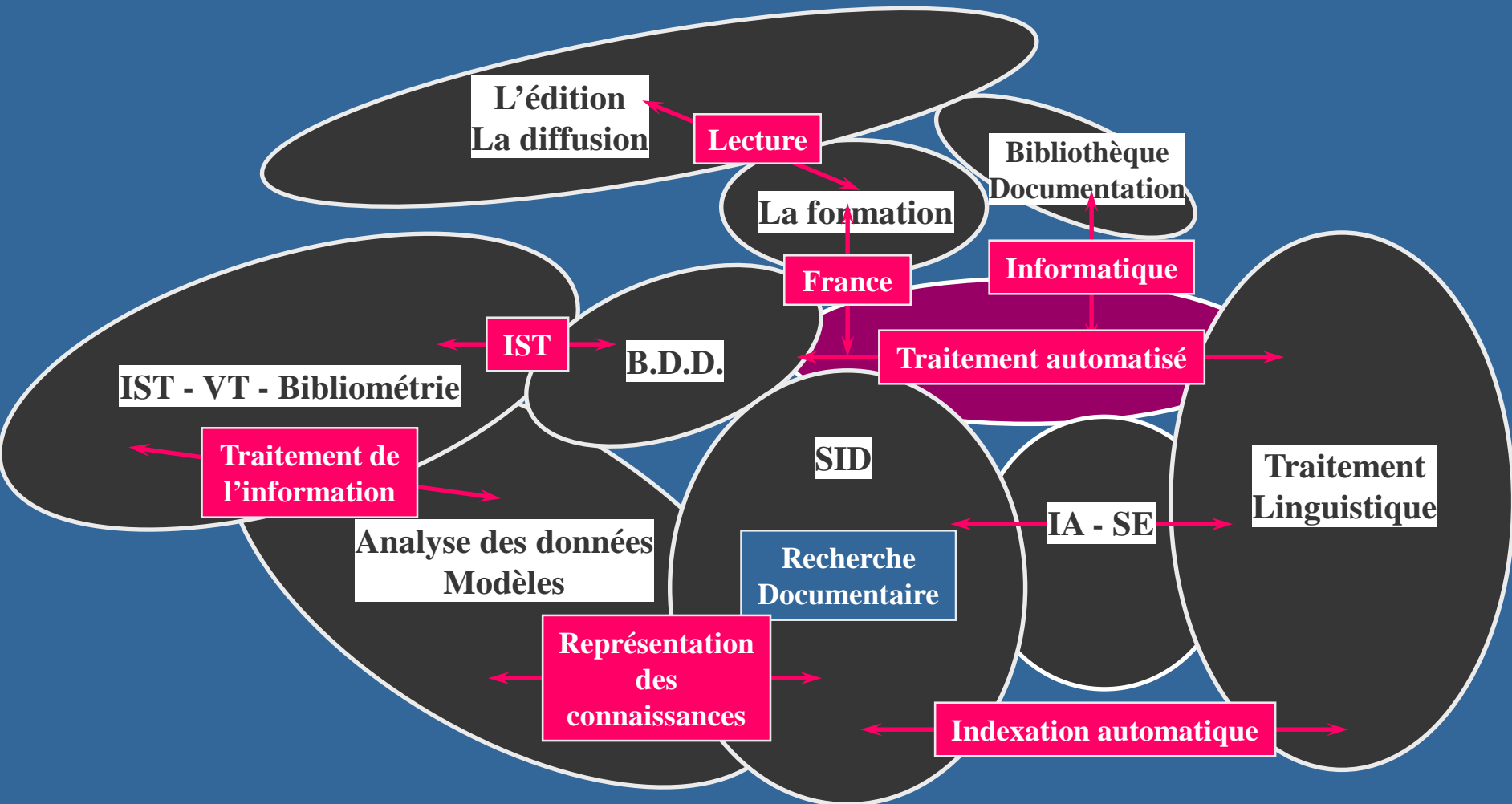
Légende des paires	
(red line)	info triviale [ 4; 5]
(green line)	info utile [ 2; 3]
(grey line)	bruit stat. [ 1; 1]



