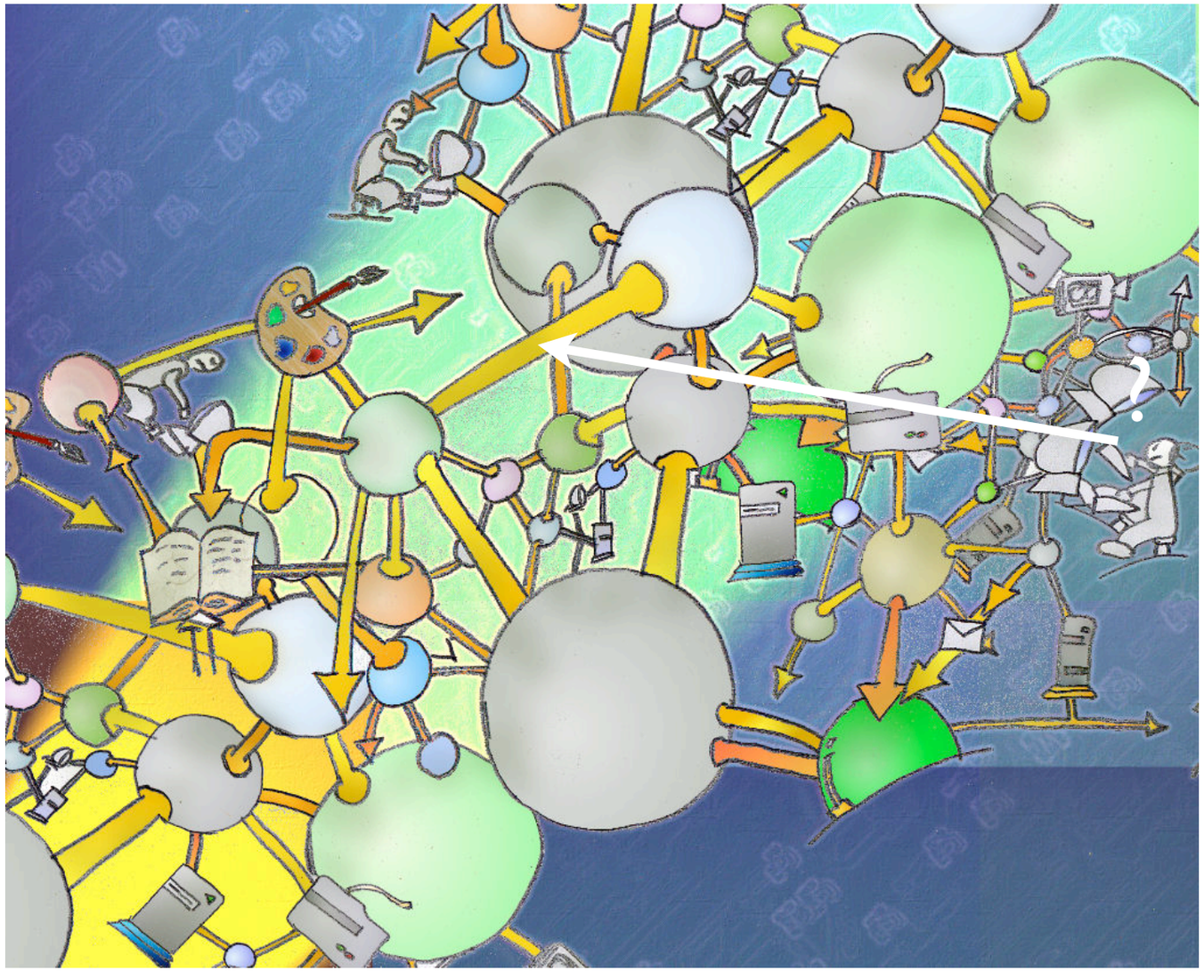
Learning in hyperlinked environments

"There are finer fish in the sea that have ever been caught," Irish proverb



Marco Gori
University of Siena (Italy)

ECIR-2007 Rome, 5 April 2007

Outline

- Beyond link analysis
- Diffusion learning
- Impact on information retrieval
- Conclusions

Beyond Link Analysis

ECIR-2007 Rome, 5 April 2007

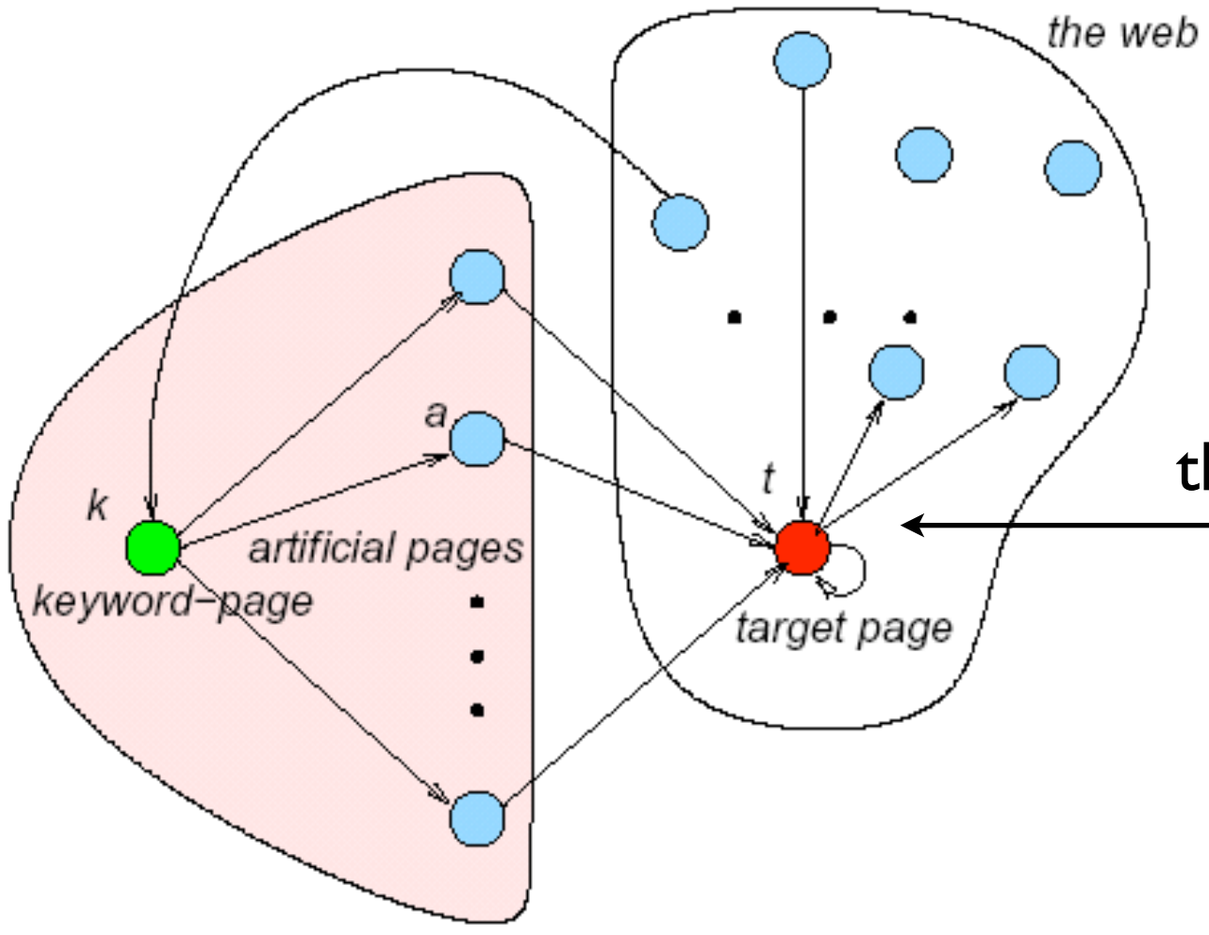
Link Analysis

By analysing how pages link to each other, determine what a page is about and whether that page is deemed to be “relevant” (does it deserve a ranking boost?)

- PageRank (Page and Brin), HITS (Kleimberg), TrustRank (Gyongy, Garcia-Molina, and Pedersen)

topological analysis only!

