

L'intelligence économique outil stratégique pour l'entreprise

Professeur Bernard DOUSSET

dousset@irit.fr <http://atlas.irit.fr>

Institut de Recherche en Informatique de Toulouse (IRIT)

Equipe Systèmes d'Information Généralisés (SIG)

Université Paul Sabatier (Toulouse III)

Plan de la présentation

Historique

Définitions et buts

La veille stratégique

Le multidimensionnel

Les classifications

Architecture

Interactivité

Travail en équipe

Principe général

Contribution

en analyse textuelle

en analyse exploratoire

en géostratégie

en dessin de graphes

Conclusion

bilan

perspectives

Définitions et buts

Veille stratégique

Secteurs d'activité

Connaissance

Informations élaborées

analyser & interpréter

diffusion restreinte

synthétiser

valider et diffuser l'information exogène

élaborer

choisir

Informations utiles

Scénarios

Stratégie

Cycle de l'intelligence économique

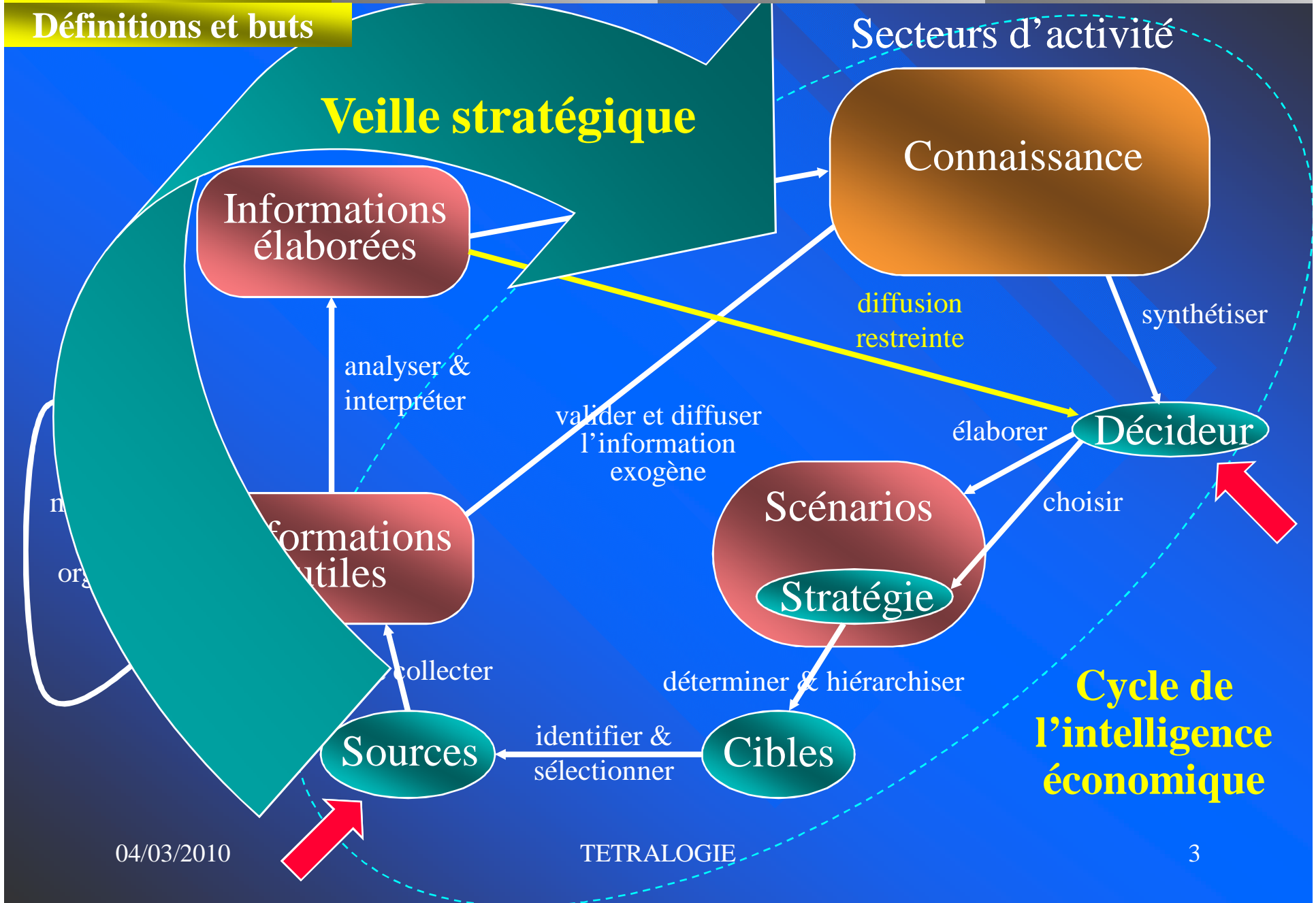
collecter

déterminer & hiérarchiser

identifier & sélectionner

Sources

Cibles



Définitions et buts

■ Etudier l'environnement et l'évolution

- Des marchés
- Des coopérations
- Des alliances
- Des fusions
- Des acquisitions
- Des participations
- Des implantations
- De l'innovation
- Des ruptures
- Des transferts de technologie
- Des dépôts de brevets
- Des équipes de recherche
- De la terminologie
- Des sources d'information
- Des publications
- De la mode
- De la publicité
- Des appels d'offres

■ Depuis toutes les sources d'information électronique

– Bases bibliographiques

- Web of science (SCI)
- Science direct
- Pascal, Francis (CNRS)
- Chemical abstract , Biosis
- Current Contents, Inspec,
- Cnki, Cqvip (Chine), ...

– Serveurs spécialisés

- Dialog, Stn, Questel, Derwent, ...

– Presse

- Factiva, Journaux électroniques
- Agences: Afp, Reuters, ...

– Internet

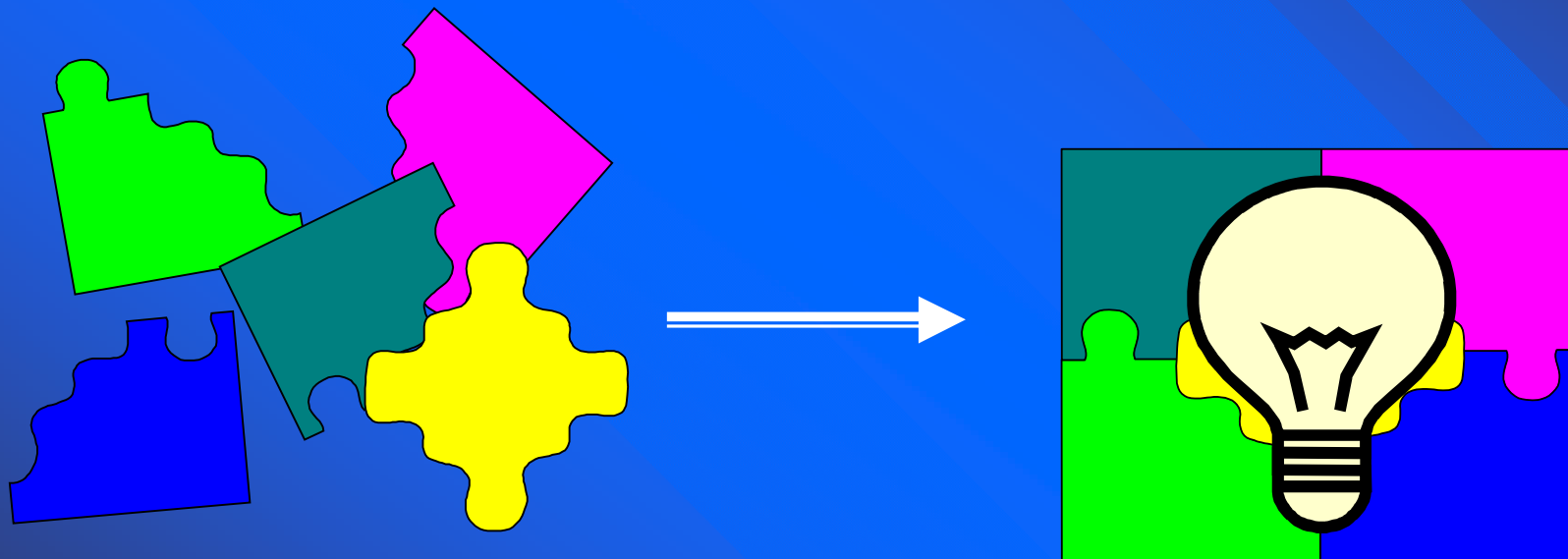
- Brevets : Uspto, Esp@cenet
- Bio: Pubmed, Biospace
- Pages web, sites web
- Blogs, news-groups
- Flux RSS, streaming, ...

– Intranet

- SI propriétaire/SGBD
- Data warehouse
- Indexations
- Web-logs
- Mails, Streams, ...