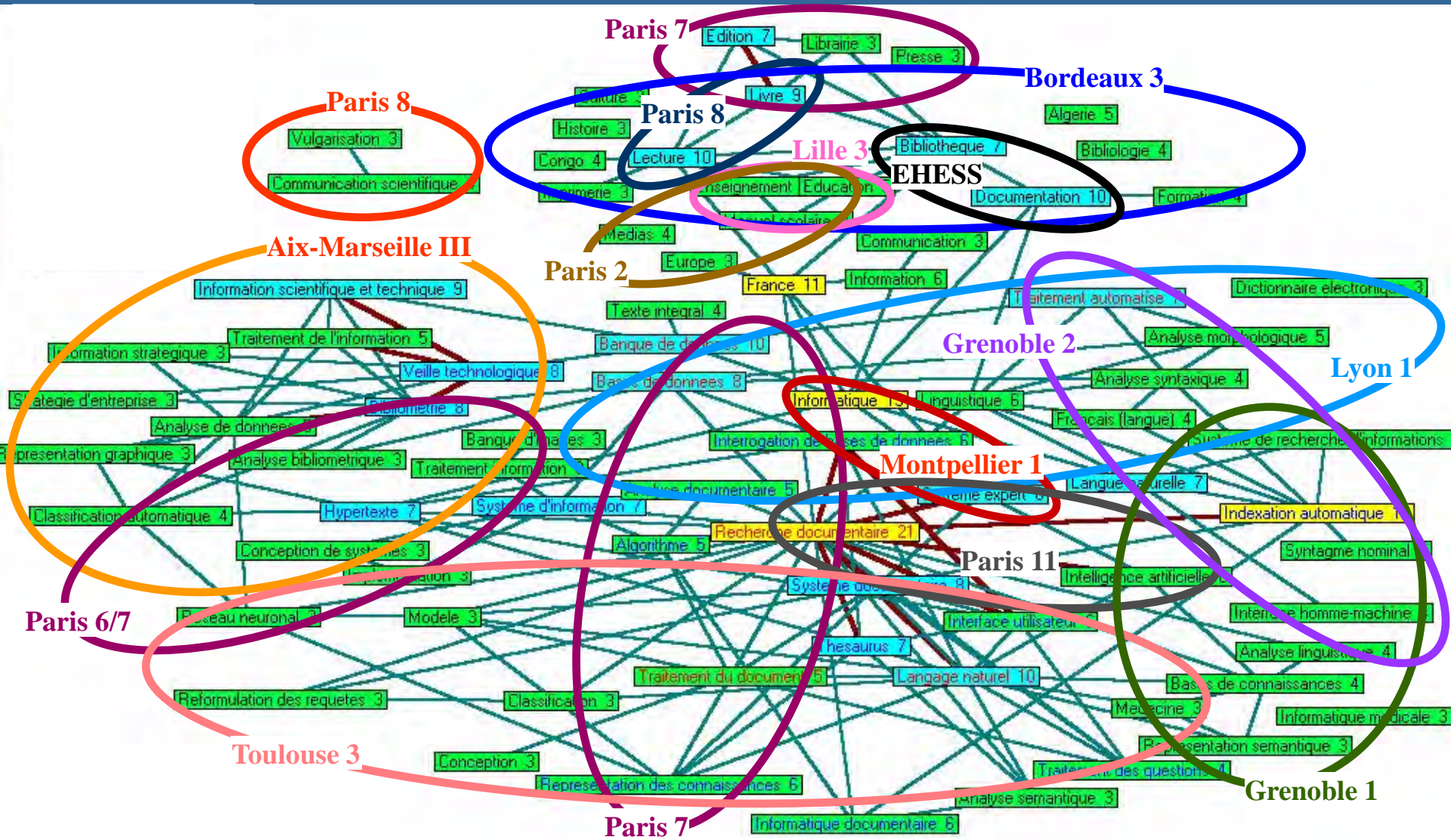
# Le réseau thématique des Sciences de l'Information



