

Le Futur de la Visualisation d'Information

Jean-Daniel Fekete

Projet in|situ
INRIA Futurs

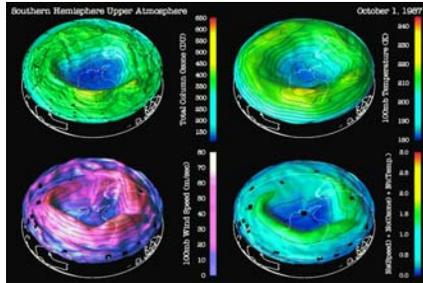
La visualisation d'information

1. Présentation

2. Bilan

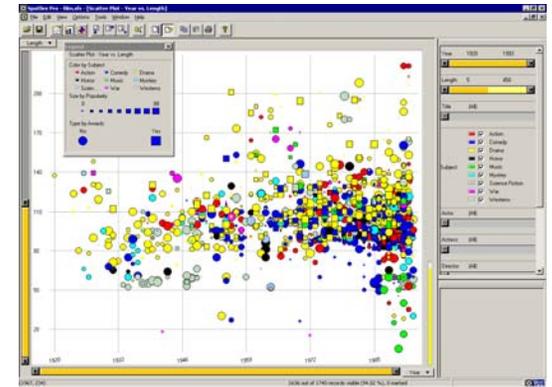
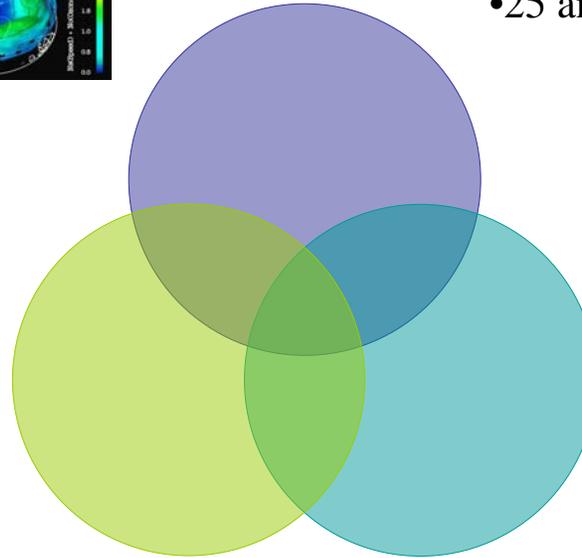
3. Perspectives

Visualisation : 3 domaines



Visualisation scientifique

- Sous communauté de l'Informatique Graphique
- 25 ans d'histoire

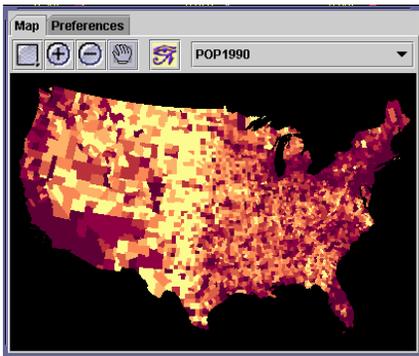


Visualisation d'information

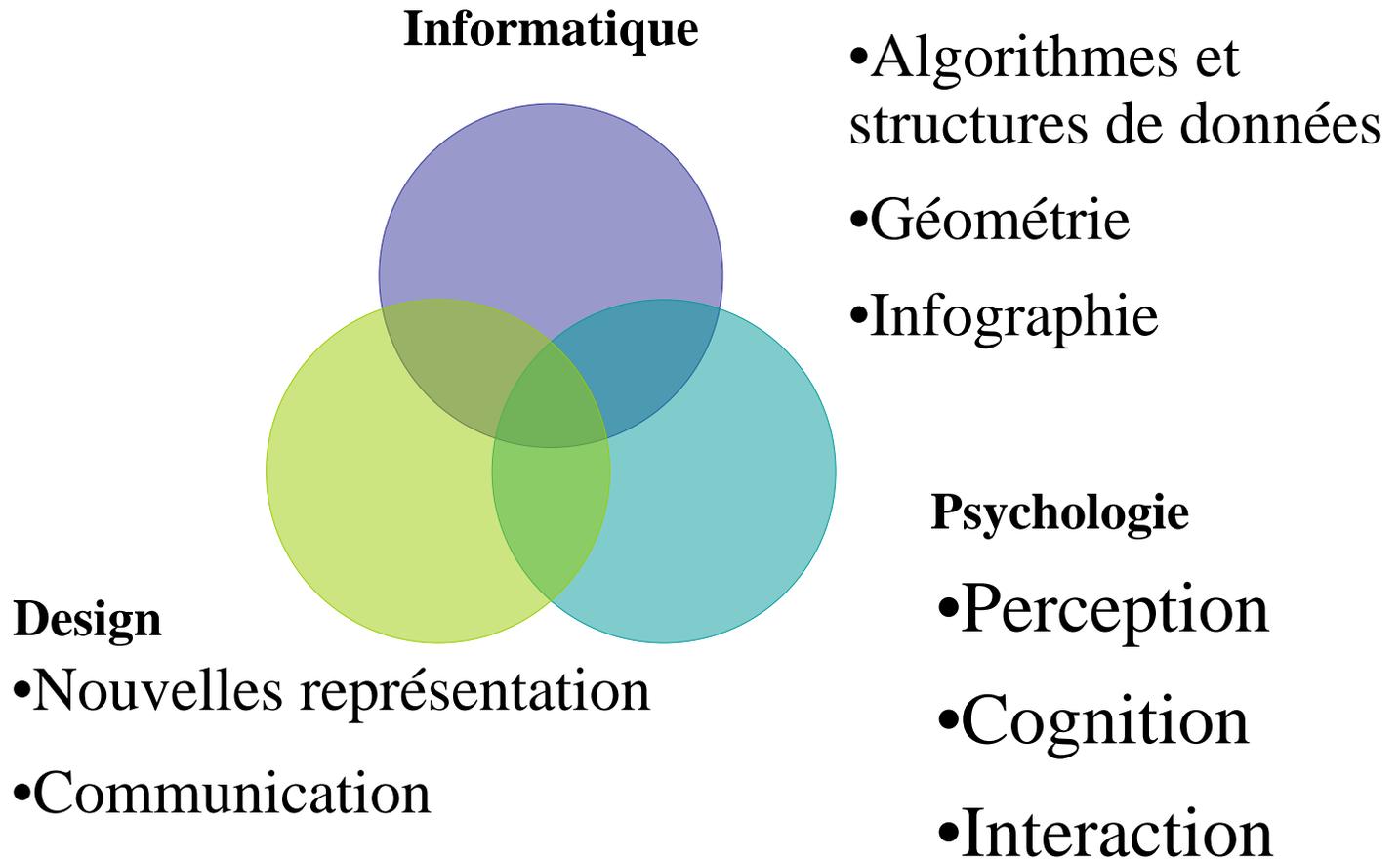
- Sous communauté de l'Interaction Homme-Machine
- 15 ans d'histoire

