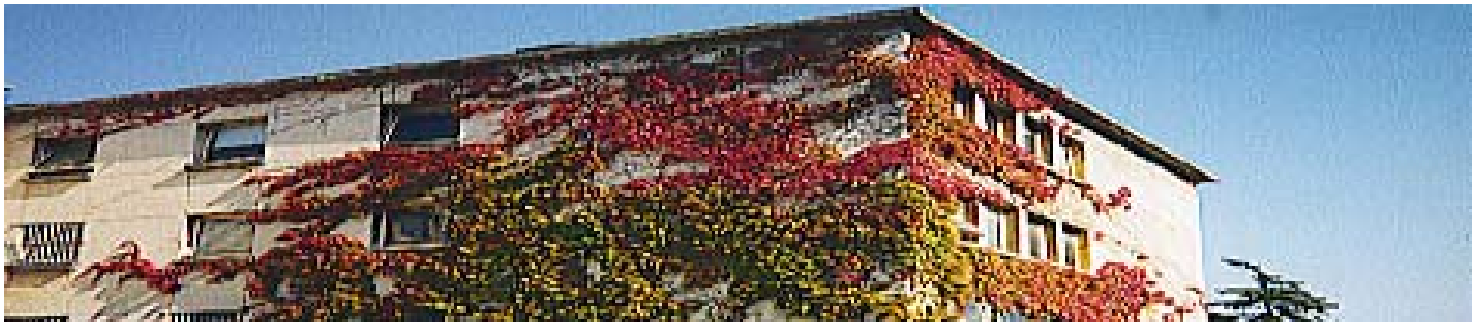


La visualisation d'information : une méthode pour explorer des données empiriques et les comprendre

Jean-Daniel Fekete

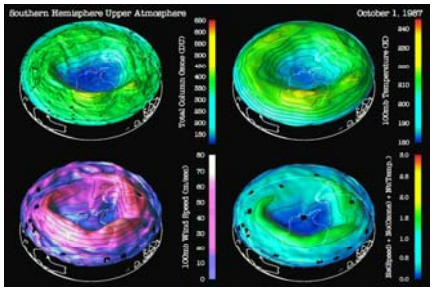
Equipe AVIZ www.aviz.fr

INRIA Saclay – Île-de-France

Plan

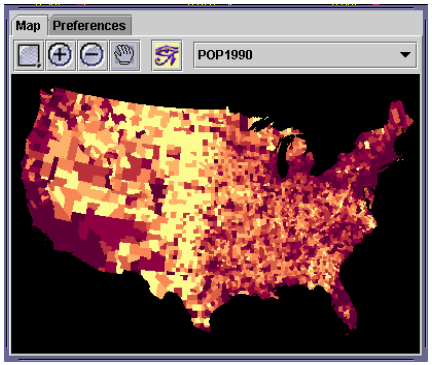
- Fondements du domaine
- Exemples de visualisations
- Systèmes complexes
  - Exécution de programme
  - Réseaux sociaux
  - Solveurs de contraintes
- Perspectives

Visualisation : 3 domaines



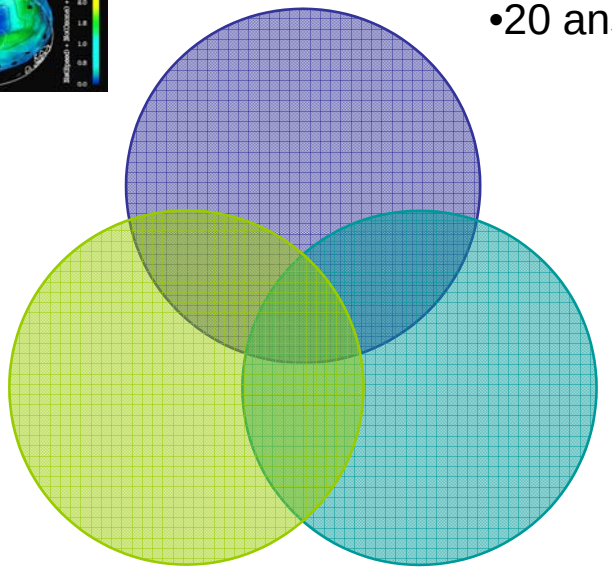
Visualisation scientifique

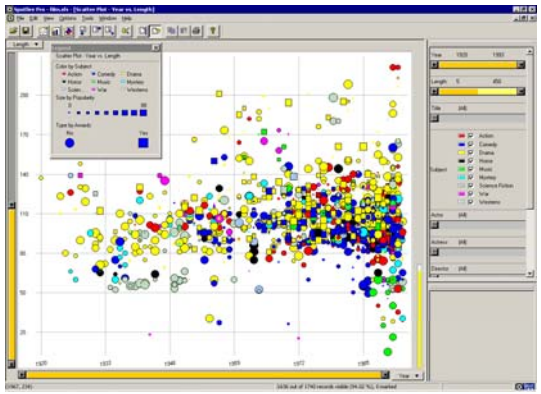
- Sous communauté de l'Informatique Graphique
- 20 ans d'histoire



Cartographie

- Communauté à part entière
- 2000 ans d'histoire





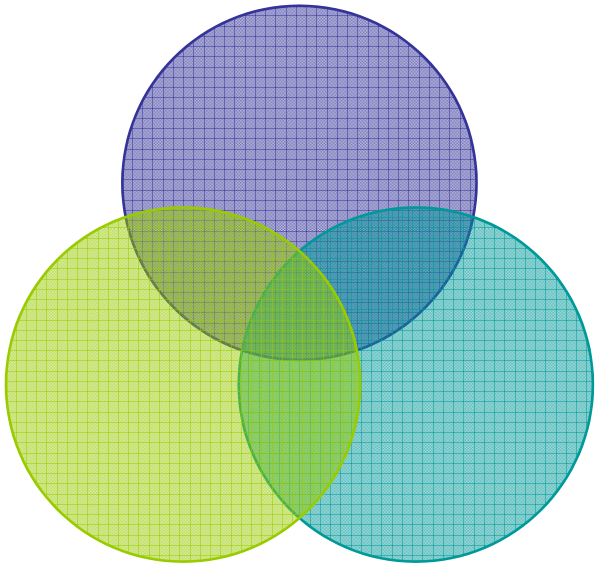
Visualisation d'information

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

Visualisation : 3 disciplines

Informatique

- Algorithmes et structures de données
- Géométrie
- Infographie



Design

- Nouvelles représentation
- Communication

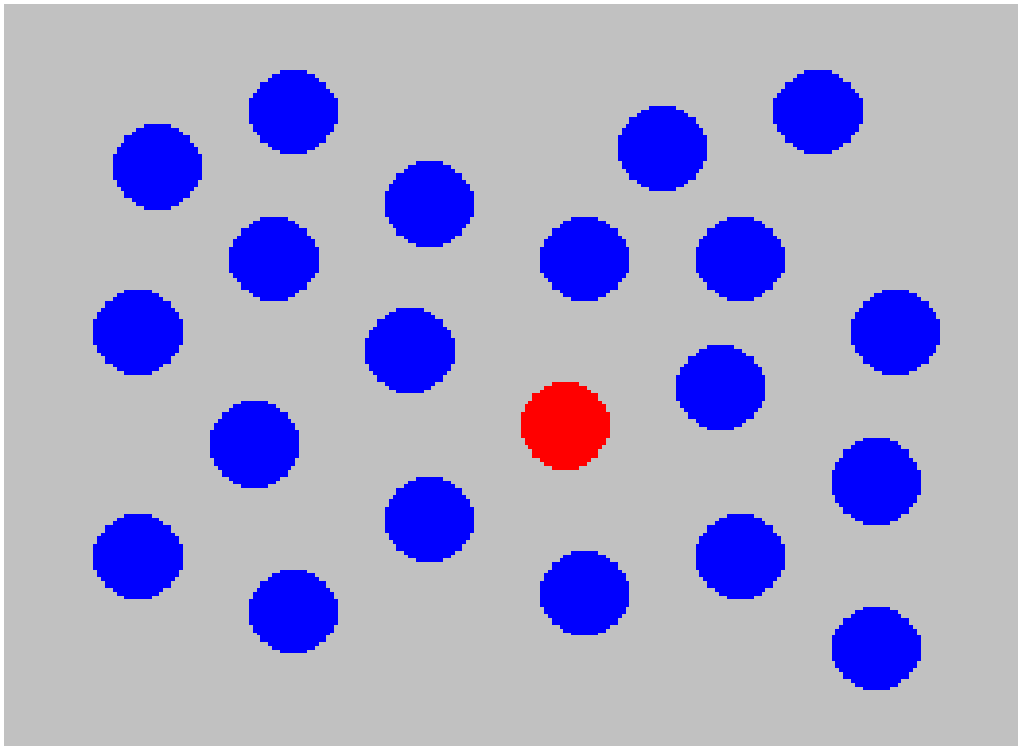
Psychologie

- Perception
- Cognition
- Interaction

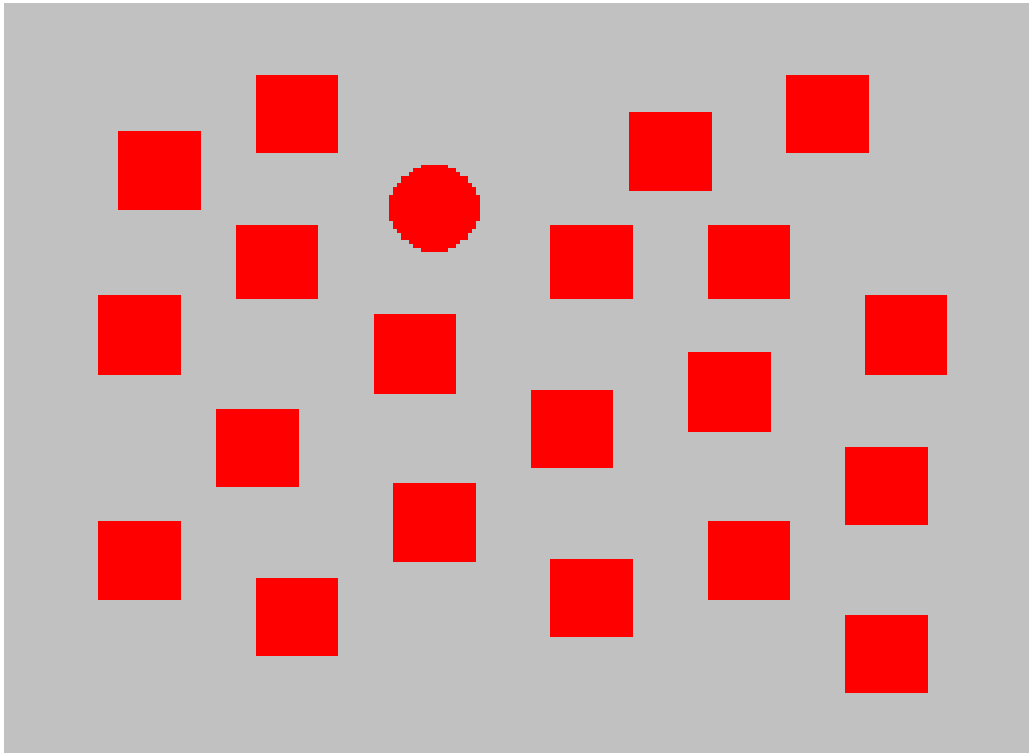
Principes de la visualisation d'information

- 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 ?

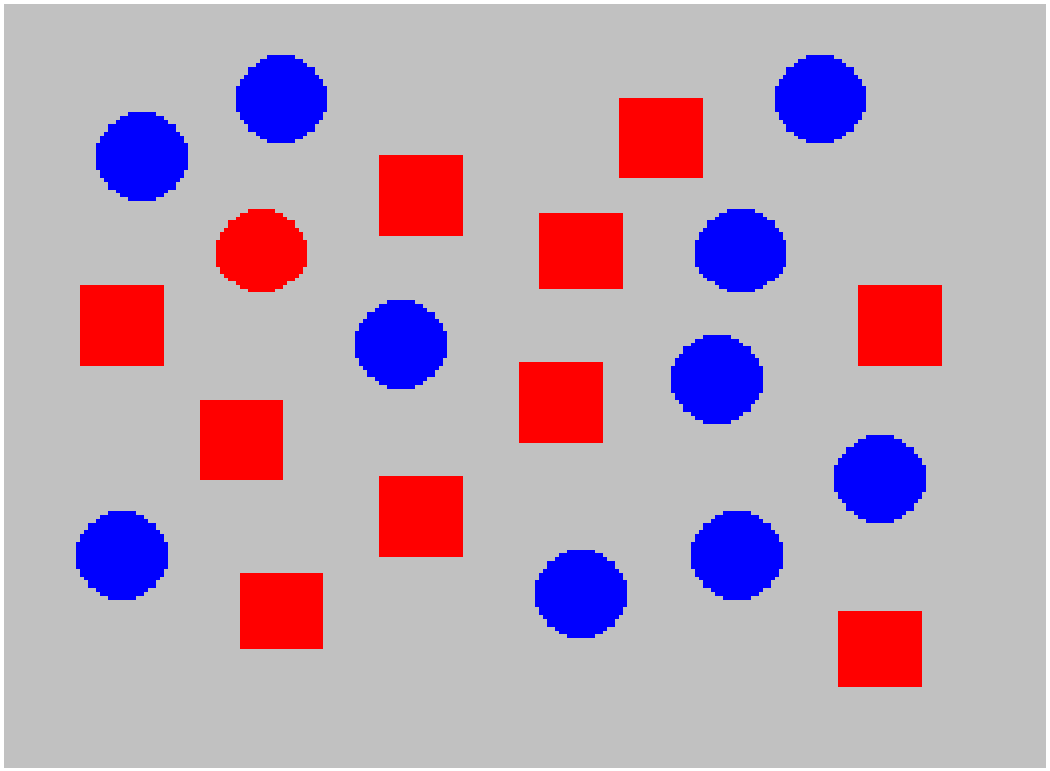
Perception préattentive (1)



Perception préattentive (2)



Perception préattentive (3)



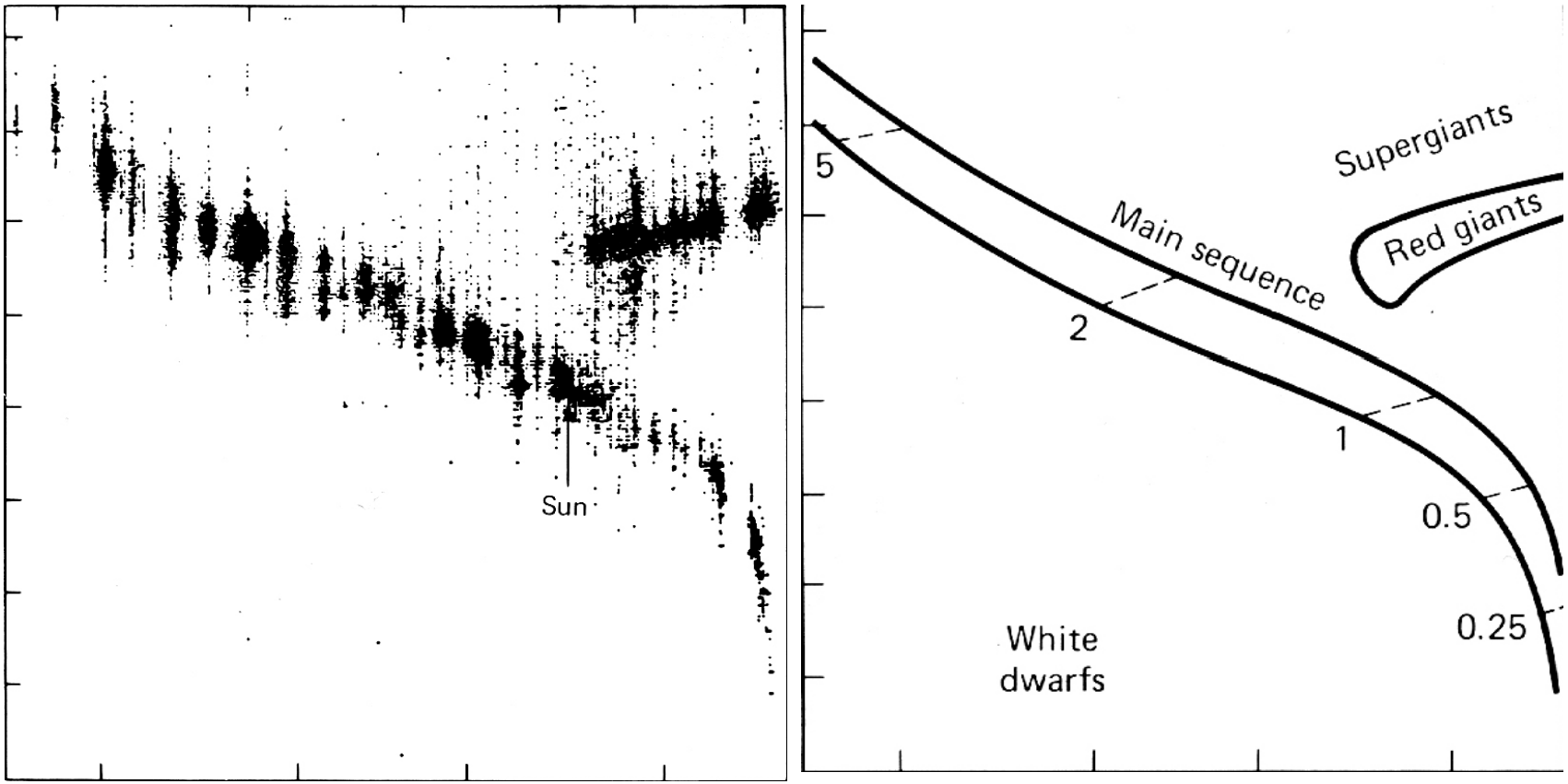
Perception préattentive : théorie

- 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
- Excellents 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



Visualisation d'information : Définition

Représentation graphique compacte ET

- Interface utilisateur pour
- Manipuler un grand nombre d'items (10^2 – 10^6)
- éventuellement extraits d'un ensemble de données beaucoup plus grand.

