The mathematics of Ranking



Brussels, October 15, 2008

Belgian Mathematical Society

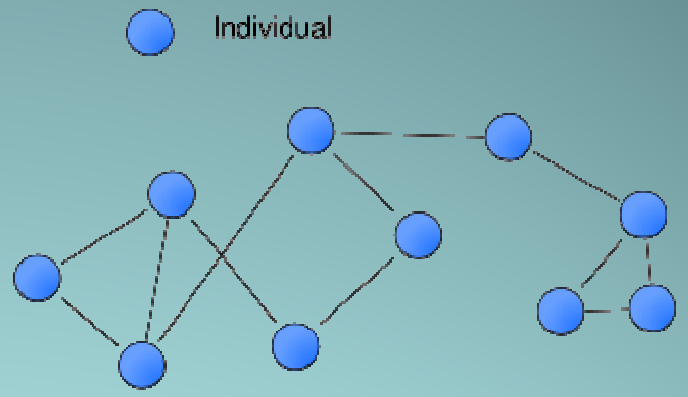
National Committee of Mathematics

A network

- Individuals
- Websites
- Persons
- Papers
- Scientists
- Blogs
- News items

- Links
- In/out-links
- Social relation
- In/out-citations
- Collaboration
- References
- Same readers

Etc.....



Which individual is the most important?

How do we RANK the individuals?