Cartographie

- Communauté à part entière
- 2000 ans d'histoire



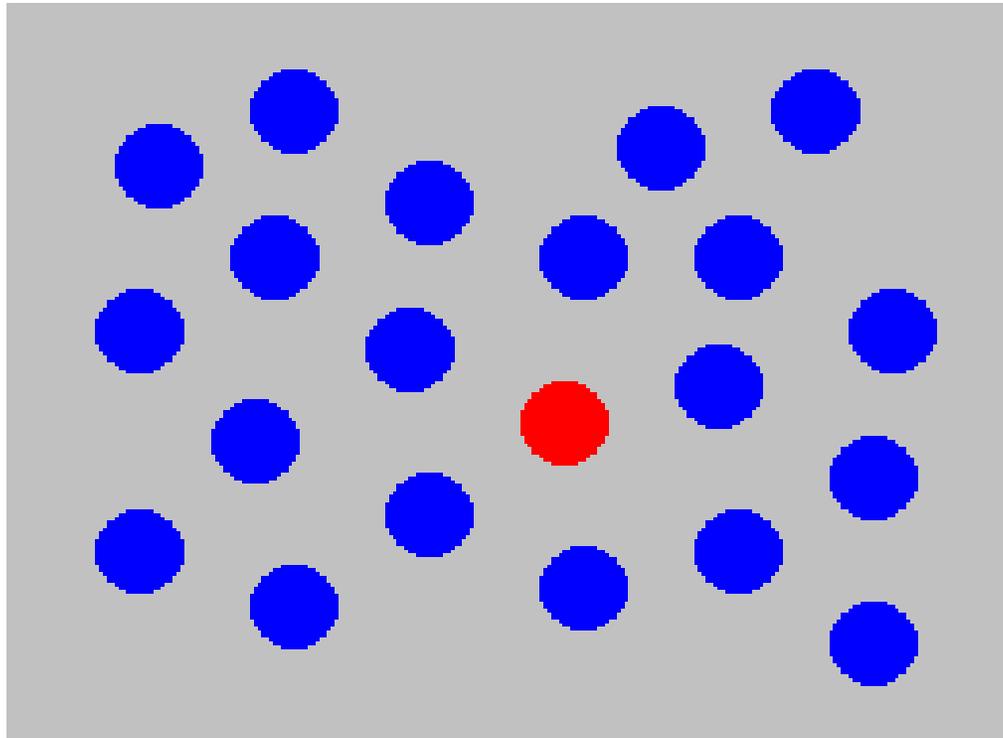
Visualisation : 3 disciplines



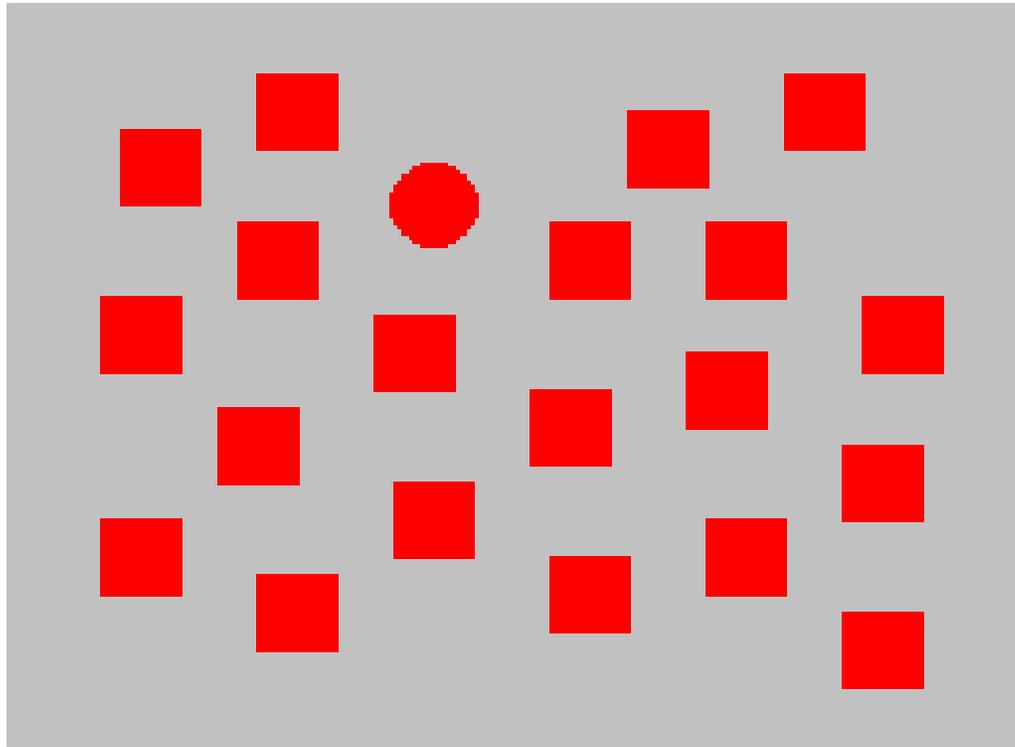
Principes de la visualisation d'information

- L'œil et la perception humaine sont remarquablement adaptés à la reconnaissance de motifs visuels
- La transformation de données abstraites en information visuelle permet d'utiliser cette aptitude
- Parmi toutes les représentations possibles, seules quelques-unes « fonctionnent » :
 - il faut les trouver et les répertorier
- La psychologie nous donne une base d'explication : la perception préattentive (Triesman, 85)
 - Sans effort
 - D'un coup d'œil
 - En temps constant
- Êtes-vous préattentifs ?

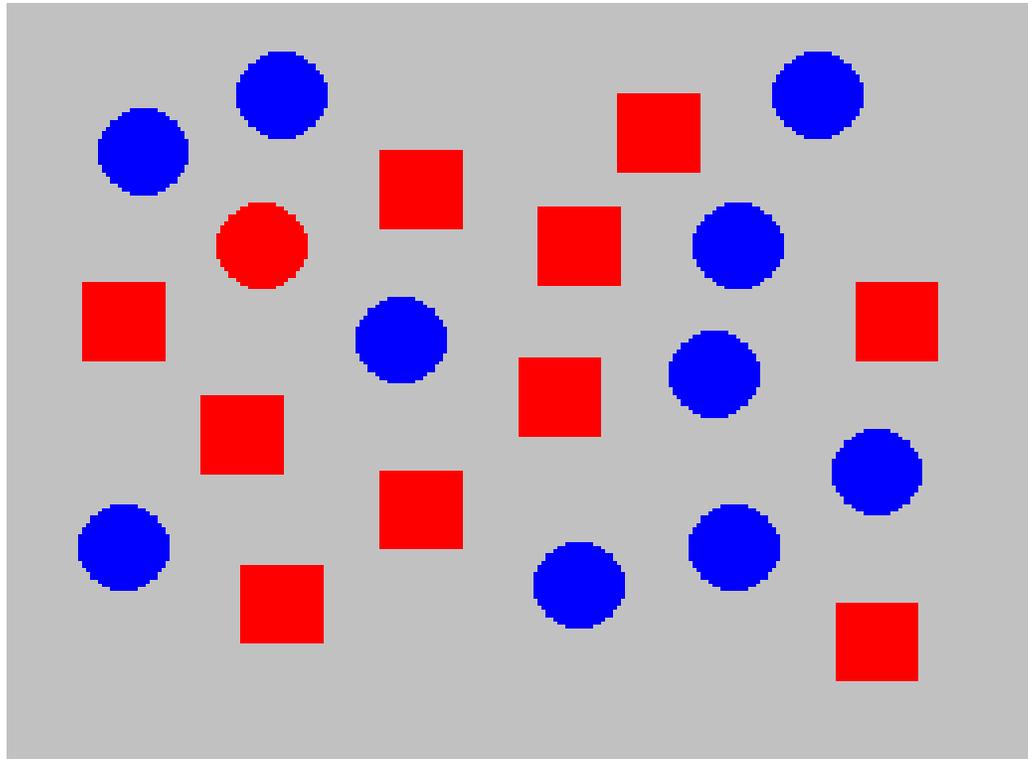
Perception préattentive (1)



Perception préattentive (2)