Permet aux utilisateurs de

- faire des découvertes
- prendre des décisions ou
- trouver des explications

Concernant des

- motifs visuels (tendances, clusters, points aberrants, …)
- des groupes d'items ou
- des items individuels.

Exemples de visualisations

- Série de valeurs
- Treemaps
- Matrices d’Adjacence
- Exécution d’un programme impératif
- Exécution d’un solveur de contraintes

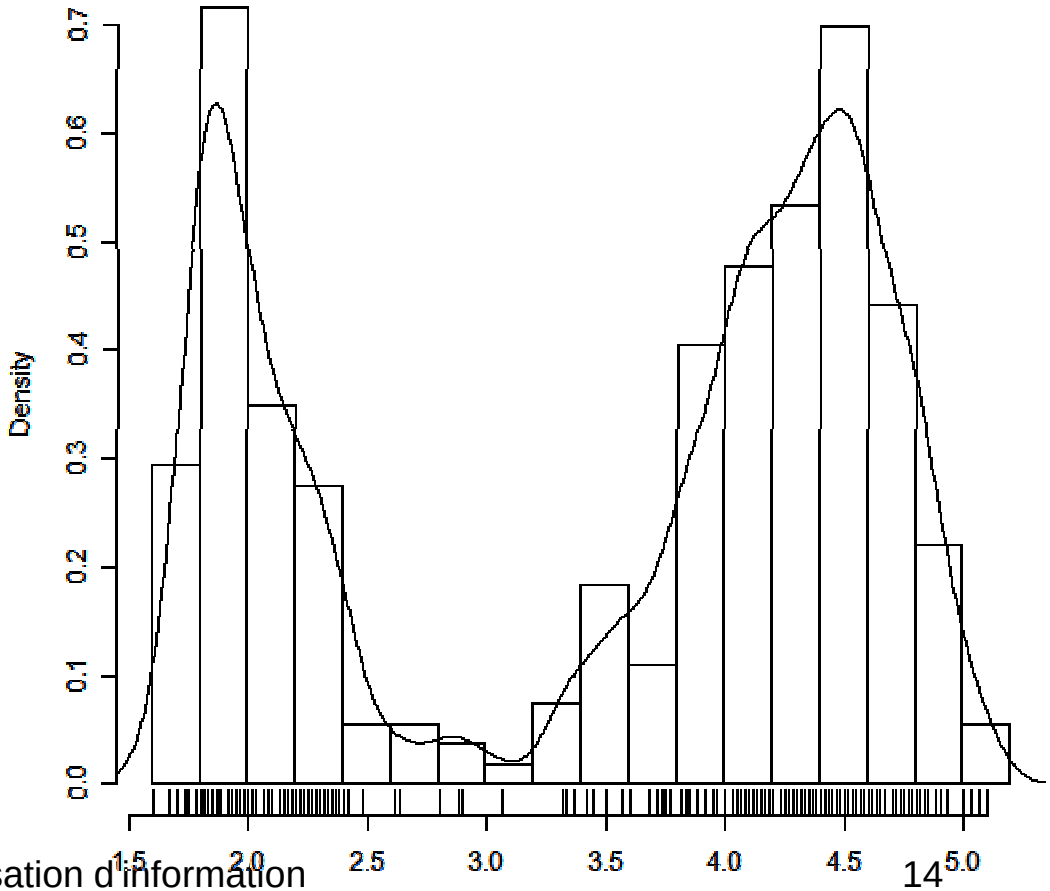
Stats traditionnelles vs. EDA

- Problème
- Données
- Modèle
- Analyse
- Conclusion

- Problème
- Données
- Analyse
- Modèle
- Validation ?
- Conclusion

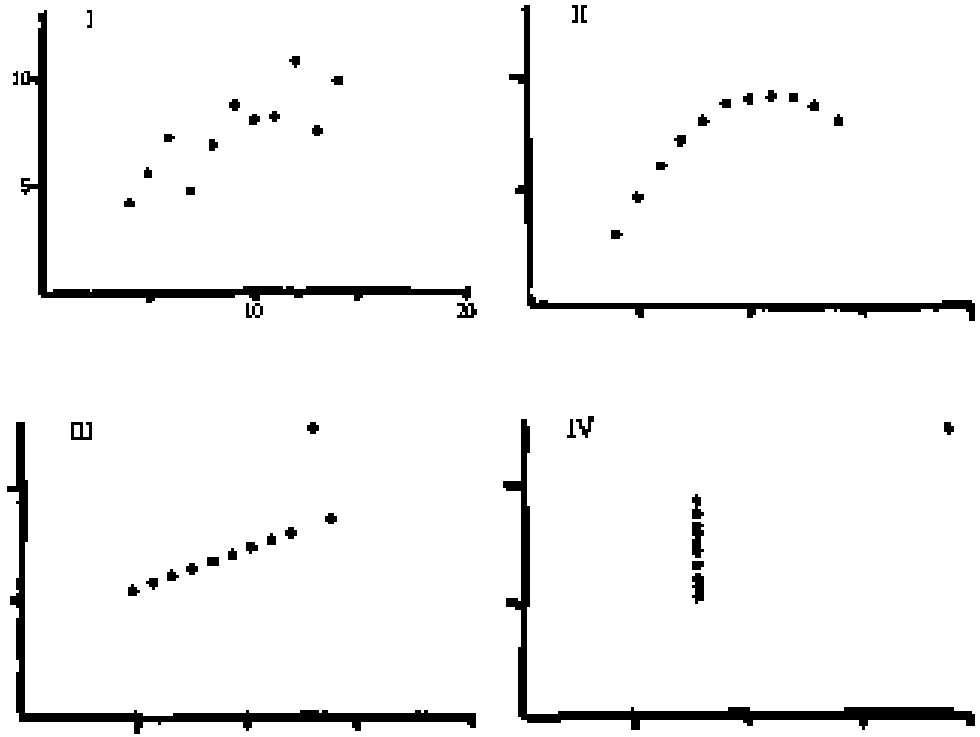
Série de valeurs

- Que contient ma série de valeurs ?
  - 272 valeurs
  - Minimum = 1.6
  - Maximum = 5.1
  - Moyenne = 3.488
  - Erreur standard = 1.141371
- Difficile à dire



Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



N:	11.0
Moyenne des X :	9.0
Moyenne des Y :	7.5
standard error of slope estimate:	0.1
sum of squares:	110.0
regression sum of squares:	27.5
residual sum of squares of Y:	13.8
correlation coefficient:	0.8
r squared:	0.7
regression line:	Y=3+0.5X

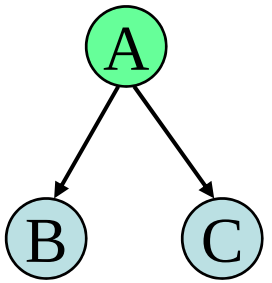
http://astro.swarthmore.edu/astro121/anscombe.html

F.J. Anscombe, "Graphs in Statistical Analysis," *American Statistician*, 27 [February 1973], 17-21

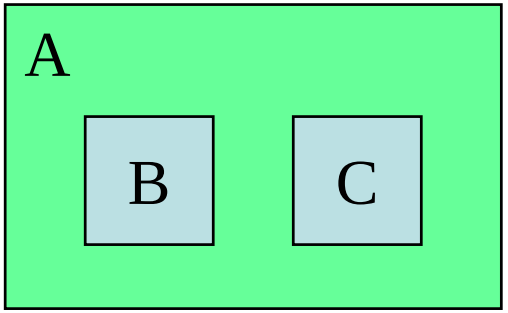
Treemaps

Visualiser les arbres : deux approches

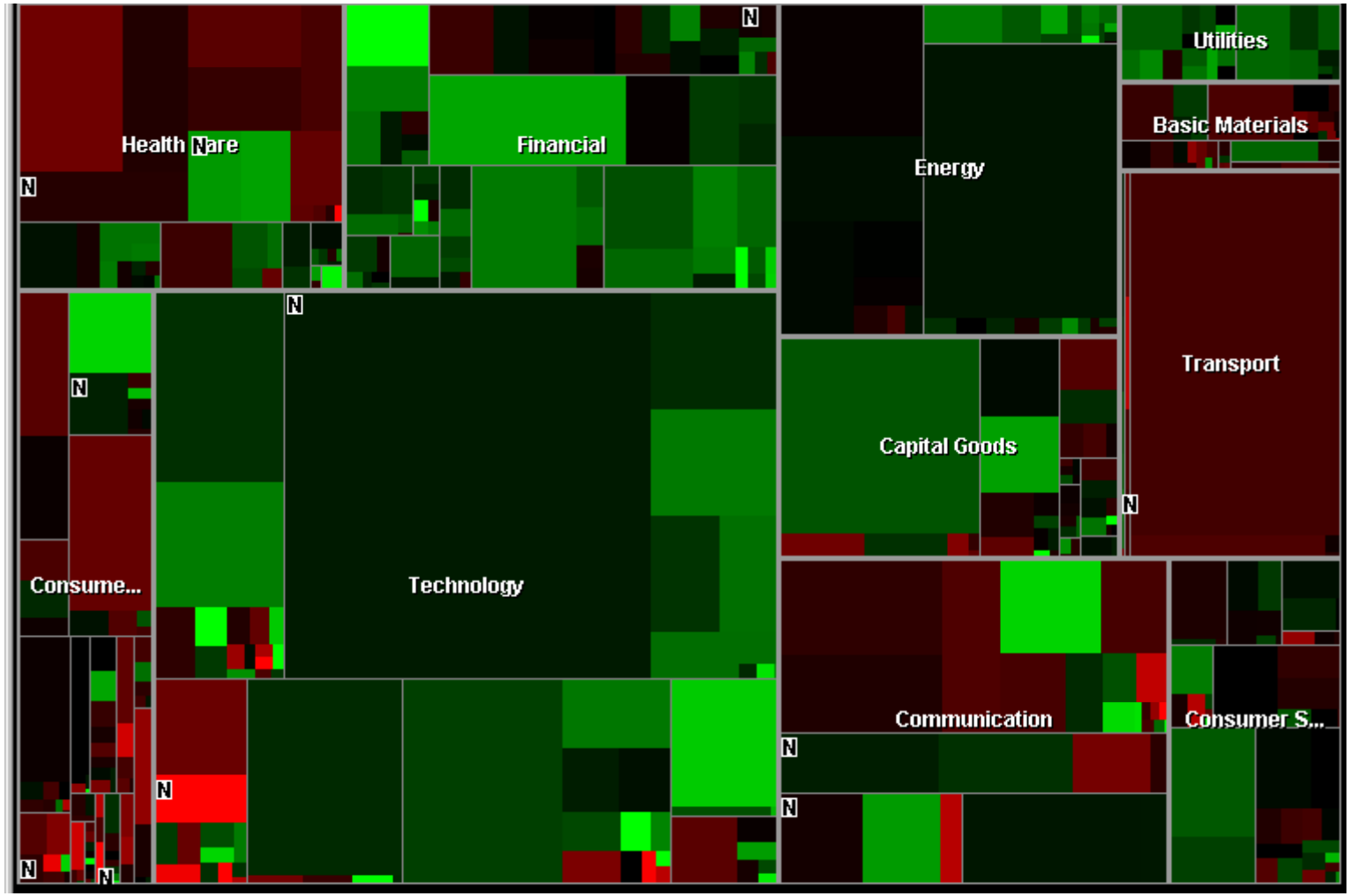
- Diagrammes nœud-lien
  - connexions



- Treemap
  - Diagramme de Venn

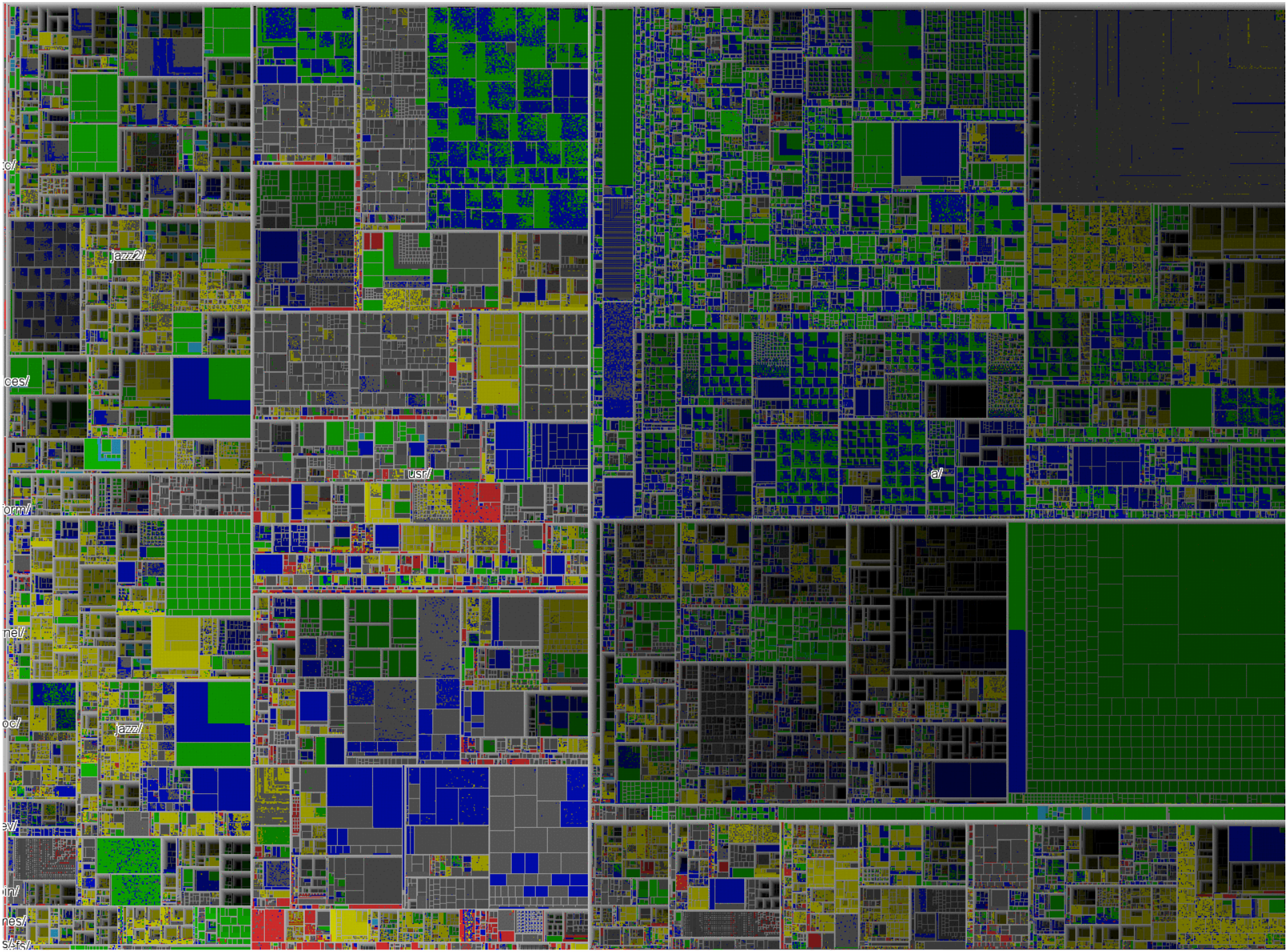


Treemap pour la bourse



Treemap d’un million de fichiers (Fekete&Plaisant 02)

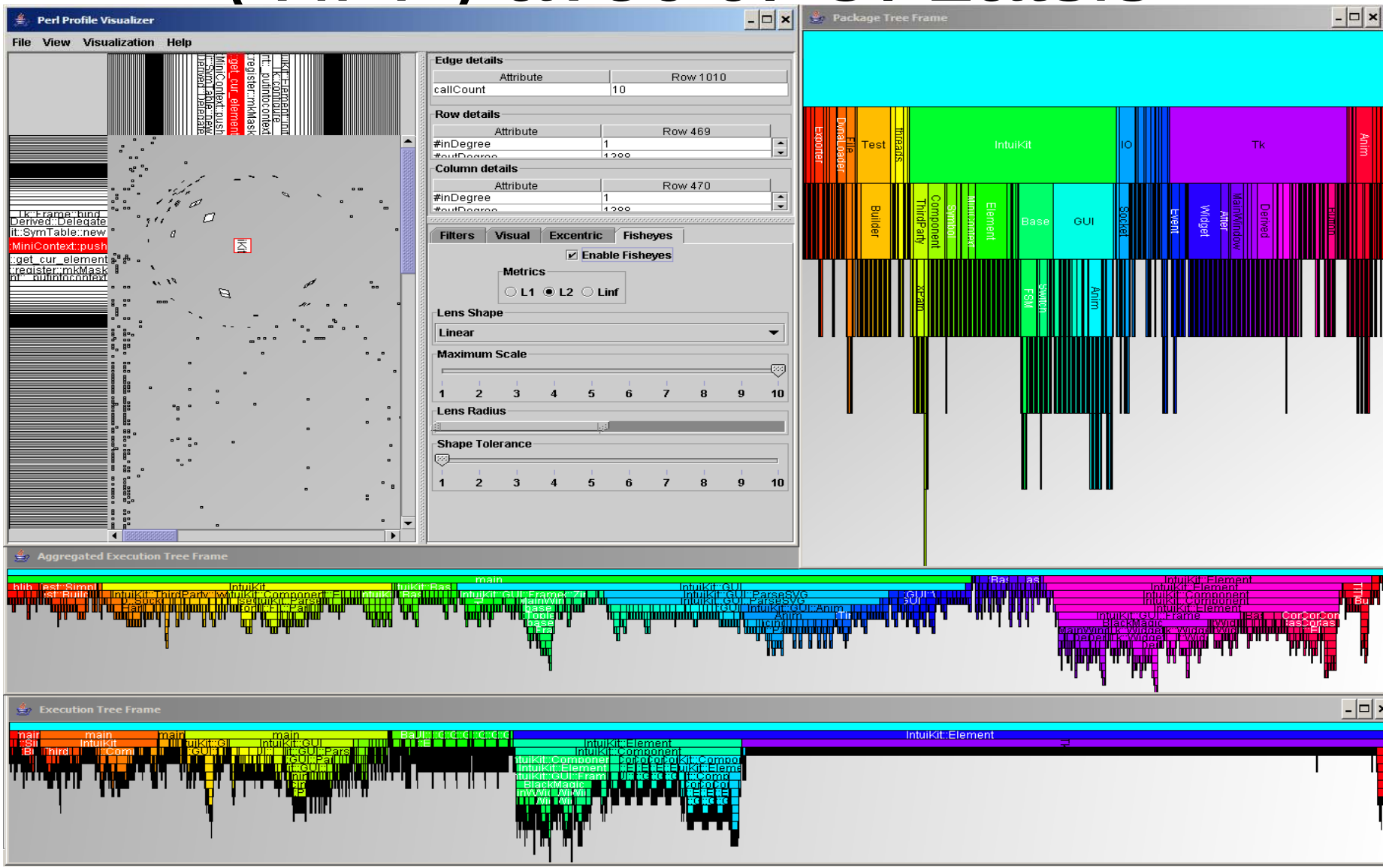
- Pour étudier les limites de la visualisation
- Limites techniques
  - Écran + cartes graphiques
- Limites perceptives
- Limites cognitives