# Cartographie des Sciences de l'Information

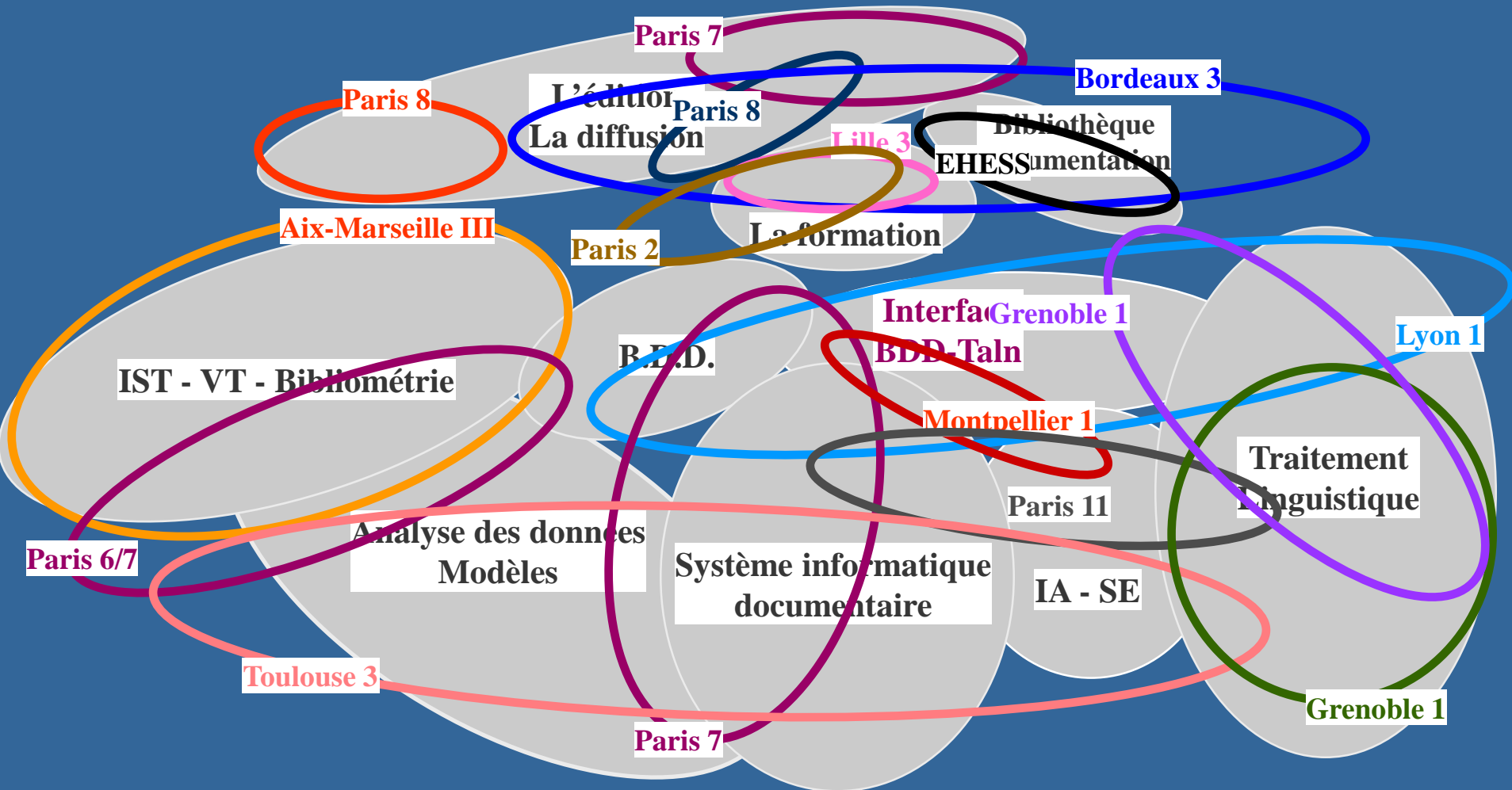


# Zones d'expertise des Universités des Sciences de l'Information





# Cartographie de l'expertise des Universités en S.I.