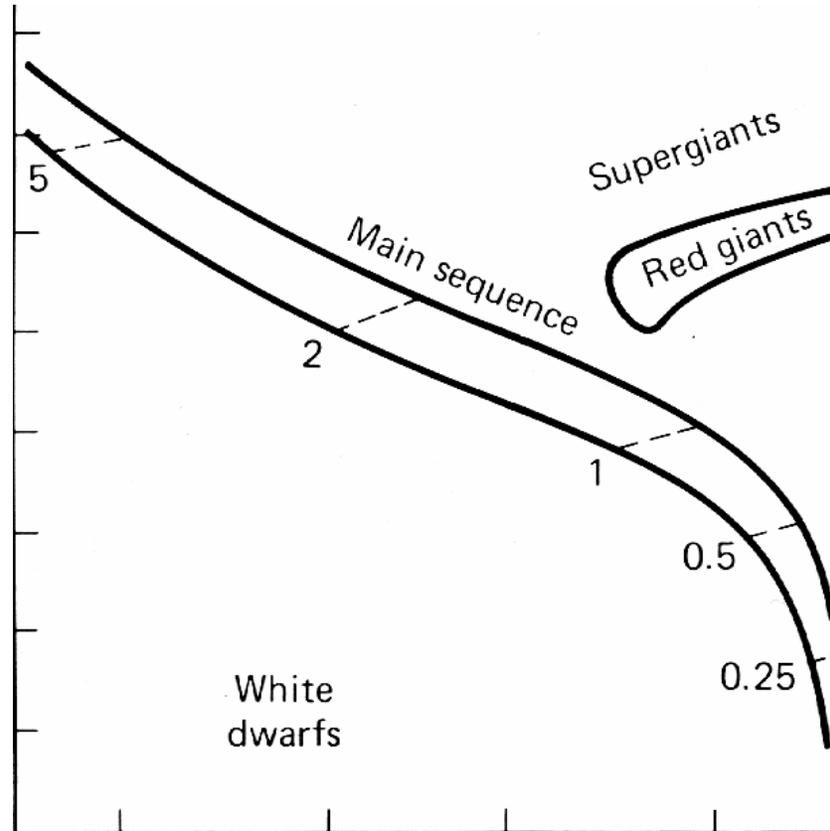
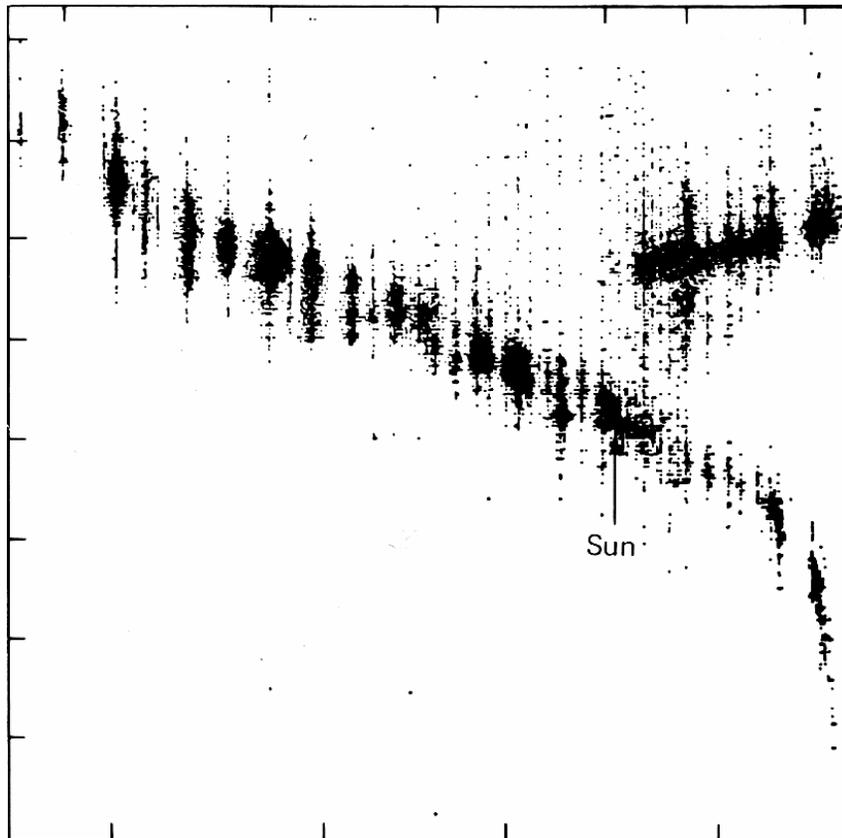
Perception préattentive (3)



Perception préattentive : théorie

- Notre système visuel de bas niveau (25 millions de cellules) fait de la reconnaissance de motif en parallèle en permanence
- Les caractéristiques préattentives sont reconnues à ce niveau
- Les autres nécessitent un parcours séquentiel !
- On a parfois besoin de données visuelles non préattentives
 - Labels/étiquettes sur les données
 - Représentations traditionnelles acceptables par les utilisateurs novices
- Excellentes théories psychologiques
 - Information Visualization: Perception for Design de Colin Ware
- Besoin de conception et réalisation de techniques qui fonctionnent
 - Recours au designer / informaticien
- Les traitements informatiques automatiques peuvent-ils faire mieux ?
 - Pas toujours