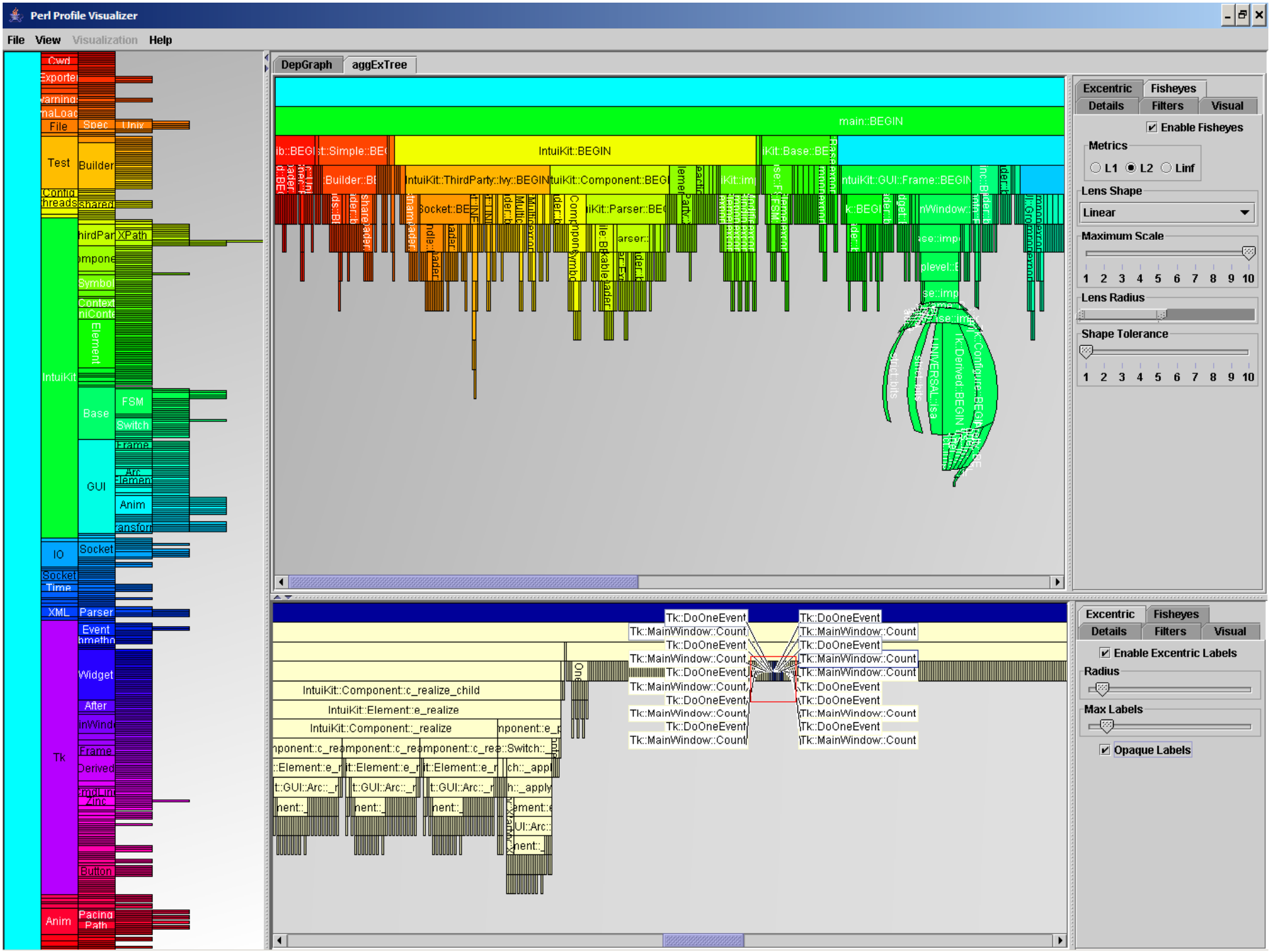


Speed: 5.917fps/0.169spf 971061 items

:root.xml.gz/a/kronos/raid/ 'a/users/hsh/src/ .ld/build.sh

Visualisation de profilage Perl (ViPP) avec J.-C. Latsis



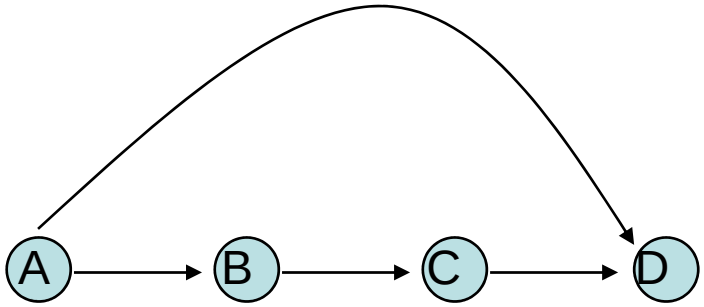

Perl Profile Visualizer
File View Visualization Help
DepGraph aggExTree
main::BEGIN
IntuiKit::BEGIN
IntuiKit::ThirdParty::Iwy::BEGIN
IntuiKit::Component::BEGIN
IntuiKit::Parser::BEGIN
IntuiKit::GUI::Frame::BEGIN
Excentric Fisheyes
Details Filters Visual
Enable Fisheyes
Metrics
L1 L2 Linf
Lens Shape
Linear
Maximum Scale
1 2 3 4 5 6 7 8 9 10
Lens Radius
Shape Tolerance
1 2 3 4 5 6 7 8 9 10
Tk::DoOneEvent
Tk::MainWindow::Count
IntuiKit::Component::c_realize_child
IntuiKit::Element::e_realize
IntuiKit::Component::_realize
Enable Excentric Labels
Radius
Max Labels
Opaque Labels

Graphes et Matrices d'Adjacence avec M. Ghoniem puis N. Henry

- Les graphes sont le plus souvent visualisés sous forme de diagrammes nœud et liens
- Est-ce toujours la meilleure représentation ?
  - Ghoniem et al. ont montrés que NON pour les graphes denses ou grands

Construction de la matrice

Graphe orienté



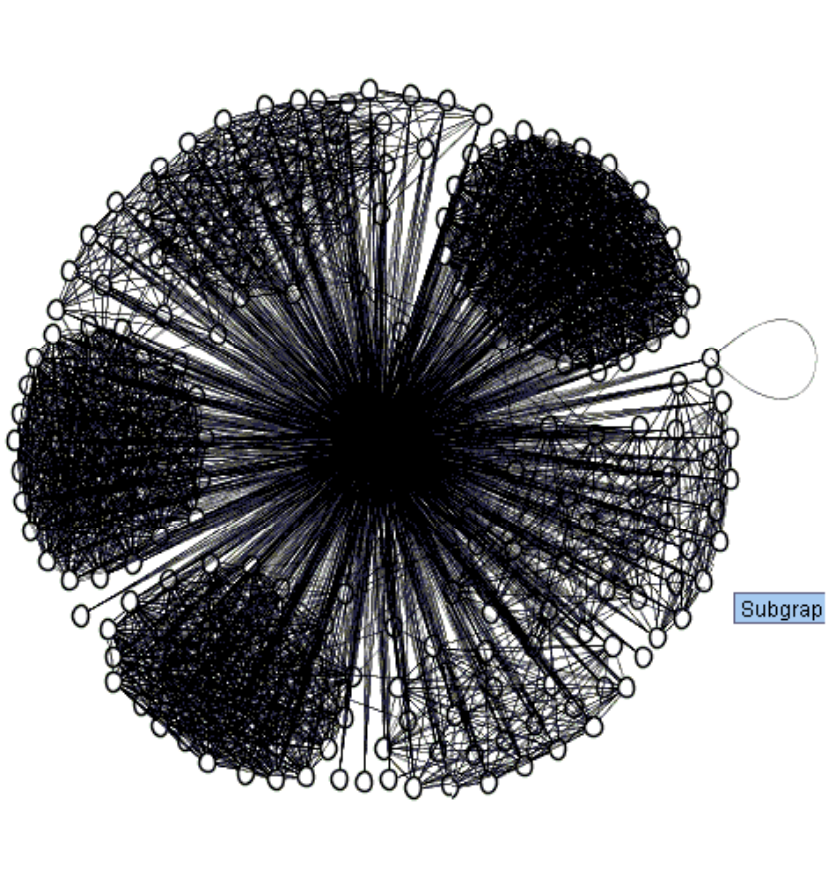
Matrice d'adjacence

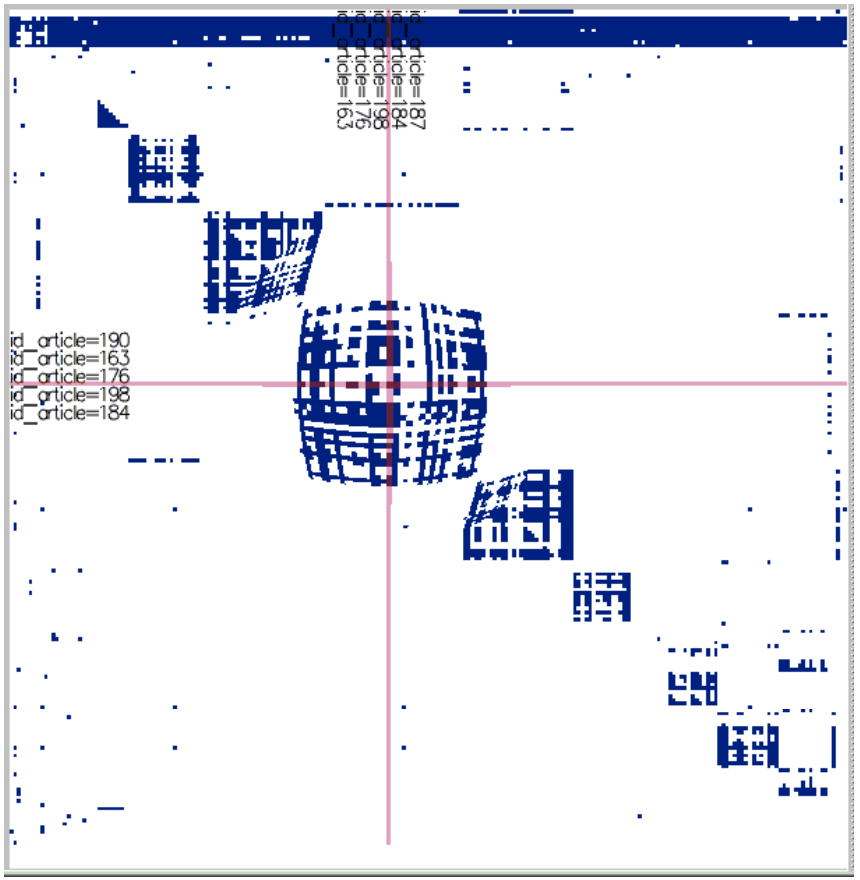
	A	B	C	D
A	0	1	0	1
B	0	0	1	0
C	0	0	0	1
D	0	0	0	0

arrivée

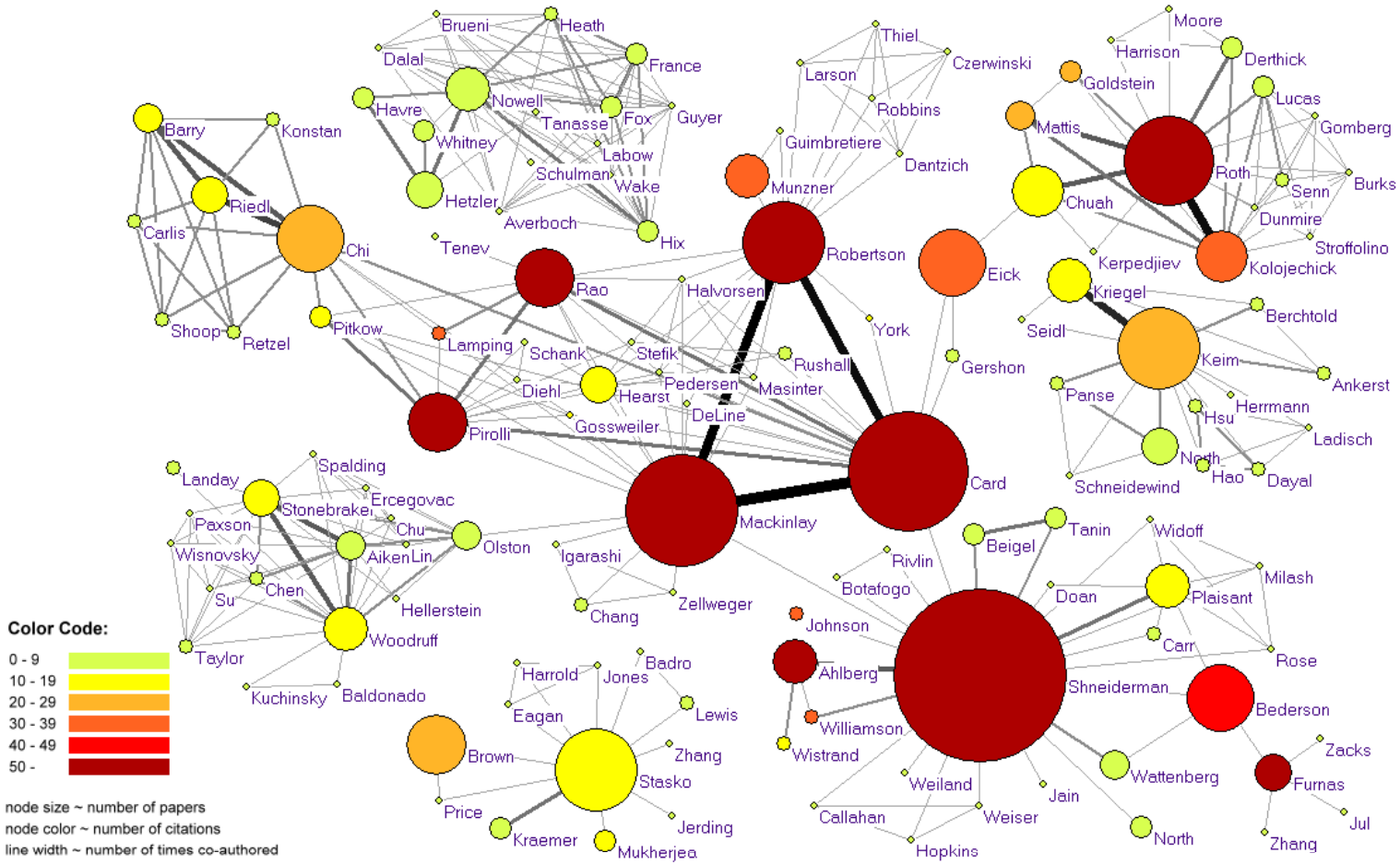
départ

Exemple de site Web



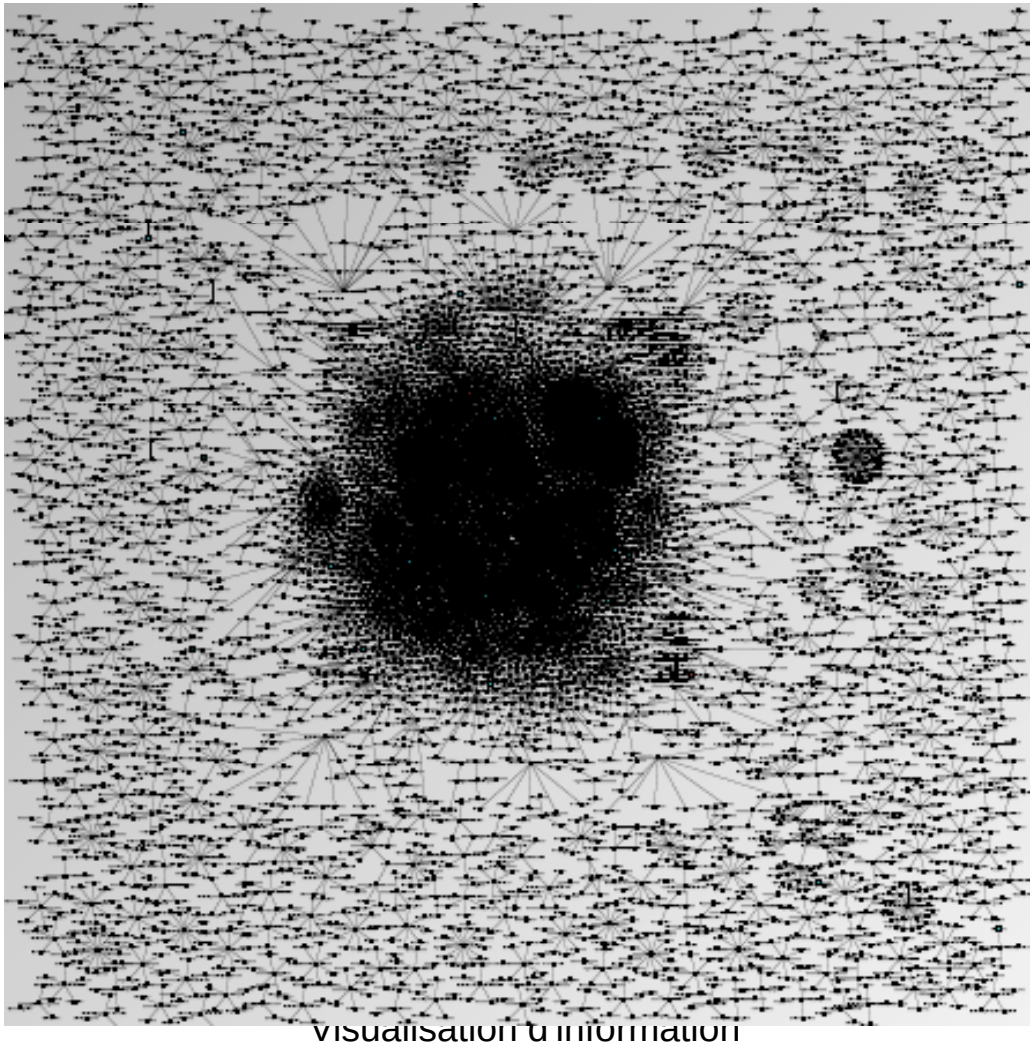


Réseaux sociaux avec N. Henry



InfoVis contest 2004 (K. Börner et al.)

