



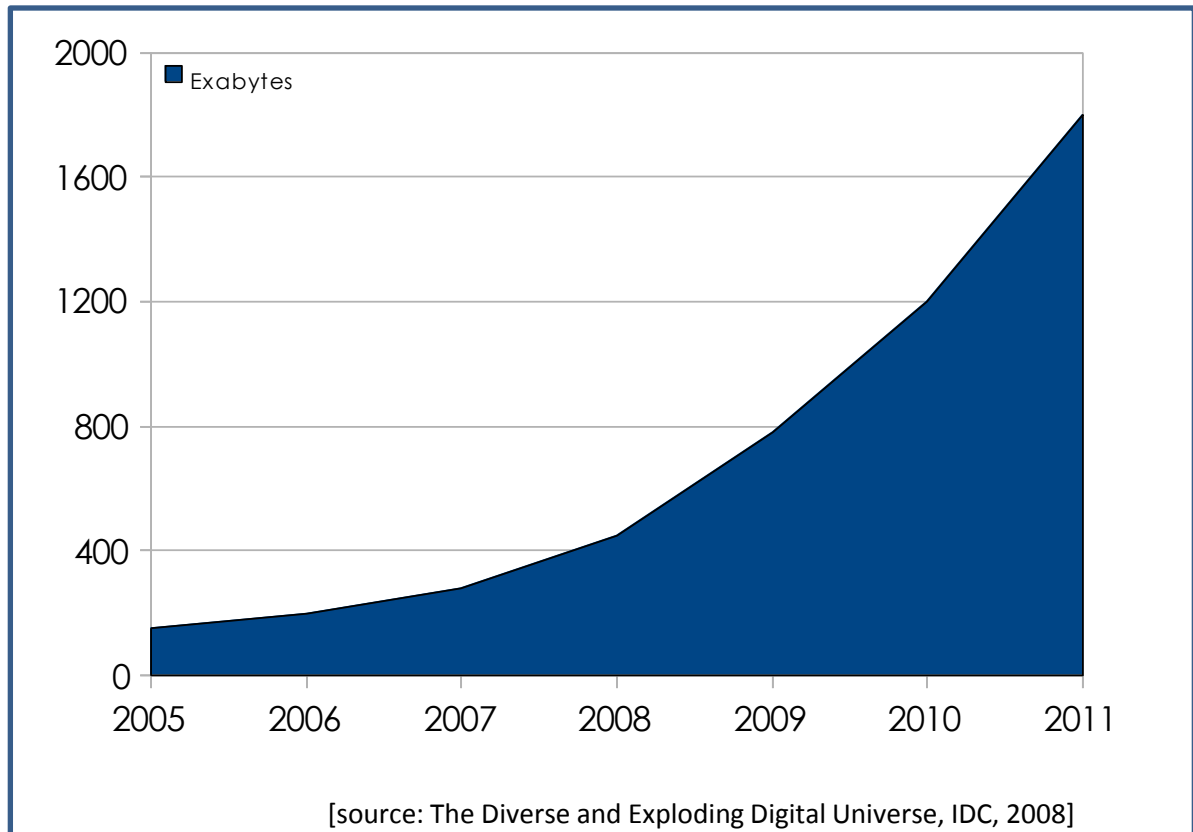
# La recherche en visualisation de données : créer un télescope pour explorer l'information

Jean-Daniel Fekete



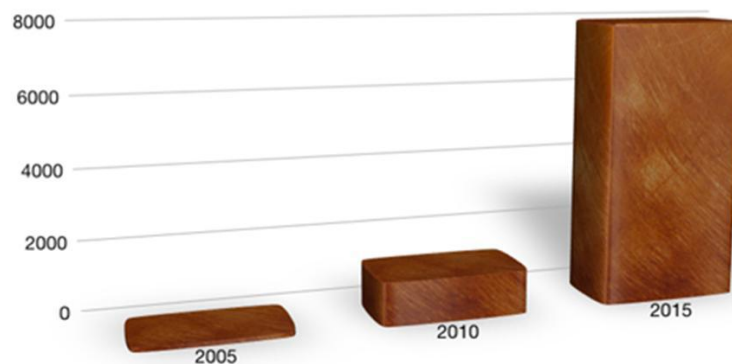
# La révolution des données

- La quantité de données croît de 30% par an depuis 1999 (SIMS, Berkeley)
- La quantité de données accessible a cru exponentiellement depuis 10 ans



# La révolution des données

- Les solutions de stockage croissent exponentiellement
- Les capteurs génèrent plus de données :
  - Pollution, logs, température, photos, vidéos, etc.
- Les données collectées sont mises en ligne
- Les données sont analysées et les analyses sont mises en ligne
- L'univers numérique explose :
  - 2007 : 281 Exa-octets (281 milliards de Giga-octets)
  - 2010 : barrière du Zeta-octet franchie
  - 2011 : 1.8 Zeta-octets



# FaceBook



# Data, Information, Knowledge, Wisdom

Where is the Life we have lost in living?

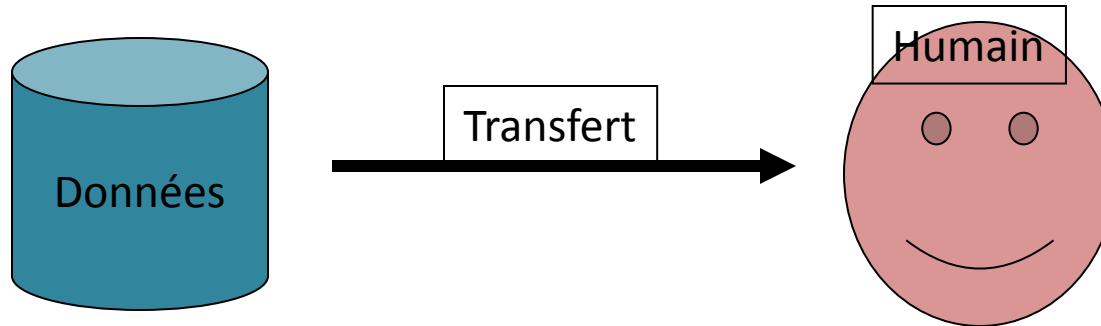
Where is the wisdom we have lost in knowledge?

Where is the knowledge we have lost in information?

-- from T.S. Eliot, "Choruses from 'The Rock'"

- Comment :
  - transformer des données en information ?
  - utiliser l'information pour prendre des bonnes décisions et apprendre de nouvelles connaissances
  - avoir des réponses à des questions fondamentales dont on n'a aucune idée aujourd'hui ?