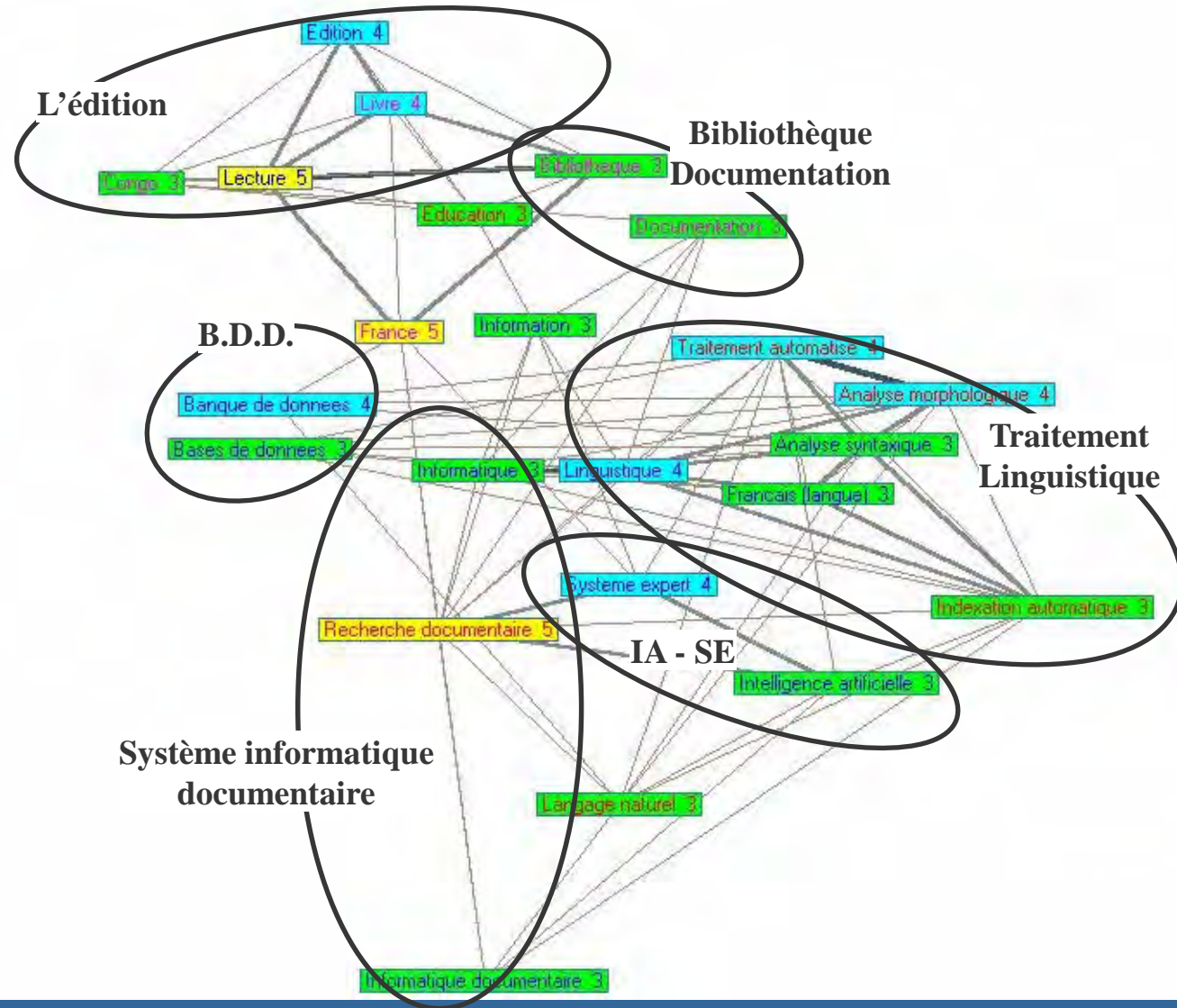


# Evolution thématique des SI : période 1

Légende des formes	
info triviale [5;5]	(yellow box)
info utile [4;4]	(blue box)
bruit stat. [3;3]	(green box)