Boosting PageRank:
Link farms and the SEO/Search Engine "war"



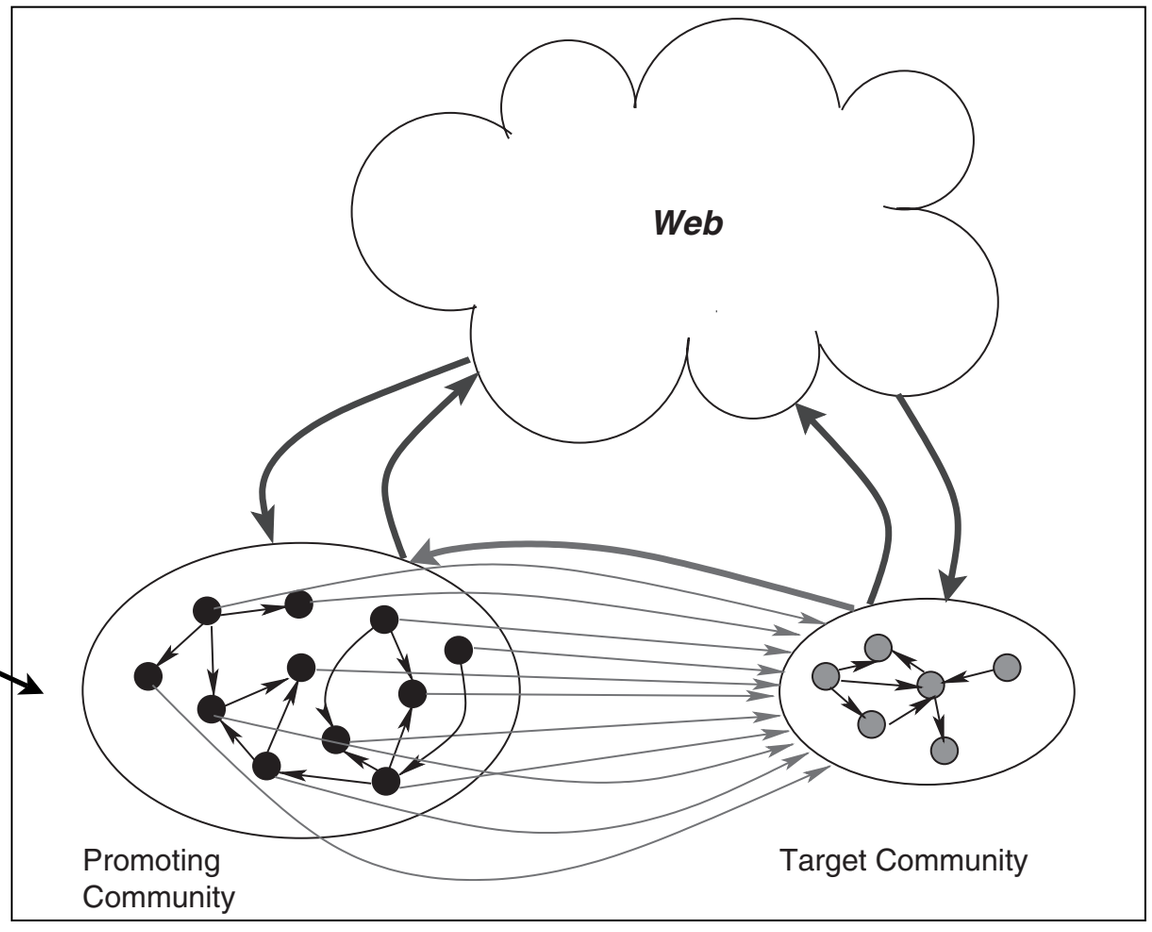
Can I combact spam
by detecting the topology of the "promoting community"?

Gori&Witten,
"The bubble of Web visibility", CACM (2005)

ECIR-2007 Rome, 5 April 2007

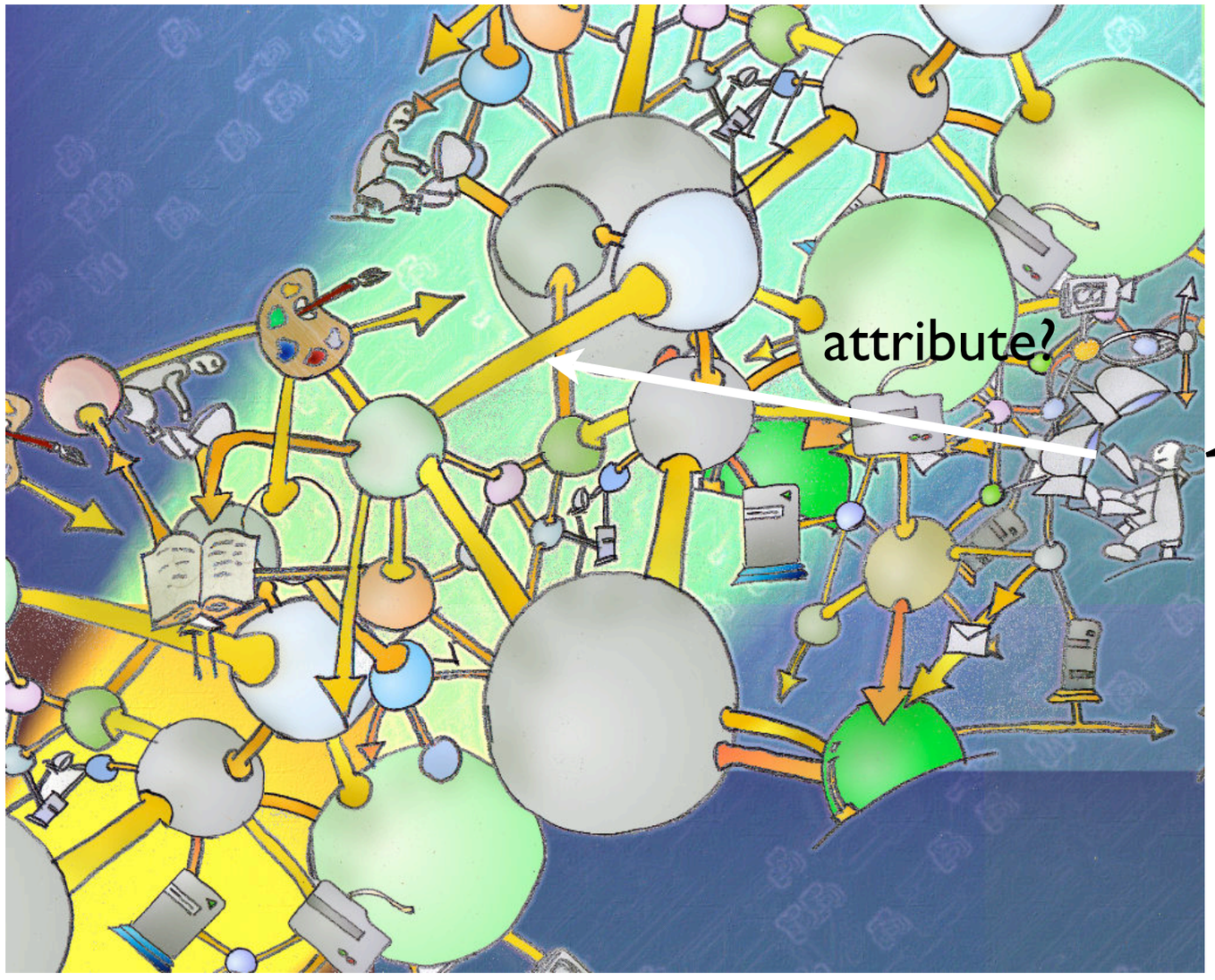
Link farms cannot be detected

Bianchini, Gori, Scarselli, ACM-TOIT, 2005 "Inside PageRank"
spam is independent
of the topology of the
promoting community!



Link analysis
is not enough!

Do I trust links?



No, I don't trust it, I need an attribute!