Diagramme de Hertzsprung Russell

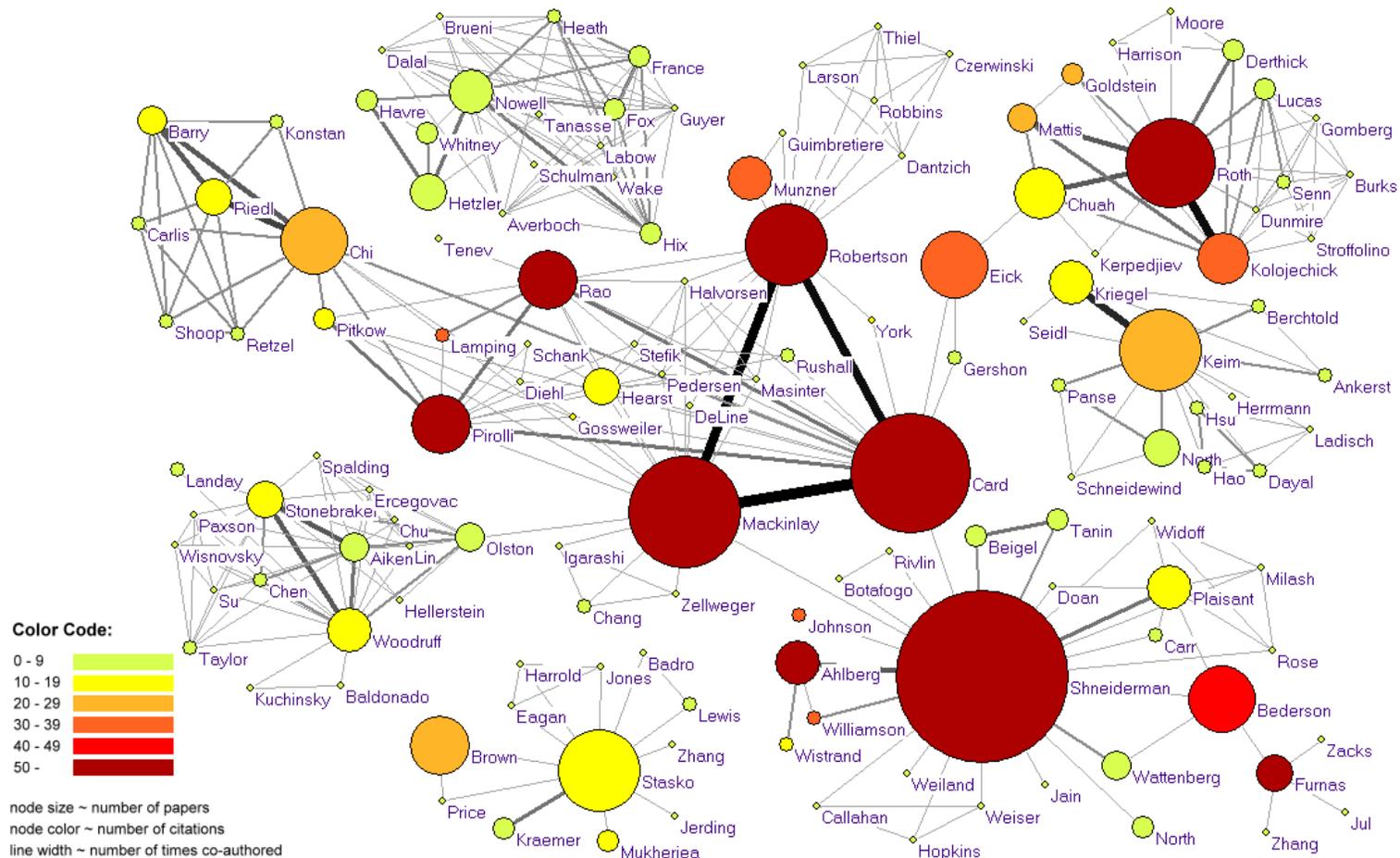


Définition

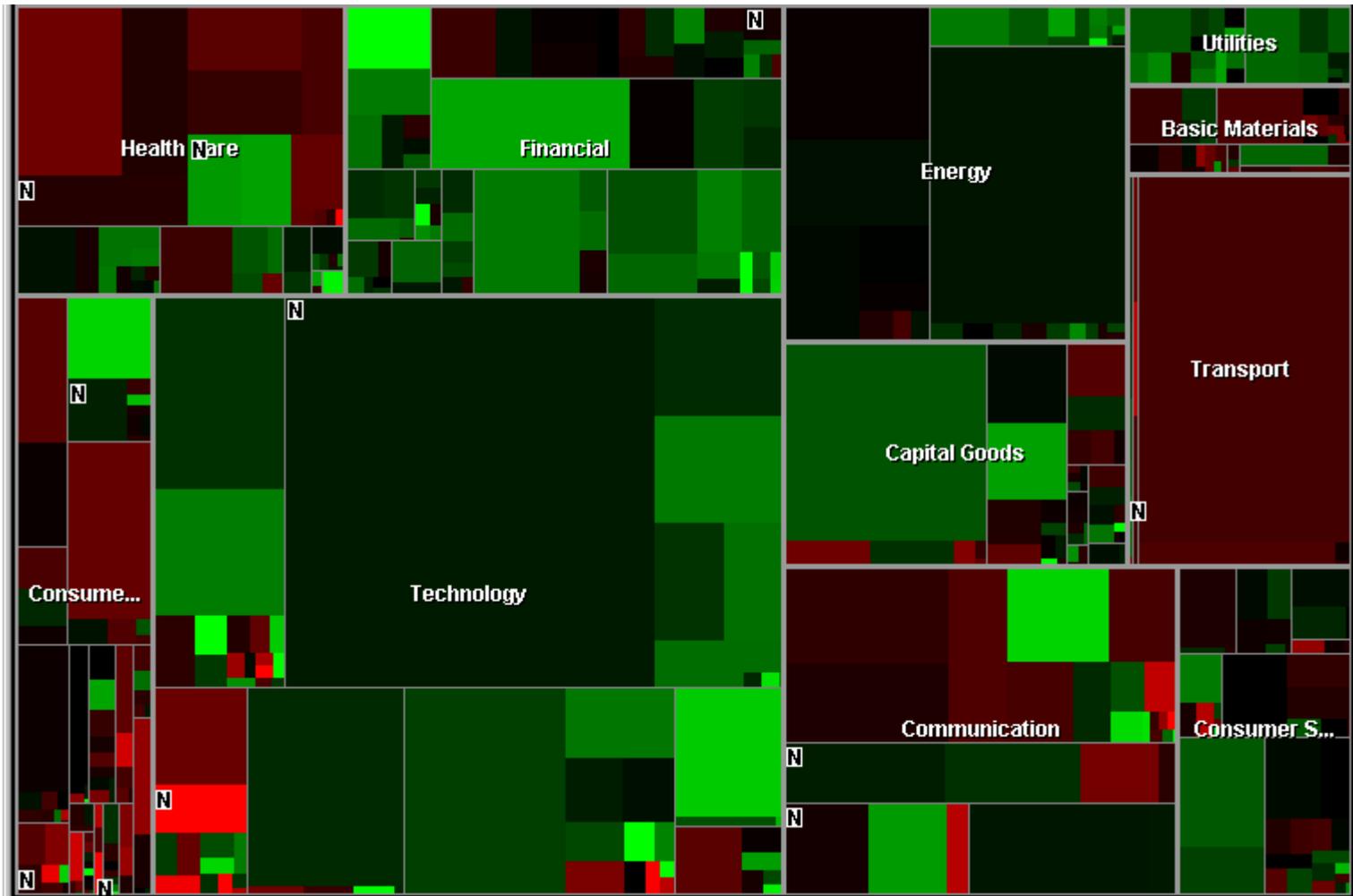
- Représentation graphique compacte et interface utilisateur pour :
 - manipuler un grand nombre d'items ($10^2 - 10^6$)
 - éventuellement extraite d'une base de donnée plus grande
- Permettant aux utilisateurs de
 - faire des découvertes
 - prendre des décisions, ou
 - trouver des explications
- sur
 - des motifs (tendances, groupes, trous, points isolés)
 - des groupes d'items
 - des items individuels
- Fouille de données visuelle
 - Données abstraites, généralement pas de représentation canonique

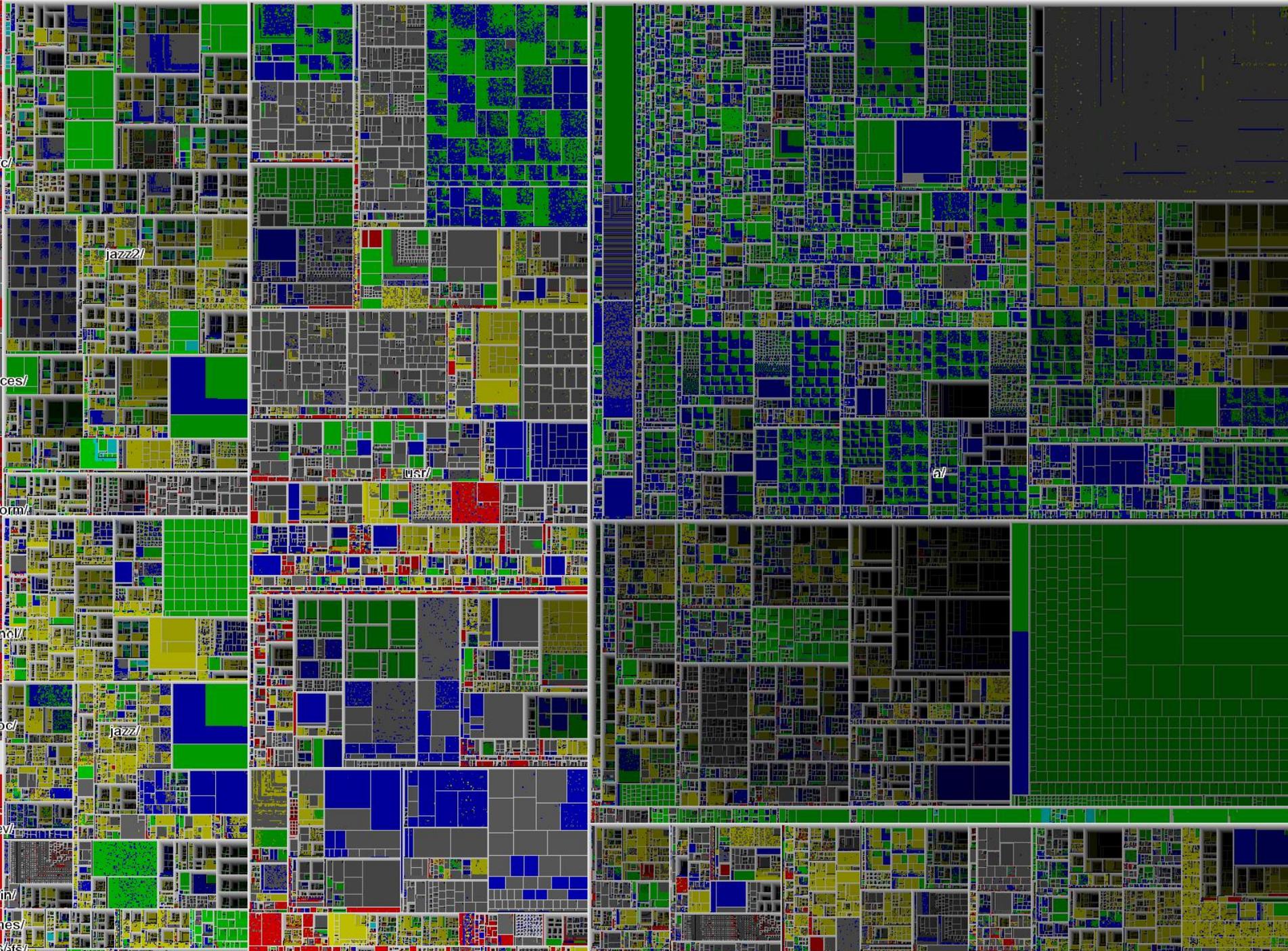
Quelques exemples

Réseaux sociaux : qui publie avec qui ?