Un réseau social en histoire… (N. Dufournaud)



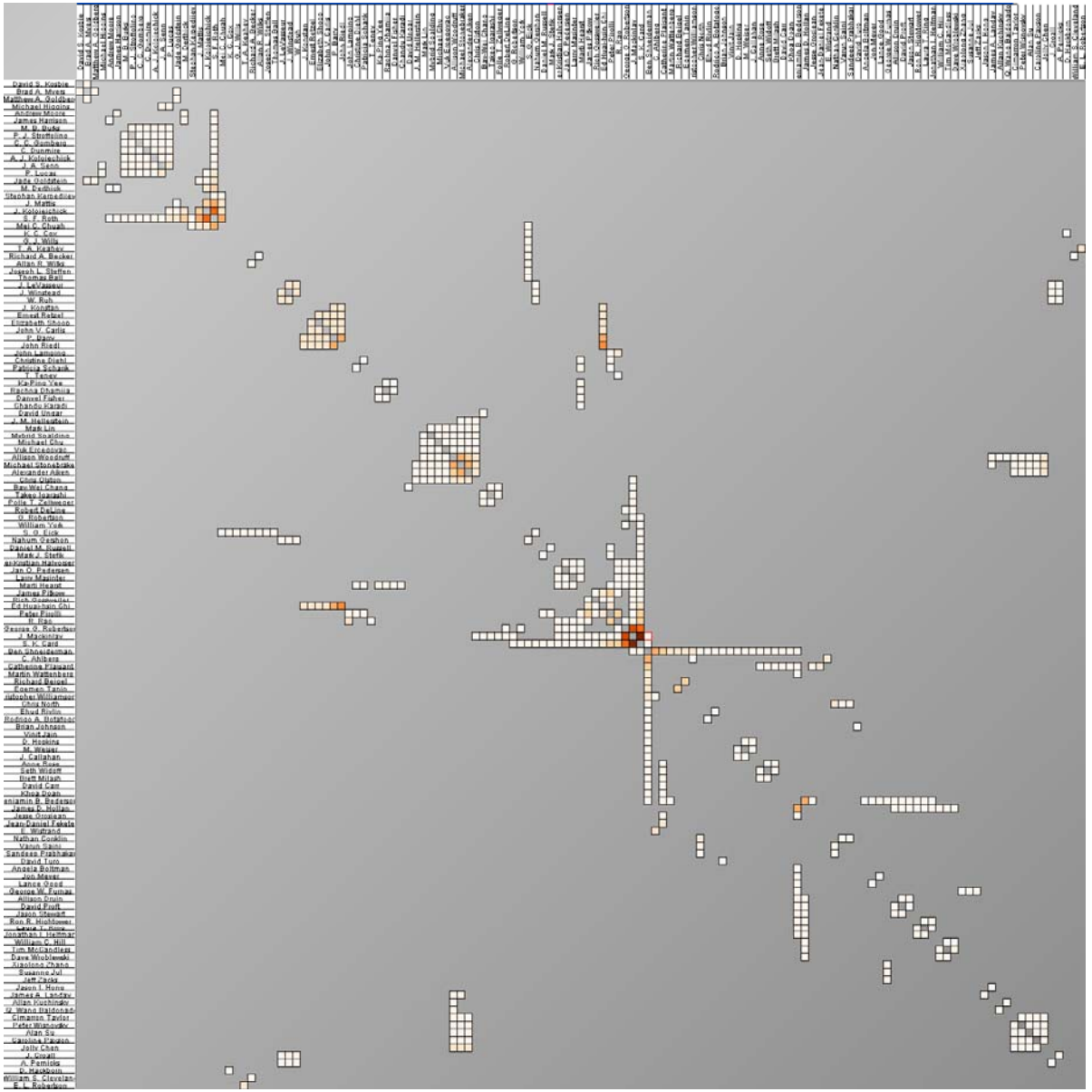
Visualisation Matricielle

Composante Principale

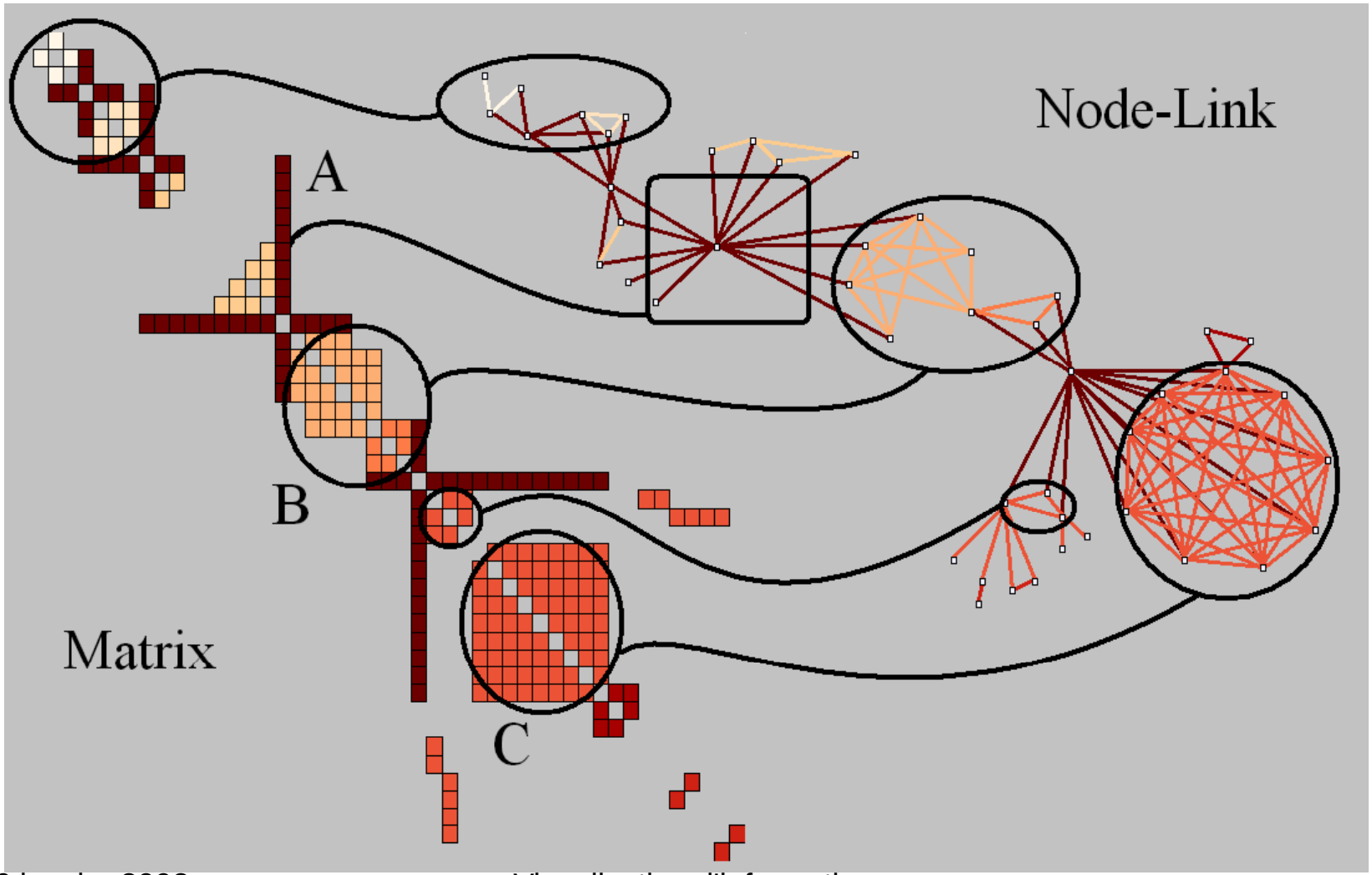
InfoVis

135 auteurs

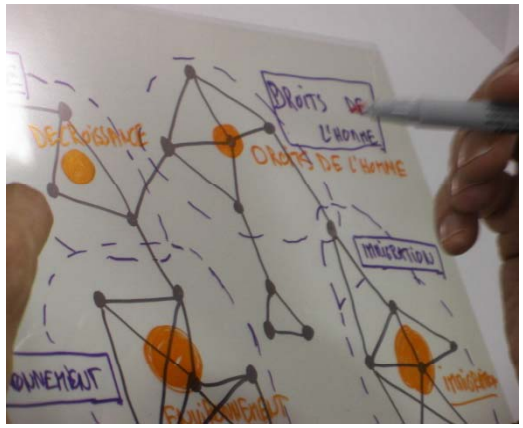
6 janvi

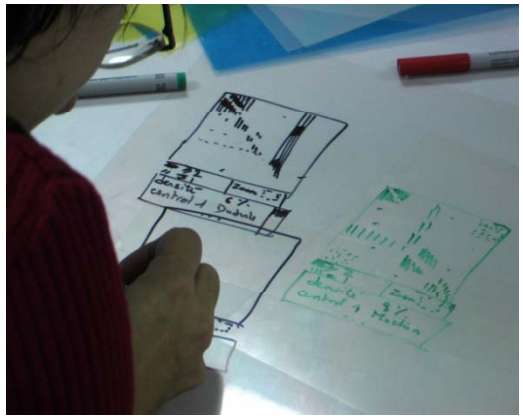


Motifs visuels



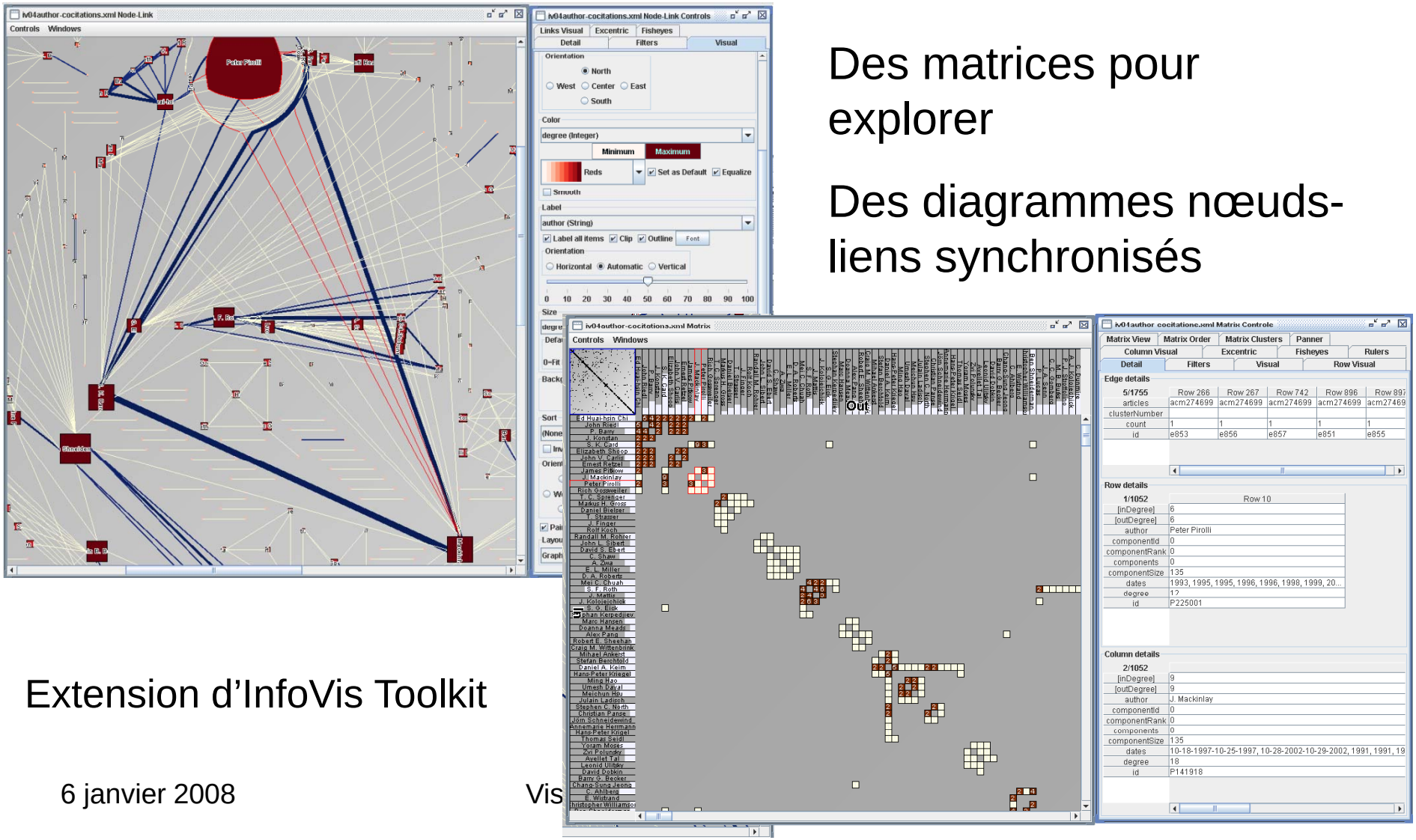
MatrixExplorer





- Explorer visuellement
- Interagir
- Conception Participative avec des historiens et sociologues