# Le problème



Web, livres, articles  
données scientifiques  
prix, liste de personnes,  
Cours de la bourse  
Informations

## Comment ?

Vision: 100 MB/s  
Audition: <100 b/s  
Haptique/tactile  
Odorat  
Goût

# Propriétés de la vision

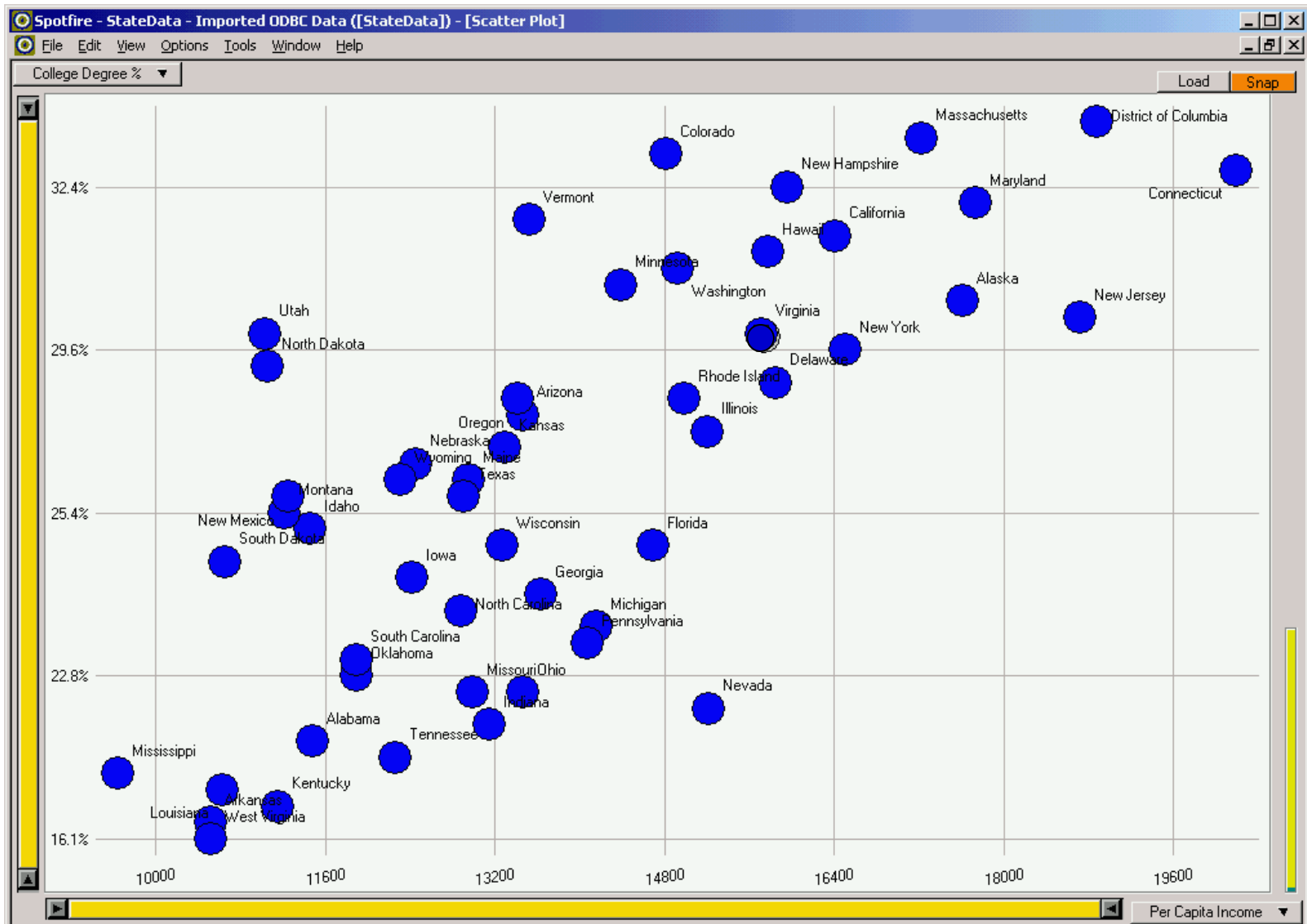
- Sens ayant la plus grande bande passante
- Rapide, parallèle
- Reconnaissance de formes
- Pré-attentif
- Etend les capacités cognitives et mémorielles
- On pense visuellement

# Utilisons la vision !

Table - StateData ()			Load	Snap
State	College Degree %	Per Capita Income		
Alabama	20.6%	11486		
Alaska	30.3%	17610		
Arizona	27.1%	13461		
Arkansas	17.0%	10520		
California	31.3%	16409		
Colorado	33.9%	14821		
Connecticut	33.8%	20189		
Delaware	27.9%	15854		
District of Columbia	36.4%	18881		
Florida	24.9%	14698		
Georgia	24.3%	13631		
Hawaii	31.2%	15770		
Idaho	25.2%	11457		
Illinois	26.8%	15201		
Indiana	20.9%	13149		
Iowa	24.5%	12422		
Kansas	26.5%	13300		
Kentucky	17.7%	11153		
Louisiana	19.4%	10635		
Maine	25.7%	12957		
Maryland	31.7%	17730		
Massachusetts	34.5%	17224		
Michigan	24.1%	14154		
Minnesota	30.4%	14389		
Mississippi	19.9%	9648		
Missouri	22.3%	12989		
Montana	25.4%	11213		
Nebraska	26.0%	12452		
Nevada	21.5%	15214		
New Hampshire	32.4%	15959		
New Jersey	30.1%	18714		
New Mexico	25.5%	11246		
New York	29.6%	16501		
North Carolina	24.2%	12885		
North Dakota	28.1%	11051		
Ohio	22.3%	13461		
Oklahoma	22.8%	11893		
Oregon	27.5%	13418		
Pennsylvania	23.2%	14068		
Rhode Island	27.5%	14981		
South Carolina	23.0%	11897		
South Dakota	24.6%	10661		
Tennessee	20.1%	12255		
Texas	25.5%	12904		
Utah	30.0%	11029		
Vermont	31.5%	13527		
Virginia	30.0%	15713		
Washington	30.9%	14923		
West Virginia	16.1%	10520		
Wisconsin	24.9%	13276		
Wyoming	25.7%	12311		