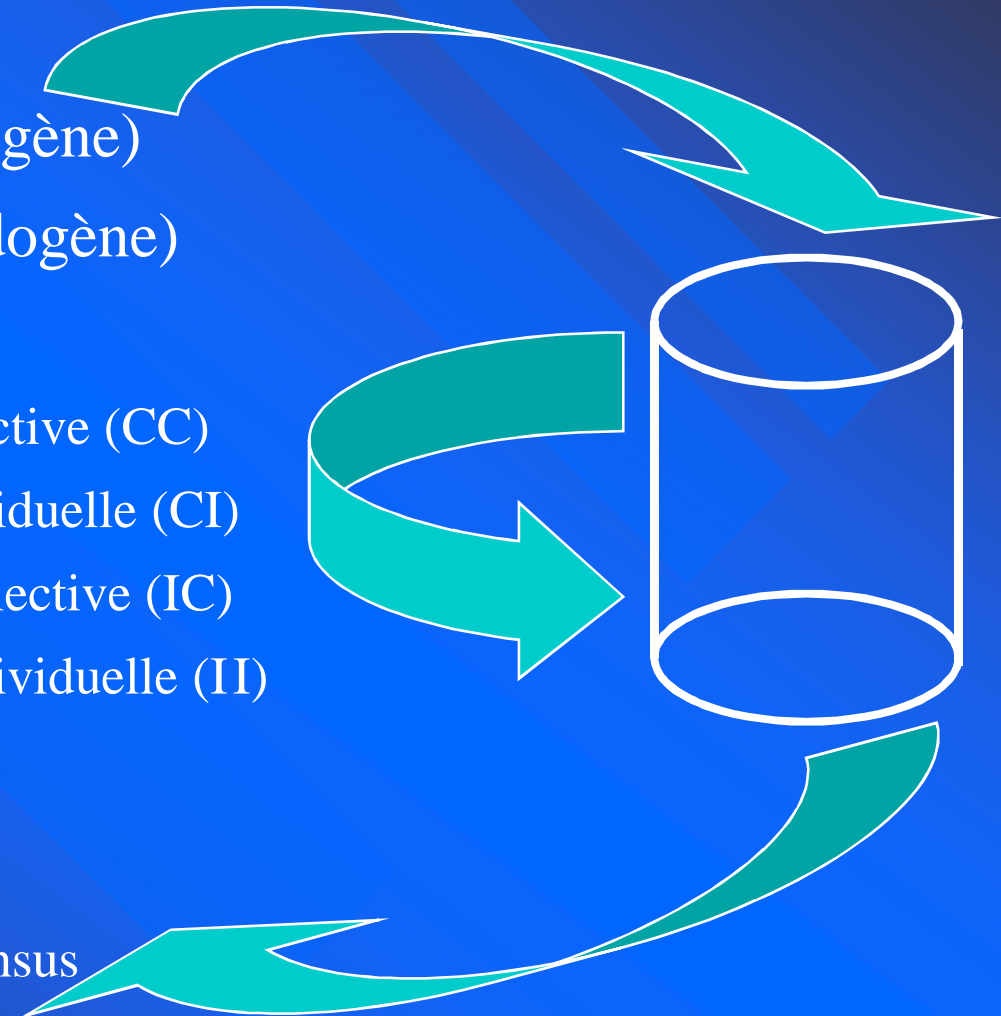
■ Information explicite

■ Information implicite



Définitions et buts

- l'information explicite (exogène)
- l'information implicite (endogène)
- Maturité de l'information
 - L'information consciente collective (CC)
 - L'information consciente individuelle (CI)
 - L'information inconsciente collective (IC)
 - L'information inconsciente individuelle (II)
- La terminologie émergente
- Les concepts émergents
 - cohérence, simultanément, consensus



La veille stratégique

- Le processus de veille stratégique comporte 5 phases :
 - la sélection des données utiles :
 - terminologie, dates, acteurs (auteurs, organismes, pays...)
 - la préparation des données :
 - pré-traitements : nettoyages [Zipf49], synonymies [Porter80]
 - transformations : modèle de représentation des données [Salton89]
 - l'analyse des données:
 - classifications, règles d'associations, séquences, ...
 - l'interprétation et l'évaluation :
 - basées en grande partie sur les visualisations
 - l'exploitation et donc **la prise de décision**

■ Les analyses multidimensionnelles

- Fondements : Spearman & Pearson dès 1904
- Analyse canonique et ACP : [Hotteling35]
- Analyse des correspondances : [Hirschfeld35] et [Guttman41]
- Analyse procustéennes : [Schonemann65]
- En France
 - Analyse Factorielle des Correspondances : [Benzécri73]
 - Analyse de deux tableaux : [Lafosse85]
 - Compléments en analyse procustéenne : [Fichet87]

■ Classifications

- Taxonomie par CAH [Sokal63]
- Centres mobiles [Forgy65]
- Nuées dynamiques [Diday71]
- Partitionnements
 - Itératifs : Minimisation de la coupure [KL70], [FM82]
 - Spectraux : VP de la matrice de Laplace [Hall70], [Hagen91]
 - Multi-niveaux : regroupement + itératif **KMETIS** [Karypis98]
 - Stochastiques : **Markov CLustering** [Van Dongen00]

Pourquoi un système interactif?

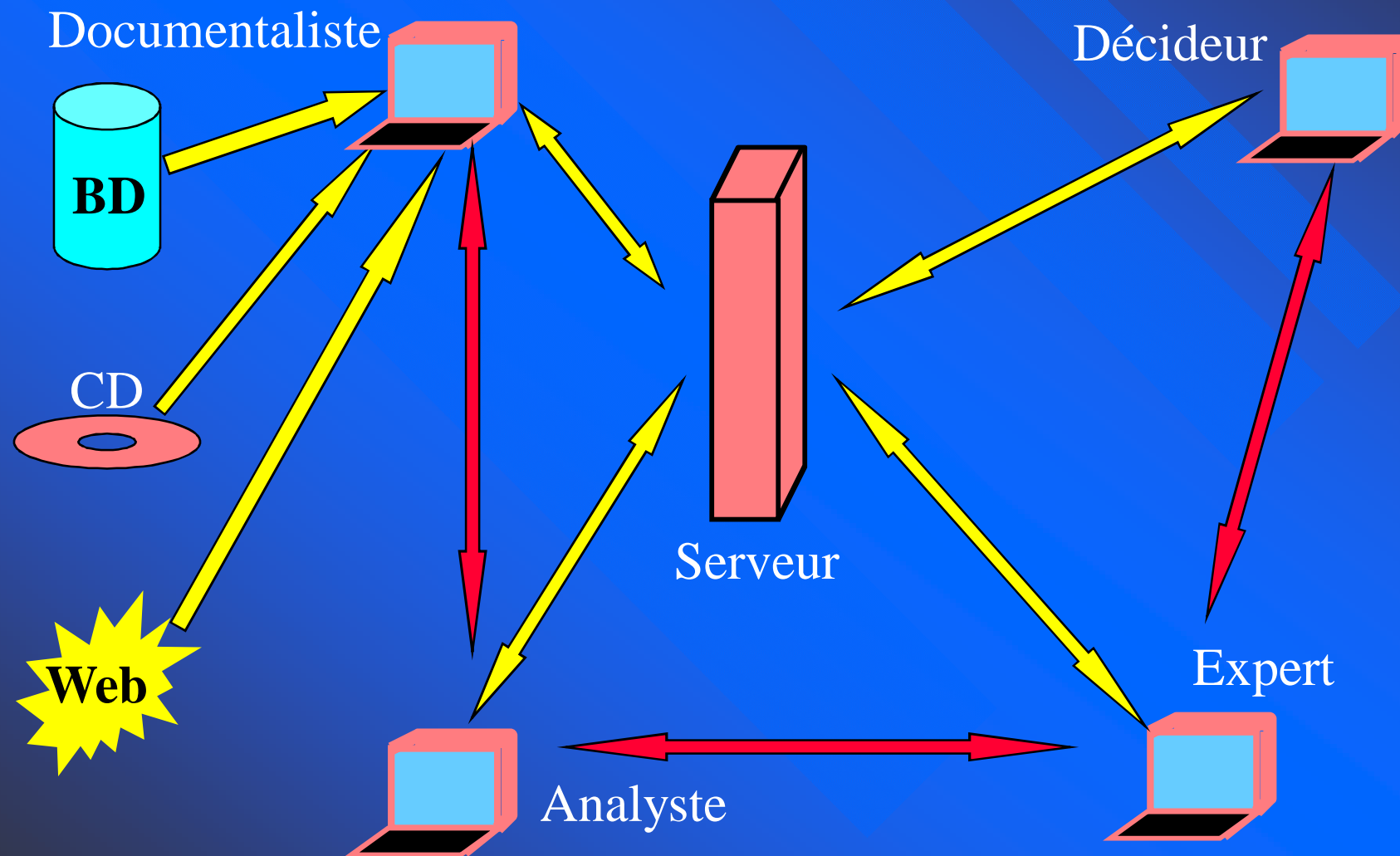
- En Amont de l'analyse l'interactivité sert à :
 - choisir les sources d'information
 - mettre au point les équations de recherche
 - évaluer les corpus (volume, pertinence, bruit)
 - choisir et valider le format optimal de sortie
 - superviser la collecte du corpus
 - contrôler les dictionnaires (forme et volume)
 - choisir les filtres (+ et -, thématiques, granularité, ...)
 - valider les synonymies et l'indexation (Multi-termes)

Pourquoi un système interactif?

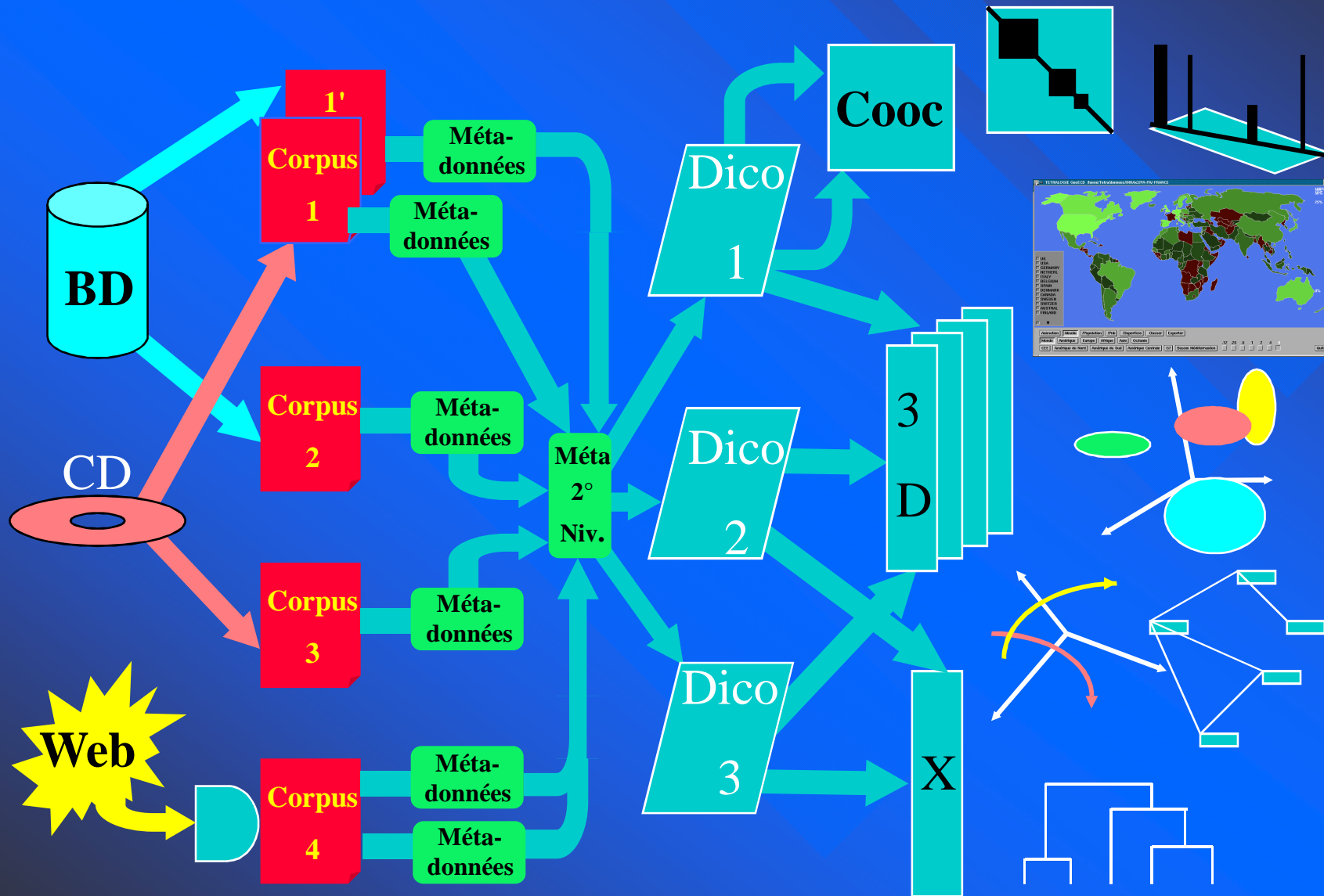
- Pendant l'analyse elle permet de
 - sélectionner et contrôler les matrices de croisement
 - choisir les méthodes à déployer
 - piloter ces méthodes (localement ou à distance)
 - extraire les connaissances grâce aux visualisations
 - choisir les vues à conserver
 - consulter les **experts** à distance
 - restituer l'analyse au **décideur**
 - éventuellement mettre l'analyse en ligne via XPlor