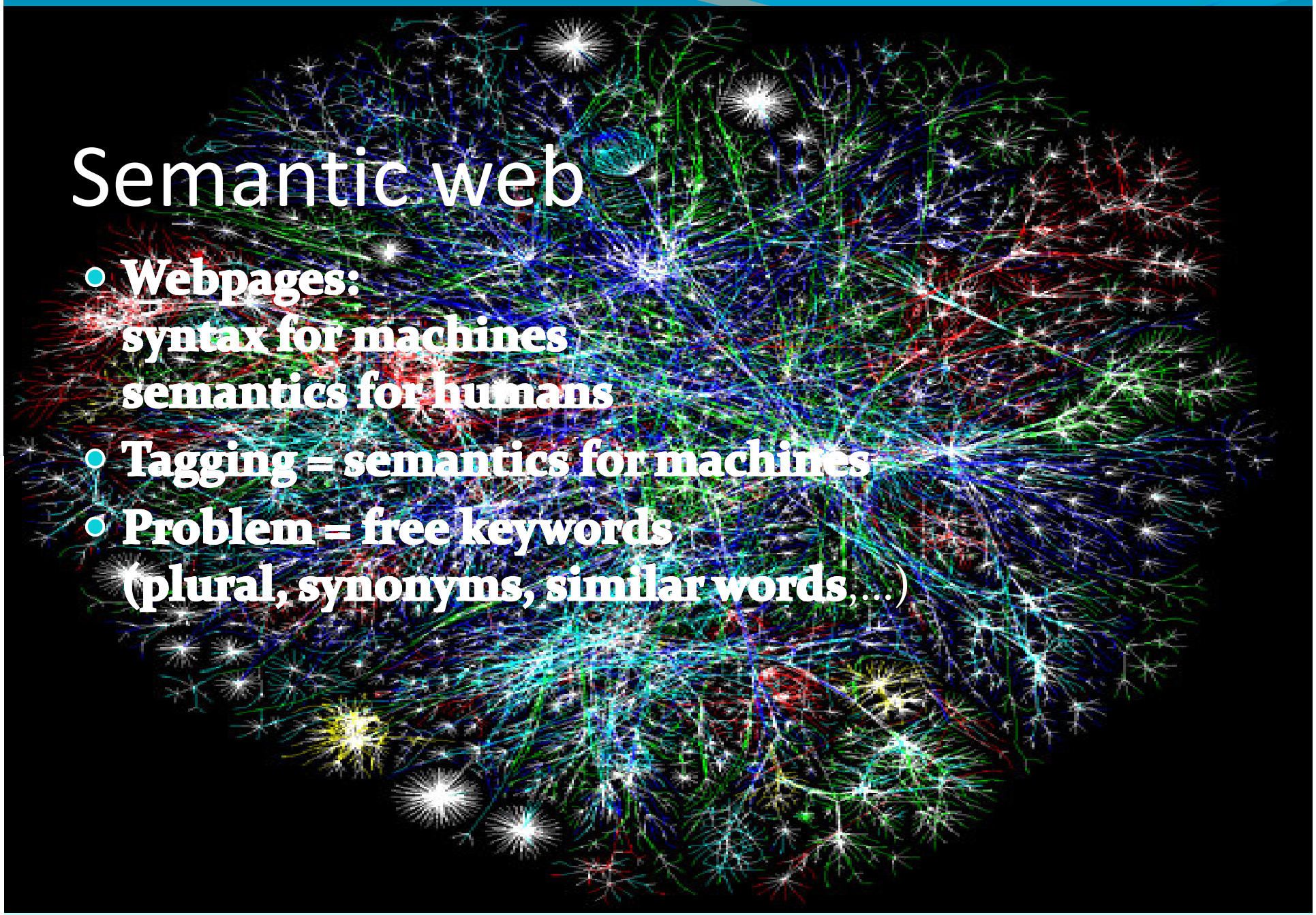
Google's PageRank



- Based on web-democracy
- Number of in- and out-links
- The out-links of pages with many in-links are more important
- A HUGE eigenvalue problem
- Dominant eigenvector = PageRank