Treemap pour la bourse

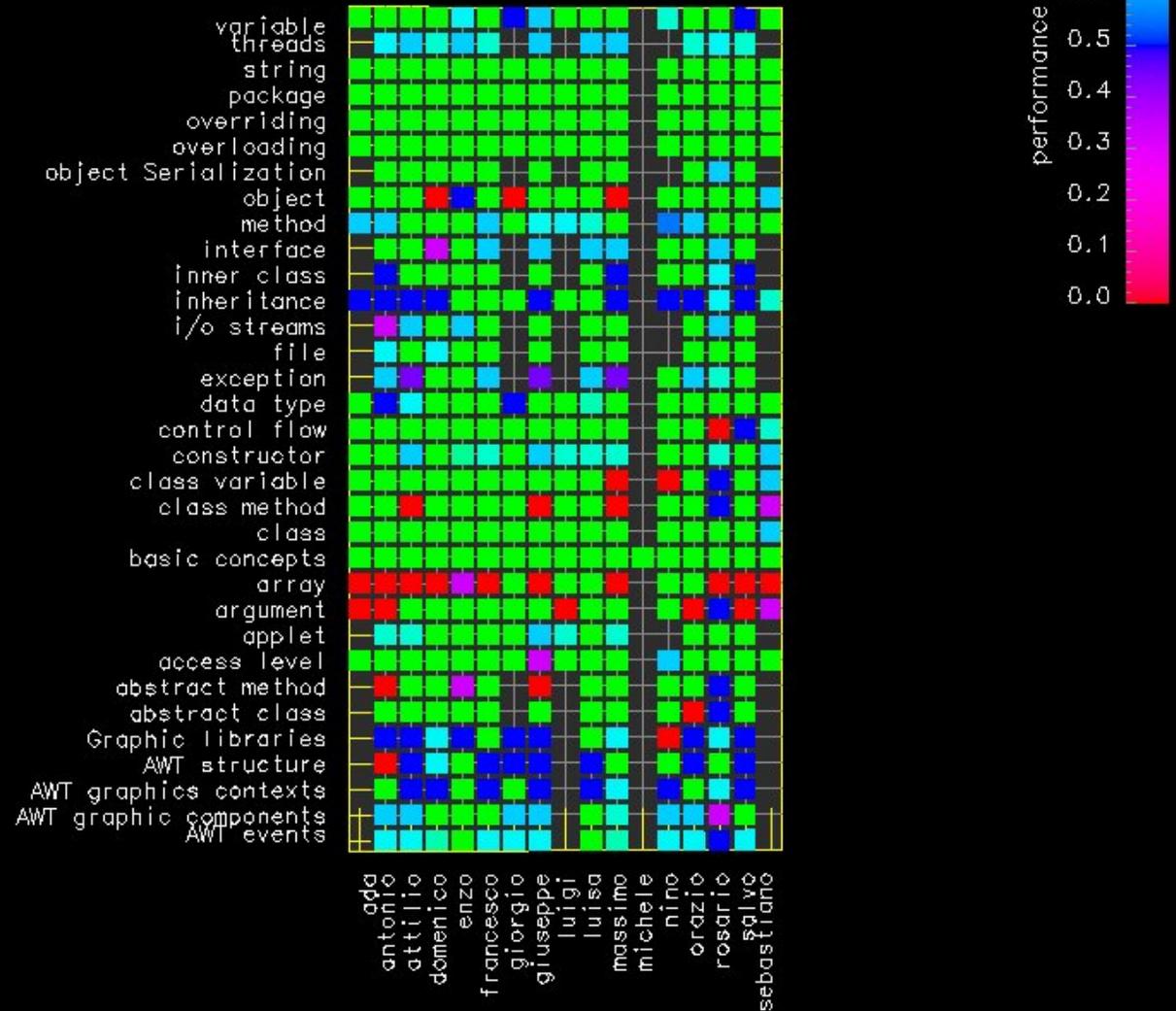




CourseVis

Riccardo
Mazza

- Moodle/
SIGMO



Cognitive matrix

Bilan

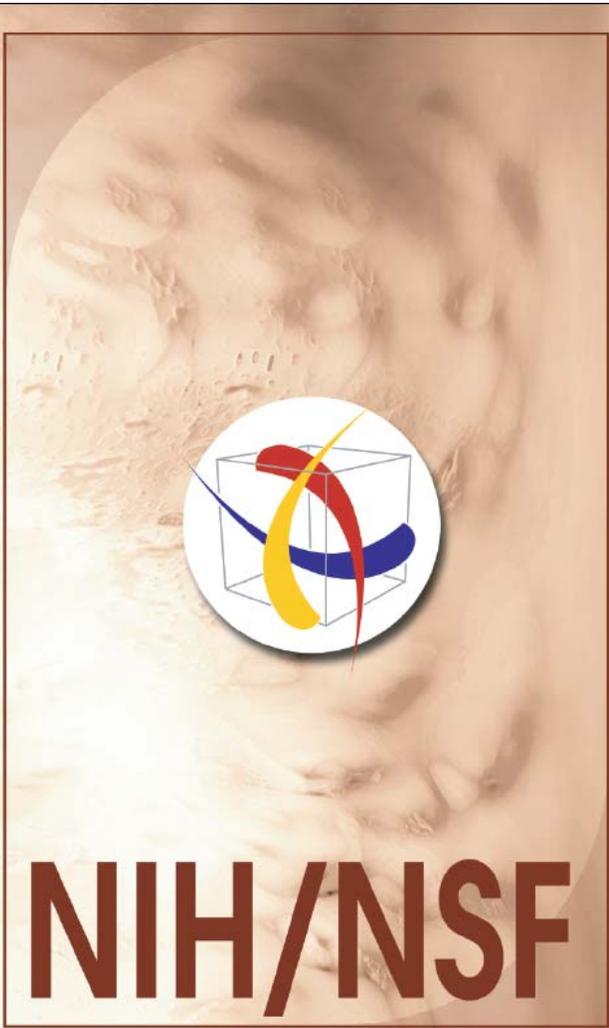
- Domaine en net développement
- Académique :
 - Symposium InfoVis depuis 12 ans
 - Nouvelles conférences : EuroVis, IV, APVIS, VAST
 - Articles à CHI chaque année
- Industriel :
 - Plusieurs sociétés en vivent : SpotFire, MacroFocus, Human-IT, Tableau Software, Hive Group, etc.
 - Plusieurs sociétés le développement : IBM, Microsoft, ILOG, etc.

Bilan

- Plusieurs classification des visualisations
 - Bertin + Mackinlay
 - Par structure de données
 - Un début de théorie
 - Perception préattentive
 - Un modèle architectural et des outils
 - Ed. CHI 97 (InfoVis, Prefuse, Improvise, Tulip)
 - Des méthodes d'évaluation
 - Contrôlées
 - Benchmarks
- 1D - Linéaires
 - 2D - Cartes
 - 3D - Scènes
 - Multidimensionnelle
 - Temporelle
 - Arbres
 - Réseaux
 - Documents
 - Algorithmes

Perspectives

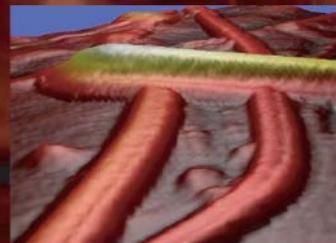
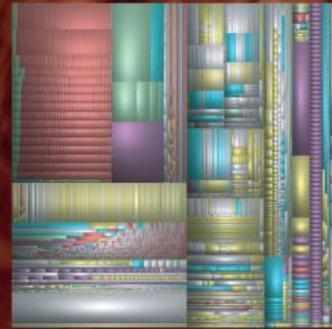
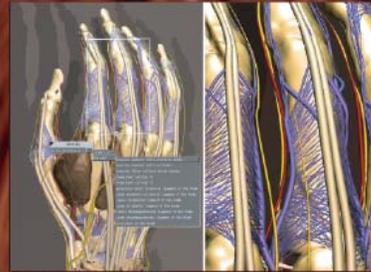
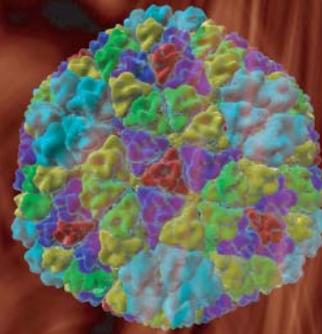
- Naissance d'un nouveau domaine
 - Visual Analytics (Analyse Visuelle)
- Gestion de masses de données
 - Agrégation / échantillonnage / réduction des dimensions
 - Méthodes multi-échelles
 - Gestion de données évoluant temporellement
- Améliorer les architectures et les interactions
- Améliorer les méthodes d'évaluation
 - Taxonomie des tâches par données
 - Évaluations composant / système / environnement
 - Évaluations longitudinales
- Améliorer la visibilité du domaine
 - Apprendre à lire les visualisations



NIH/NSF

VISUALIZATION RESEARCH CHALLENGES

JANUARY 2006



Analyse Visuelle

- Née après le 11 septembre aux USA
- Ensemble de méthodes permettant de raisonner sur des données afin de prendre des décisions
 - Recueil de données
 - Traitement analytique (Data Mining)
 - Visualisation pour l'aide à l'analyse
 - Aide à la poursuite d'hypothèses
 - Visualisation pour la présentation des résultats

J. J. Thomas, K. A. Cook (editors), Illuminating the Path: The Research and Development Agenda for Visual Analytics, IEEE Press, August 2005, <http://nvac.pnl.gov/agenda.stm>