Learning in hyperlinked environments

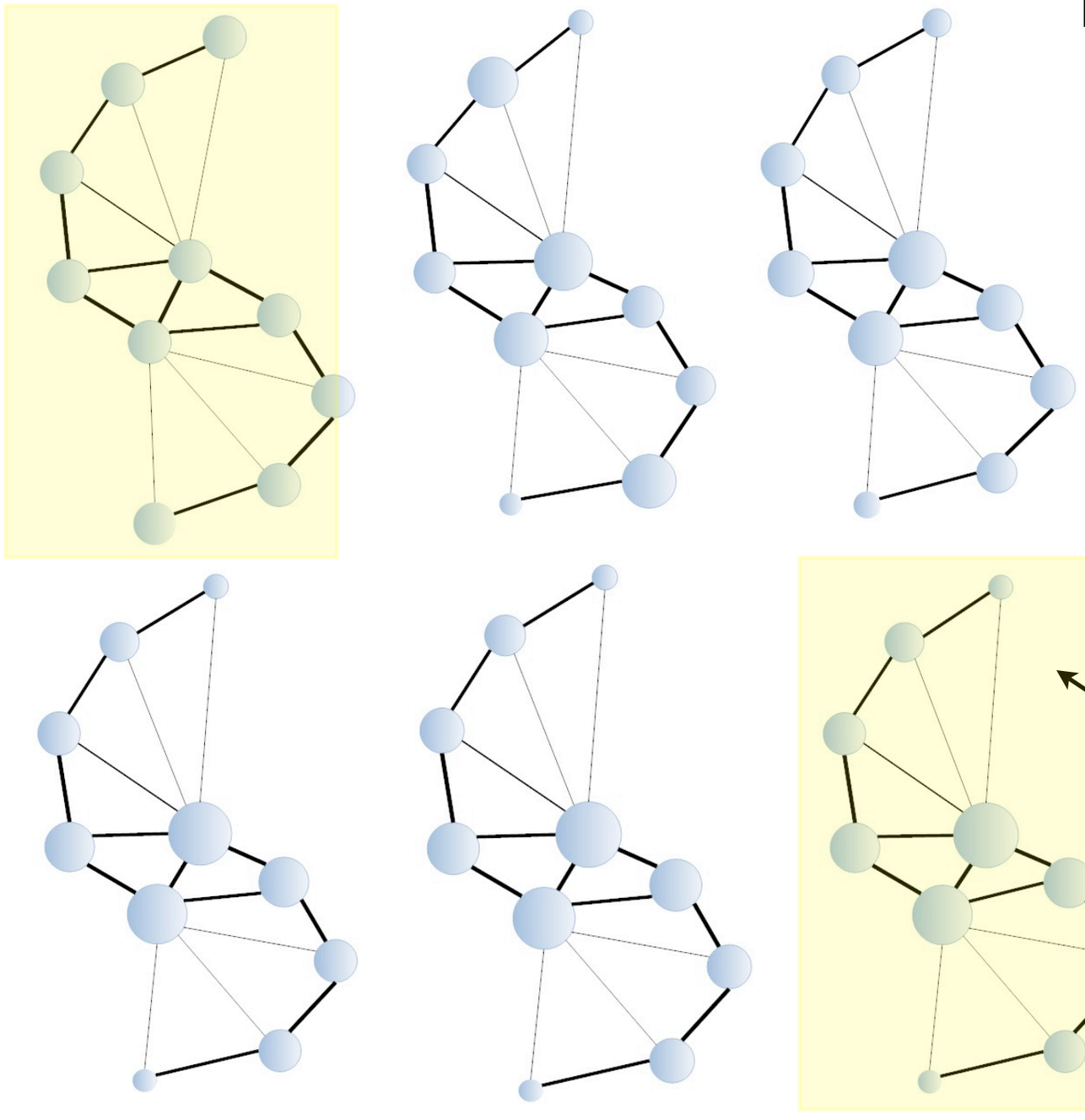
looks for methods to discover appropriate attributes for links (categorical or numerical)

Diffusion Learning

ECIR-2007 Rome, 5 April 2007

Diffusion machines

init: same rank



PageRank: A noticeable example

$$x_p = d \sum_{q \in pa[p]} \frac{x_q}{h_q} + (1-d)$$

$$\boldsymbol{x} = d\boldsymbol{W}\boldsymbol{x} + (1-d)\mathbb{1}_N$$

$$\boldsymbol{x}(t) = d\boldsymbol{W}\boldsymbol{x}(t-1) + (1-d)\mathbb{1}_N$$

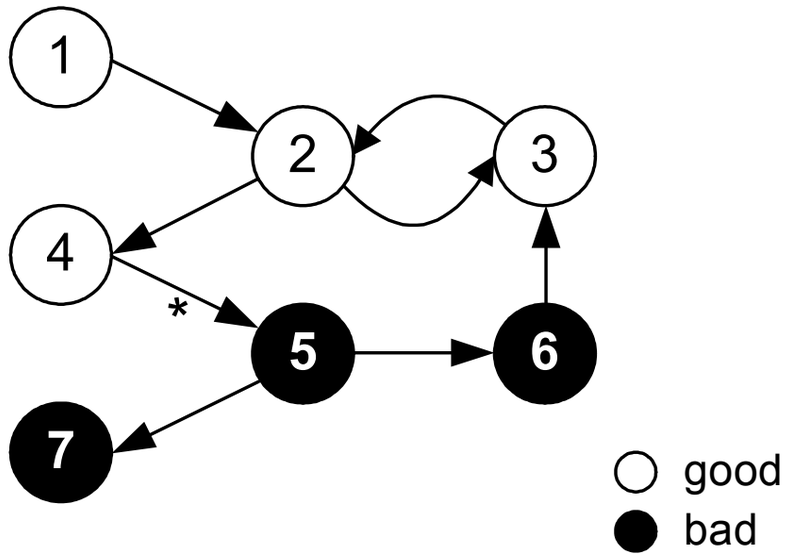
Social nets

random walk

the efficiency issue

stationary distribution

TrustRank



Invoke the "oracle"

propagate from "good" pages

begin

// evaluate seed-desirability of pages

(1) $\mathbf{s} = \mathsf{SelectSeed}(\ldots)$

// generate corresponding ordering

(2) $\sigma = \mathsf{Rank}(\{1,\ldots,N\},\mathbf{s})$

// select good seeds

(3) $\mathbf{d} = \mathbf{0}_N$

for $i = 1$ to L do
  if $\mathsf{O}(\sigma(i)) == 1$ then
    $\mathbf{d}(\sigma(i)) = 1$

// normalize static score distribution vector

(4) $\mathbf{d} = \mathbf{d}/|\mathbf{d}|$

// compute TrustRank scores

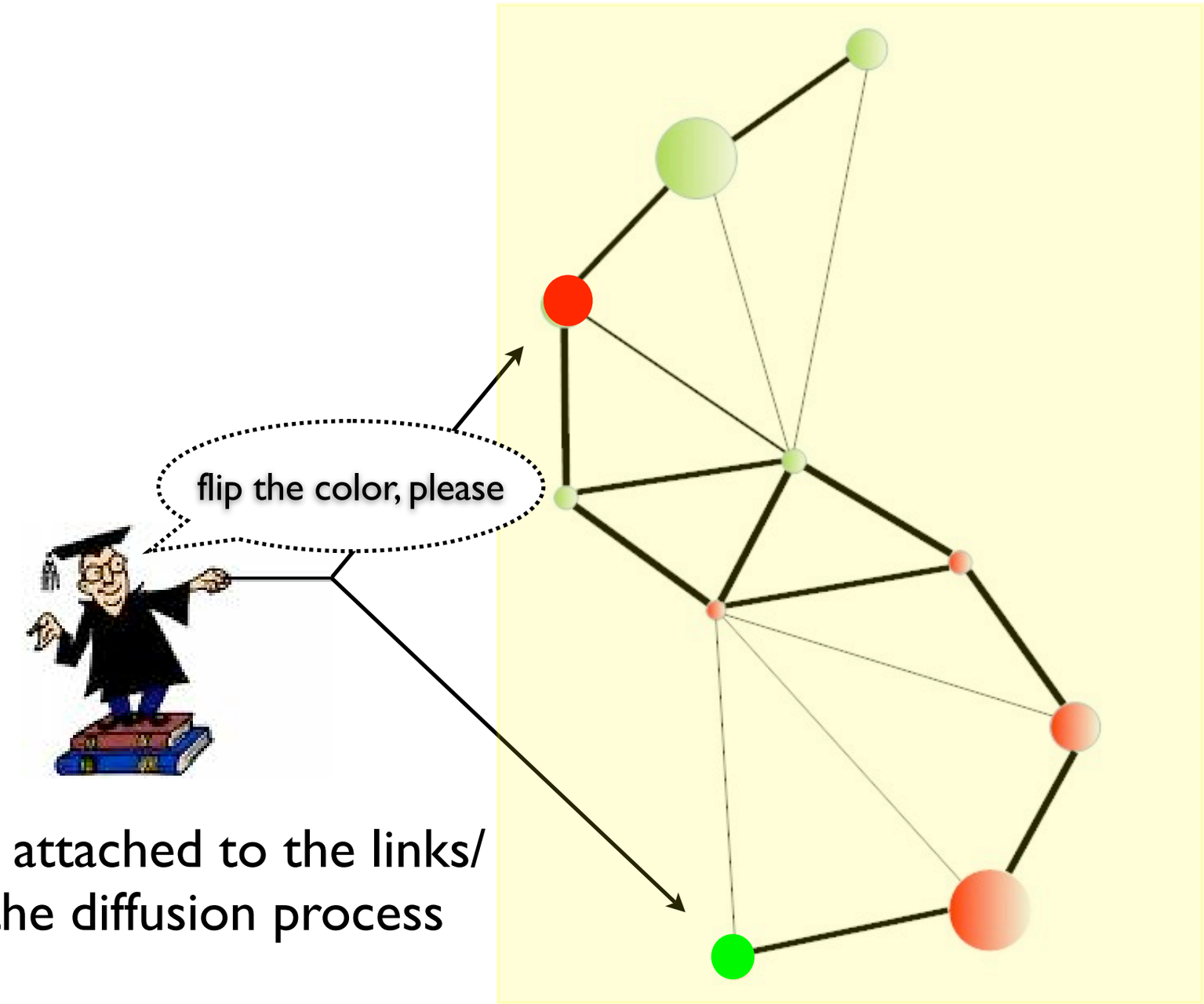
(5) $\mathbf{t}^* = \mathbf{d}$

for $i = 1$ to M_B do
  $\mathbf{t}^* = \alpha_B \cdot \mathbf{T} \cdot \mathbf{t}^* + (1-\alpha_B) \cdot \mathbf{d}$