Semantic web



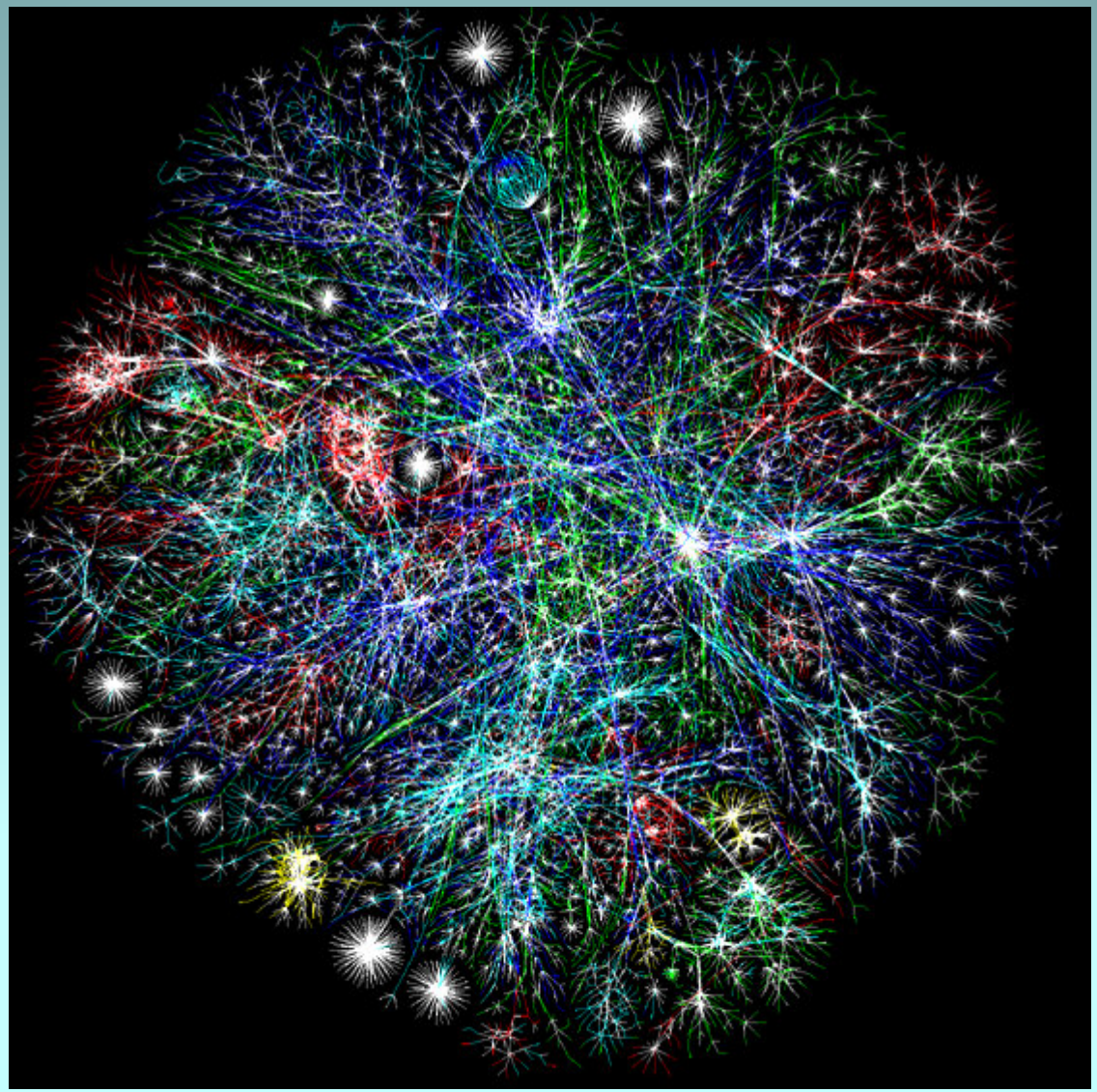
- **Webpages:**
  syntax for machines
  semantics for humans
- **Tagging = semantics for machines**
- **Problem = free keywords**
  (plural, synonyms, similar words,…)