Travail en équipe

■ Connexions entre acteurs de la veille



Principe général



Dictionnaire + ou -

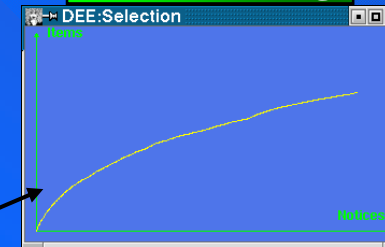
Tétralogie V7.fr Calcul des fréquences absolues Fichier : HUN

	Frq	000	Liste de mots	Synonymes	Charge?
MTM : MT	1	0			
Base : BA	1	0	BA		
Tout_NO : NO	1	0			
Titre : TI	1	0	-Vide		
titre : Ti	1	0			
RP : RP	1	0			
Adresse : AD	1	0			
Organisme : OR	1	0	OR	OR	
Email : EM	1	0	E-mails		
Journal : JN	1	0		JN	
Date : DP	1	0	Periodes	Periodes	
JS : JS	1	0			
Keyword : DE	1	0	DE	DE	
Classif : CL	1	0			
CC : CC	1	0	CC-Alpha		
Auteur_l : AL	1	0	AL-2	AL	
Auteur_c : AC	1	0			
Ville : VI	1	0			
Pays : PA	0	0	Pays	Pays	
pays : Pa	1	0			
Resume : AB	1	0			

EXECUTER QUITTER

■ Extraction des dictionnaires

Courbe de charge



Dictionnaire de synonymes

file:/home/Tetra/donnees/season/DE.Syn - KWrite

Fichier Édition Signets Outils Configuration Aide

ADJUSTMENT COSTS . ADJUSTMENT COST
 ADVERSE SELECTION . ADVERSE-SELECTION
 ASSETS PRICING . ASSET PRICING
 ASYMMETRIC ADJUSTMENTS . ASYMMETRIC ADJUSTMENT
 AUSTRALIAN . AUSTRALIA
 CONFIDENCE INTERVALS . CONFIDENCE INTERVAL
 COST UTILITY ANALYSIS . COST-UTILITY ANALYSIS
 COST-UTILITY ANALYSIS . COST-UTILITY ANALYSIS
 DAY-OF-THE-WEEK EFFECTS . DAY-OF-THE-WEEK EFFECT
 EM ALGORITHM . EM-ALGORITHM
 ERROR-CORRECTION . ERROR CORRECTION
 ERROR CORRECTION MODELS . ERROR-CORRECTION MODELS
 FISCAL ADJUSTMENTS . FISCAL ADJUSTMENT
 FORECASTING COMPETITIONS . FORECASTING COMPETITION
 FORWARD RISK ADJUSTED MEASURE . FORWARD-RISK-ADJUSTED ME

Lig : 1 Col : 0 INS NORM

en analyse textuelle

■ Détection des multi-termes

Dictionnaire de mots vides

Dictionnaire de multi-termes

Dictionnaire de synonymes

Seuil de détection statistique

Traitement syntaxique

Traitement morphologique

Traitement statistique

04/03/2010

TETRAL

```

MTM: 11
MTM: NEURAL-NET APPROACH MEDIUM-TERMSHORT-TERM-COORDINATION CAPACITY-TRANSFORMER TAP RADIAL DISTRIBUTION
TI: Neural-nets approach to medium-term/short-term coordination of capacitors and transformer
AU: Saric-AT; Calovic-MS
AA: Cacak Coll. of Eng., Kragujevac Univ., Serbia
ED: Gubina-A
SO: Proceedings of the 11th International Conference on Power System Automation and Control.
p.103-9
PY: 1999
RT: Conference-Paper
CD: Proceedings of the International Power System Automation and Control Conference (PSAC '99)
Slovenia
English

MTM: PAPER CONSIDER PROBLEM COORDINATED REACTIVE-POWER-COMPENSATION VOLTAGE-REGULATION VARVOLT RADIAL
MTM: MECHANIC SWITCH SHUNT-CAPACITY TRANSFORMER VOLTAGE REGULATOR SHORT-TERM MEDIUM-TERM OPERATION
MTM: OBJECTIVE-FUNCTION REPRESENTED MINIMIZATION SYSTEM LOSSES
MTM: KEEP NODE VOLTAGE SPECIFIED LIMITS
MTM: OVERALL ALGORITHM TWO-LEVEL TYPE
MTM: MUTUALLY DEPENDENT STAGE MEDIUM SHORT-TERM-PLANNING
MTM: MEDIUM-TERM PLANNING FORECASTED SOURCE NODE VOLTAGE MONTHLY BUS LOADS
MTM: CLASSICAL DYNAMIC-PROGRAMMING-ALGORITHM SOLVE OPTIMA MONTHLY STATES-REGULATING RESOURCE SWITCHED
MTM: STATES RESOURCE SEASONAL ADJUSTMENT-DATA STAGE SHORT-TERM-PLANNING
MTM: FINAL OPTIMAL STATES-REGULATING RESOURCE DETERMINED
MTM: APPLICATION METHOD VERIFIED REAL DISTRIBUTION SYSTEM
MTM: SHORT-TERM DAILY VARVOLT-COORDINATION-PROBLEM
MTM: VARIABLE REACTIVE POWER SWITCH CAPACITY TAP POSITION UNDER LOAD TAP CHANGING ULTC TRANSFORMERS
    
```

■ Croisements 2D

Filtre positif ou négatif

Dictionnaire de synonymes

Matrice carrée

Matrice asymétrique

Courbe de charge

	MT	BA	TI	RP	AD	OR	EM	JN	DP	JS	DE	CL	CC	AU	VI	PA	AB	Frq	%00	Filtres	Synonymes	Charge?
MTM : MT																		1	0	MT-5	MT	
Base : BA																		1	0			
Titre : TI																		1	0			
RP : RP																		1	0			
Adresse : AD																		1	0			
Organisme : OR																		1	0	OR-0	OR	
Email : EM																		1	0			
Journal : JN																					JN	
Date : DP																					Per	
JS : JS																						
Descript : DE																						DE
Classif : CL																						
CC : CC																						
Auteur : AU																		1	0	AU-1	AU	
Ville : VI																		1	0	VI-1	VI	
Pays : PA																		1	0	PA	PA	
Resume : AB																		1	0			

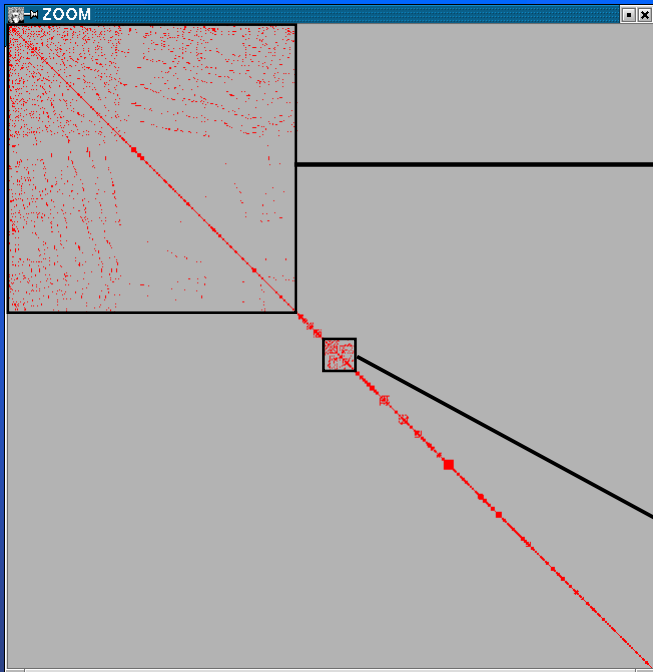
en analyse exploratoire

- pour le traitement des matrices
 - proposition de plusieurs algorithmes de tris
 - génération supervisée de matrices de croisement
 - tableur 3D adapté aux matrices de grande taille (zooms 2 et 3D)
- pour les analyses multidimensionnelles
 - visualisations interactives en 3D et 4 D
 - synchronisation de cartes locales ou distantes
 - visualisation de trajectoires et rotations procustéennes (AFCM)
- pour les classifications
 - arbres hiérarchiques interactifs avec exportation des classes
 - partitionnement de graphes, graphes de classes
 - segmentation pour les cartes géographiques