MatrixExplorer

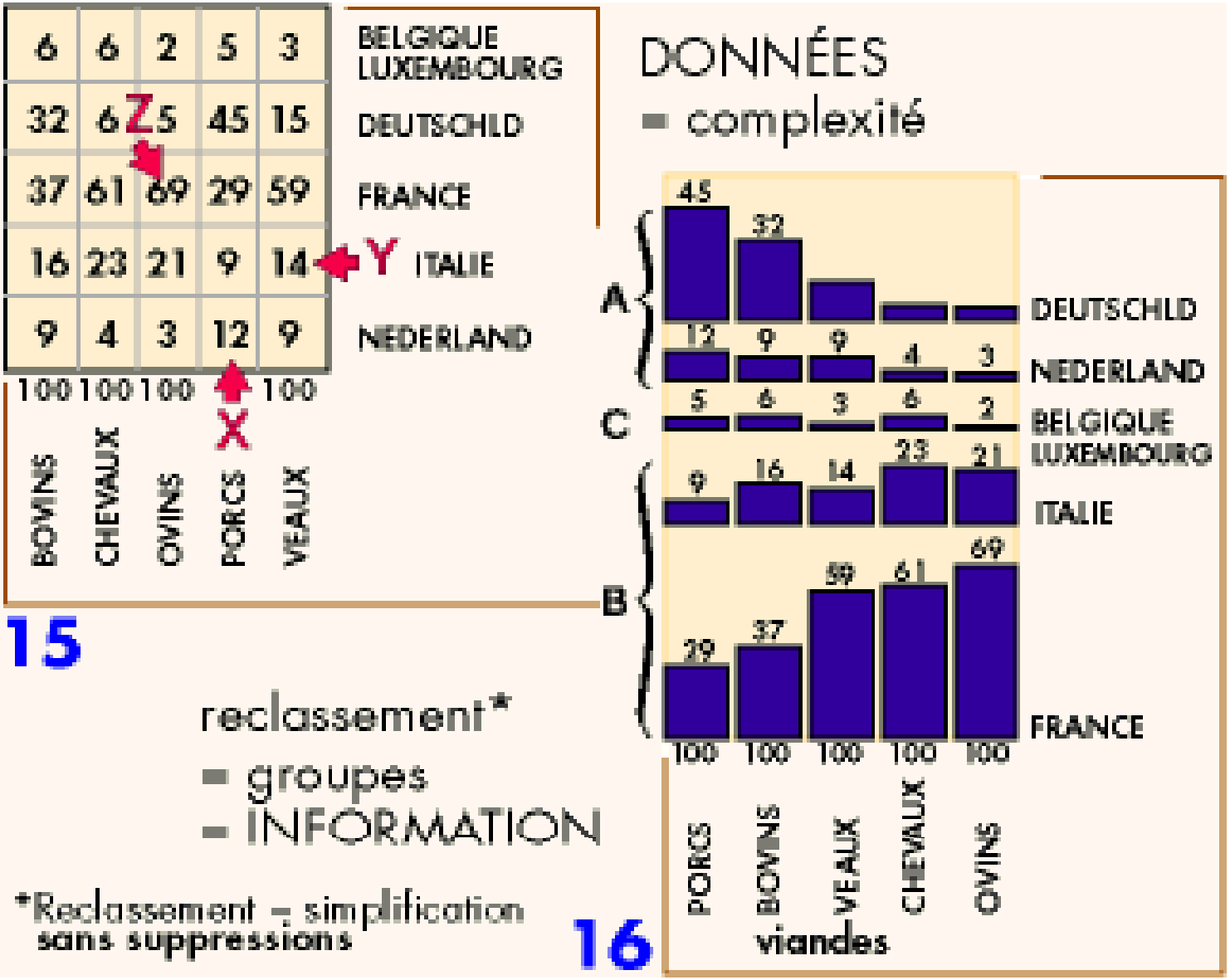


Des matrices pour explorer

Des diagrammes nœuds-liens synchronisés

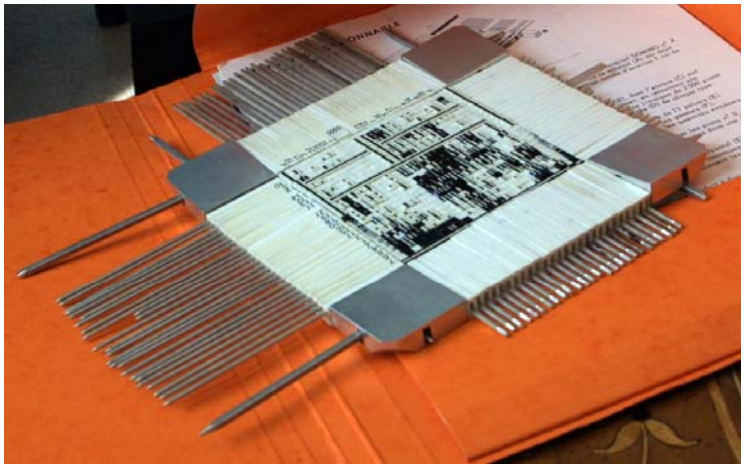
Extension d'InfoVis Toolkit

Réordonner pour comprendre



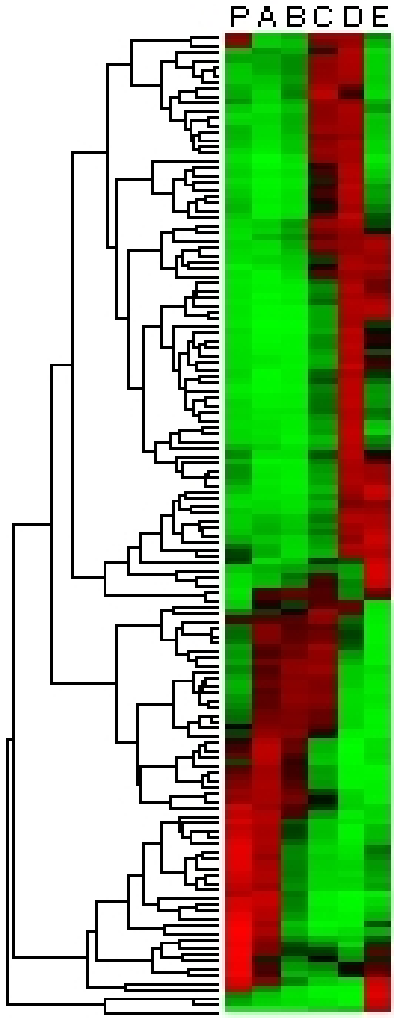
Réordonner une matrice

- Algorithmes automatiques
  - Optimisation d’une fonction objectif
    - Problème des 1 consécutifs
    - Sériation de graphes
  - Heuristiques
  - Block modeling ou partitionning
- Interaction



Réordonner une matrice

- Méthodes issues de bio-informatiques
  - Clustering Hiérarchique + Sériation
  - Voyageur de commerce
- Minimum de paramétrage
- Rapides et robustes
- Appliquées sur des tables d’expression de gènes



PermutMatrix

Matrice d’adjacence

- Les matrices d’adjacence de réseaux sociaux sont creuses
- Utiliser la matrice des distances (plus court chemins)

Calcul composantes connexes

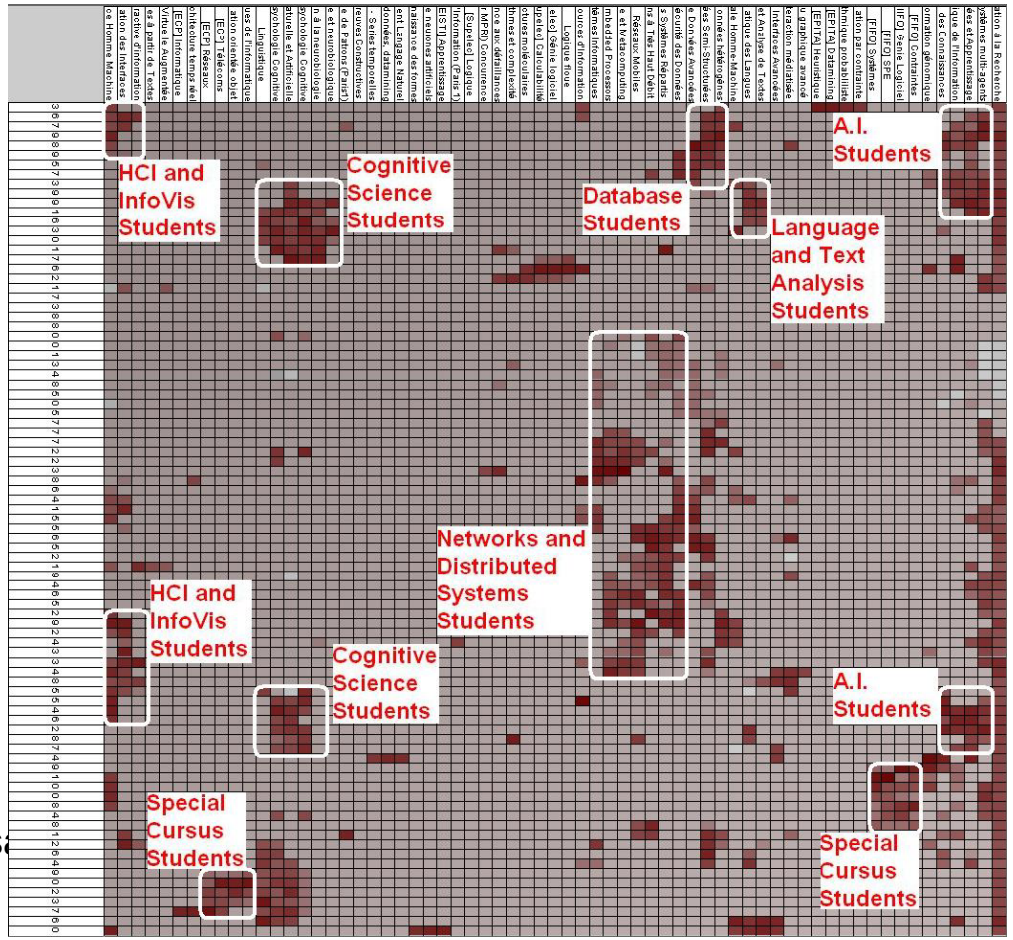
Pour chacune

Matrice des distances

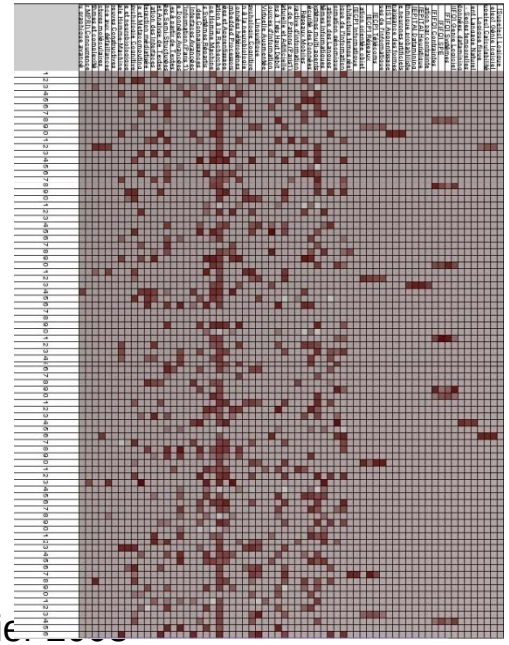
Algorithme de réordonnancement

Resultats?