return $\mathbf{t}^*$

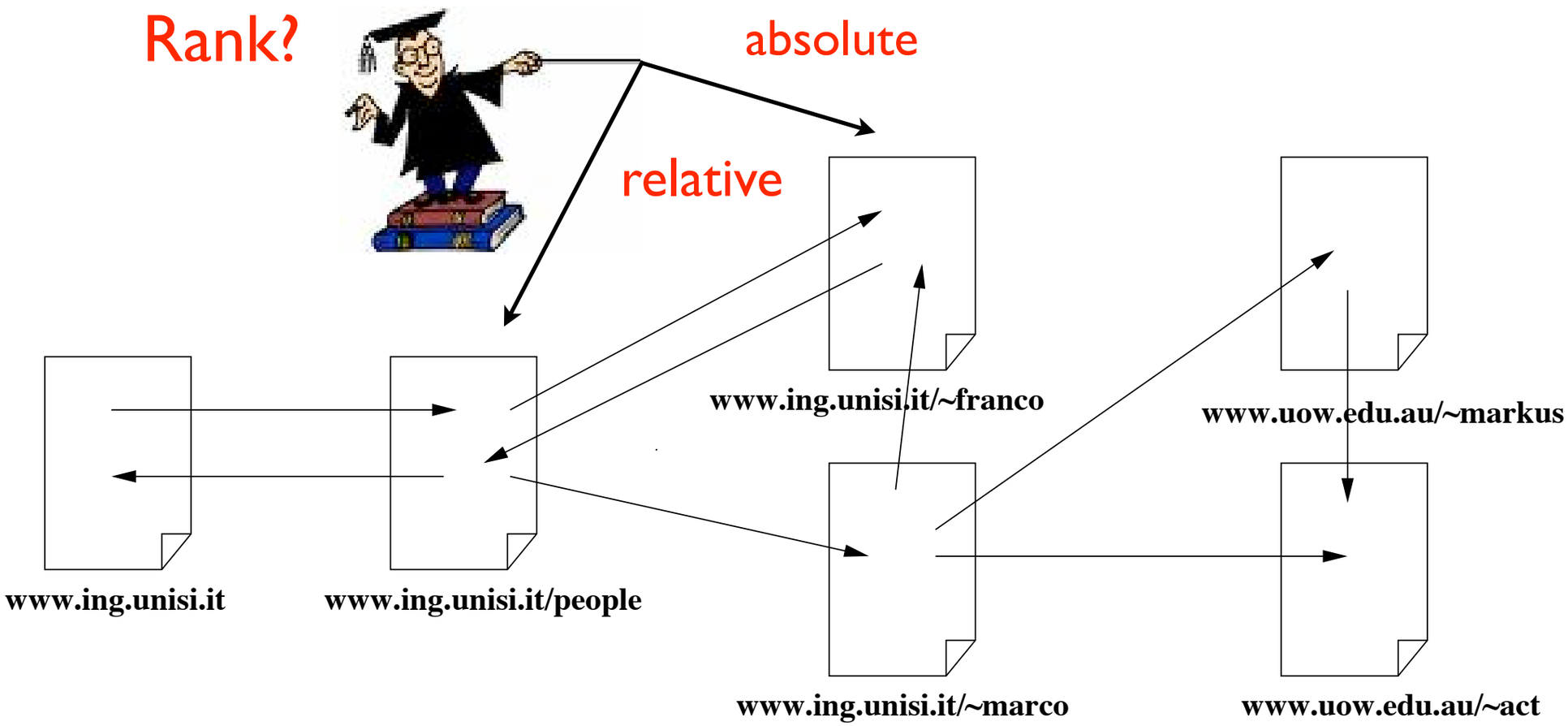
end

Diffusion Learning machines



the parameters attached to the links/ nodes affect the diffusion process

Learning to Rank Web Pages



Set a *supervision value* (topology & content)

we put supervision on "some pages" and ask inference on the rest ...
this is based on a learning machine that attaches attributes to links (trust)

ECIR-2007 Rome, 5 April 2007

Deeper Inside Diffusion ... and Learning

- Diffusion machines to compute functions on graphs
- Variational approach: machines induced by target tracking and regularization

ECIR-2007 Rome, 5 April 2007

Re-foundation of Calculus

(see e.g. A. Bensoussan & J.-L. Menaldi)

Difference Equations on Weighted Graphs (with A. Bensoussan). (*Journal of Convex Analysis (Special issue in honor of Claude LeMarechal), 12 (2005), pp. 13-44.*)

- Notions of boundary, Hilbert space of functions, norms & semi-norms (Sobolev spaces)
- Extentions of differential operators ... Green's formula
- Variational problems, Harmonic functions, Dirichlet & Neumann' problems
- Relationships with random walk

Harmolodic Functions

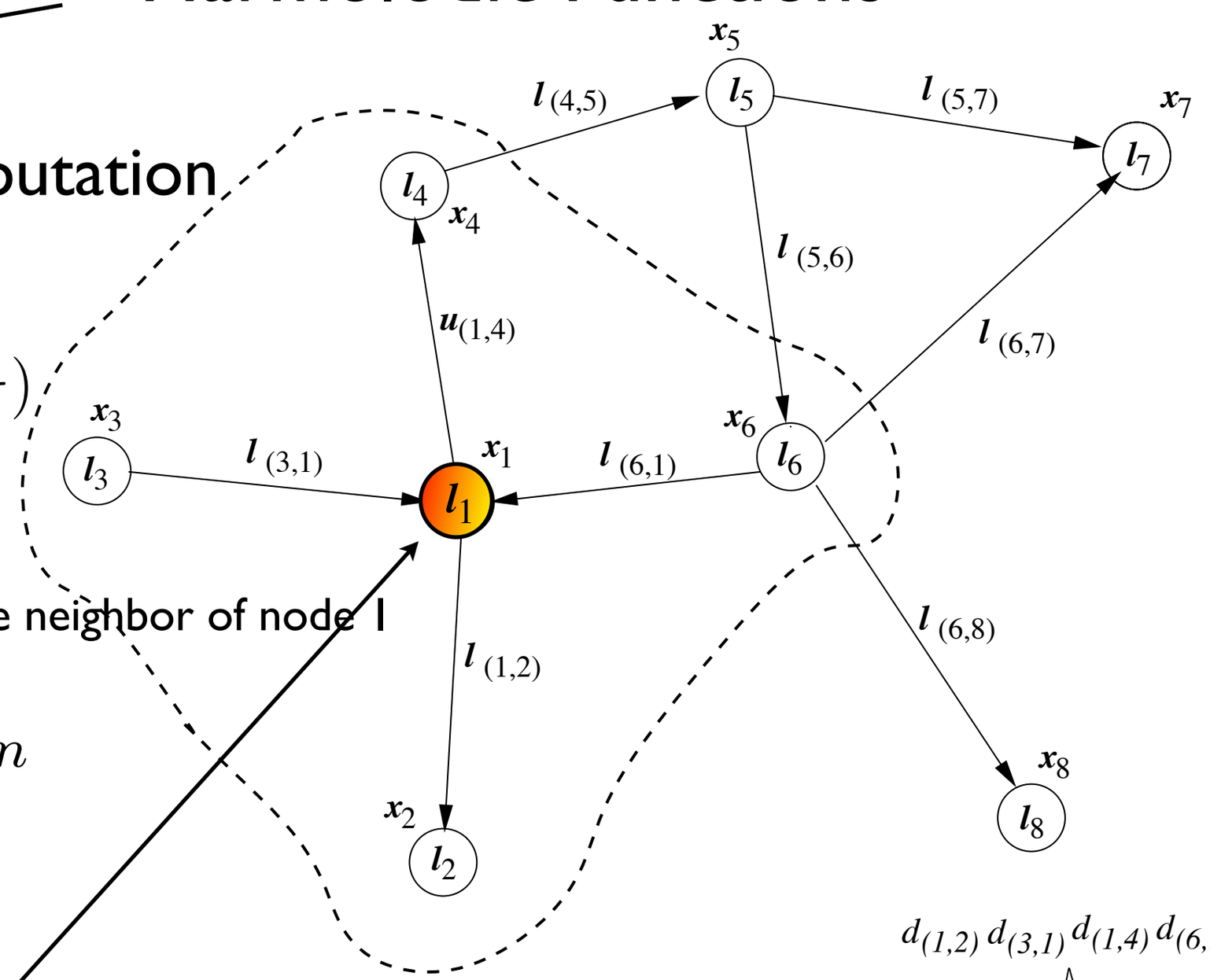
jazz saxophonist
Ornette Coleman

Local computation

$$\boldsymbol{x} = F_{\boldsymbol{w}}(\boldsymbol{x}, \boldsymbol{l})$$

$$\boldsymbol{o} = G_{\boldsymbol{w}}(\boldsymbol{x}, \boldsymbol{l}_{\boldsymbol{N}})$$