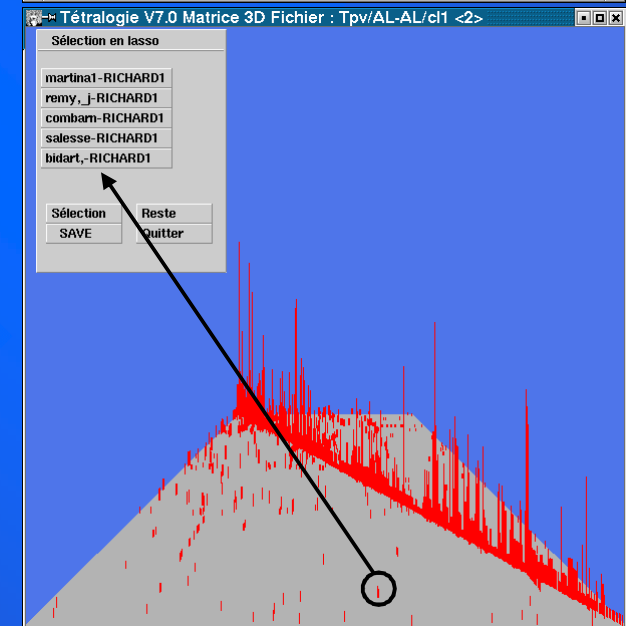
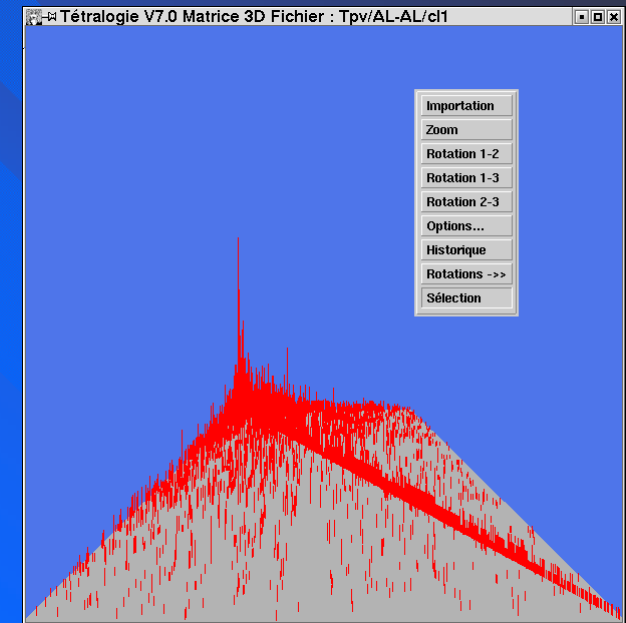
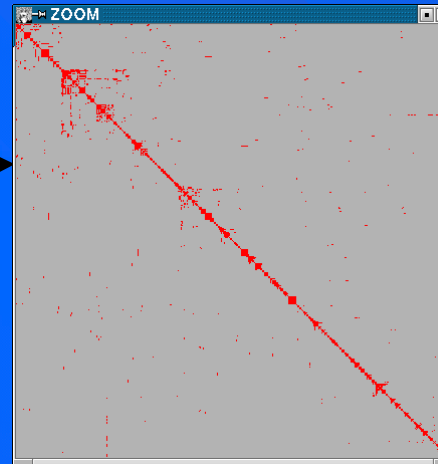
en analyse exploratoire

■ Algorithmes de tris de matrices

Tri par classes de connexité



Tri par blocs diagonaux



en analyse exploratoire

■ Analyse en composantes principales (ACP)

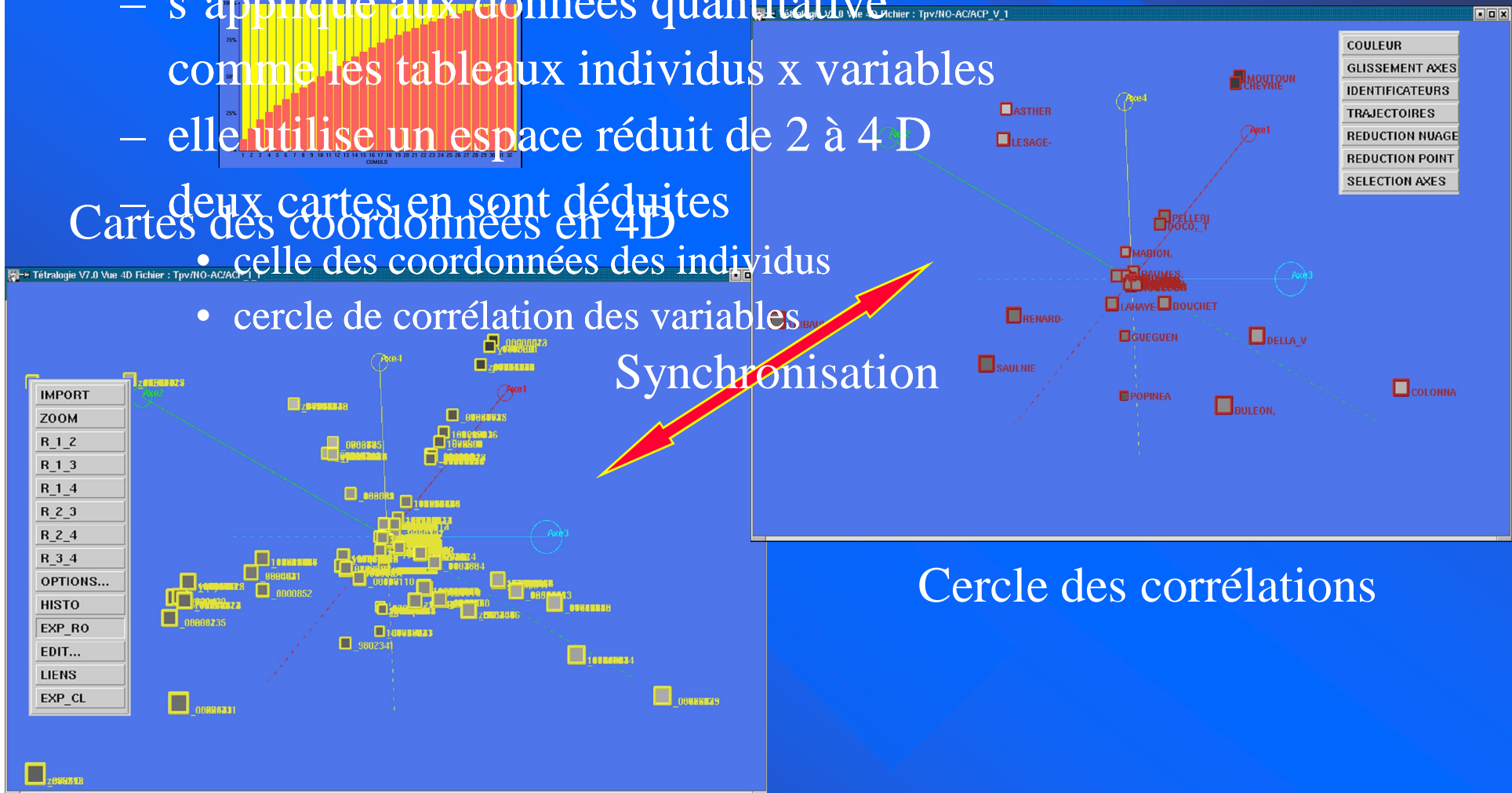
- s'applique aux données quantitative
- elle utilise un espace réduit de 2 à 4D

deux cartes en sont déduites

Cartes des coordonnées en 4D

- celle des coordonnées des individus
- cercle de corrélation des variables

Synchronisation



Cercle des corrélations

en analyse exploratoire

■ Analyse factorielle des correspondances multiple (AFCM)

- nous l'appliquons aux matrices de cooccurrences 3D

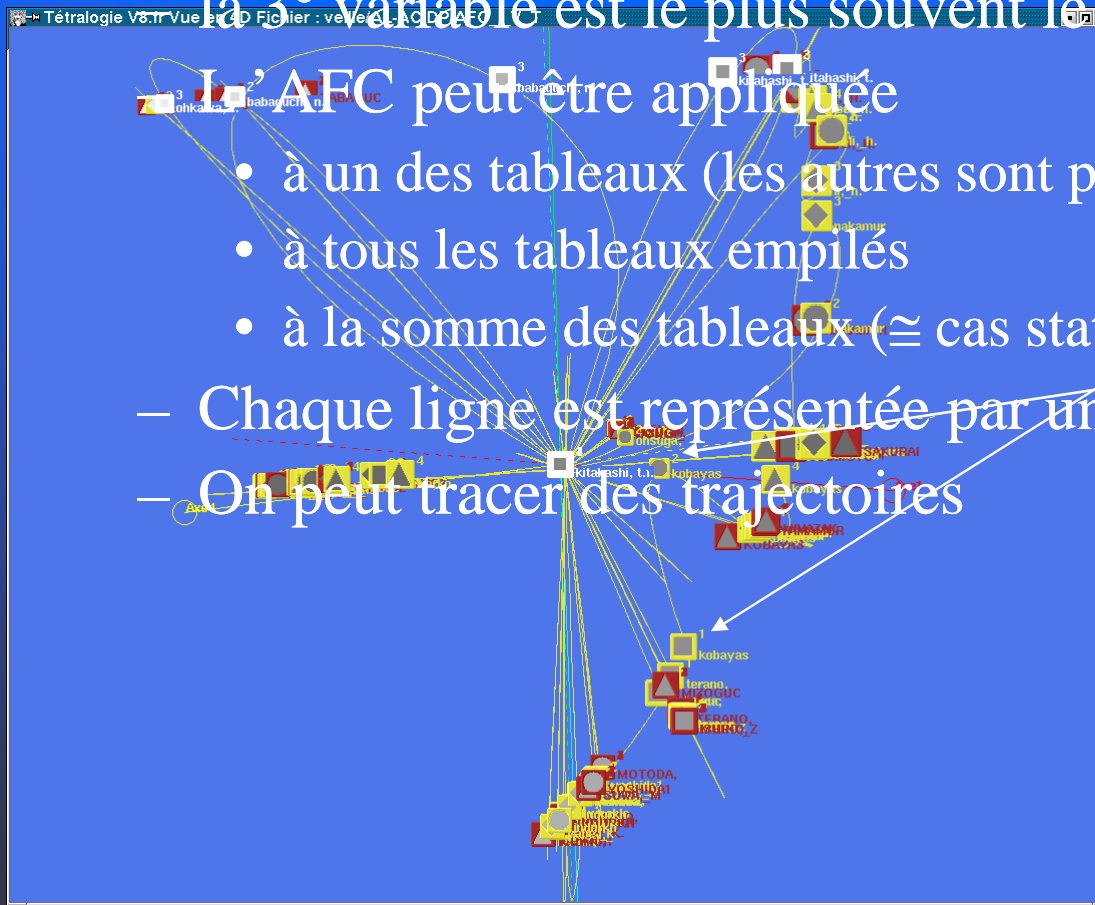
la 3^o variable est le plus souvent le temps