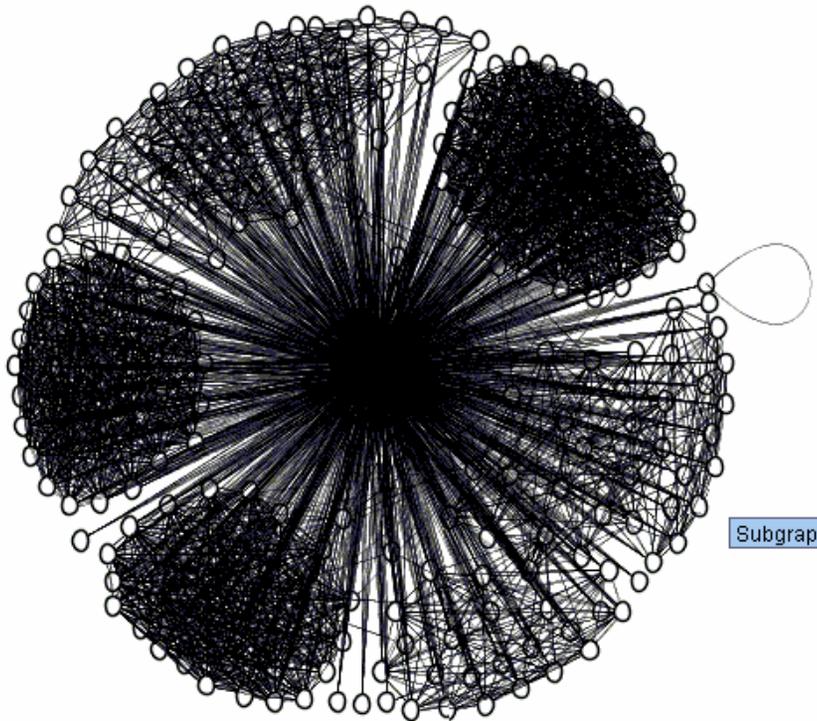
Gestion de masses de données

- Notre œil n'a que 25 millions de cônes + bâtonnets
 - On ne peut pas aller au-delà
- Comment utiliser la visualisation pour des données plus grandes ?
 - Échantillonnage
 - Agrégation
 - Réduction de dimensions
 - Navigation
 - Interaction
 - Animation
 - Méthodes multi-échelles
- Les systèmes de visualisation ne gèrent pas plus du million de données

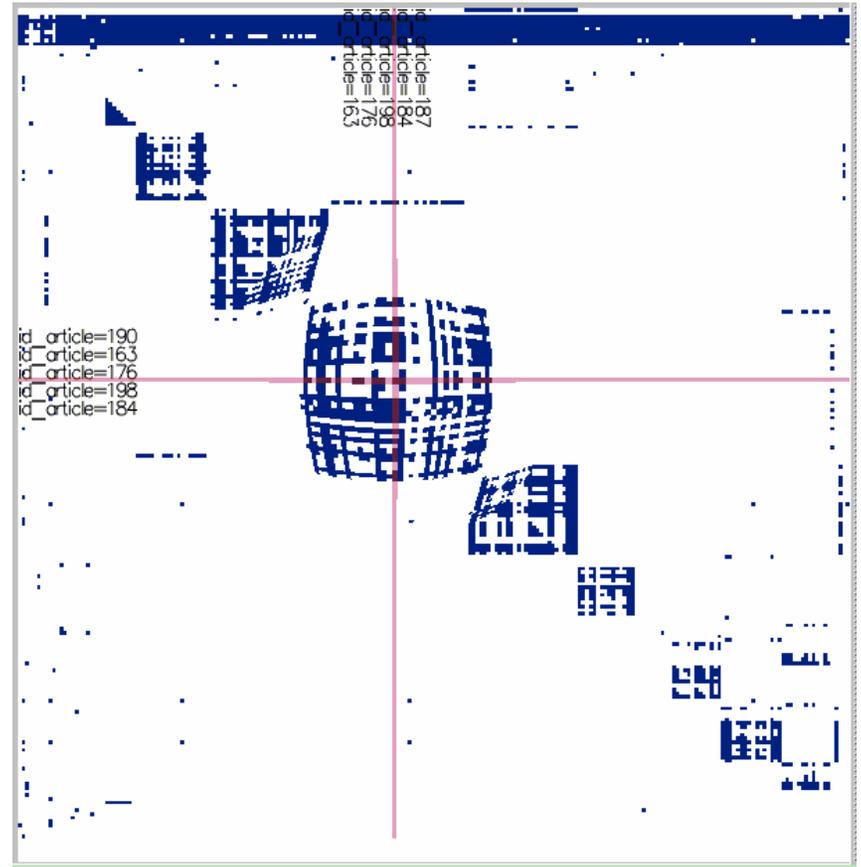
Exemple de passage à l'échelle

- La visualisation de graphes
- Représentation traditionnelle
 - nœuds et liens
- Autre représentation possible
 - Matrice d'adjacence
- Avant 2003, personne n'avait de doute
 - Les matrices sont illisible !

Exemple de site Web

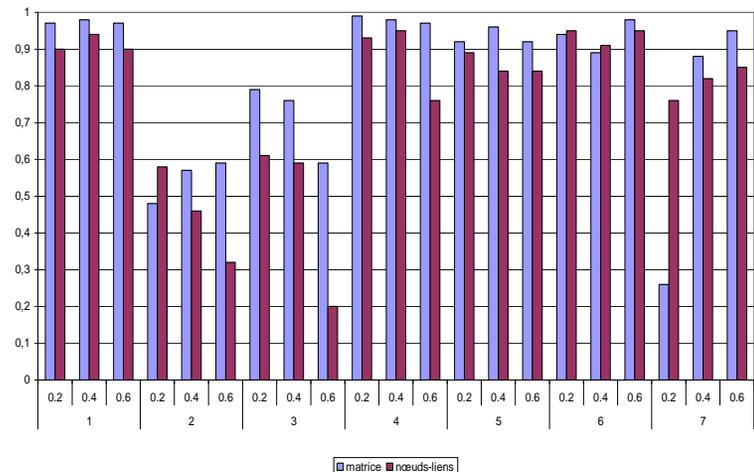
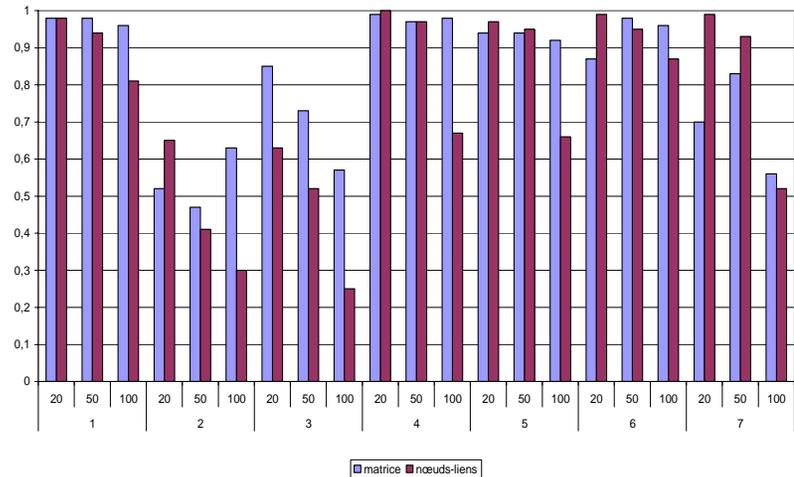