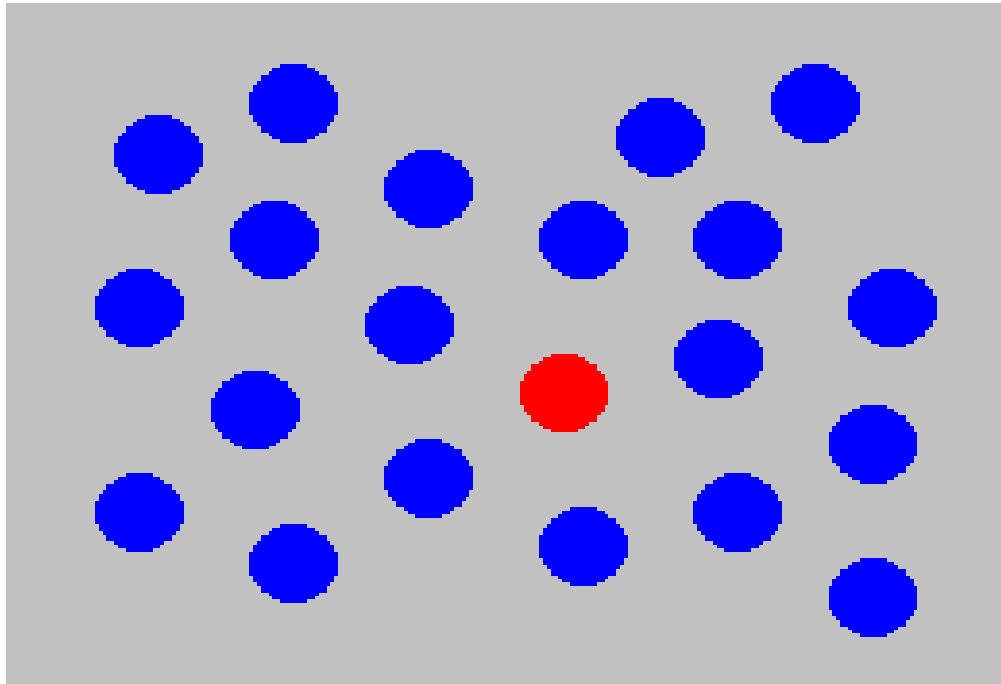
# Utilisons la vision !



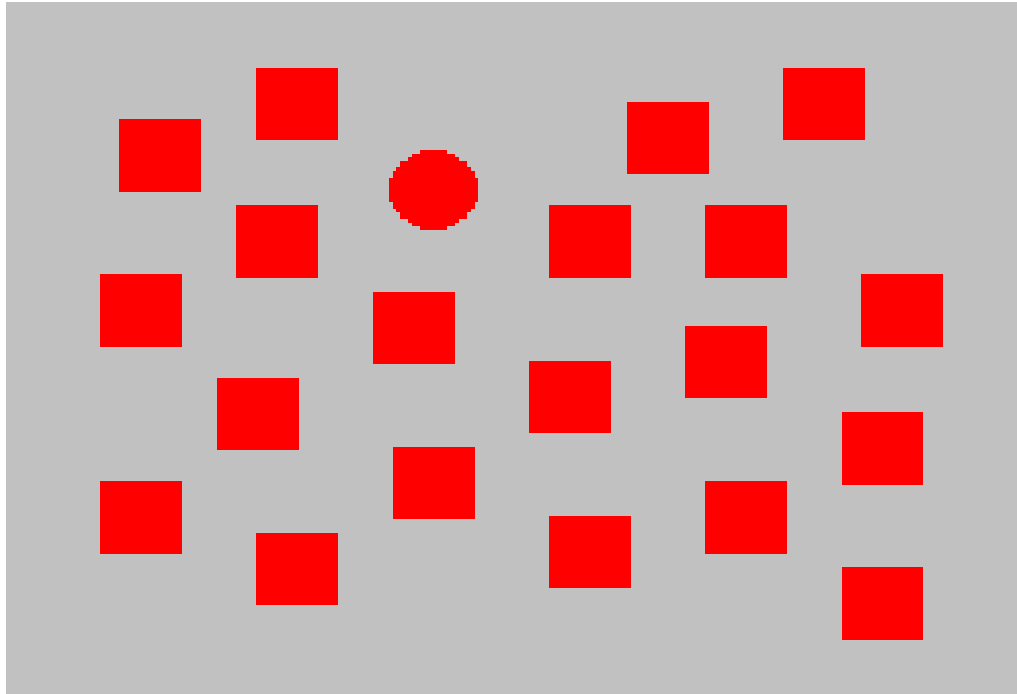
# Principes de la visualisation d'information

- 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)



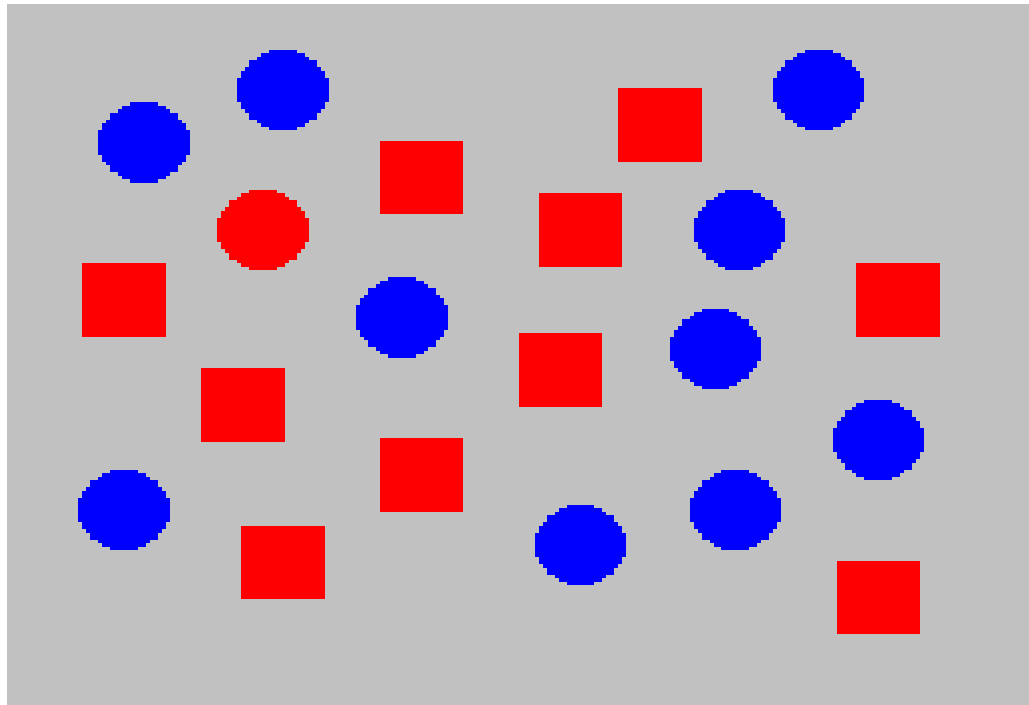
# Perception préattentive (2)