TSP Order

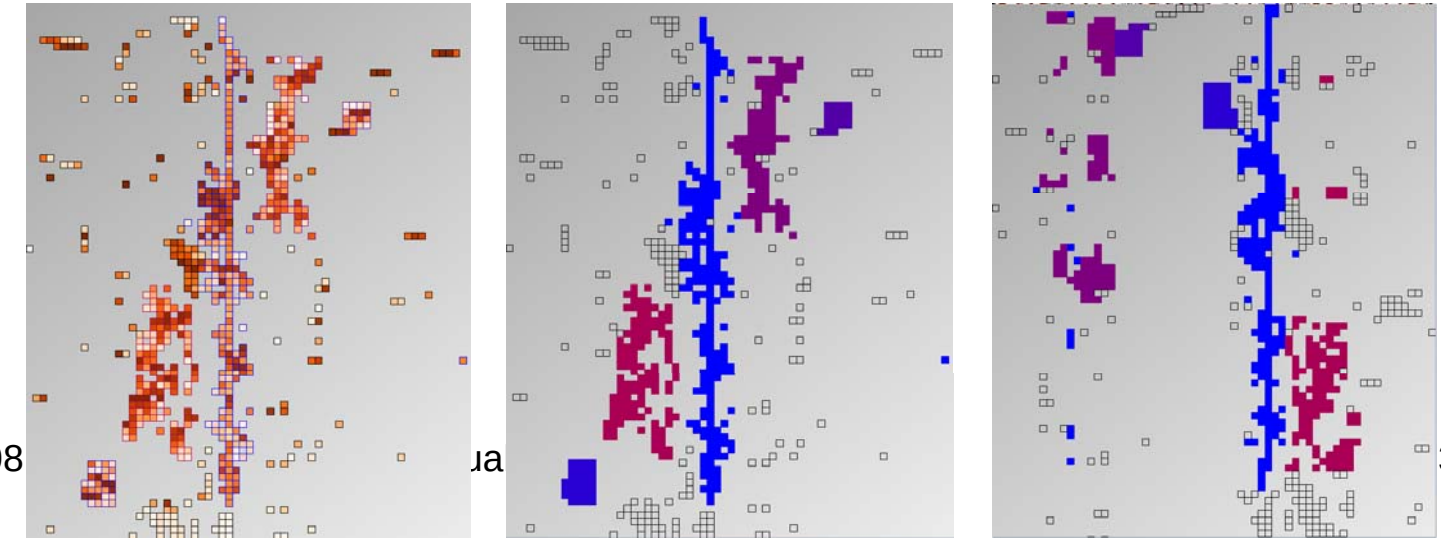


Alphabetical Order



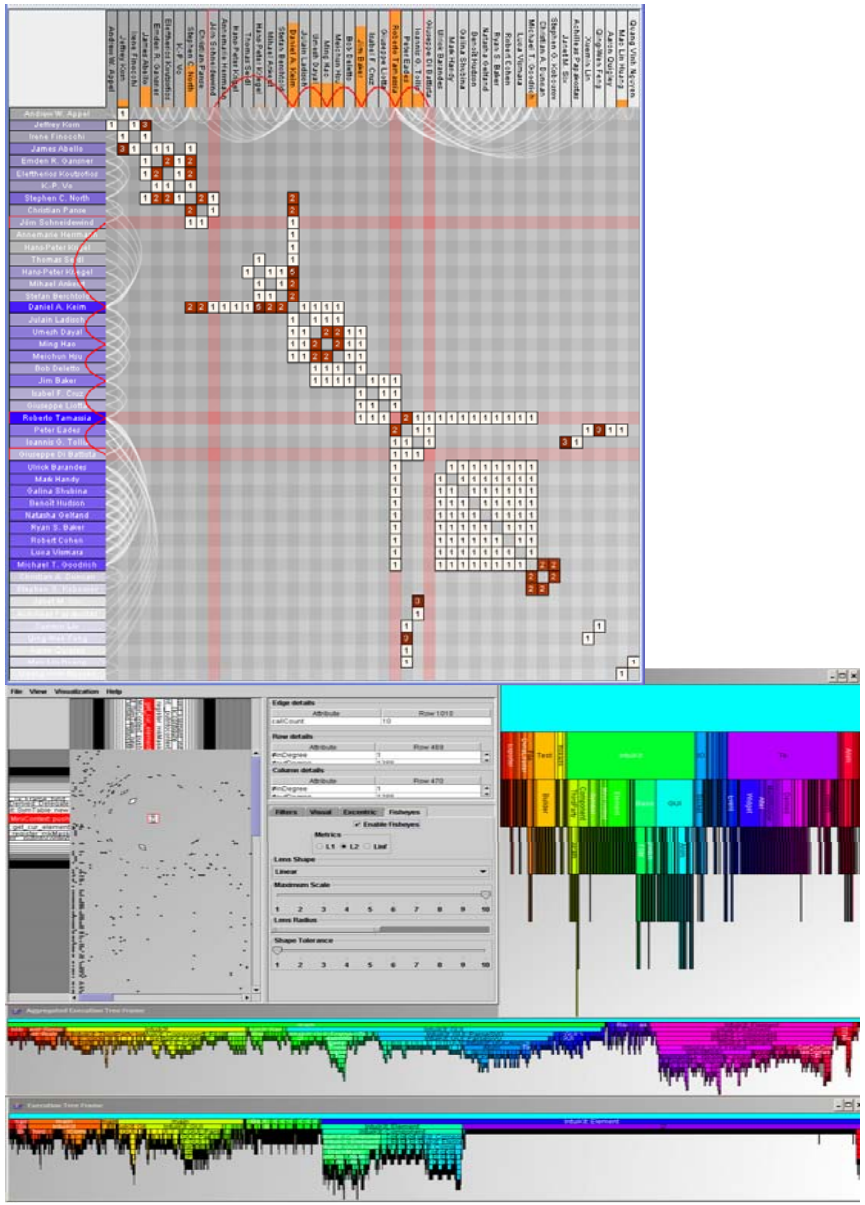
Confronter, Trouver un consensus

- Processus
  - Ordonner
  - Grouper
  - Réordonner
  - Trouver les invariants

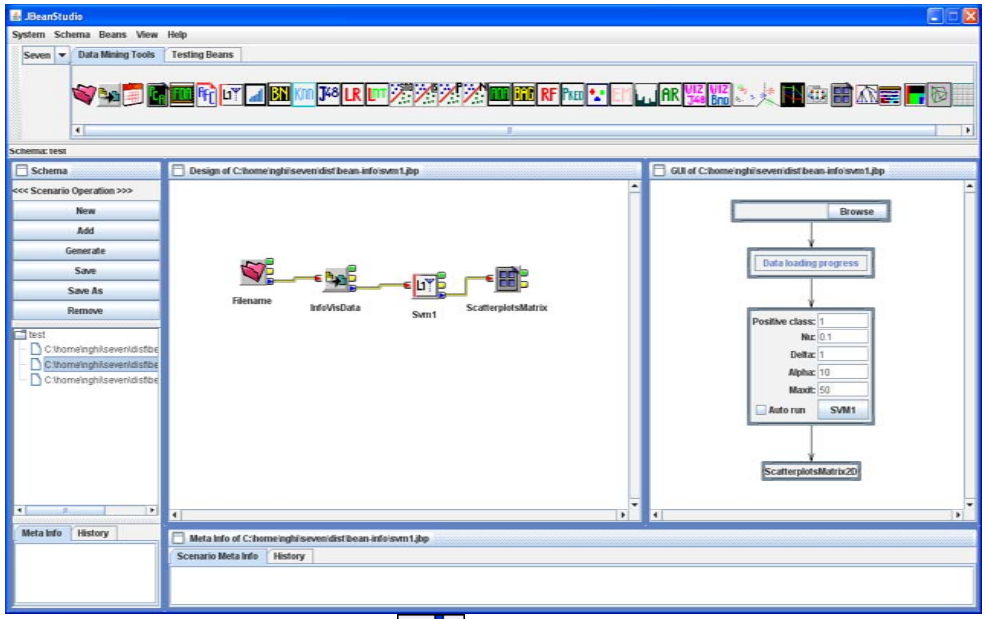


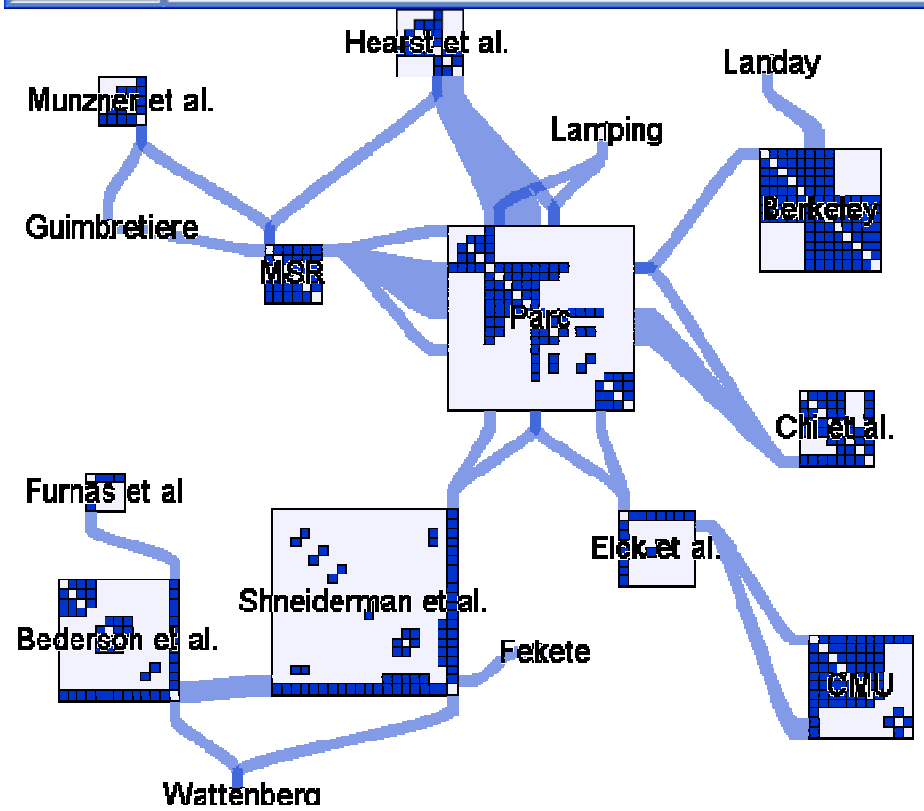
Multi-scale representations

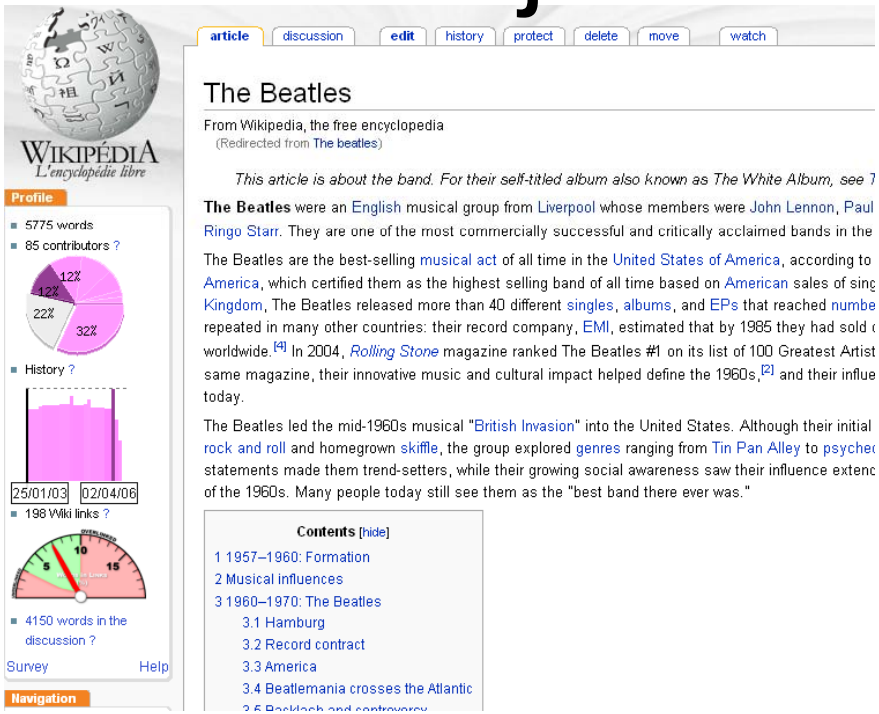
- Multi-scale representations of Large Networks
- Multi-scale representations of Time-Series and Logs

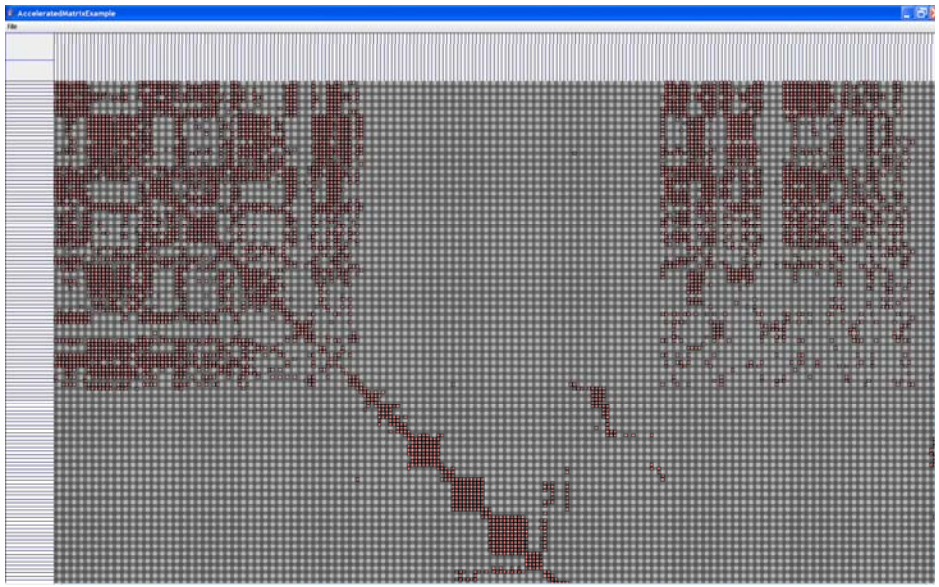


Projects


JBeanStudio
System Schema Beans View Help
Data Mining Tools
Testing Beans
Schema
Design of C:\home\nghi\seven\dist\bean-info\svm1.jbp
GUI of C:\home\nghi\seven\dist\bean-info\svm1.jbp
Filename
InfoVisData
Svm1
ScatterplotsMatrix
Browse
Data loading progress
Positive class:
Auto run
SVM1
ScatterplotsMatrix20
Meta Info
History
Scenario Meta Info

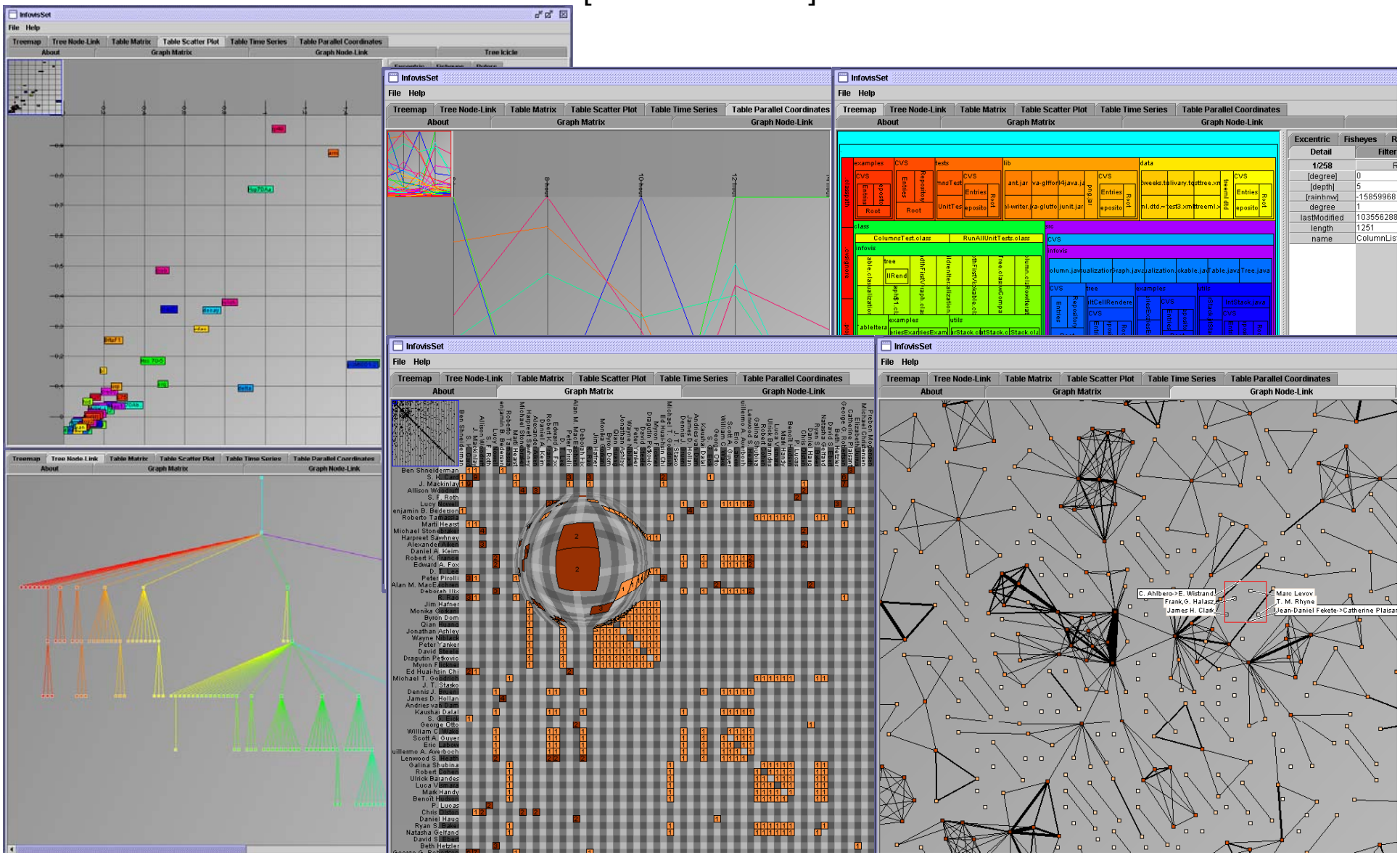

Hearst et al.
Landay
Munzner et al.
Lamping
Guimbretiere
Berkeley
MSR
Parc
Chi et al.
Furnas et al
Elek et al.
Shneiderman et al.
Bederson et al.
Fekete
CMU
Wattenberg


article
discussion
edit
history
protect
delete
move
watch
The Beatles
From Wikipedia, the free encyclopedia
(Redirected from The beatles)
This article is about the band. For their self-titled album also known as The White Album, see T
The Beatles were an English musical group from Liverpool whose members were John Lennon, Paul
Ringo Starr. They are one of the most commercially successful and critically acclaimed bands in the
The Beatles are the best-selling musical act of all time in the United States of America, according to
America, which certified them as the highest selling band of all time based on American sales of sing
Kingdom, The Beatles released more than 40 different singles, albums, and EPs that reached numbe
repeated in many other countries: their record company, EMI, estimated that by 1985 they had sold
worldwide.[4] In 2004, Rolling Stone magazine ranked The Beatles #1 on its list of 100 Greatest Artist
same magazine, their innovative music and cultural impact helped define the 1960s,[2] and their influe
today.
The Beatles led the mid-1960s musical "British Invasion" into the United States. Although their initial
rock and roll and homegrown skiffle, the group explored genres ranging from Tin Pan Alley to psyche
statements made them trend-setters, while their growing social awareness saw their influence extend
of the 1960s. Many people today still see them as the "best band there ever was."
Contents [hide]
1 1957–1960: Formation
2 Musical influences
3 1960–1970: The Beatles
3.1 Hamburg
3.2 Record contract
3.3 America
3.4 Beatlemania crosses the Atlantic
Profile
5775 words
85 contributors ?
12%
12%
22%
32%
History ?
25/01/03
02/04/06
198 Wiki links ?
4150 words in the discussion ?
Survey
Help
Navigation



InfoVis Toolkit (ivtk.sf.net)

[IEEE InfoVis 04]



Constraints Programming

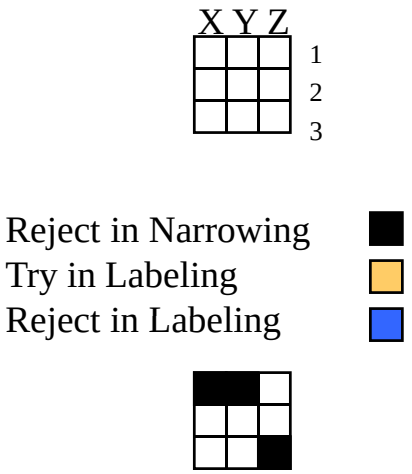
- Declarative programming where you describe the properties of the solutions
- The resolution is carried out by a solver
- Very useful for industrial problems involving optimization, scheduling, resource allocation, sequencing, timetabling…

Constraints Programming

- A few languages implement this paradigm:
  - Prolog IV (PrologIA), Chips Prolog (Cosytec), GNU Prolog (INRIA)
  - ILOG Solver (ILOG)
  - Choco (Bouygues)
  - PaLM (EMN)
- Live research topic and active industrial field

Example of resolution (1/3)

- 3 Phases:
  - **declaration**, **narrowing** and **labeling**
- Declaration
  - 3 variables : $X \in [1..3]$, $Y \in [1..3]$, $Z \in [1..3]$
  - 3 constraints : c_1: $X \neq Y$; c_2: $X \geq Y$; c_3: $Y > Z$
- Narrowing Phase
  - $c_3 \Rightarrow Y \neq 1$ and $Z \neq 3$
  - $c_2 \Rightarrow X \neq 1$

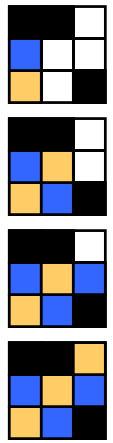


Example of Resolution (2/3)

- Declarations (reminder)
  - 3 variables : $X \in [2..3]$, $Y \in [2..3]$, $Z \in [1..2]$
  - 3 constraints : c_1: $X \neq Y$; c_2: $X \geq Y$; c_3: $Y > Z$
- Labeling Phase
  - Try: $X = 2$
    - c_1 is valid ; $c_2 \Rightarrow Y \neq 3$
  - $X = 2$ and $Y = 2$ violates c_1 : Failure
  - Backtrack over X=2: Try $X = 3$

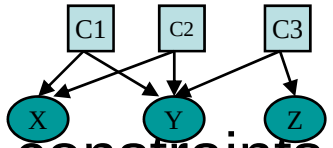
Example of Resolution (3/3)

- Declarations (reminder)
  - 3 variables : $X \in [2..3]$, $Y \in [2..3]$, $Z \in [1..2]$
  - 3 constraints : c_1: $X \neq Y$; c_2: $X \geq Y$; c_3: $Y > Z$
- Labeling (continued)
  - Try : $X = 3$
  - $c_1 \Rightarrow Y \neq 3$ I.e. $Y = 2$; wake up c_3
  - $c_3 \Rightarrow Z \neq 2$
  - $X = 3$; $Y = 2$; $Z = 1$ is a solution of the problem.



Visualizing co-Activity Graphs

- Networks of entities interacting with one another
  - The network of constraints
  - The network of variables
  - The mixed network of variables and constraints



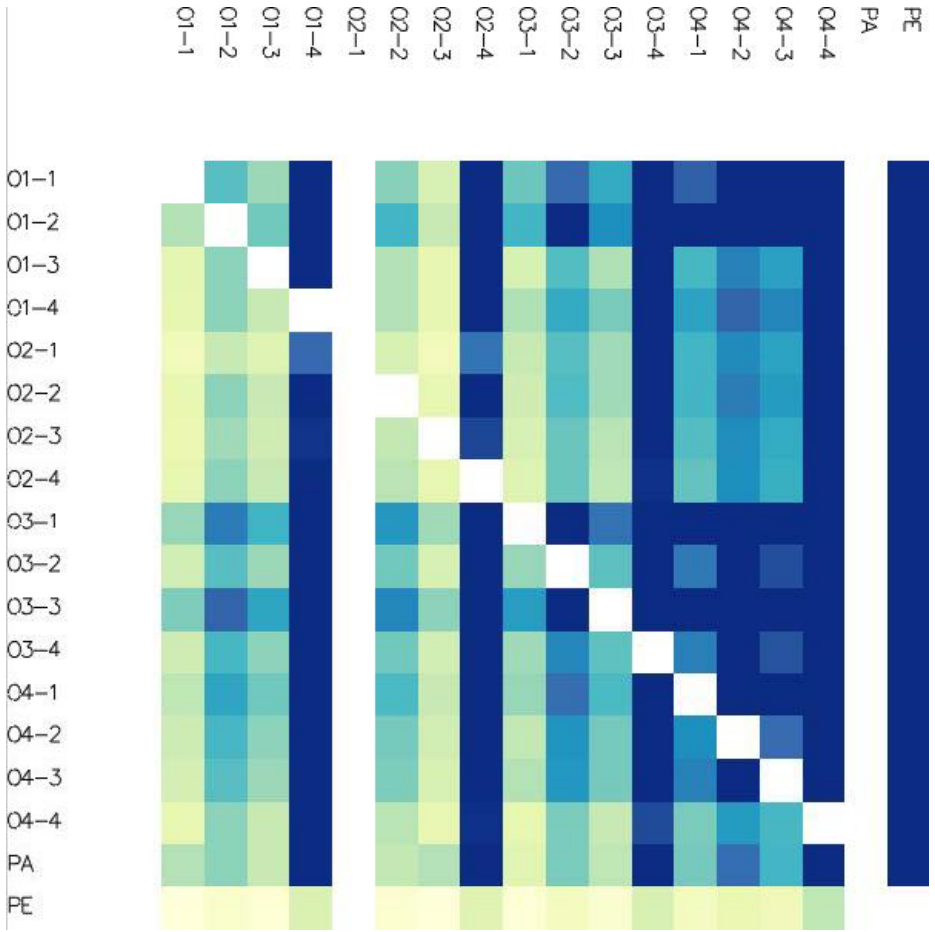
- Edge weight / color reflects activity
- Main features
  - These networks are potentially large and dense
  - These networks evolve over time
    - Edges may be added or removed
    - Edge weight i.e. activity varies over time