contraction



the neighbor of node l

$$\varphi_{\boldsymbol{w}}(\boldsymbol{G}, n) = \boldsymbol{o}_n$$

harmolodic function

$$\boldsymbol{x}_1 = f_{\mathrm{w}}\,(\boldsymbol{l}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_6, \boldsymbol{l}_{(1,2)}, \boldsymbol{l}_{(3,1)}, \boldsymbol{l}_{(1,4)}, \boldsymbol{l}_{(6,1)}, \boldsymbol{l}_2, \boldsymbol{l}_3, \boldsymbol{l}_4, \boldsymbol{l}_6, \boldsymbol{0}, \boldsymbol{1}, \boldsymbol{0}, \boldsymbol{1})$$

$d_{(1,2)}\ d_{(3,1)}\ d_{(1,4)}\ d_{(6,1)}$

$x_{ne[1]}$ $l_{co[1]}$ $l_{ne[n]}$

parametric dependence

coding the direction

Diffusion Machines

$$\boldsymbol{x}_n = \sum_{u \in \mathrm{ne}[n]} h_{\boldsymbol{w}}(\boldsymbol{l}_n, \boldsymbol{l}_{(n,u)}, \boldsymbol{x}_u, \boldsymbol{l}_u), \quad n \in \boldsymbol{N}$$

Something familiar ... $h_{\boldsymbol{w}}(\boldsymbol{l}_n, \boldsymbol{l}_{(n,u)}, \boldsymbol{x}_u, \boldsymbol{l}_u) = \boldsymbol{A}_{n,u}\boldsymbol{x}_u + \boldsymbol{b}_n$

$$\frac{\partial x(t)}{\partial t} = \chi \mathcal{L} x(t) + \beta \qquad \mathcal{L}x = b$$

Poisson equation

$$\mathcal{L}(h_{\boldsymbol{w}})\ \boldsymbol{x} = \boldsymbol{x} - \sum_{u \in \mathrm{ne}[n]} h_{\boldsymbol{w}}(\boldsymbol{l}_n, \boldsymbol{l}_{(n,u)}, \boldsymbol{x}_u, \boldsymbol{l}_u)$$

$$\frac{\partial \boldsymbol{x}(t)}{\partial t} = \mathcal{L}(h_{\boldsymbol{w}})\ \boldsymbol{x}(t)$$

convergence to a stationary configuration

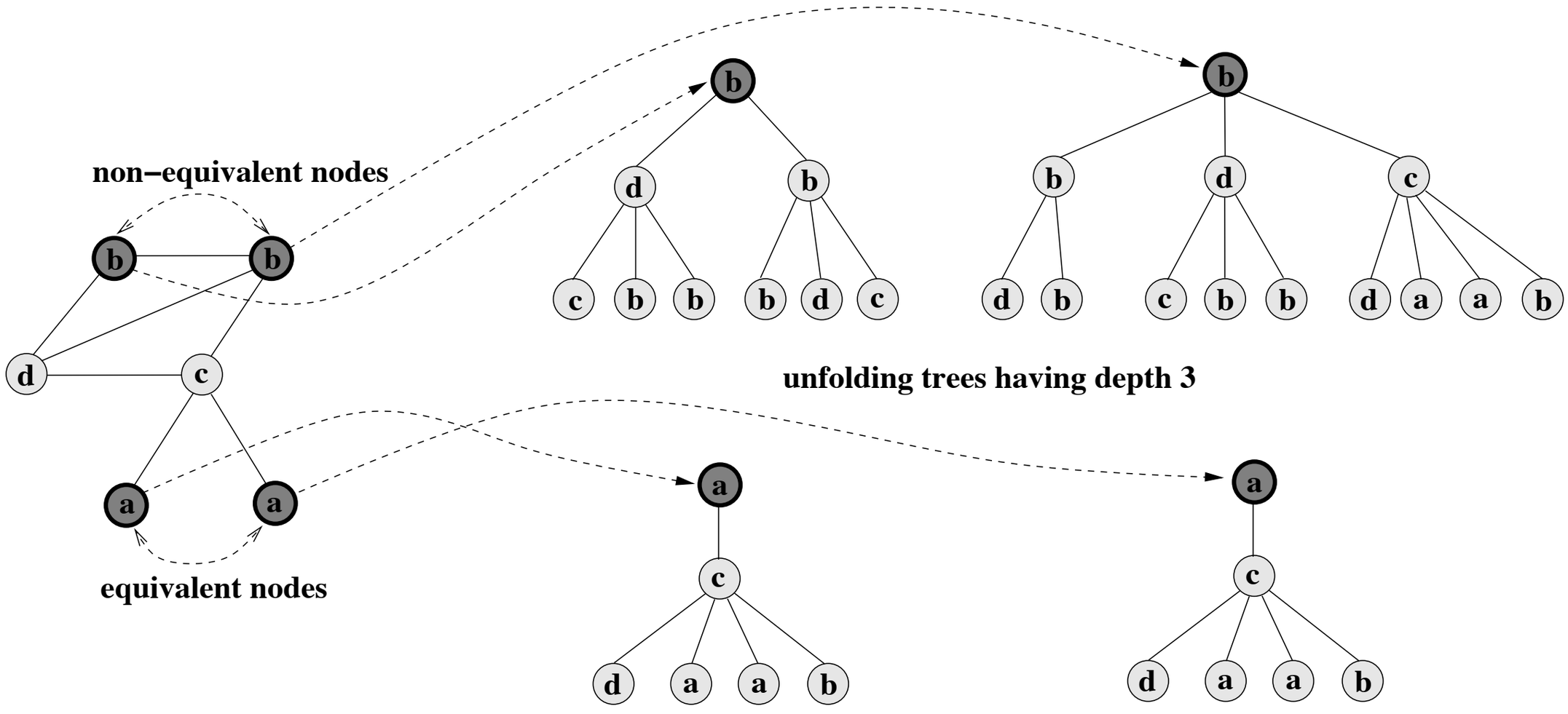
Representation Power of Harmolodic Functions

Crucial question

What are we missing with DLM (computing harmolodic functions)?

ECIR-2007 Rome, 5 April 2007

Unfolding Equivance



"Graph fibrations, graph isomorphism, and PageRank," Boldi et al
unfolding equivalence is equivalent to fibration prime graph
... crucial for the isomorphism problem!