# Perception préattentive

- Beaucoup de caractéristiques visuelles peuvent être perçues préattentivement :
  - Orientation de ligne/blob, longueur, épaisseur, taille, courbure, cardinalité, terminaisons, intersections, inclusion, teinte, clignotement, direction de mouvement, profondeur stéréoscopique, indices 3D, direction de l'éclairage
- Problèmes :
  - Les caractéristiques préattentives interfèrent entre elles
    - On pensait même que toutes les caractéristiques préattentives étaient incompatibles entre elles
  - Les caractéristiques sont préattentives dans certaines limites
    - 7 couleurs max (Healey, 96)
    - 2 ou 3 formes
    - Etc.

# Perception préattentive (3)



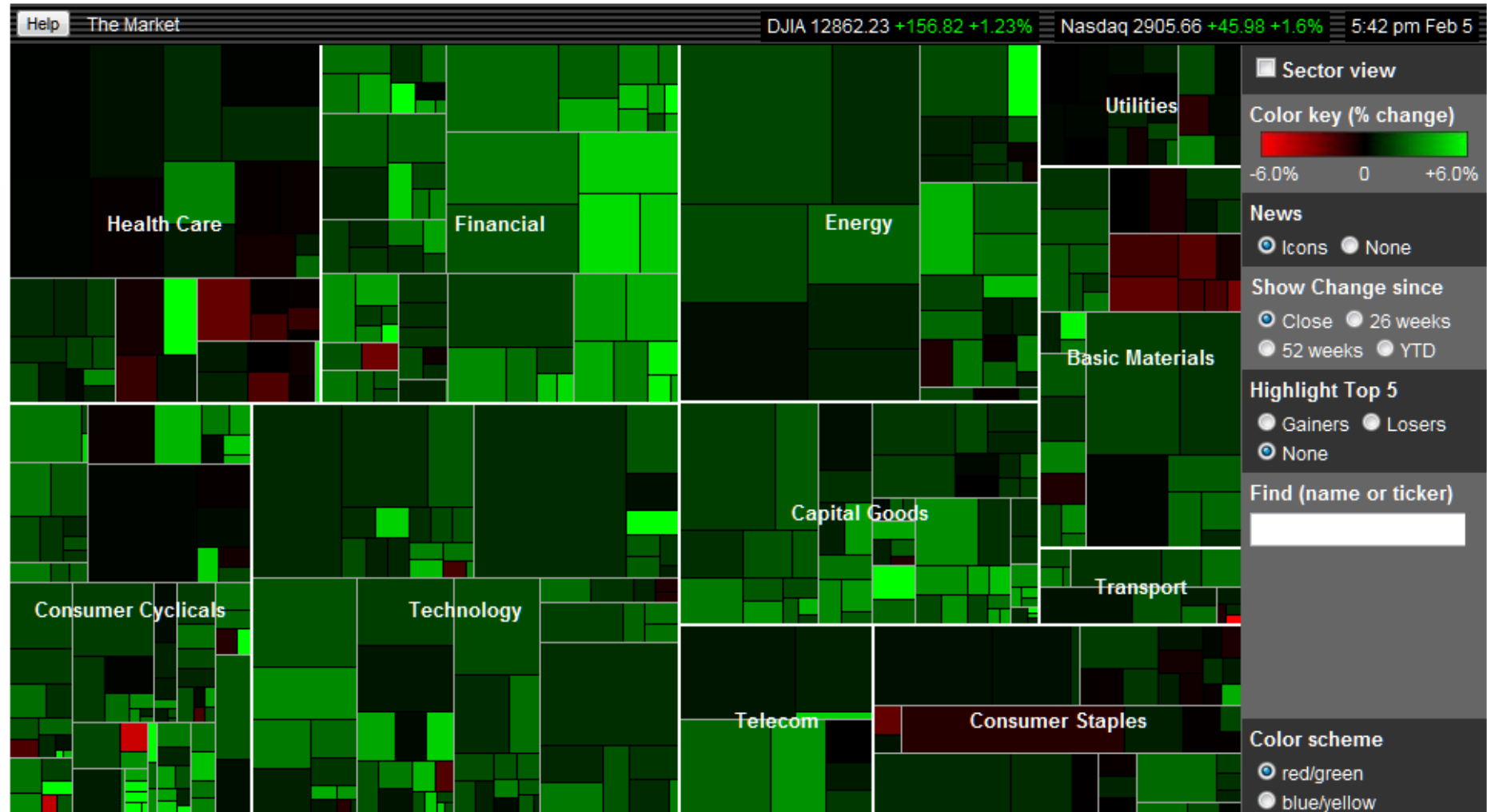
# Visualisation : exemples

<http://www.smartmoney.com/map-of-the-market/>

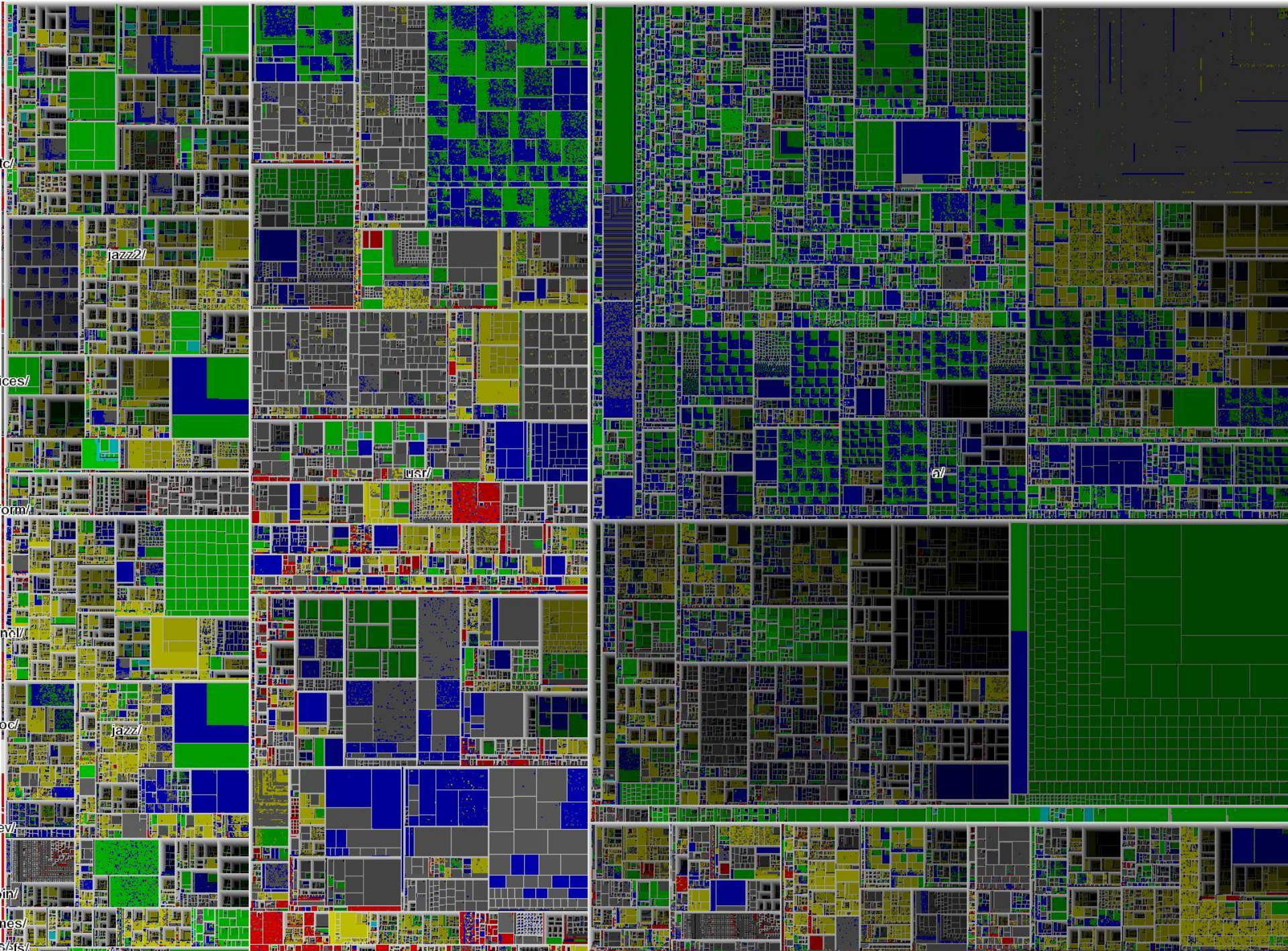
## Map of the Market

SmartMoneySelect

Upgrade [here](#) to access the **Market Map 1000** and search 1,000 companies with enhanced capabilities.