Visualizing co-Activity Graphs

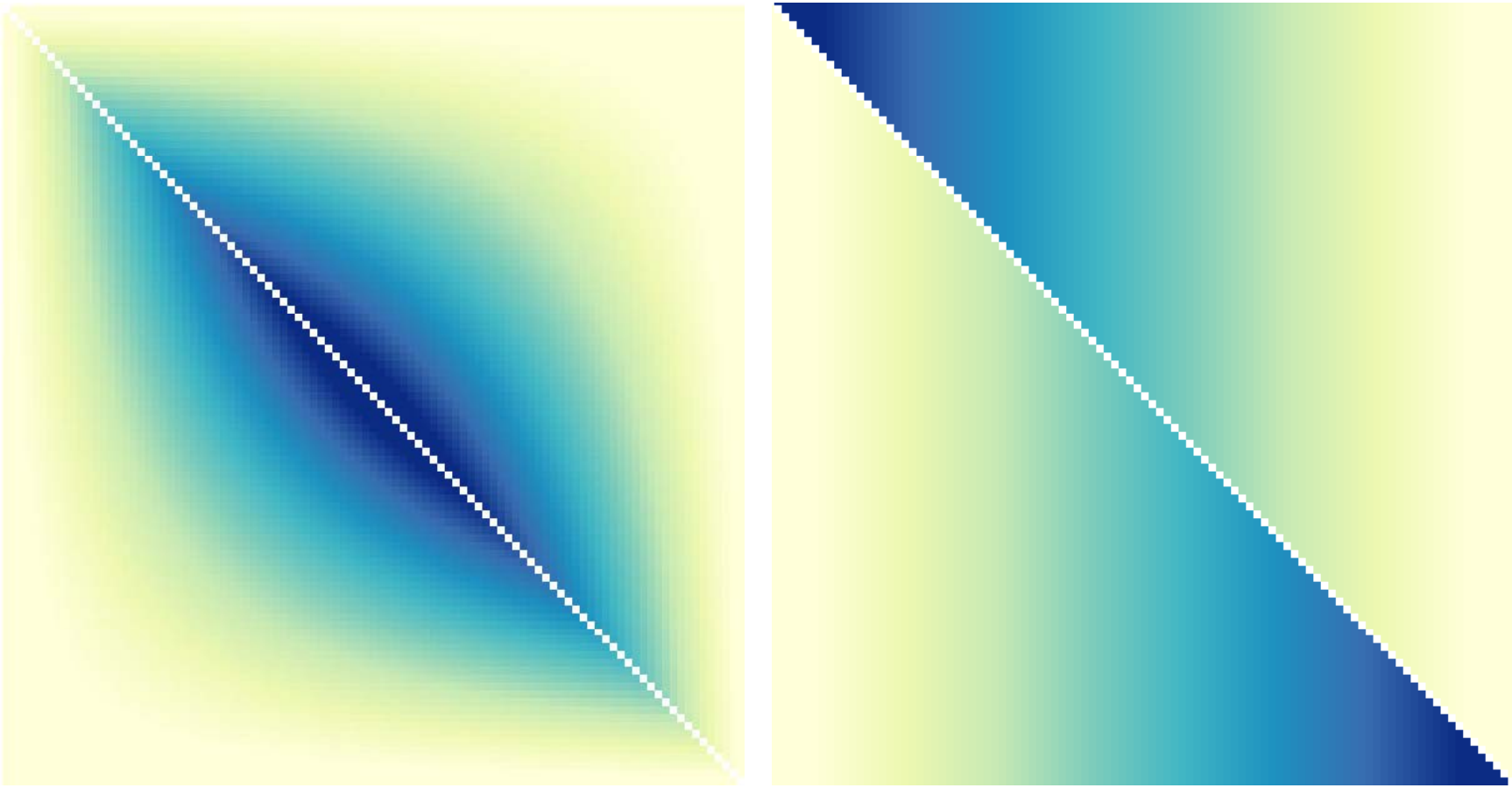
- Matrix-based visualizations are more readable than node-link diagrams
  - Clear overview
  - Orderable
  - Scales well
  - Takes advantage of its duality

Animation of Activity Graphs

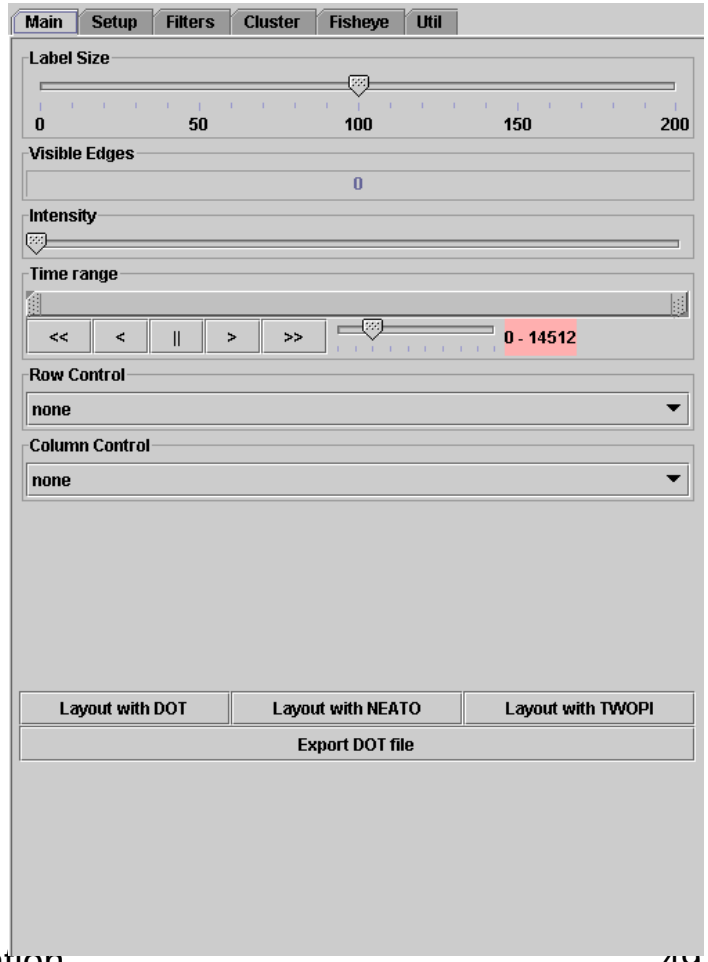
- Color coding of activity
- Full history on edges
- Time filtering

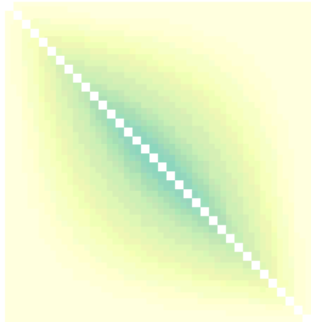


Simple Problems: Sorting



Animation





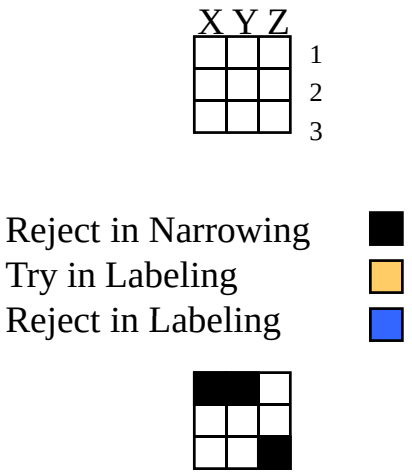
MG - NJ - JDF

Explanation Based Activity

- An explanation contains a set of constraints that justify a decision taken by the solver
- An explanation involves directly a subset of the constraints, and indirectly a subset of the variables
- Throughout the resolution, the sequence of explanations provides information about the real-time behavior of the solver.

Example of explanations (1/3)

- 3 Phases:
  - **declaration**, **narrowing** and **labeling**
- Declaration
  - 3 variables : $X \in [1..3]$, $Y \in [1..3]$, $Z \in [1..3]$
  - 3 constraints : c_1: $X \neq Y$; c_2: $X \geq Y$; c_3: $Y > Z$
- Narrowing Phase
  - $c_3 \Rightarrow Y \neq 1$ and $Z \neq 3$
  - $c_2 \Rightarrow X \neq 1$
- Explanations
  - X: c_2, Y: c_3, Z: c_3

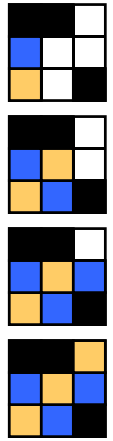


Example of Resolution (2/3)

- Declarations (reminder)
  - 3 variables : $X \in [2..3]$, $Y \in [2..3]$, $Z \in [1..2]$
  - 3 constraints : c_1: $X \neq Y$; c_2: $X \geq Y$; c_3: $Y > Z$
- Labeling Phase
  - Try: $X = 2$
    - c_1 is valid ; $c_2 \Rightarrow Y \neq 3$
  - $X = 2$ and $Y = 2$ violates c_1 : Failure
  - Backtrack over X=2: Try $X = 3$
  - X=3 is added as a new constraint c_4

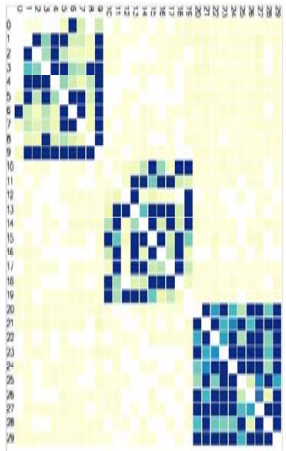
Example of Resolution (3/3)

- Declarations (reminder)
  - 3 variables : $X \in [2..3]$, $Y \in [2..3]$, $Z \in [1..2]$
  - 3 constraints : c_1: $X \neq Y$; c_2: $X \geq Y$; c_3: $Y > Z$
- Labeling (continued)
  - Try : $X = 3$
  - $c_1 \Rightarrow$ $Y \neq 3$ I.e. $Y = 2$; wake up c_3
  - $c_3 \Rightarrow$ $Z \neq 2$
  - $X = 3$; $Y = 2$; $Z = 1$ is a solution of the problem.
- Explanations
  - X: c_2 c_4, Y: c_3, Z: c_3



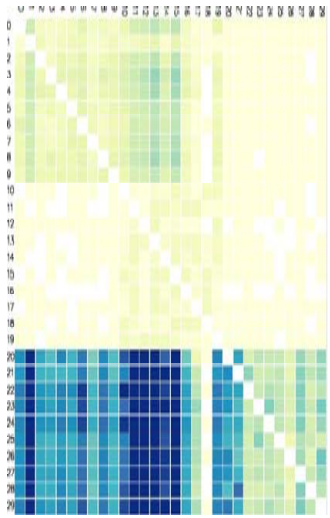
Investigations on Problem Structure

- 3 sets of variables with strong internal bonds and loose external bonds
- At the end of the propagation phase, the structure is visible
- Weak links do not clutter the view



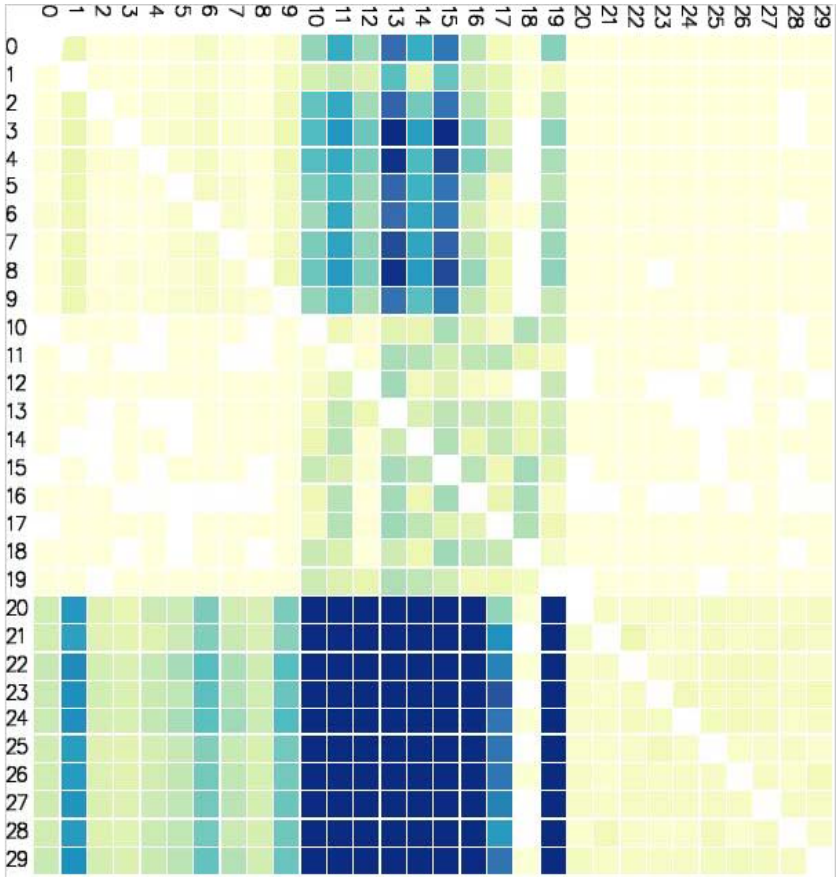
Investigations on Problem Structure

- The mindom heuristic is run for 2 minutes and interrupted
- The first and second variable set has a great impact on the third set
- The solver is misled by weak bonds into early bad decisions



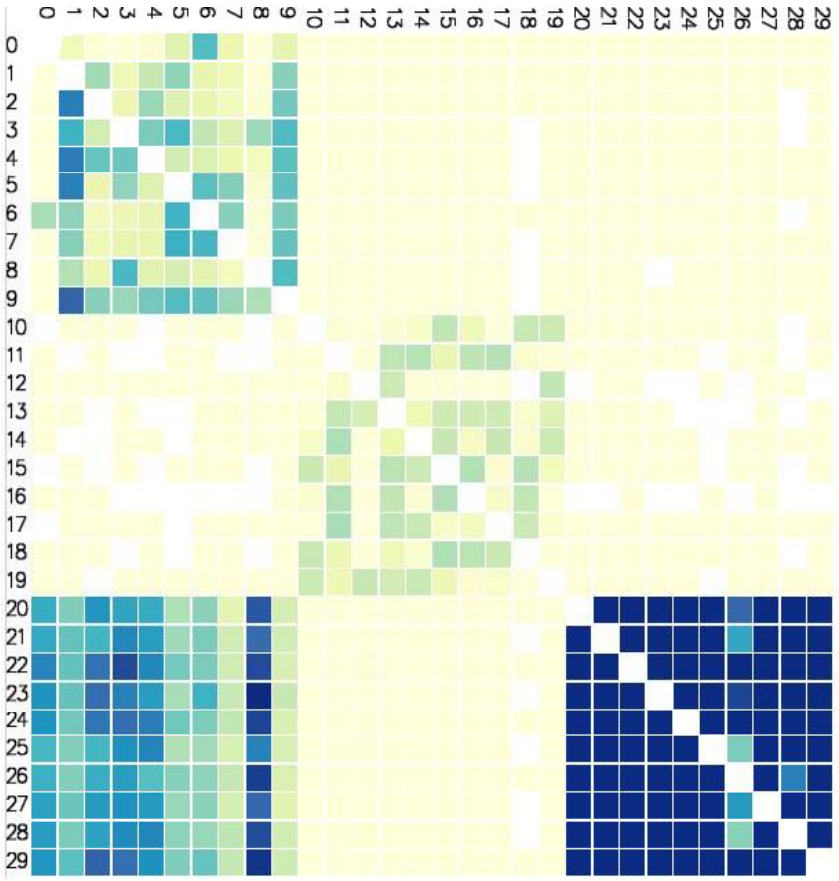
Investigations on Problem Structure

- A normalized view
- Early infrequent decisions appear in dark
- Early poor decisions involve the second set of variables



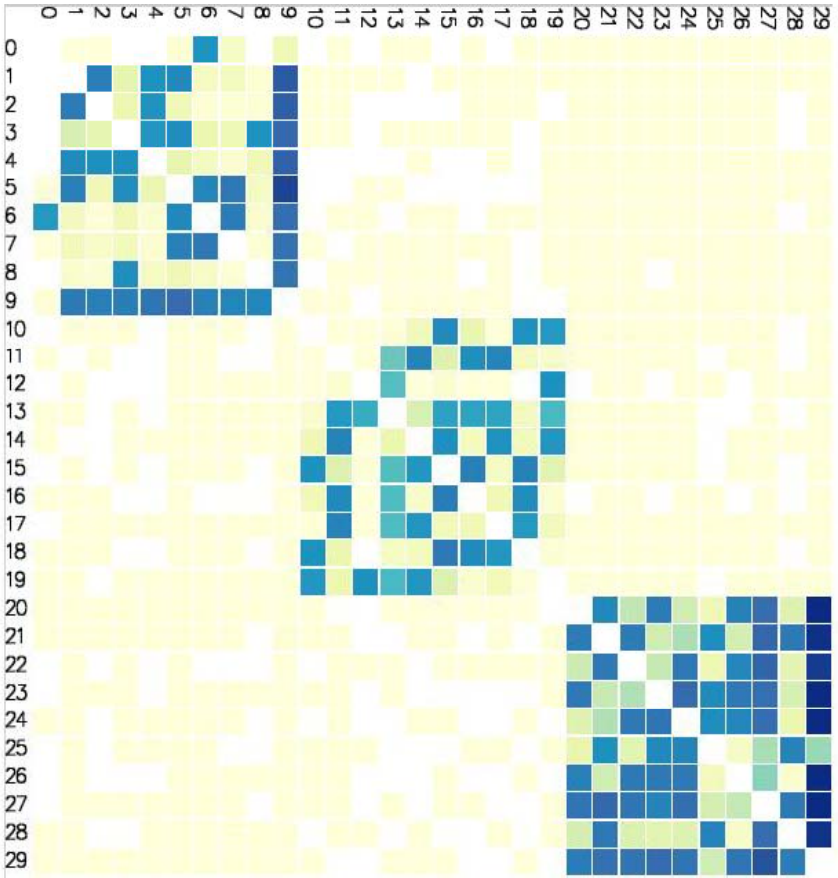
Investigations on Problem Structure

- We discard the impact of early decisions gradually
- The solver is stuck in a combinatorial enumeration involving sets #1 and #3 due to a bad decision taken on set #2