L'AFC peut être appliquée

- à un des tableaux (les autres sont projetés)
- à tous les tableaux empilés
- à la somme des tableaux (\cong cas statique)

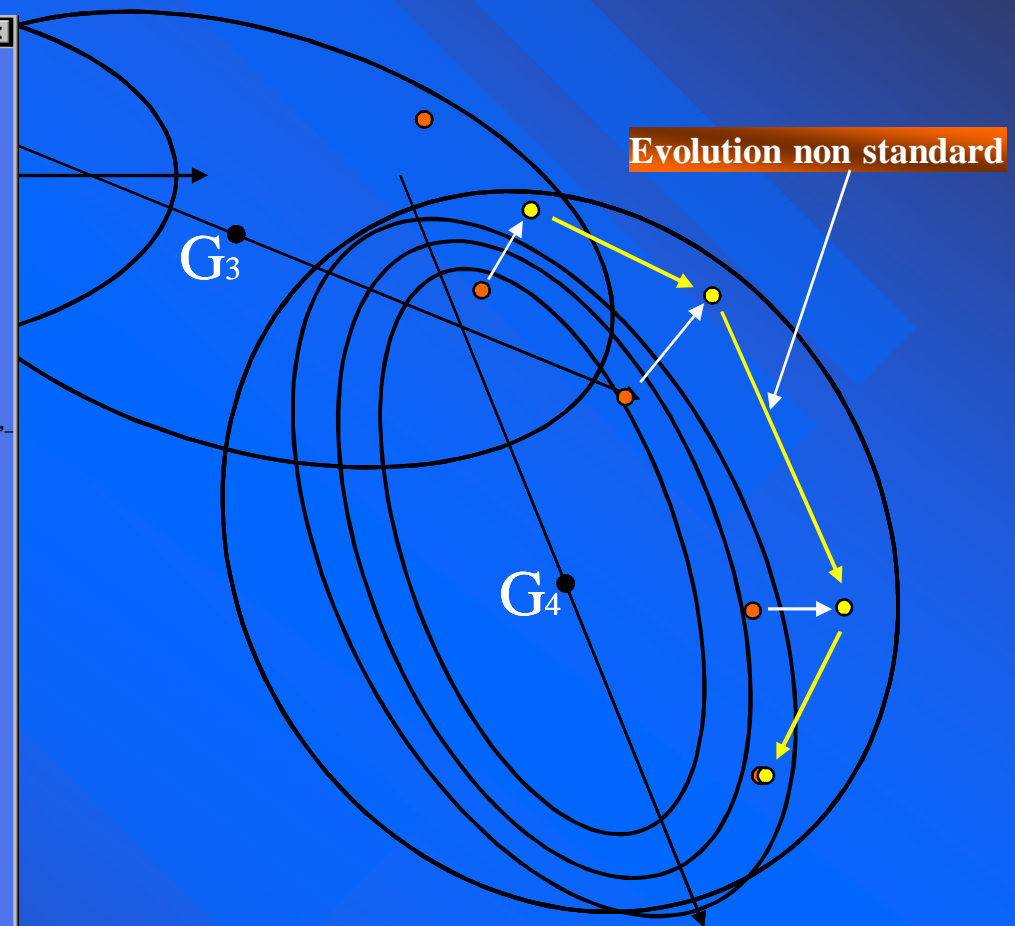
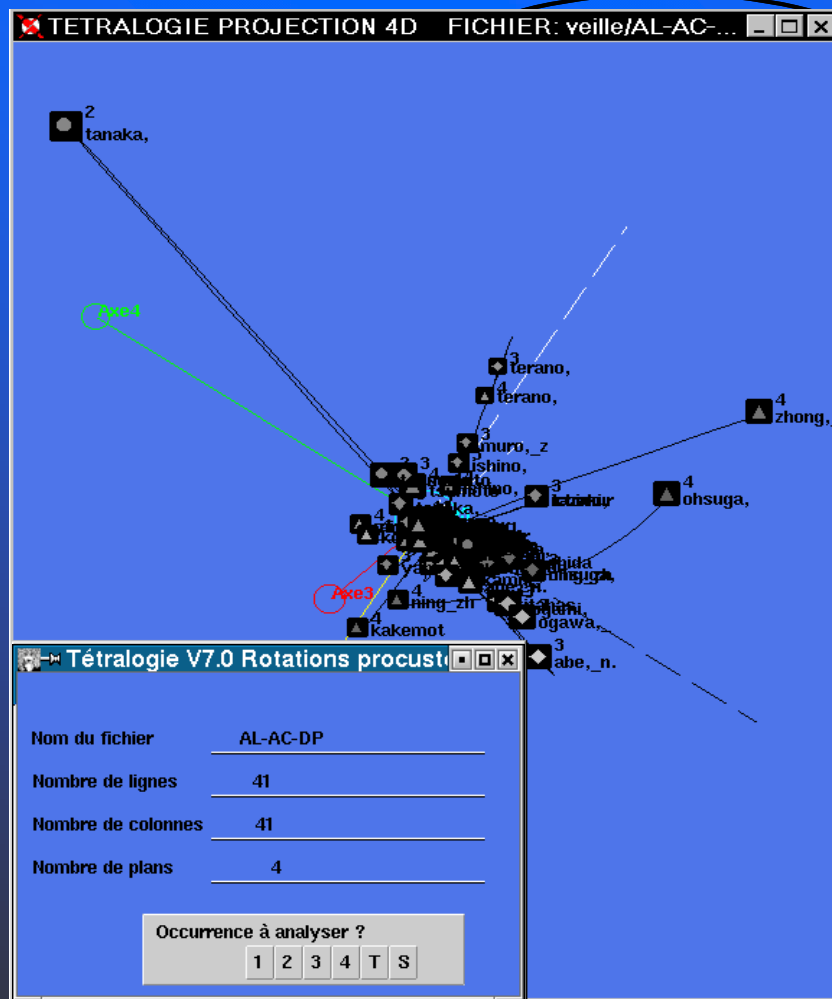
- Chaque ligne est représentée par une série de points

- On peut tracer des trajectoires



	CERCONE	TSUMOTO	ZIARKO,	TANAKA,	JIawei_	KOBAYAS	YAMAMUR	MIYAZAK	NAKAMUR
1 cercone	2								
2 tsumoto		1							
3 ziarko,			1						
4 tanaka,				2					
5 kobayas						2			
6 jiawei_									
7 miyazak									
8 yamamur									
9 nakamur									
10 hasegaw									
	CERCONE	TSUMOTO	ZIARKO,	TANAKA,	JIawei_	KOBAYAS	YAMAMUR	MIYAZAK	NAKAMUR
1 cercone	1				1				
2 tsumoto		1							
3 ziarko,			2						
4 tanaka,				1					
5 kobayas						4			
6 jiawei_	1								
7 miyazak									
8 yamamur									
9 nakamur									1
10 hasegaw									
	CERCONE	TSUMOTO	ZIARKO,	TANAKA,	JIawei_	KOBAYAS	YAMAMUR	MIYAZAK	NAKAMUR
1 cercone	14		3	6	4				
2 tsumoto		6							
3 ziarko,	3		16	7					
4 tanaka,			6	7					
5 kobayas						3	2	1	
6 jiawei_	4				10				
7 miyazak							1	1	
8 yamamur							2	2	1
9 nakamur									2
10 hasegaw									
	CERCONE	TSUMOTO	ZIARKO,	TANAKA,	JIawei_	KOBAYAS	YAMAMUR	MIYAZAK	NAKAMUR
1 cercone	12		3		1				
2 tsumoto		6	1	4					1
3 ziarko,	3		8						
4 tanaka,			4	4					1
5 kobayas				1		2	1	1	1
6 jiawei_	1				6				
7 miyazak							1	2	2
8 yamamur							1	2	2
9 nakamur									1
10 hasegaw									

■ Analyse procustéenne



en analyse exploratoire

■ Classification ascendante hiérarchique (CAH)

méthode non supervisée

$$d_e(A,B) = \sqrt{\sum_{i=1, n} (x_i - y_i)^2}$$

$$d_M(A,B) = \text{Max}_{k,l} \{d_e(x_k, y_l)\}$$

$$d_\mu(A,B) = \frac{\sum_k \sum_l d_e(x_k, y_l)}{|A| |B|}$$

— on a le choix de la distance (euclidienne, max, sigma, ...)