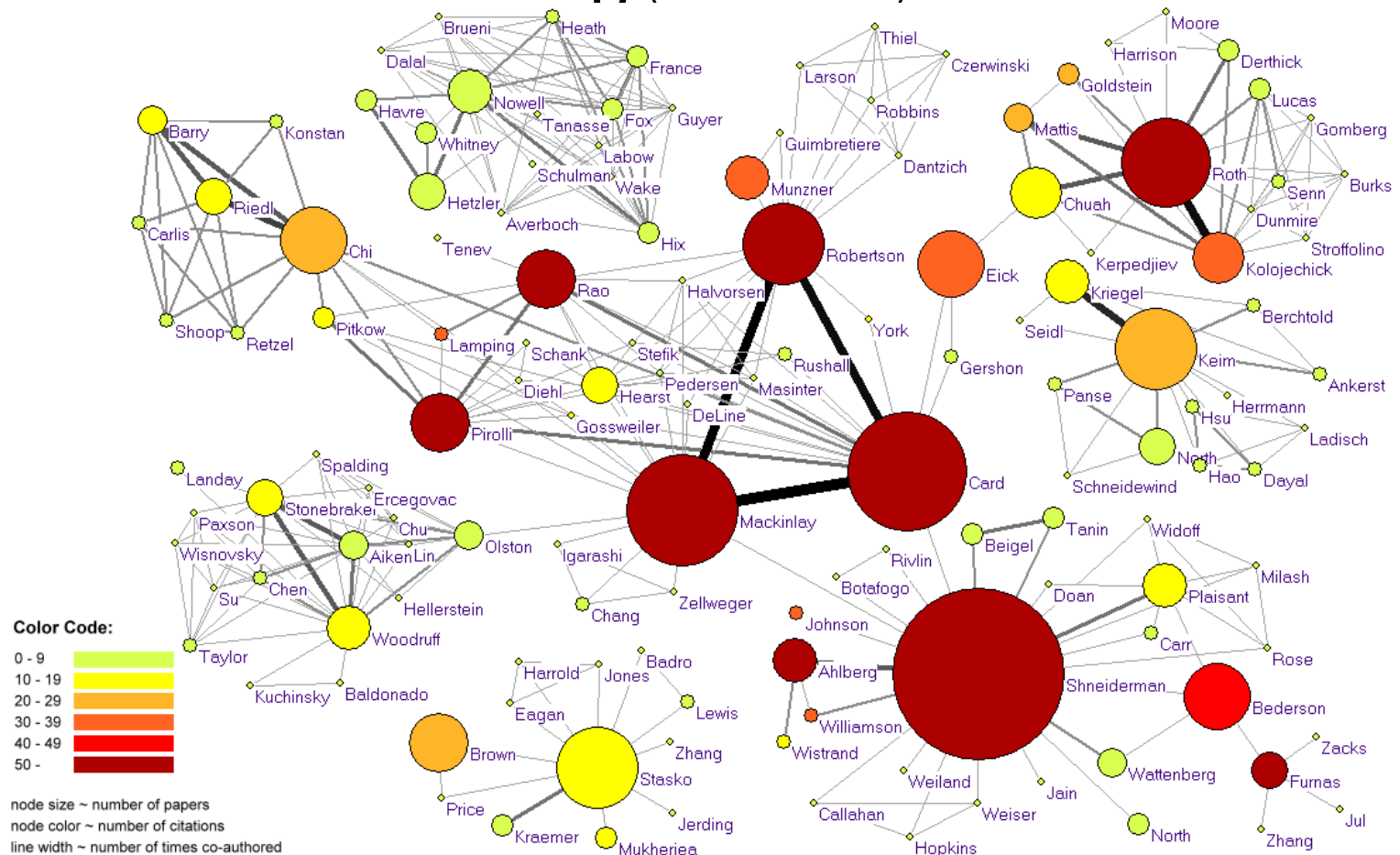




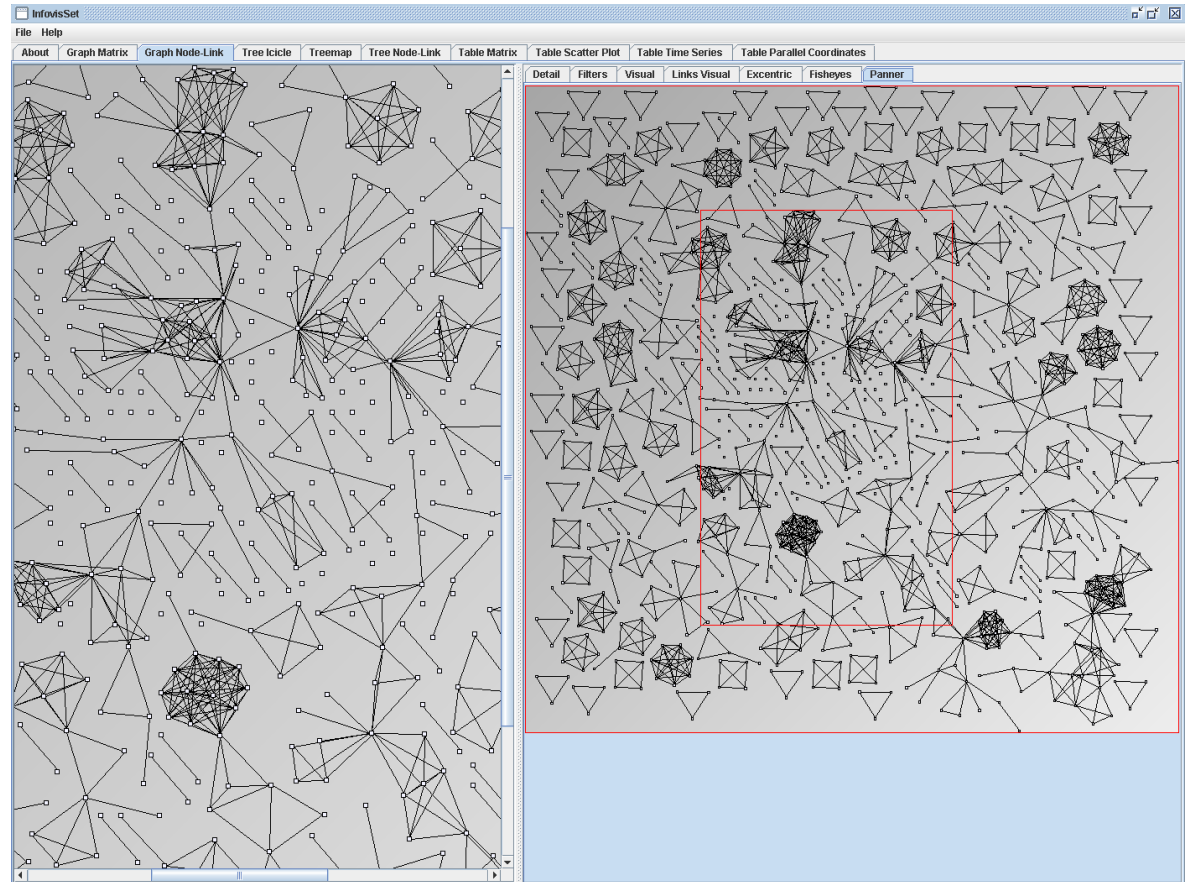
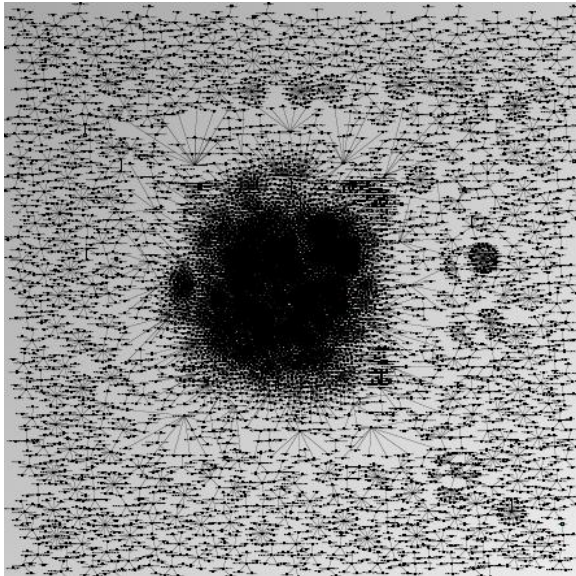