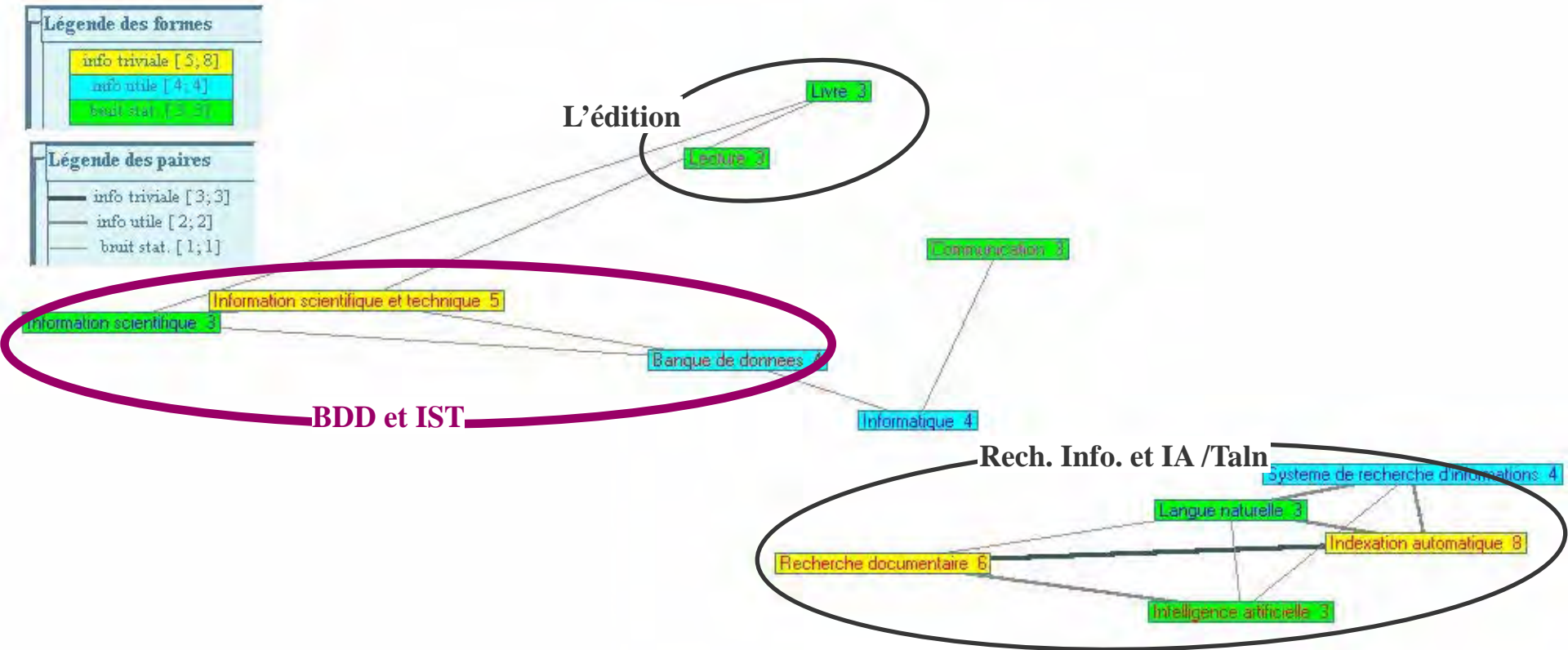
  

Légende des paires	
info triviale [3;3]	(thick black line)
info utile [2;2]	(medium black line)
bruit stat. [1;1]	(thin black line)