Folksonomy

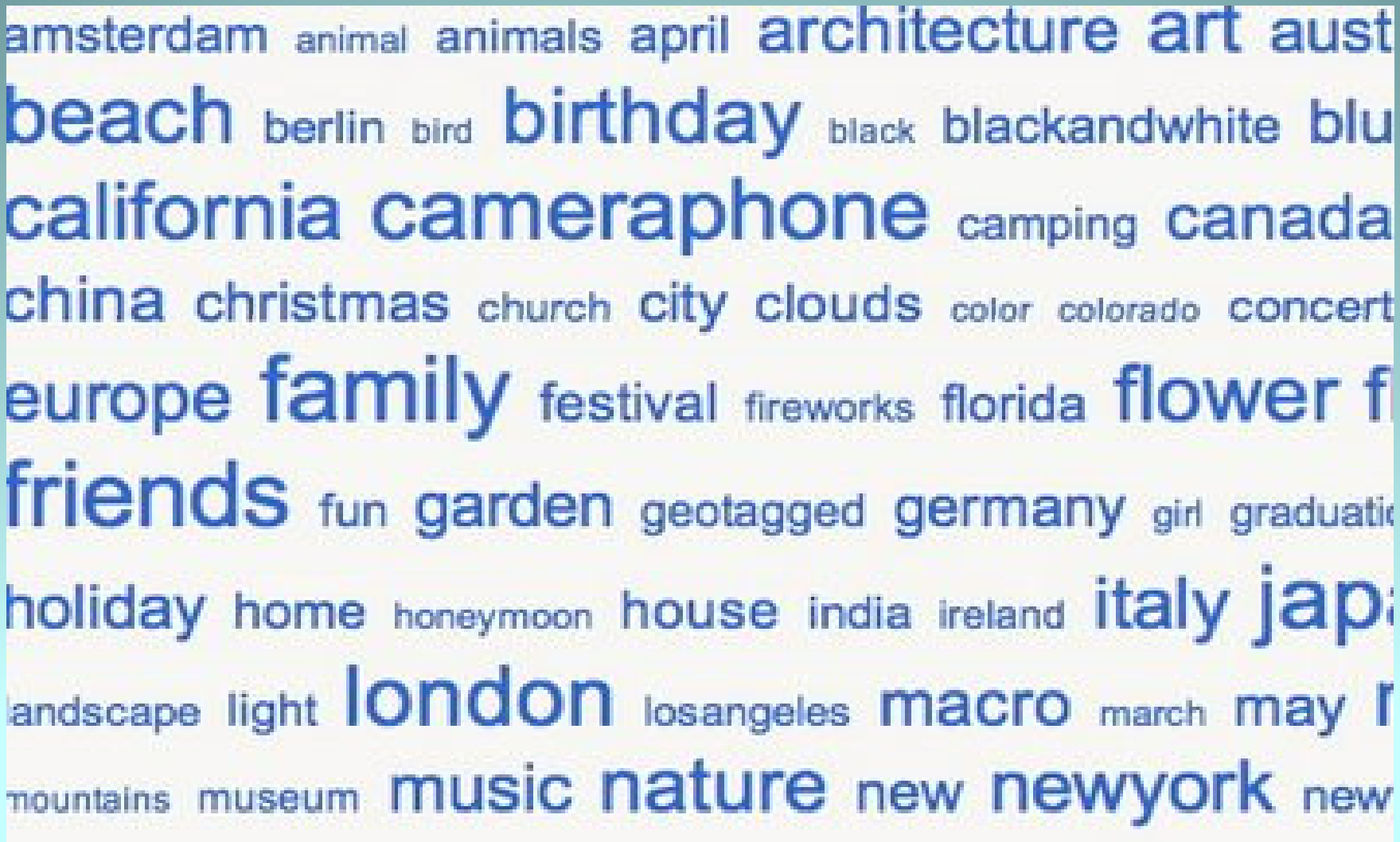


- Folk + taxonomy
- Multi-dimensional graphs
- **Example: web shopping**
  Customers buying similar items are grouped
  - Customers refer to items
  - Customers refer to customers
  - Customers refer to groups
- Personalized advertizing
- Customer finds items he did not know he was looking for
- Requires ranking of keywords

Folksonomy



Folksonomy


amsterdam animal animals april architecture art aust
beach berlin bird birthday black blackandwhite blu
california cameraphone camping canada
china christmas church city clouds color colorado concert
europe family festival fireworks florida flower f
friends fun garden geotagged germany girl graduatio
holiday home honeymoon house india ireland italy jap
landscape light london losangeles macro march may
mountains museum music nature new newyork new

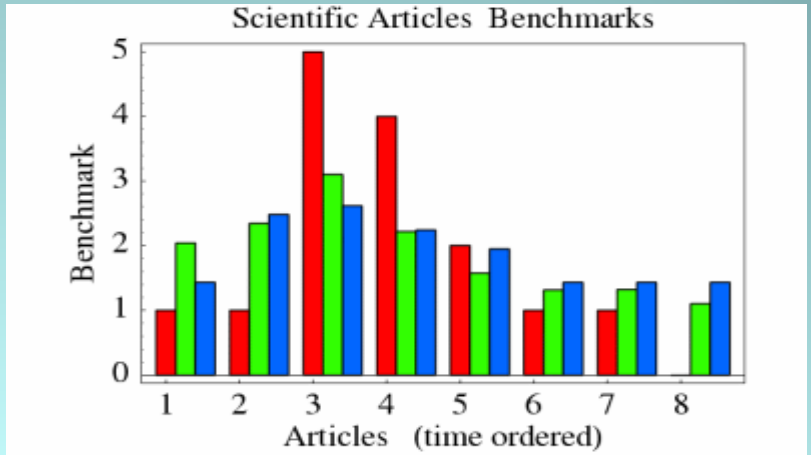
Ranking

- Is important for web, for search engines, for end-users, for business,…
- Applications in data mining
- It defines te pecking order
- Etc…



Bibliometrics, informetrics, scientometrics

- Relations among the individuals is less important
- "Sources" produce "items"
  - Papers -> citations
  - Researcher -> papers
  - Product -> buyers
  - …



- How to quantify the success of the sources? Model the productivity, and rank them.
- Impact factors, h-index, g-index, etc…

This symposium

- 10h00 **Hugo Zaragoza:**
  Ranking text in search of knowledge and wealth
- 11h00 **Bettina Berendt:**
  Ranking - Use and Usability
- 12h00 Lunch
- 13h30 **Paul Van Dooren:**
  Some graph optimization problems in data mining
- 14h30 **Leo Egghe:**
  Lotkaian informetrics and applications to social networks
- 15h30 Coffee
- 16h00 **Gianna M. Del Corso:**
  Evaluating scientific products by means of citation-based models

Thanks to
SRN Fundamental Methods and Techniques in Mathematics
BMS Belgian Mathematical Society

It’s all a matter of ranking