# Des réseaux sociaux de grande taille

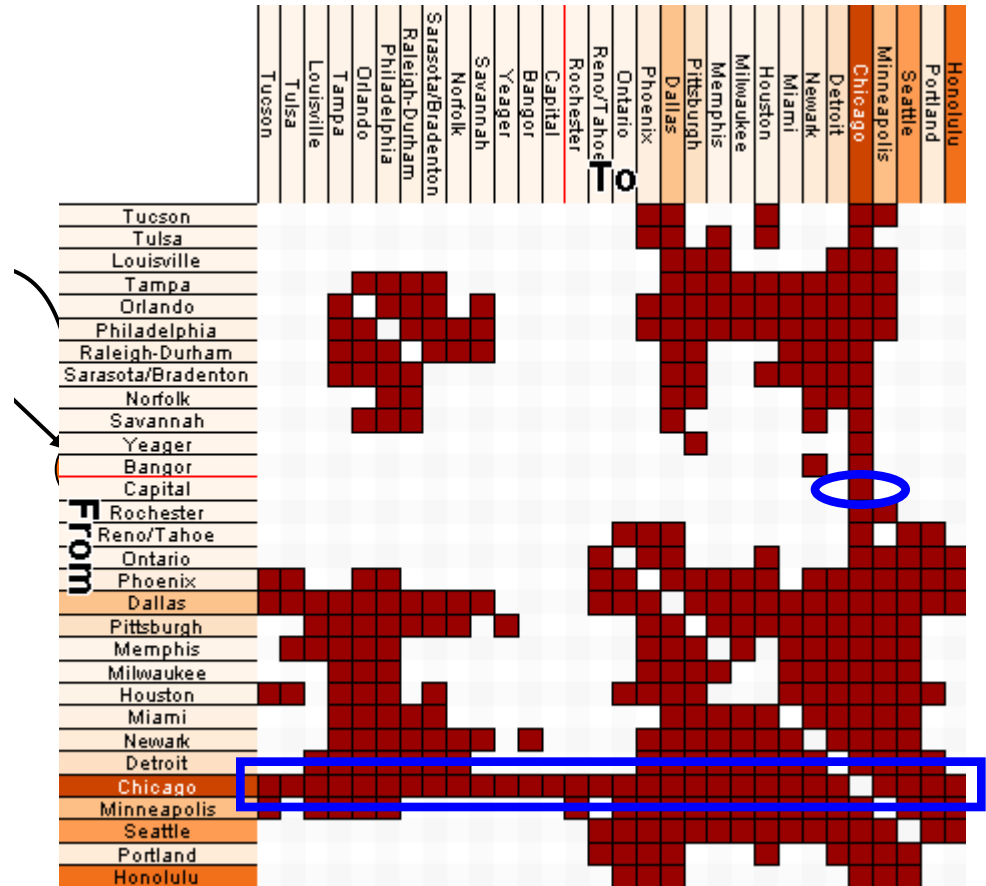
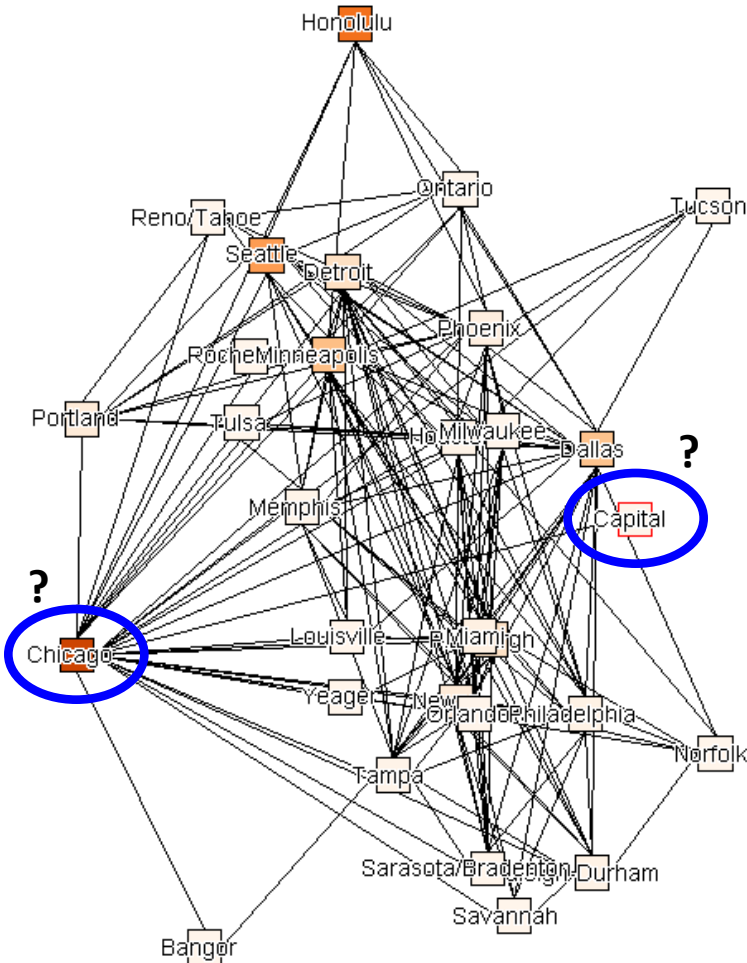
- InfoVis Co-authoring (K. Börner et al.)