ECIR-2007 Rome, 5 April 2007

Unfolding-Equivalence

Definition 3 *A function $l : \mathcal{G} \times \mathcal{N} \rightarrow \mathrm{R}^m$ is said to preserve the unfolding equivalence on $\mathcal{G}$, if $n \sim u$ implies $l(\boldsymbol{G}, n) = l(\boldsymbol{G}, u)$, for any nodes n, u of $n\mathcal{G}$. The set of functions that preserve the unfolding equivalence on $\mathcal{G}$ will be denoted by $\mathcal{F}(\mathcal{G})$.*

Theorem 4 APPROXIMATION BY POSITIONAL DLMS
Let $\mathcal{D}$ be a domain that contains positional graphs. For any measurable function $\tau \in \mathcal{F}(\mathcal{D})$ preserving the unfolding equivalence, any norm $\|\cdot\|$ on $\mathbb{R}^m$, any probability measure P on $\mathcal{D}$, and any reals $\varepsilon, \mu, \lambda$, where $\varepsilon > 0$, $0 < \lambda < 1$, $0 < \mu < 1$, there exist two continuously differentiable functions f and g such that

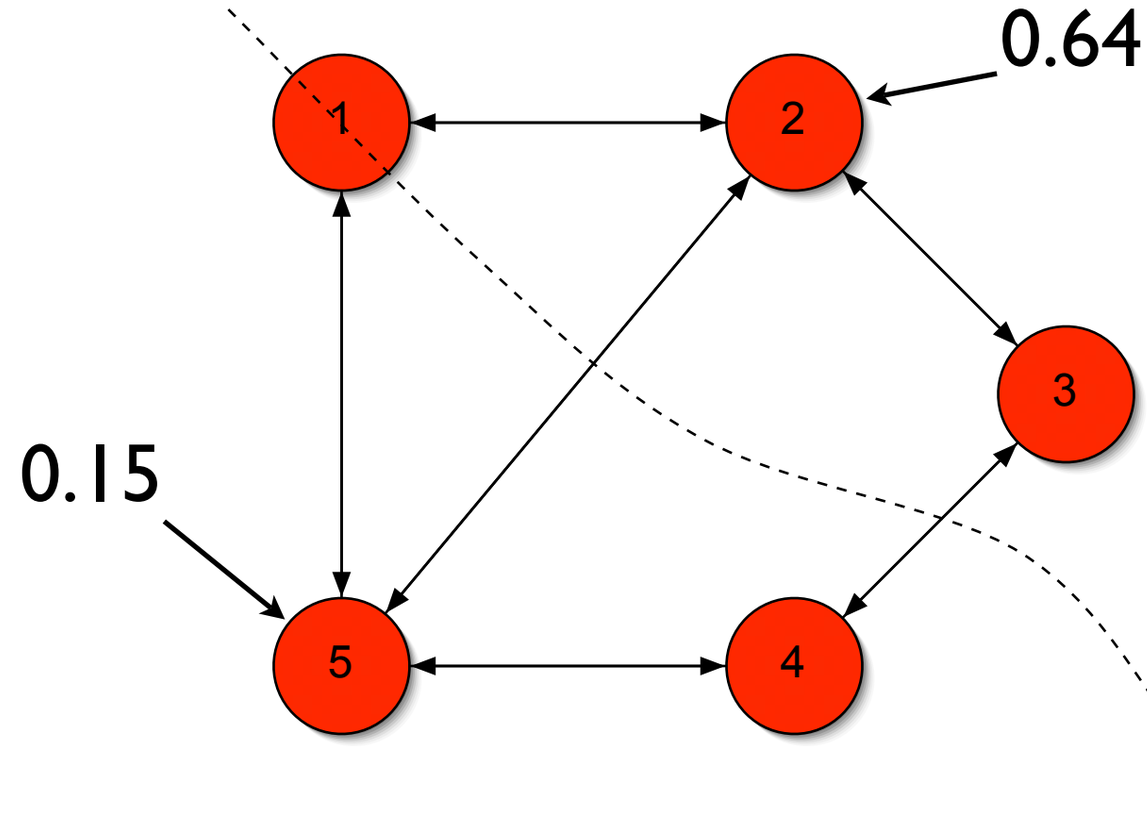
$$\begin{aligned} \boldsymbol{x}_n &= f(\boldsymbol{l}_n, \boldsymbol{l}_{\mathrm{co}[n]}, \boldsymbol{x}_{\mathrm{ne}[n]}, \boldsymbol{l}_{\mathrm{ne}[n]}) \\ \boldsymbol{o}_n &= g(\boldsymbol{x}_n, \boldsymbol{l}_n),\ n \in \boldsymbol{N}\ , \end{aligned}$$

the global transition function F is a contraction with contracting constant μ, the state dimension is $s = 1$, and the corresponding harmolodic function defined by $\varphi(\boldsymbol{G}, n) \doteq \boldsymbol{o}_n$ satisfies the condition

$$P(\|\tau(\boldsymbol{G}, n) - \varphi(\boldsymbol{G}, n)\| \geq \varepsilon) \leq 1 - \lambda$$

What Cannot Be Calculated?

Warning!
The symmetry constraints



Only 3 different values ...

Impossible!
The function doesn't preserve unfolding equivalence

Diffusion Learning Machines (DLM)

$$\begin{aligned} \boldsymbol{x}_n(t+1) &= f_{\boldsymbol{w}}(\boldsymbol{l}_n, \boldsymbol{l}_{\mathrm{co}[n]}, \boldsymbol{x}_{\mathrm{ne}[n]}(t), \boldsymbol{l}_{\mathrm{ne}[n]}) \\ \boldsymbol{o}_n(t) &= g_{\boldsymbol{w}}(\boldsymbol{x}_n(t), \boldsymbol{l}_n), \end{aligned}$$

learning parameters

discrete-time version

$$\frac{\partial \boldsymbol{x}(t)}{\partial t} = \mathcal{L}(h_{\boldsymbol{w}})\, \boldsymbol{x}(t)$$

The dynamic evolution of DLM yields harmolodic functions ...
convergence to the fixed point of the map

ECIR-2007 Rome, 5 April 2007

Semi-Supervised Framework

quadratic error w.r.t. the target

$$e_{\boldsymbol{w}} = \sum_{i=1}^{q} (\boldsymbol{t}_i - \varphi_{\boldsymbol{w}}(\boldsymbol{G}, n_i))^2$$

entropy-based measure for ranking:
Burges et al - RankNet, ICML05

$$C_{ij} \equiv C(o_{ij}) = -\bar{P}_{ij} \log P_{ij} - (1 - \bar{P}_{ij}) \log (1 - P_{ij})$$

$$P_{ij} \equiv \frac{e^{o_{ij}}}{1 + e^{o_{ij}}}$$

From "B&W to color TV"

traditional machine learning framework

towards graphical domains

1 MAIN

2 initialize $\boldsymbol{w}$;

3 $\boldsymbol{x}$=FORWARD($\boldsymbol{w}$);

4 **repeat**

5 $\frac{\partial e_{\boldsymbol{w}}}{\partial \boldsymbol{w}}$=BACKWARD($\boldsymbol{x}, \boldsymbol{w}$);

6 $\boldsymbol{w}=\boldsymbol{w} - \lambda \cdot (\frac{\partial e_{\boldsymbol{w}}}{\partial \boldsymbol{w}})'$;

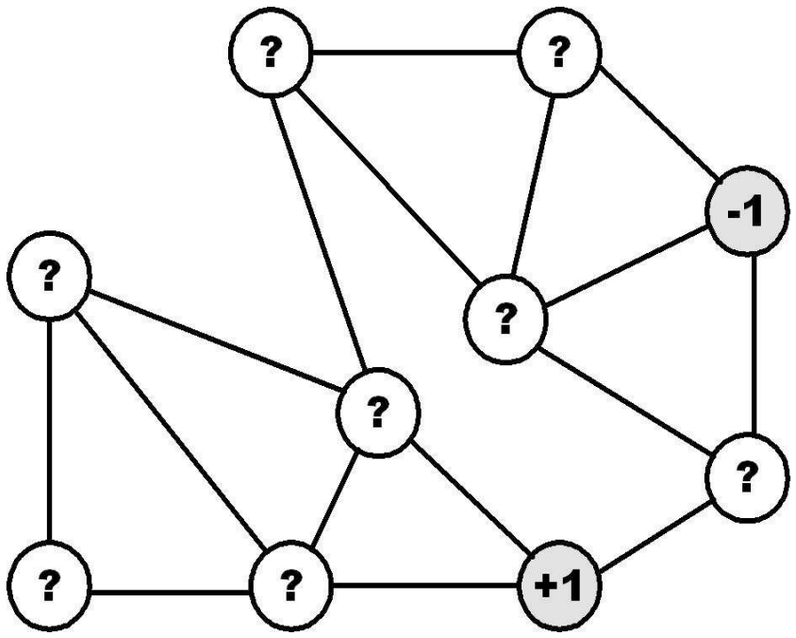
7 $\boldsymbol{x}$=FORWARD($\boldsymbol{w}$);

8 **until** the stopping criterion is met;

9 return $\boldsymbol{w}$;