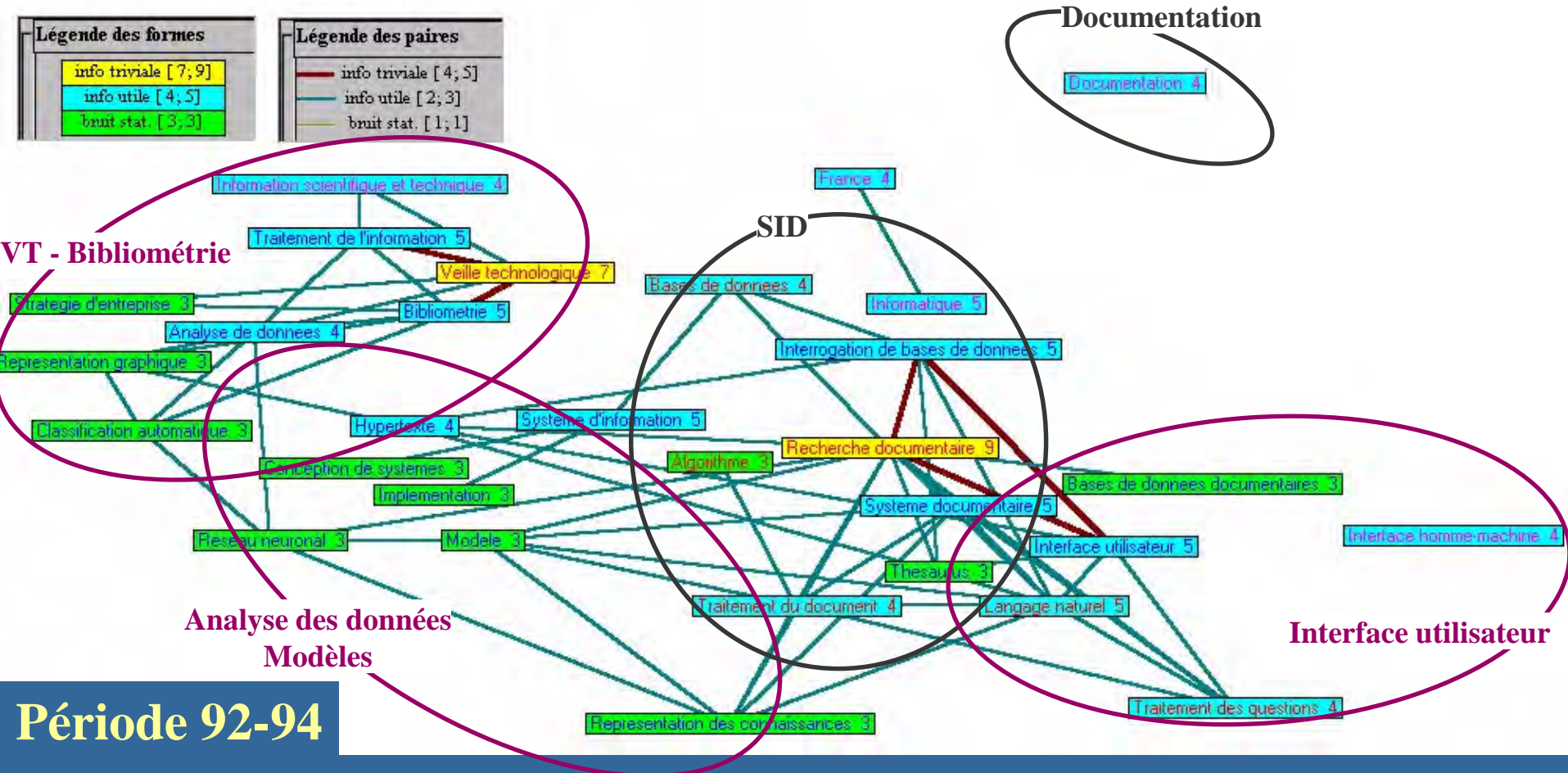
Période 86-88

# Evolution thématique des SI : période 2



Période 89-91

# Evolution thématique des SI : période 3



# Conclusions

- Premiers résultats qui restent à affiner
- Complémentarité des différentes méthodes
- Analyse des thèses n'est qu'une vision partielle de l'activité de recherche
- Fiabilité des données insatisfaisante