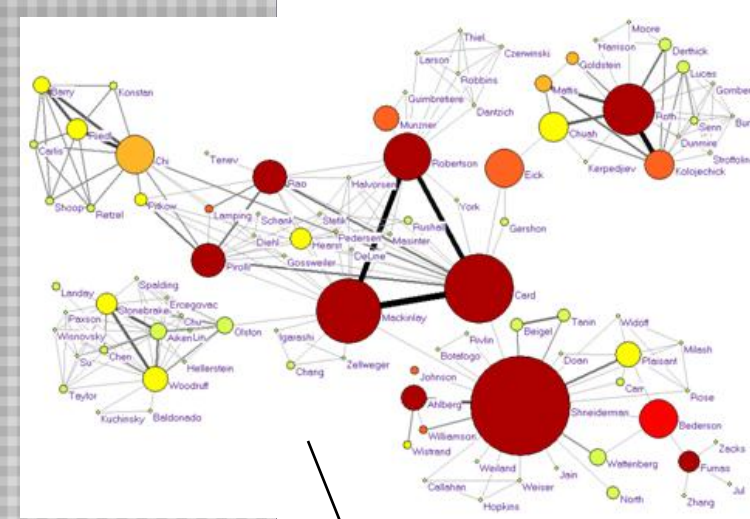
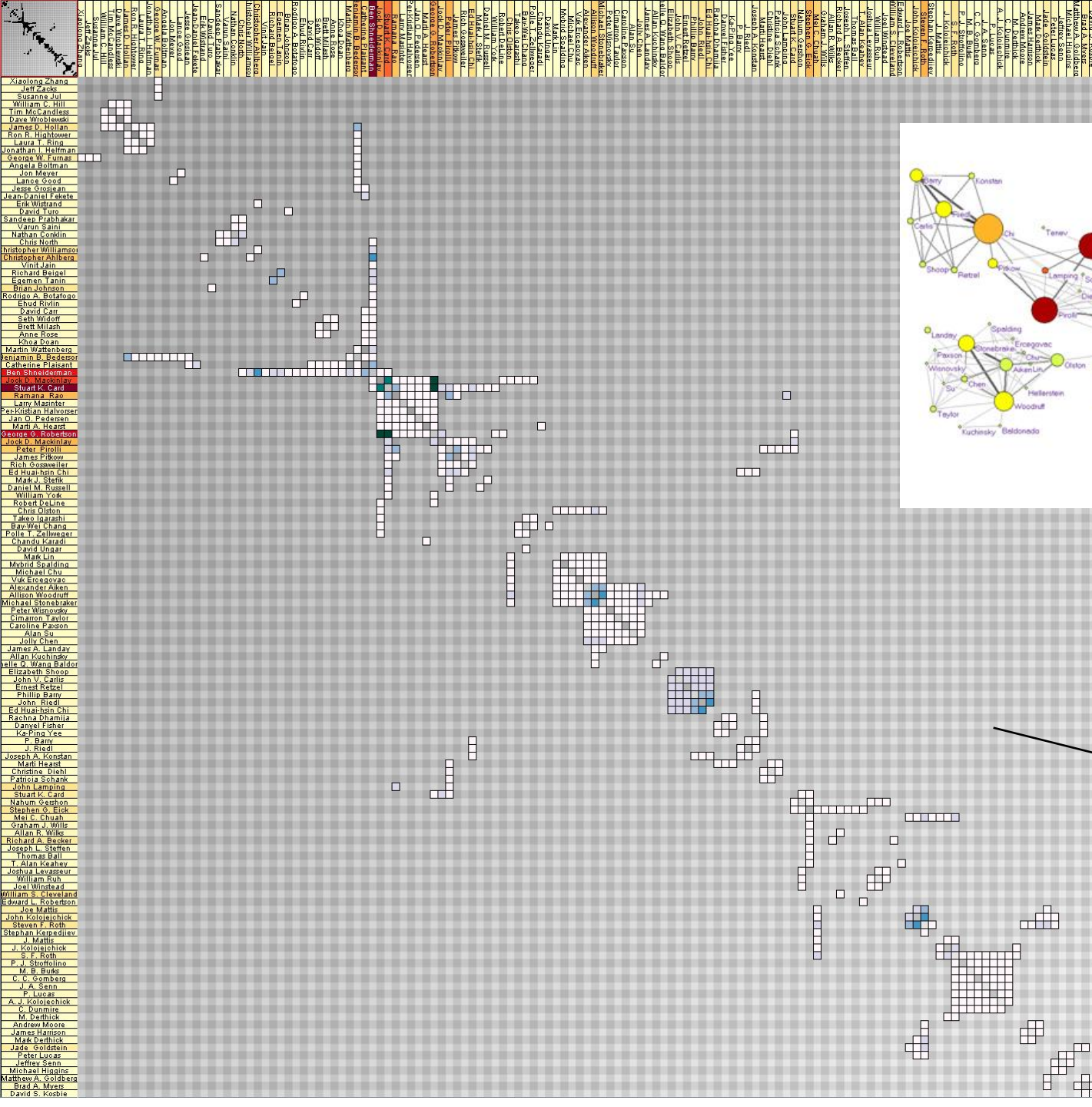
# En réalité, lorsqu'on visualise les données brutes ...