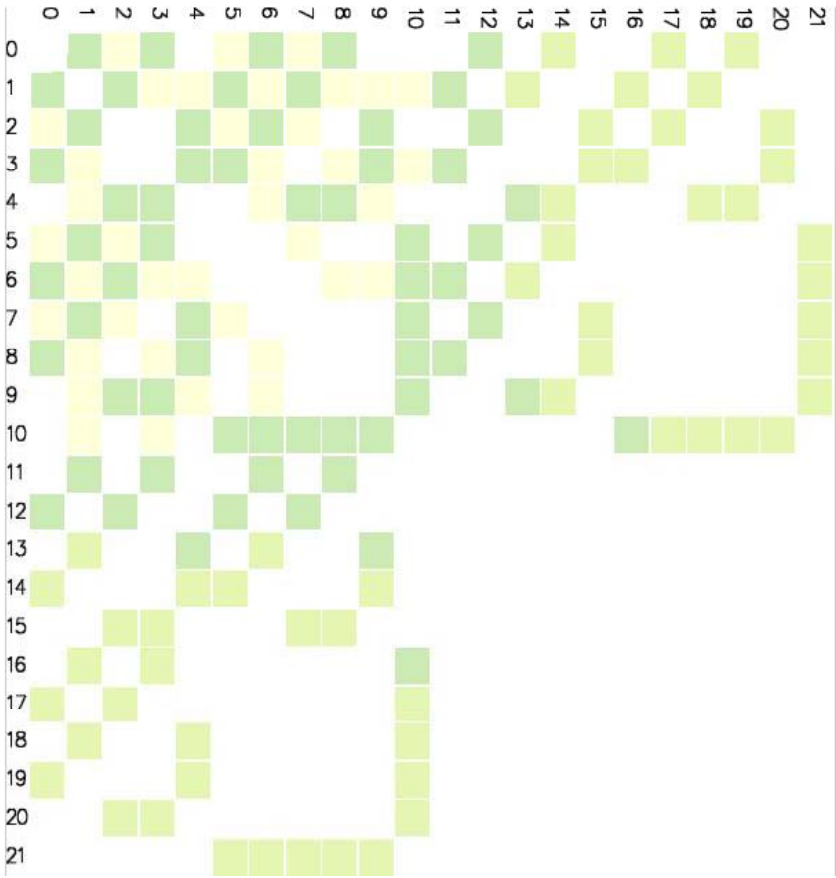
Investigations on Problem Structure

- Adapted mindom heuristic
- Revocation of constraints inducing abnormal activity levels
- The problem is solved instantaneously



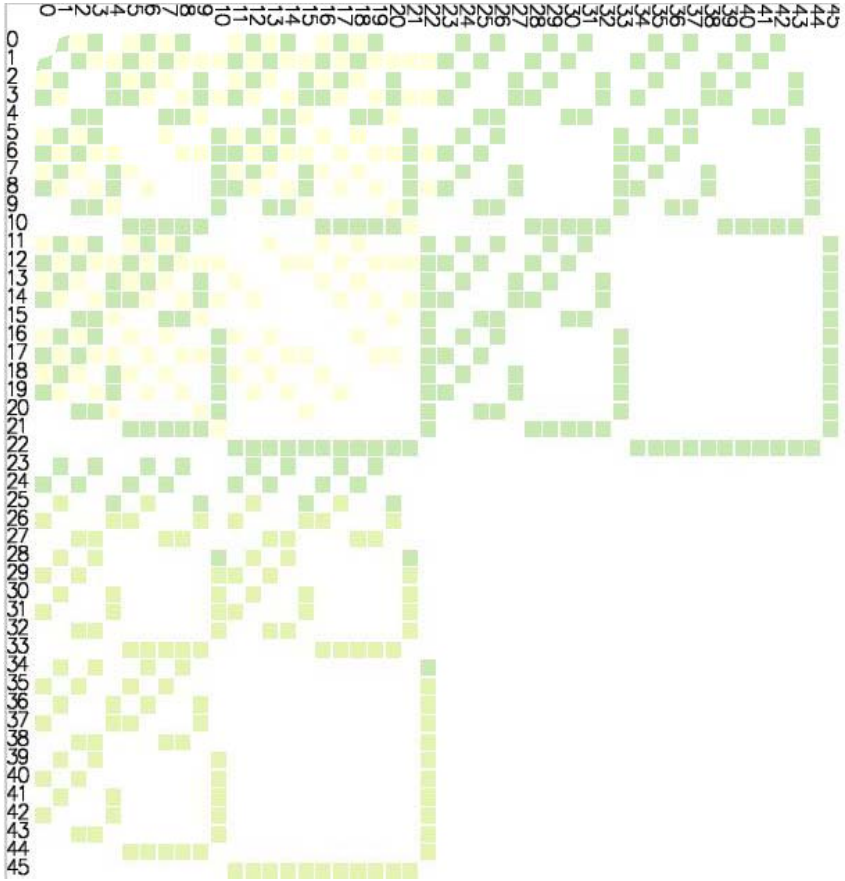
Complex Problems: Graph Coloring

- Mycielski 4 instance
- Variables 11 to 21 are not related, which is made obvious by the use of matrices
- The problem can be split in a principal and secondary problems



Complex Problems: Graph Coloring

- Mycielski 5 instance
- The structure is recurrent
- An incremental resolution is likely to be effective



Complex Problems: Graph Coloring

- The k-insertion problem (k = 1) based on a Mycielsky 4 instance
- An incremental solution is likely to be a good option

Conclusion

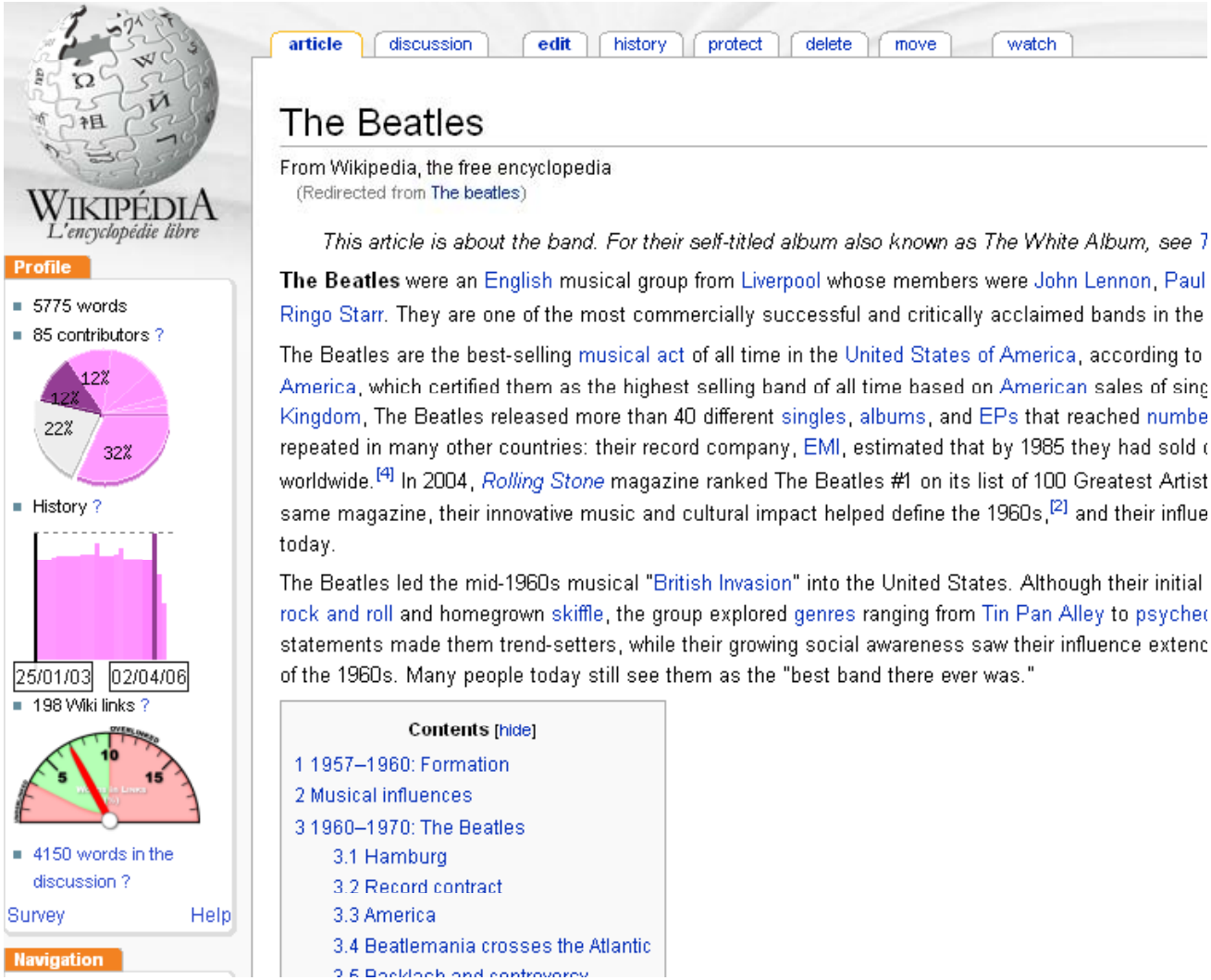
- Matrix-based representations are effective at monitoring large dense dynamic networks
- They help understand the behavior of constraint-oriented solvers on various problems
- A controlled experiment is yet to be carried out to assess its added value in different contexts involving constraint programmers
- This work begs to be extended to other application domains involving co-activity networks

WikipediaViz: Article Quality Assessment

Stéphane Huot & Jean-Daniel Fekete

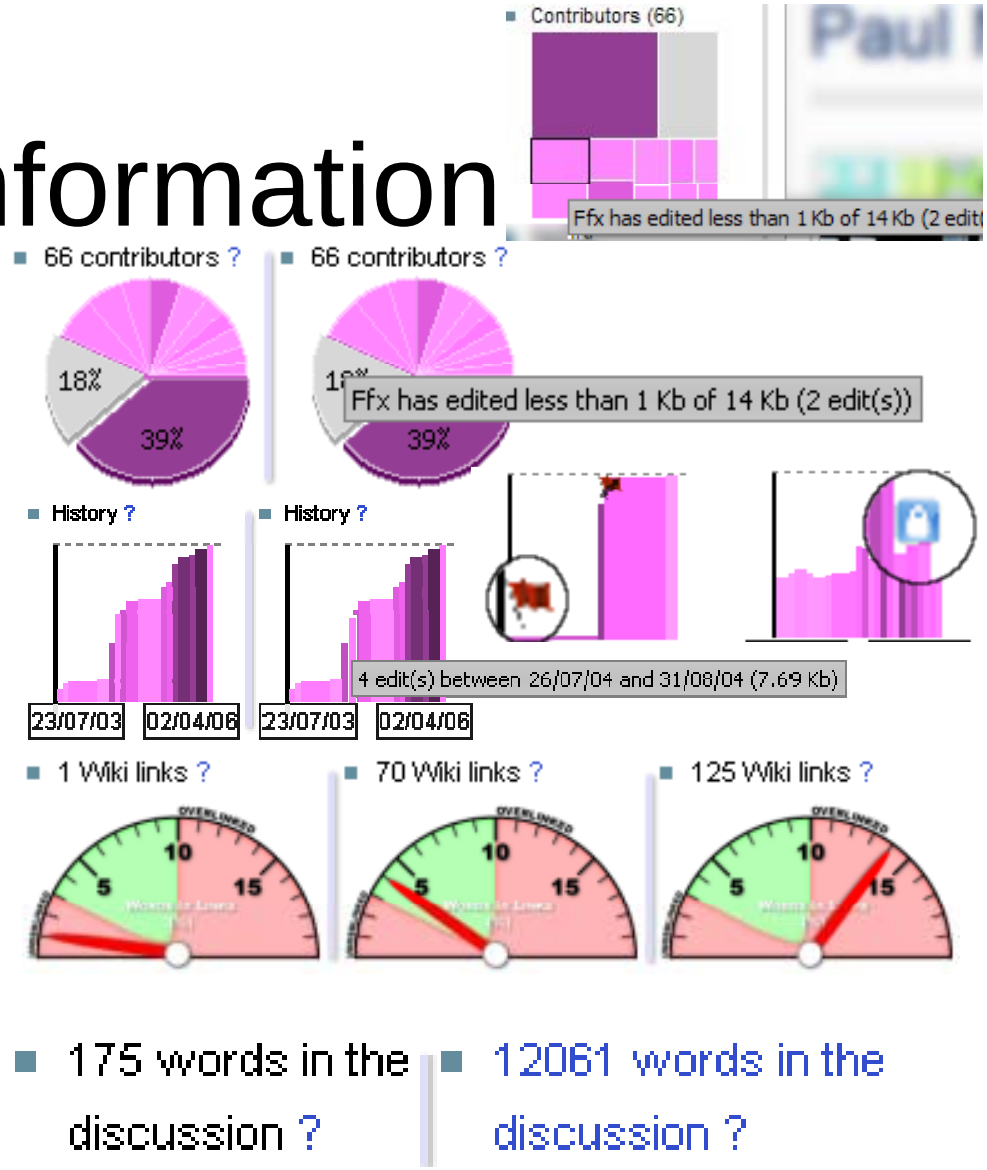
- Wikipedia is very popular for good reasons
  - Ranked #10 in 7 years of existence
  - Free encyclopedia that anyone can edit
  - Good coverage, accuracy and up-to-dateness
- But
  - unverified information
  - Vandalisme
- Occasional Wikipedia Readers are not aware

Solution: Inform and Show



Added Information

- **Authors Contributions**
- **Article Timeline**
- **Internal Links Meter**
- **Discussion Length and Activity Indicator**



Conclusion

- Awareness of the problem is **social**
  - **No technical solution**
- Awareness of the metrics can be improved by visualization
- Simple non-obtrusive visualizations
  - Alternative to full-screen full-attention visualizations

Bibliographie

- N. Elmqvist, T-N. Do, H. Goodell, N. Henry and J-D. Fekete**. ZAME: Interactive Large-Scale Graph Visualization**. In Proceedings of the IEEE Pacific Visualization Symposium 2008, April 2008. IEEE Press. to appear.
- N. Elmqvist, N. Henry, Y. Riche and J-D. Fekete**. Mélange: Space Folding for Multi-Focus Interaction**. In Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems, April 2008. ACM Press. to appear.
- C. Plaisant, J-D. Fekete and G. Grinstein. **Promoting Insight-Based Evaluation of Visualizations**: From Contest to Benchmark Repository. IEEE Transactions on Visualization and Computer Graphics, 14(1):120–134, 2008.
- N. Henry and J-D. Fekete. **MatLink: Enhanced Matrix Visualization for Analyzing Social Networks**. In C. Baranauskas, P. Palanque, J. Abascal and S. D. J. Barbosa, editors, Human-Computer Interaction – INTERACT 2007, volume 4663 of LNCS, pages 288–302, 2007. Springer. (Brian Shackel Award).
- N. Henry, J-D. Fekete and M. J. McGuffin. **NodeTrix: a Hybrid Visualization of Social Networks**. IEEE Transactions on Visualization and Computer Graphics, 13(6):1302-1309, 2007.
- N. Henry, H. Goodell, N. Elmqvist and J-D. Fekete. **20 Years of four HCI conferences: A Visual Exploration**. International Journal of Human-Computer Interaction — Special issue in honor of Ben Shneiderman's 60th birthday: Reflections on Human-Computer Interaction, 23(3):239–285, 2007.
- C. Appert and J-D. Fekete. **OrthoZoom Scroller: 1D Multi-Scale Navigation**. In Proceedings of ACM CHI 2006 Conference on Human Factors and Computing Systems, pages 21–30, April 2006. ACM Press.
- N. Henry and J-D. Fekete. **MatrixExplorer: a Dual-Representation System to Explore Social Networks**. IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization 2006), 12(5):677-684, September-October 2006.
- M. Ghoniem, J-D. Fekete and P. Castagliola. **Readability of Graphs Using Node-Link and Matrix-Based Representations: Controlled Experiment and Statistical Analysis**. Information Visualization Journal, 4(2):114–135, 2005.
J-D. Fekete. **The InfoVis Toolkit**. In Proceedings of the 10th IEEE Symposium on Information Visualization (InfoVis 04), pages 167-174, Austin, TX, October 2004. IEEE Press.
- J-D. Fekete and C. Plaisant. **Interactive Information Visualization of a Million Items**. In Proc. IEEE Symposium on Information Visualization 2002 (InfoVis 2002), pages 117-124, Boston, USA, October 2002. IEEE Press.
- J-D. Fekete and C. Plaisant. **Excentric Labeling: Dynamic Neighborhood Labeling for Data Visualization** Proceedings of the International Conference on Human Factors in Computing Systems (CHI 99). Ed. Kate Ehrlich and William Newman, ACM *may* 1999