10 **end**

A variational approach (Zhou & Schoelkopf, 2004)



$$\arg \min_{f \in L^2(V)} \left\{ \mathcal{S}(f) + \frac{\mu}{2} \|f - y\|^2 \right\}$$

error w.r.t. target

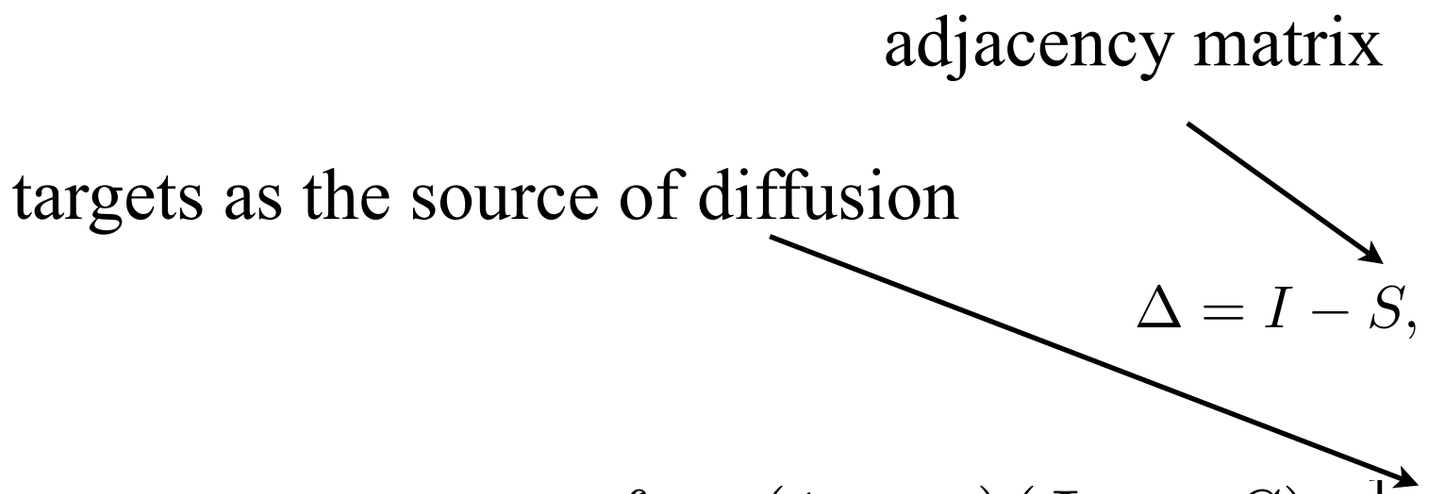
regularization term

$$\mathcal{S}(f) = \frac{1}{2} \sum_v \|\nabla_v f\|^2.$$

$$\left. \frac{\partial f}{\partial e} \right|_u = \sqrt{\frac{w(u,v)}{d(u)}} f(u) - \sqrt{\frac{w(u,v)}{d(v)}} f(v).$$

An elegant solution

Theorem 1. *The solution of the optimization problem (2.9) satisfies* $\Delta f + \mu(f - y) = 0$.



$$f = (1 - \alpha)(I - \alpha S)^{-1} y$$

$$f(v) \leftarrow \alpha(Sf)(v) + (1 - \alpha)y, \ \forall v$$

A PageRank-like solution!

One step further ...
at airgroup.unisi.it

initial heuristic value

$$\begin{cases} \varphi^e = \arg\min_{\varphi \in \mathcal{H}(V)} \left\{ S_2(\varphi, w^e) + \frac{\mu \|\varphi - y\|^2}{2} \right\} \\ w^{e+1} = \arg\min_{w \in \mathcal{H}(E)} \left\{ S_2(\varphi^e, w) + \frac{\mu \|w - \bar{w}\|^2}{2} \right\} \end{cases}$$

weight refinement

Impact on information retrieval

ECIR-2007 Rome, 5 April 2007

What is *diffusion learning* for?

- pattern classification: Hancock, Bekin, ...
- link prediction: Zhou&Schoelkopf, SRL04
- network evolution: Barabasi (Linked) et al ...
- community detection (M.B. Hasting, Physical Review 06) ...
- ...

Ranking

- Tsoi et al ACM-TOIT 2007
- Diligenti et al IJCAI05
- Scarselli et al WIC05
- Chakabarti et al KDD06

airgroup.dii.unisi.it

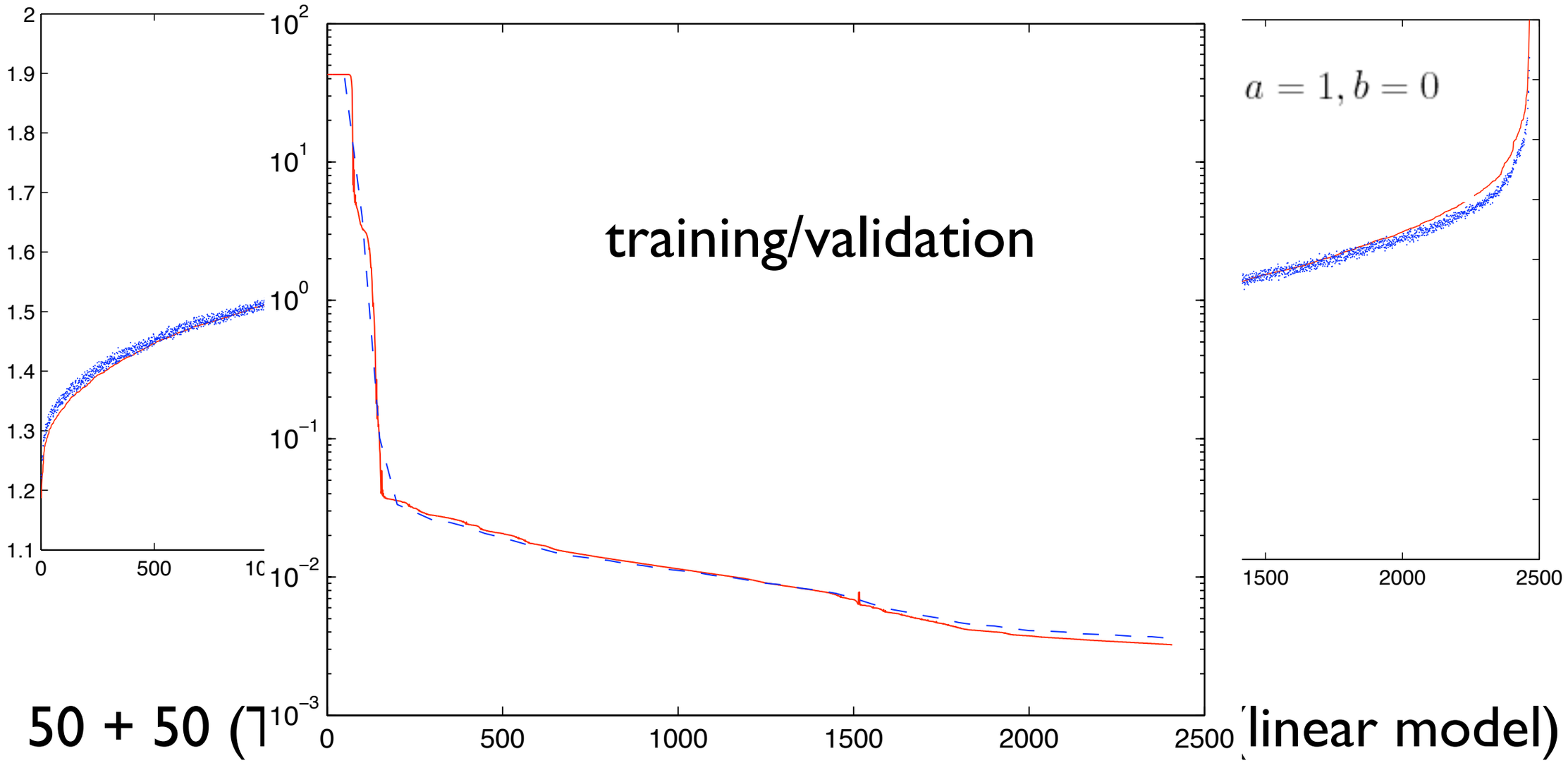
- Learning to rank papers (ACM portal)
- HITS (Kleimberg, 1998)
- PageRank-like functions

links only

links and content!