# Visualisation de réseau par matrice d'adjacence



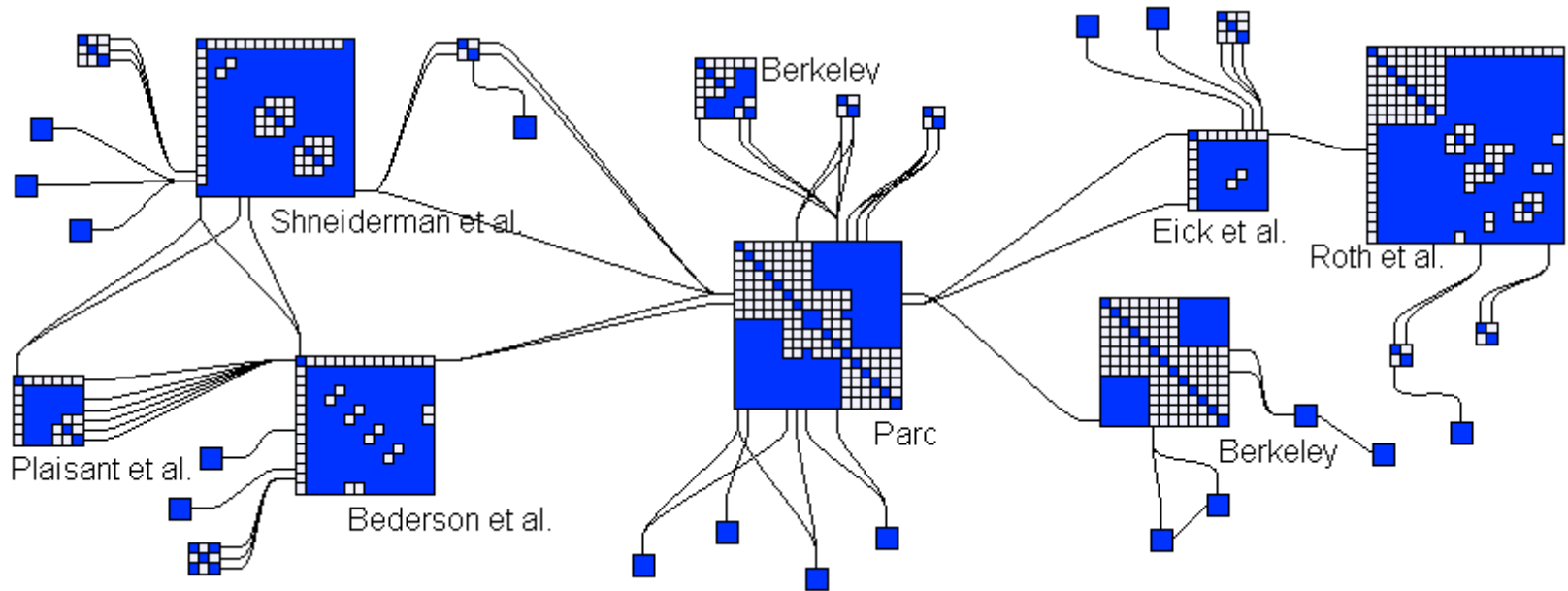
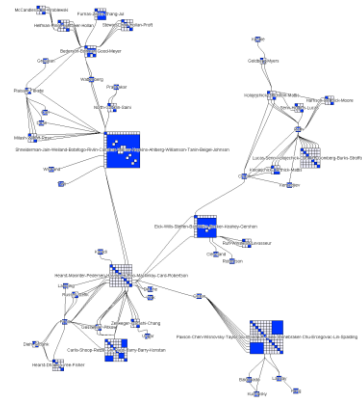




Communicate

Explore

# Plus de détails, plus de structure



Infovis Coauthorship (133 actors)



# Conclusion

- L'analyse visuelle, couplée à d'autres analyses automatiques, permet de donner du sens aux données
  - On trouve ce que l'on cherche
  - On trouve aussi ce que l'on ne cherchait pas !
- Plusieurs représentations sont possibles
  - Mais beaucoup d'autres sont inefficaces et donc nuisibles
  - Les représentations spectaculaires ne sont pas toujours efficaces
- Utilisez la visualisation de données / information
  - Explorer ses données révèle toujours des informations inattendues