Subgrap



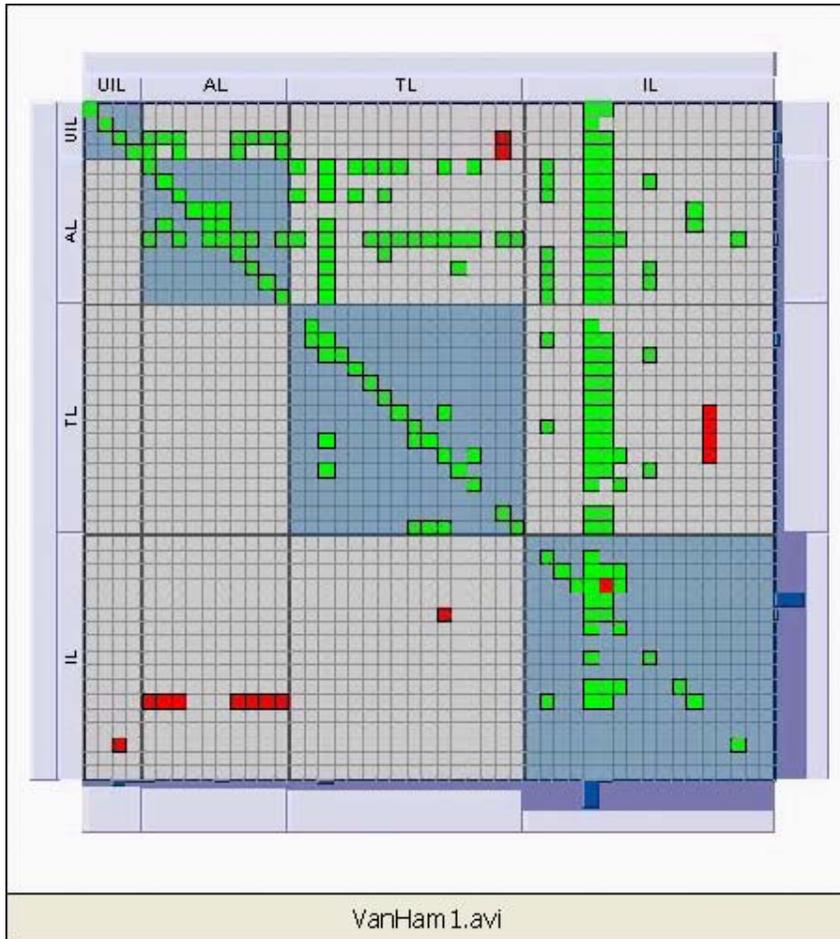
Résultats

- nœuds-liens préférable pour les petits graphes (20 sommets) peu denses
- matrice plus lisible pour les graphes denses, les graphes moyens ou grands (> 20 sommets), sauf les tâches de chemins



Hiérarchiser pour naviguer

(van Ham 03, Abello&van Ham 04)



Matrix Zoom
A Visual Interface to
Semi-External Graphs

James Abello
Frank van Ham

3 min avi
captured at 12 FPS

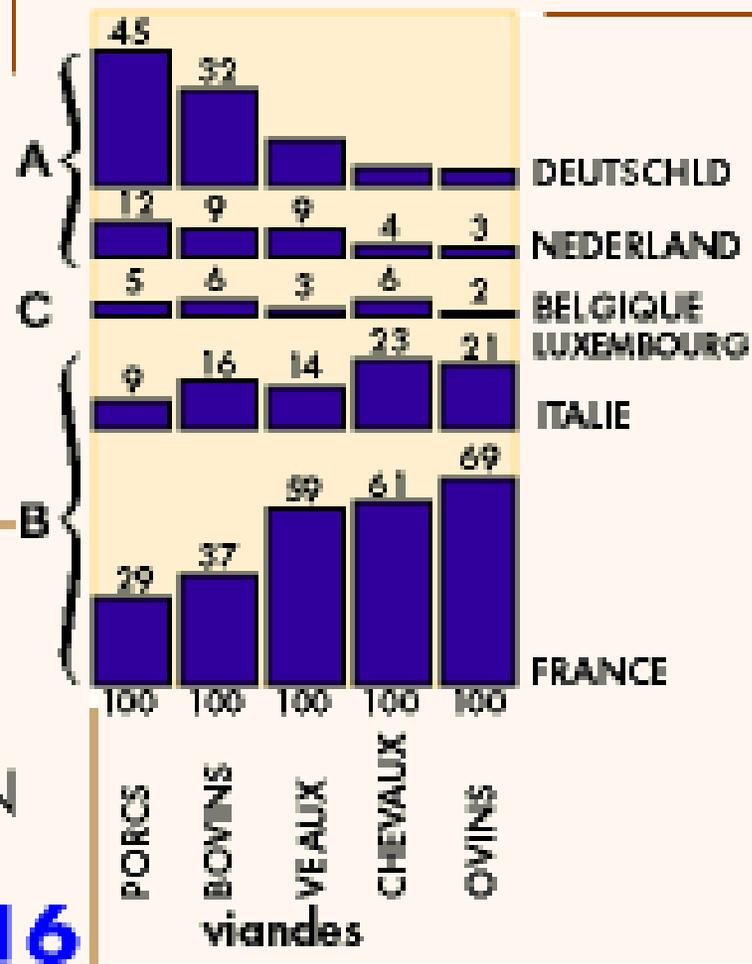
matrix zoom.avi

Réordonner pour comprendre

6	6	2	5	3	BELGIE LUXEMBOURG
32	6	5	45	15	DEUTSCHLAND
37	61	69	29	59	FRANCE
16	23	21	9	14	ITALIE
9	4	3	12	9	NEDERLAND
100	100	100	100	100	
BOVINS	CHEVAUX	OVINS	PORCS	VEAUX	