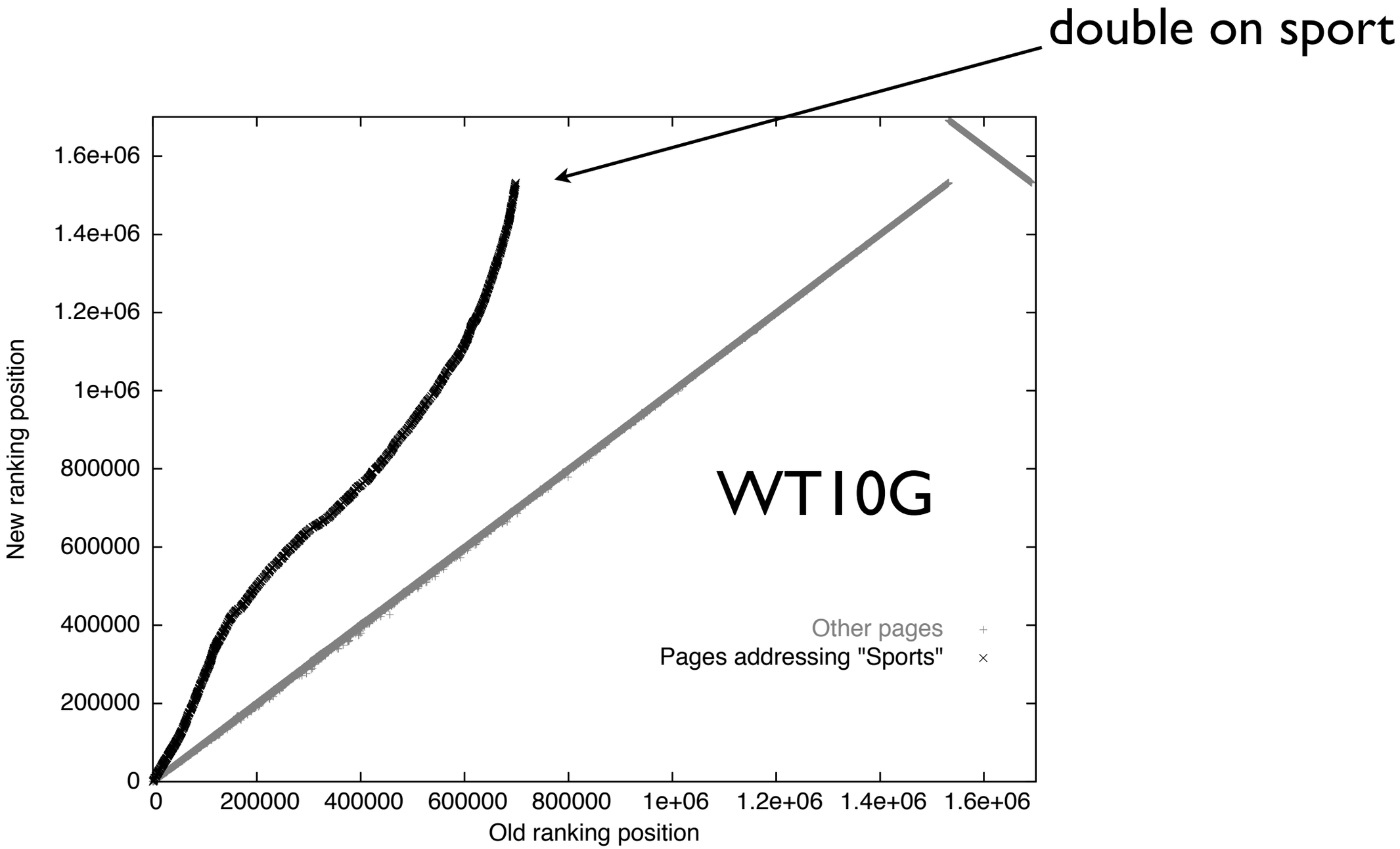
Learning PR-like Functions

Learn the PageRank (PR)
Each node has a label with 2 Booleans

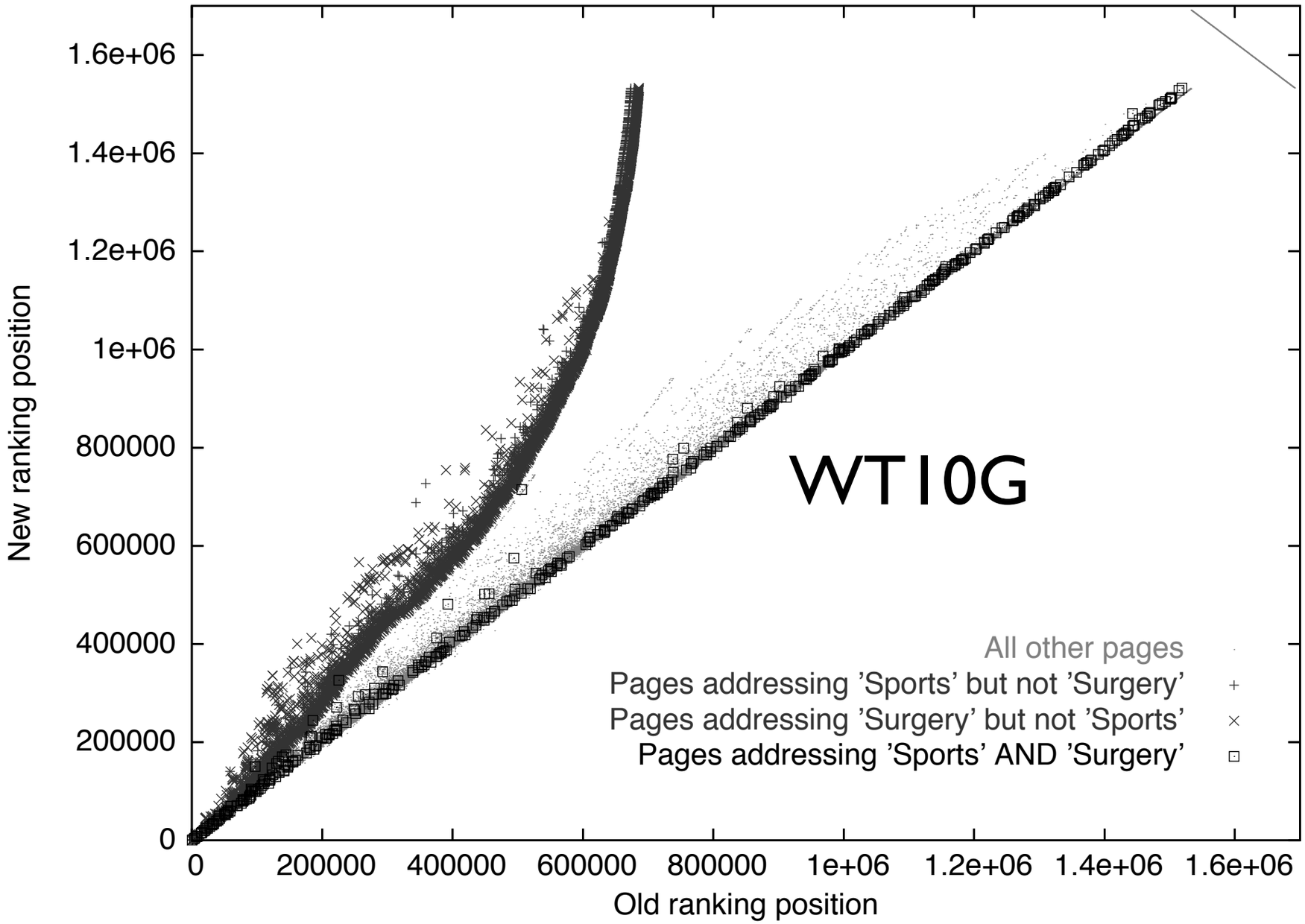


50 + 50 (linear model)

more focus on a selective topic



The “XOR”


WT10G
All other pages
Pages addressing ’Sports’ but not ’Surgery’
Pages addressing ’Surgery’ but not ’Sports’
Pages addressing ’Sports’ AND ’Surgery’
New ranking position
Old ranking position
0
200000
400000
600000
800000
1e+06
1.2e+06
1.4e+06
1.6e+06

Conclusions

- Beyond link analysis: A general framework to build functions on the Web (content and hyperlinks)
- Links with relational learning, Dagstuhl seminar (15-20 April) on "Probabilistic, Logical, and Relational Learning"
- An interesting perspective: Facing the link trust problem (Web spam, ... harder time for spammers)
- A lot of work for choosing the "right trade-off" ... see e.g. complexity issues in Domingo et al (QD-PageRank)

Acknowledgements

- Research at Univ. of Siena (Franco Scarselli, Gabriele Monfardini, Augusto Pucci, Vicenzo Di Massa)
- Markus Hagenbuchner, Linus Yong (Univ. of Wollongong, Australia)
- Ah Chung Tsoi (Monasch Univ., Melbourne, Australia)
- Hendrik Blockeel & Werner Uwents Université catholique de Louvain (Beligium)
- Paolo Frasconi (Univ. of Florence, Italy)
- Alessandro Sperduti (Univ. of Padua, Italy)
- Yoshua Bengio (Université de Montréal, Canada)

ECIR-2007 Rome, 5 April 2007

Thanks for your attention!

ECIR-2007 Rome, 5 April 2007