— on a le choix du mode d'agrégation (centre, inf, sup, moyenne)

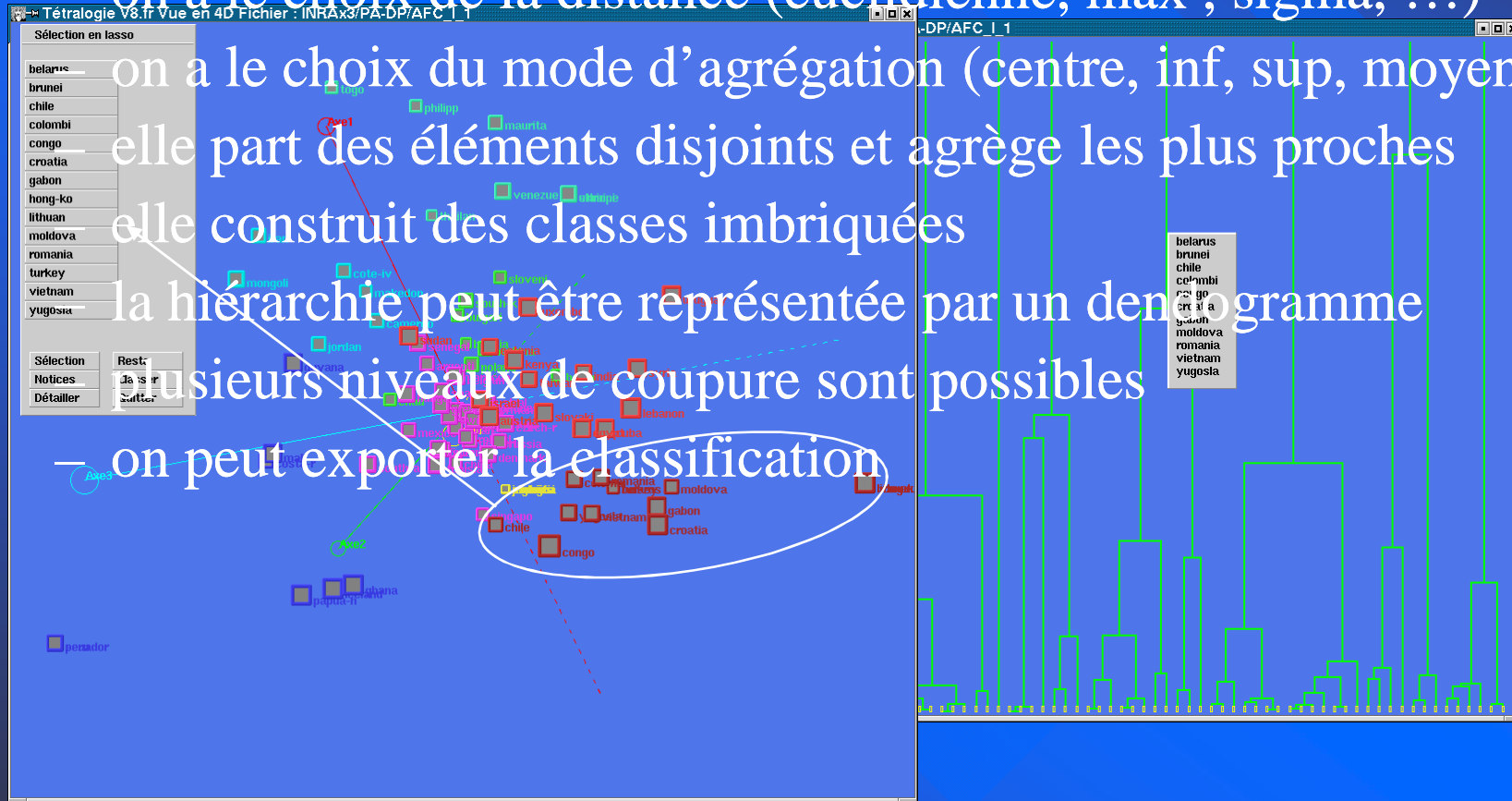
— elle part des éléments disjoints et agrège les plus proches

— elle construit des classes imbriquées

— la hiérarchie peut être représentée par un dendrogramme

— plusieurs niveaux de coupure sont possibles

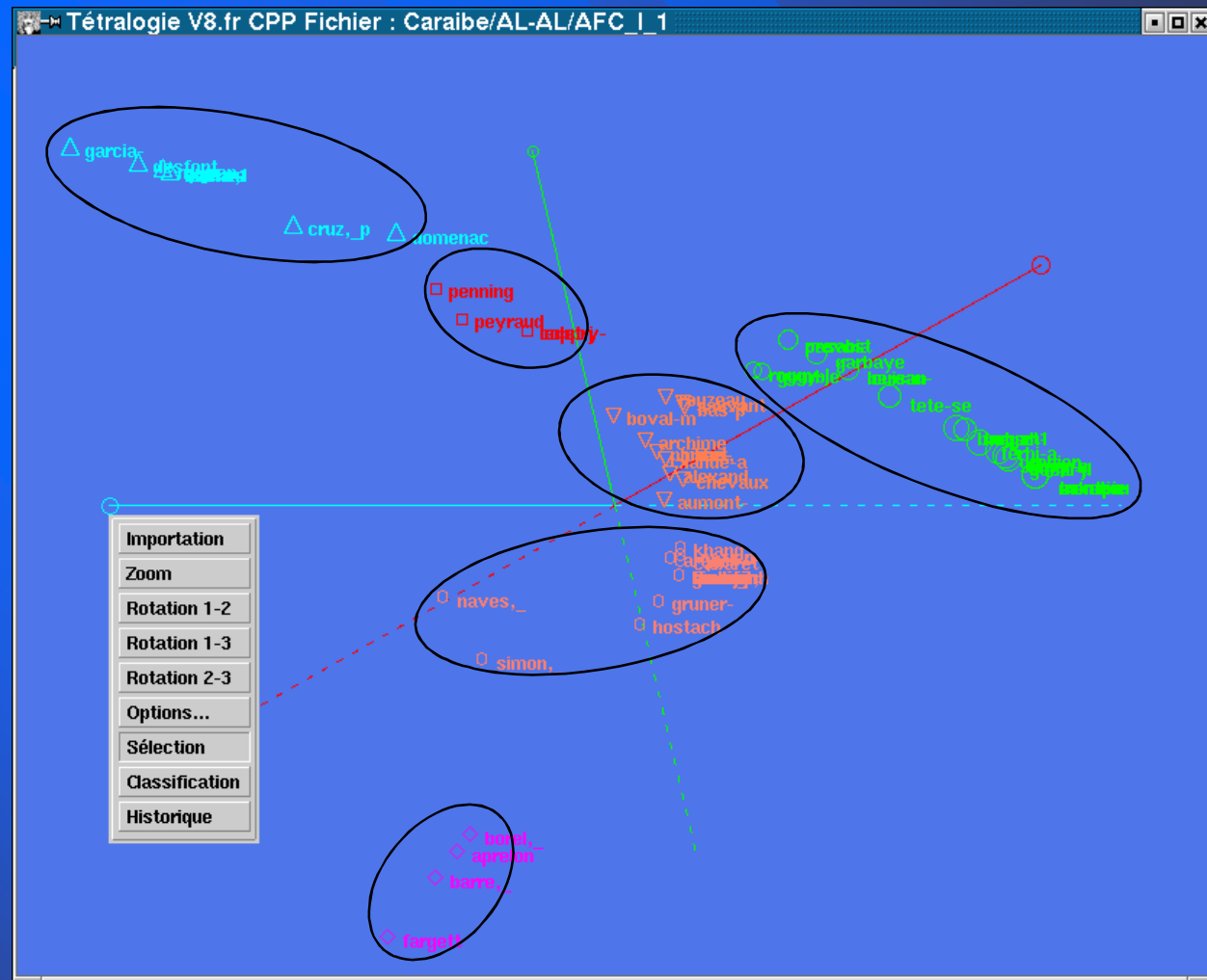
— on peut exporter la classification



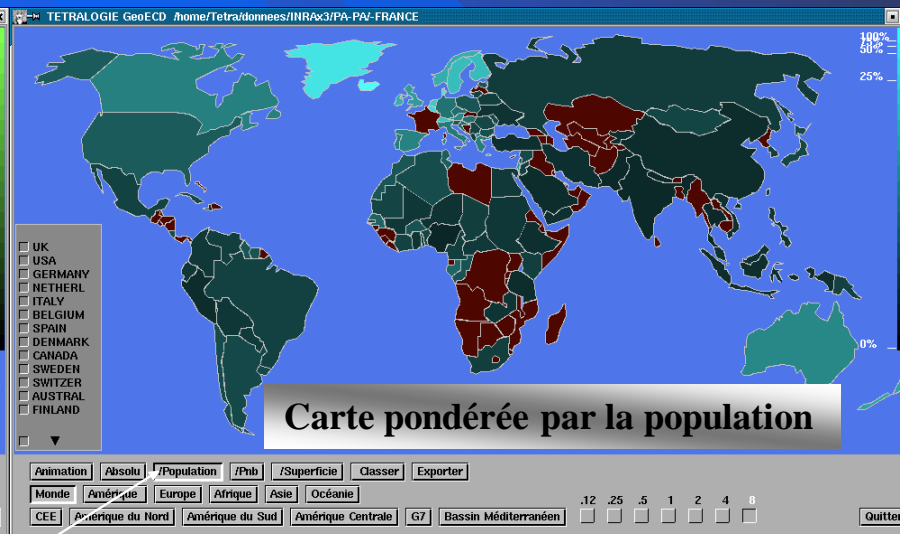
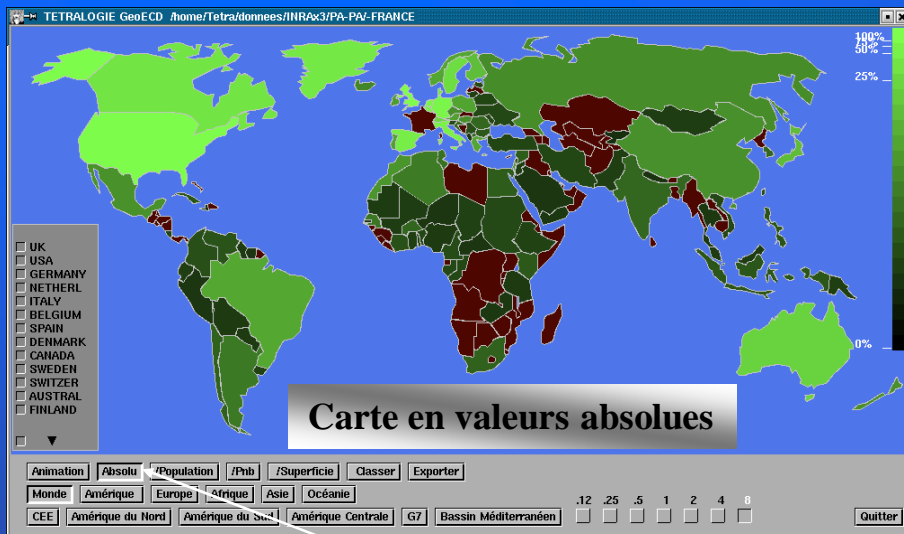
en analyse exploratoire

■ Classification par partition CPP (Centres mobiles)