DONNÉES

■ complexité



15

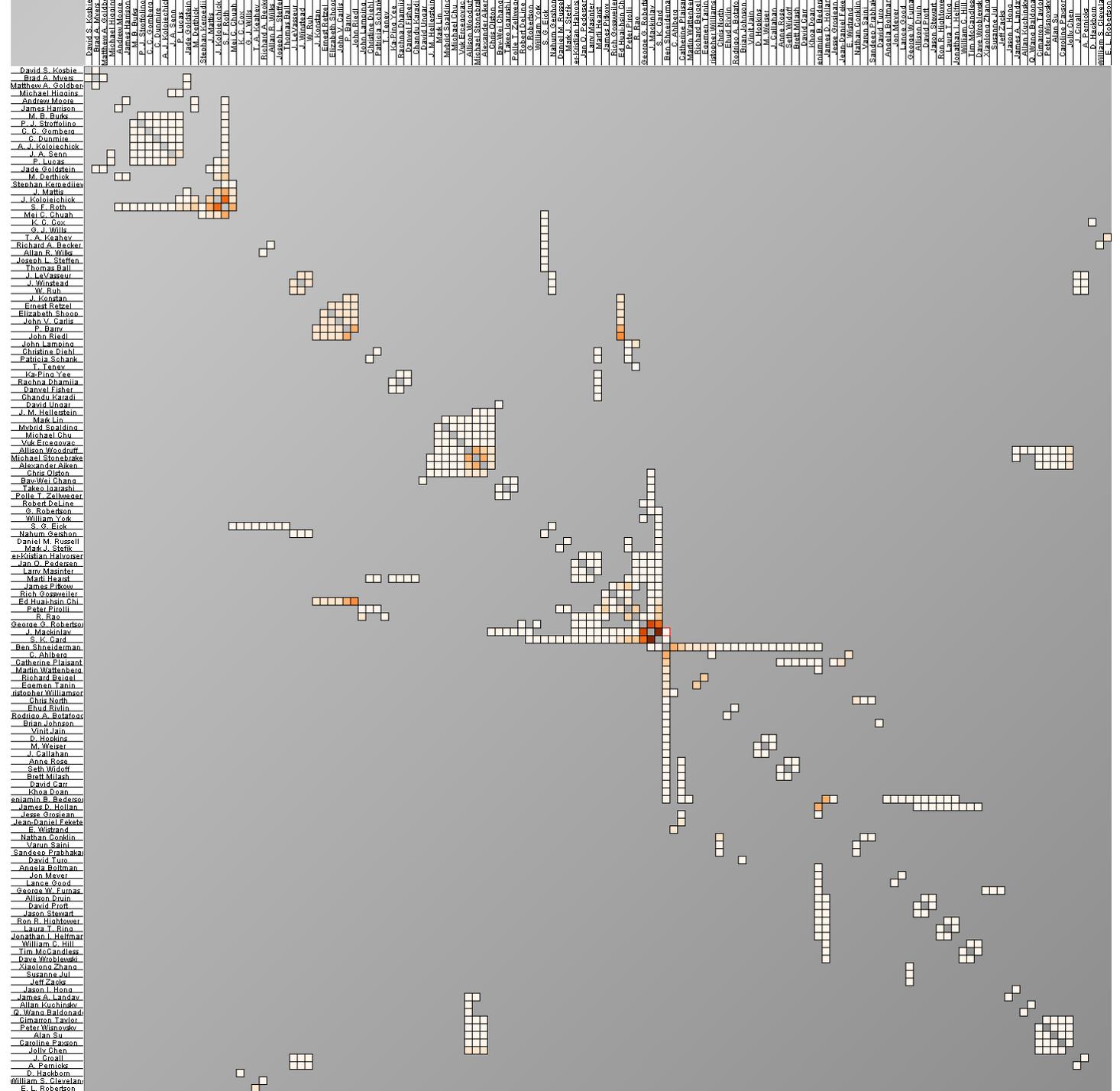
reclassement*

■ groupes

■ INFORMATION

*Reclassement = simplification sans suppressions

16



- David S. Kosbie
- Brad A. Mvaz
- Matthew A. Seldner
- Michael Higgins
- Andrew Moore
- James Harrison
- M. B. Burke
- P. J. Stoffolino
- G. G. Combsen
- C. Dunmire
- A. J. Koloschick
- J. A. Sano
- F. Lucas
- Jake Goldstein
- M. Dethick
- Stephan Kapradiev
- J. Mattis
- J. Koloschick
- S. F. Roth
- Mel C. Chuah
- G. J. Willis
- T. A. Kearney
- Richard A. Becker
- Alan F. Willis
- Joseph L. Steffen
- Thomas Ball
- J. LeVasseur
- J. Winstead
- W. Ruh
- J. Konstan
- Ernest Reitel
- Elizabeth Shoop
- John V. Carlin
- J. Bang
- John Friedl
- John Lampino
- Christine Diehl
- Patricia Schank
- T. Teney
- Ka-Ping Yee
- Rachna Dhamija
- Danual Fisher
- Chandu Khandi
- David Ungar
- J. M. Halenstein
- Mark Lin
- Mybnd Spaldino
- Michael Chu
- Vuk Elocosovac
- Allison Woodrum
- Michael Skorsbrake
- Alexander Aiken
- Chris Dickon
- Baw-Wa Chango
- Takeo Igarashi
- Felix T. Zellweger
- Robert Delina
- G. Robertson
- William York
- S. G. Eick
- Nathan Schuman
- Daniel M. Russell
- Mark J. Steink
- Christian Halverson
- Jan D. Federsan
- Lynn Mastiter
- Matt Heiser
- James Pitkow
- Rich Szoswella
- Ed Huachsin Chi
- Peter Piralli
- R. Rao
- George G. Robertson
- J. MacIntyre
- S. K. Gair
- Ben Shneiderman
- C. Ahlberg
- Catherine Flaisant
- Martin Watanabe
- Richard Biesel
- Edman Tann
- Christopher Williams
- Chris North
- Ehud Rivlin
- Rodfane A. Easton
- Erin Johnson
- Vinyl Jain
- D. Hopkins
- M. Weiser
- J. Callahan
- Anna Rose
- Seth Wildoff
- Brett Milash
- David Carr
- Khoa Doan
- Shamim B. Bafese
- James D. Hollan
- Jesse Grogan
- Jean-Daniel Fekete
- E. Wietland
- Nathan Conklin
- Vainu Saini
- Sandeep Prabhakar
- David Lute
- Angela Boltman
- Jon Meyer
- Lance Good
- George W. Furnas
- Allison Drin
- David Froot
- Jason Stewart
- Ron F. Holtzman
- Laura J. Ring
- Jonathan I. Halmar
- William C. Hill
- Tim McCardless
- Dave Wroblewski
- Xiaolou Zhang
- Suzanne Jul
- Jeff Zade
- Jason I. Hong
- James A. Landay
- Allan Kuchinsky
- D. Wang Baldwin
- Clintanton Taylor
- Peter Winkovky
- Alan Su
- Caroline Eason
- Jolly Chen
- J. Goyal
- A. Frenkel
- D. Backbin
- William S. Cleveland
- E. L. Robertson

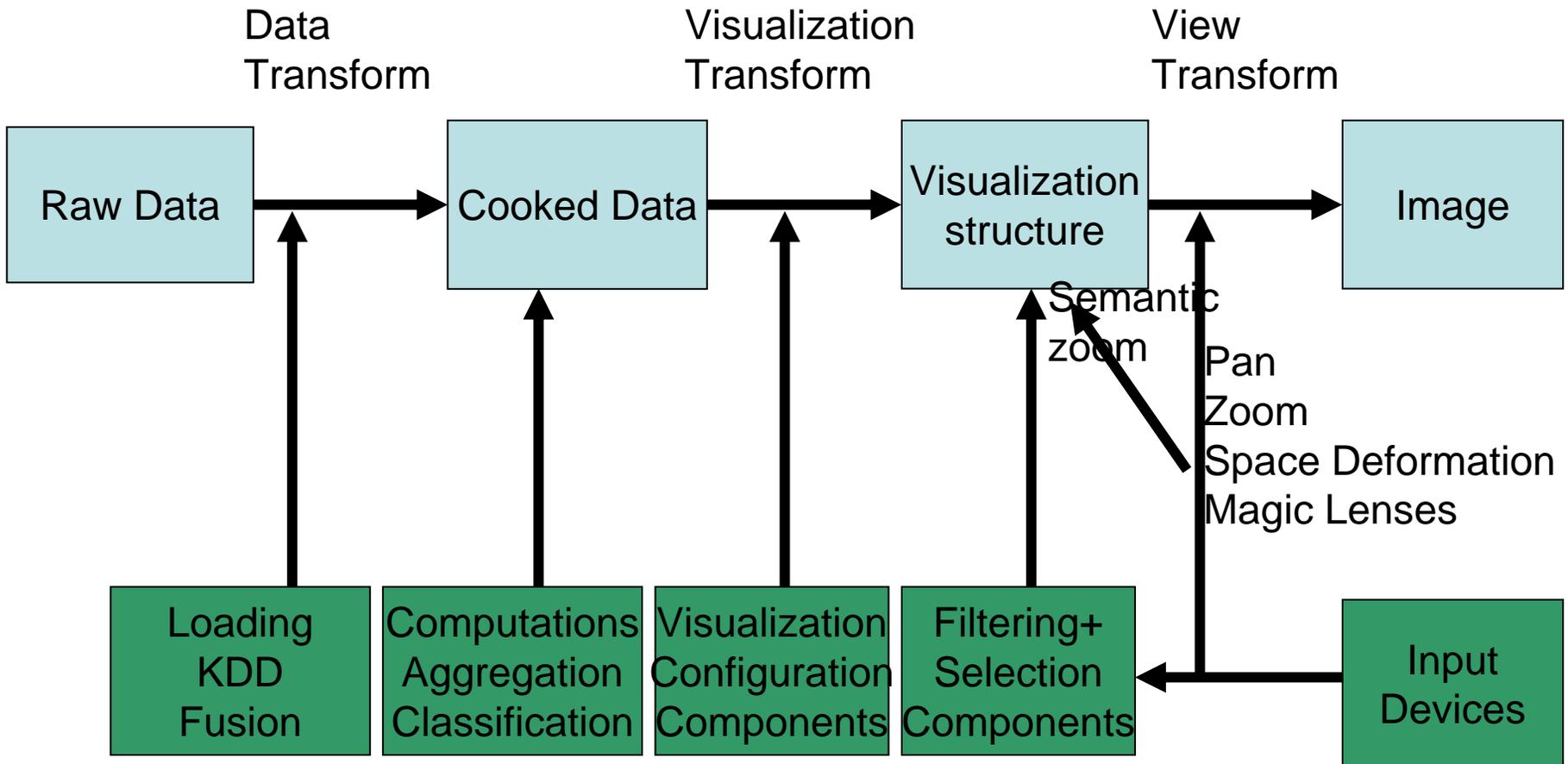
Passage à l'échelle

- Réordonner des grosses matrices
 - ACP sur un échantillon de la matrice (Harel&Koren 02)
 - Quelques secondes pour 1 million de sommets
- Hiérarchiser ensuite
- Naviguer

Architectures pour la visualisation d'information

- Améliorer les infrastructures pour gérer
 - Plus de données
 - Des données plus riches
 - Toujours dans un temps « interactif »
 - A travers le Web (asynchrone)
 - Intégrable aux applications graphiques interactives
 - Permettant l'édition des données

InfoVis Architecture



Améliorer les méthodes d'évaluation

- Le rapport NVAC distingue trois niveaux
 - Le composant
 - Le système
 - L'environnement
- On sait évaluer un composant et un peu un système, pas l'environnement
- Comment faire des l'évaluation longitudinale sur des systèmes visuels ?
- Quelles métriques de succès ?

Améliorer la visibilité du domaine

- L'analyse visuelle demande du temps
 - Pour comprendre le codage visuel
 - Pour se familiariser avec les systèmes
- On n'apprend pas ces codages à l'école
- En exposant le public à des visualisations, il apprendra

Que faire ?

- La France est dans la course
 - LaBri (Bordeaux), LIRMM (Montpellier), LRI (Orsay), FUNDP (Namur), etc.
 - ILOG
- L'industrie et la recherche sont demandeurs
 - Microsoft Dublin
 - EDF, France Telecom, etc.
 - Bioinformatique, bibliothèques numériques
- Aidez-nous à maîtriser la complexité des données
 - et à augmenter la cognition humaine