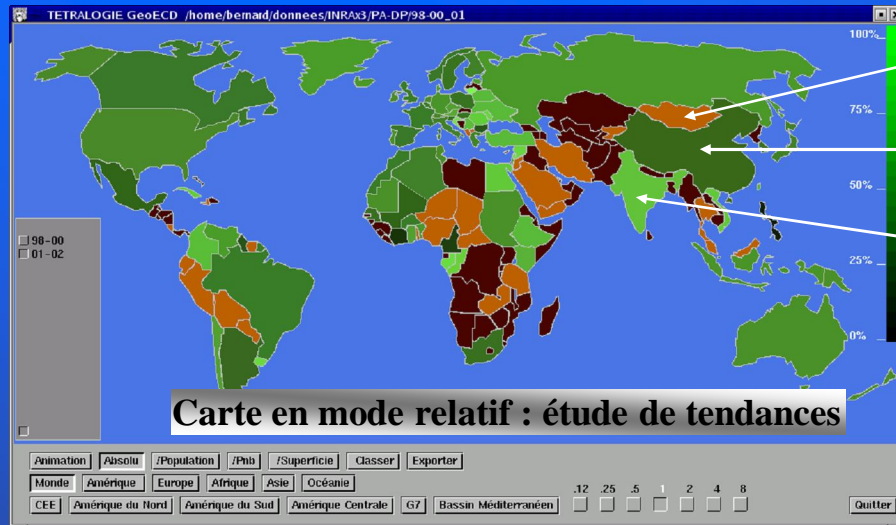
- supervisée
- k classes
- 1 représentant
- choix en 3D
- coloration
- icônes
- 4 classes
- 6 classes



■ Pondération par des données externes



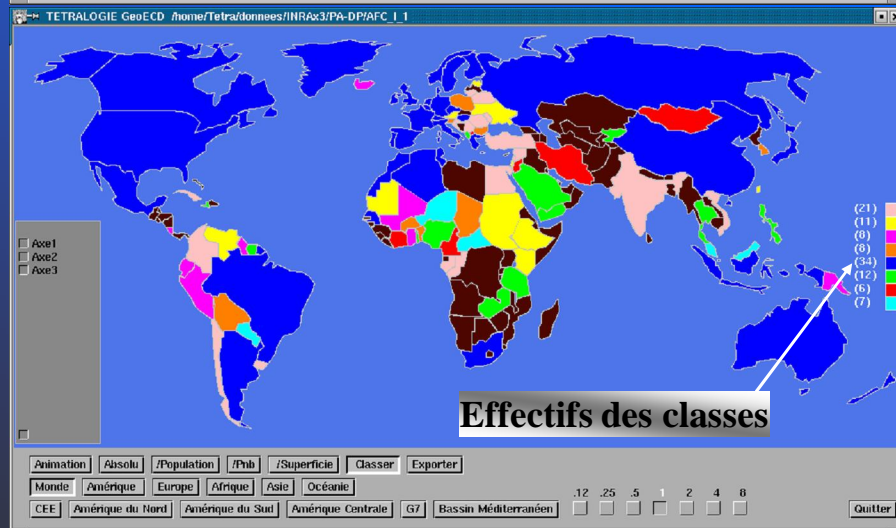
■ Segmentation et importation de classes



Pays non présents dans la dernière période

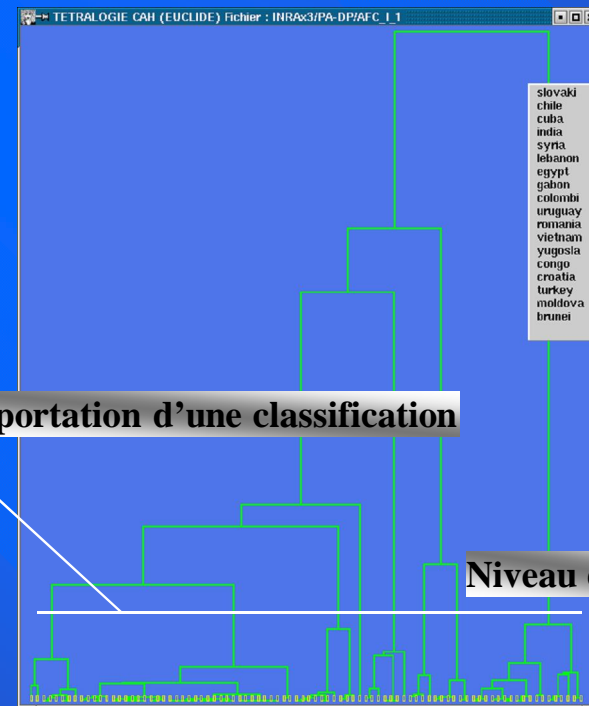
Pays en récession

Pays en forte croissance

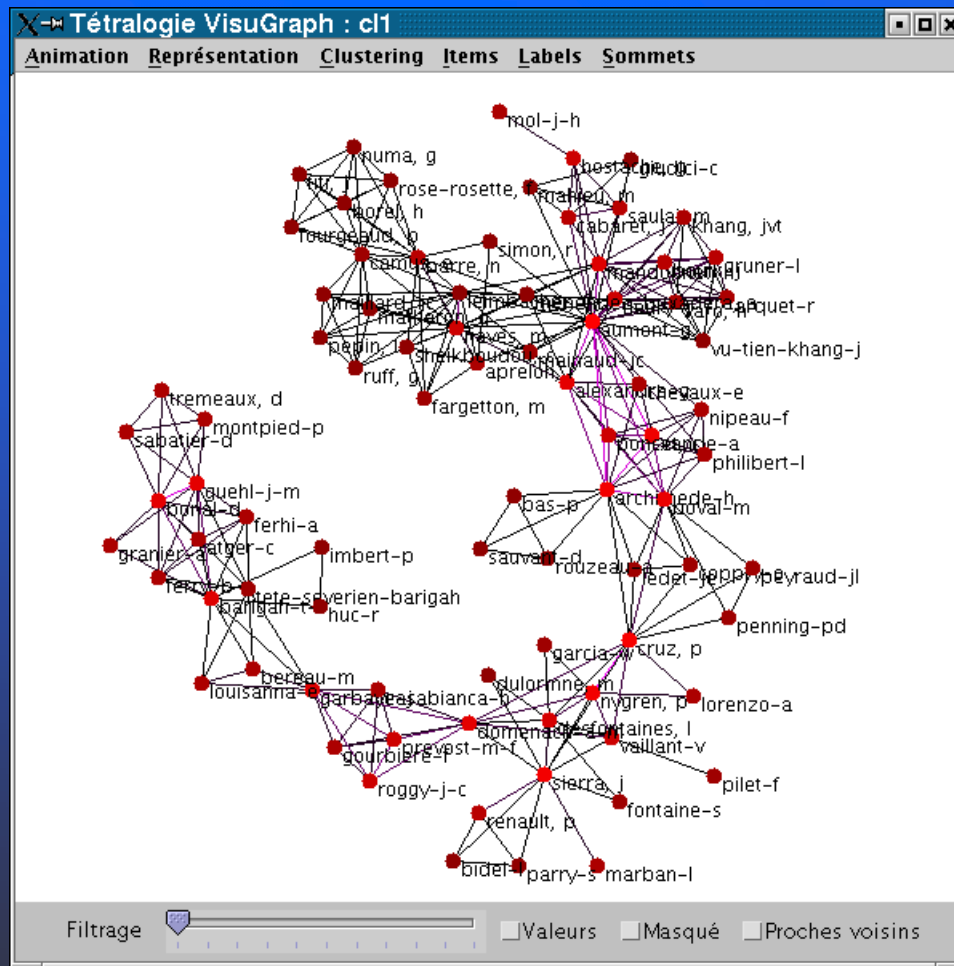


Exportation d'une classification

Niveau de coupure



Convergence rapide du placement des sommets



≈ Algorithme de [Fruchterman91]

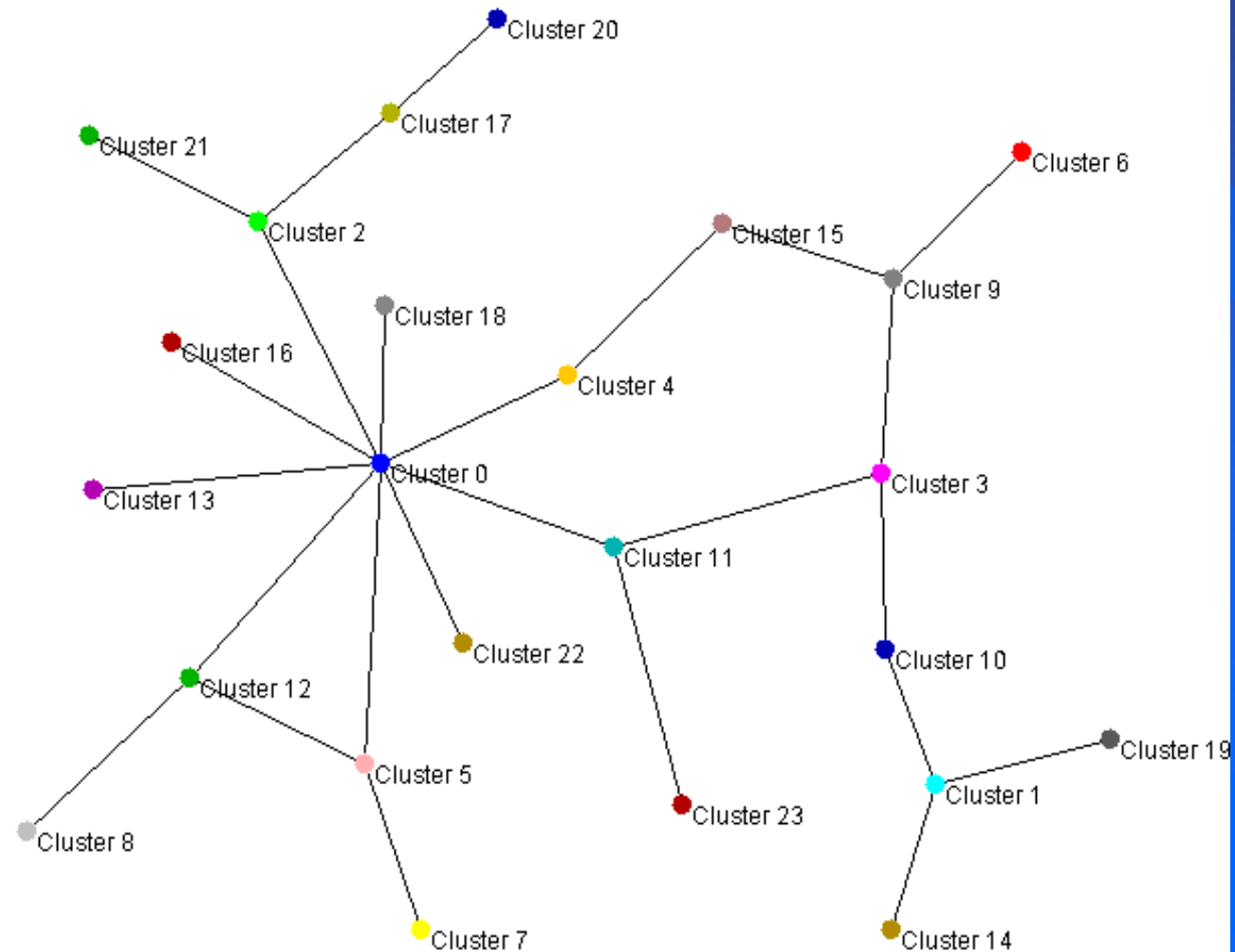
Attraction :

$$f_a(v_i, v_j) = \frac{a_{ij} \times d_{ij}^{\alpha_a}}{k}$$

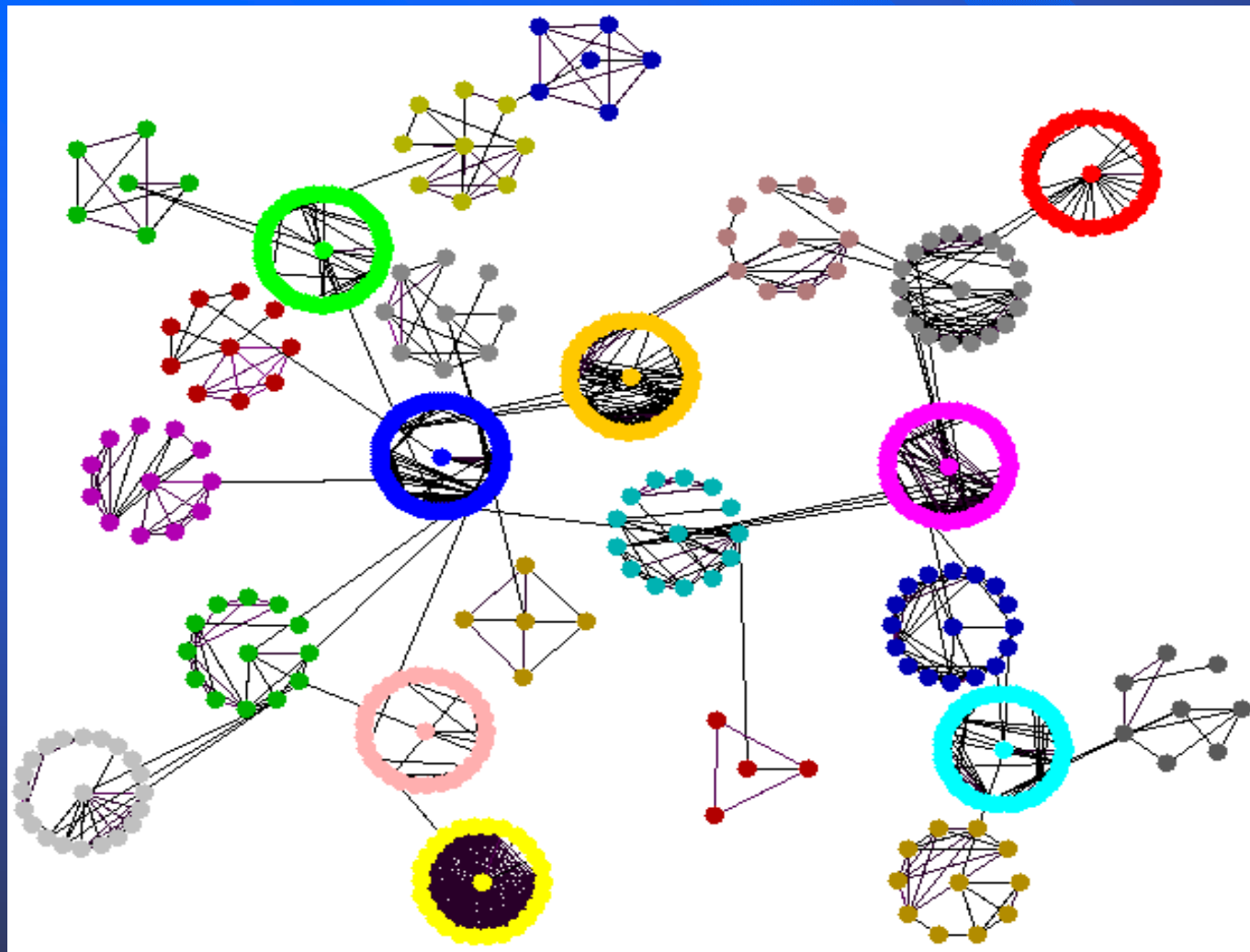
Répulsion :

$$f_r(v_i, v_j) = -\frac{k^2}{d_{ij}^{\alpha_r}}$$

Graphe de clusters



Graphe partitionné



Historique

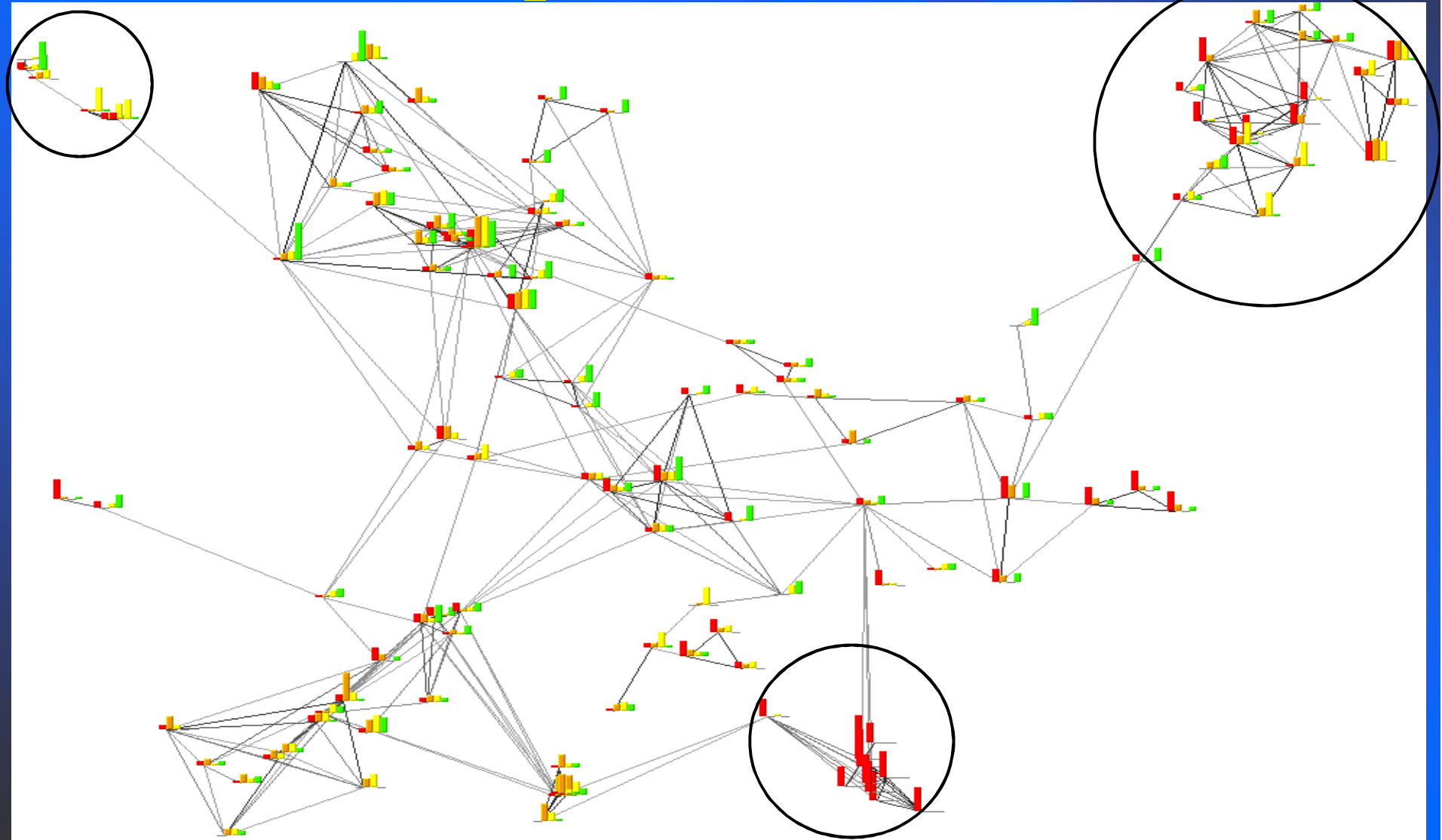
Architecture

Contribution

Conclusion

en visualisation

Graphe évolutif



- Philosophie de notre plate-forme dédiée à l'IE
 - c'est un ensemble cohérent de **prototypes** inter-opérants,
 - qui utilise un **standard unique** pour le format des données,
 - son **interface graphique** est homogène,
 - elle permet de partager ressources et méthodes **via le réseau**.
- Son utilisation en recherche
 - **support d'évaluation** de méthodes, d'outils et de produits d'IE,
 - vaste base d'exemples, à l'**échelle**, déjà analysés.
- Domaine d'application : l'IE et la veille stratégique
 - veille **scientifique** (scientométrie, indicateurs, évaluation)
 - veille **technologique** (propriété industrielle, produits, procédés)
 - veille **économique** (marchés, concurrents, substituts, ...)

Dans le cadre de l'**Intelligence économique**,
cette approche doit permettre

- le recueil, le suivi, le traitement, la diffusion ciblée et la protection de l'information stratégique
- l'accès systématique à l'information implicite
- l'émulation des compétences via les réseaux
- une culture collective interne (partage et émulation)
- une culture collective externe (logique de réseau d'acteurs)
- éventuellement , une culture proactive (influence, lobbying, contre-information/rumeurs, extraction des signaux faibles, décryptage des réseaux,...)