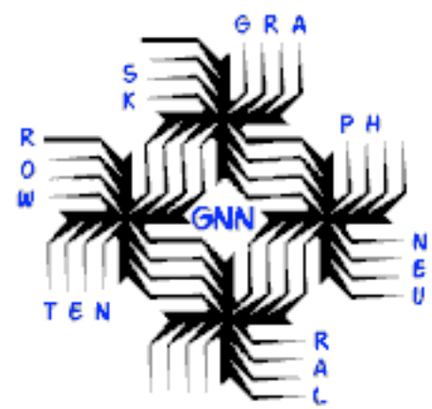
Learning Functions in Graphical Domains



Marco Gori
University of Siena (Italy)

DaMn Bertinoro 2005

Outline

- Learning functions in graphical domains (graphs in data ... graphs in functions) - motivations
- Graphical Neural Networks
- Experimental results

Paradigm Shift?

- Learning functions defined on the nodes (outputs on nodes, arcs)
- The learning environment:
  - A *unique graph* (a collection of "nodes" can always be regarded as a special disconnected graph)
  - Graphs to express the examples and the relationships between the examples

Re-foundation of Calculus

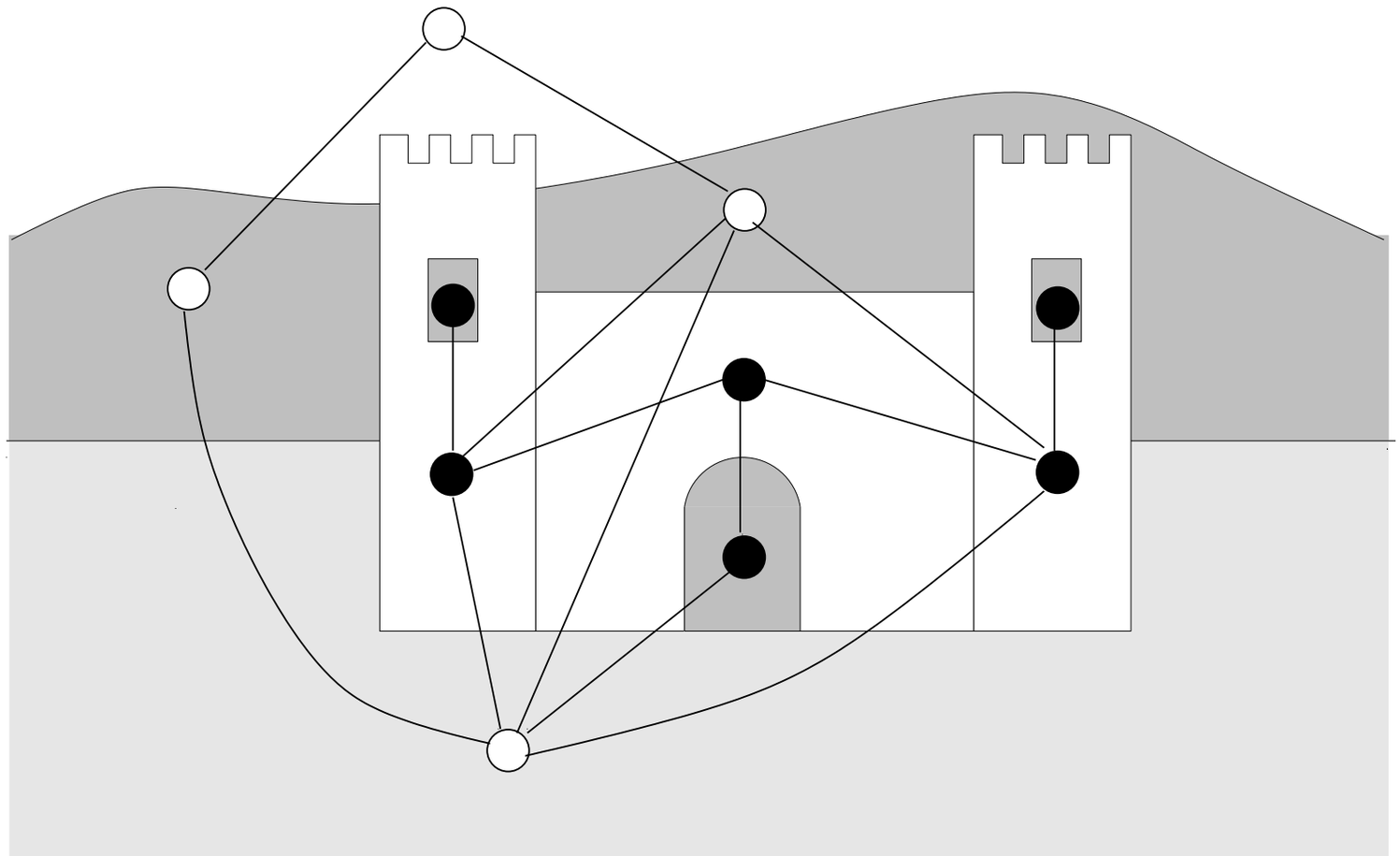
(see e.g. A. Bensoussan & J.-L. Menaldi)

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

What's a Castle?

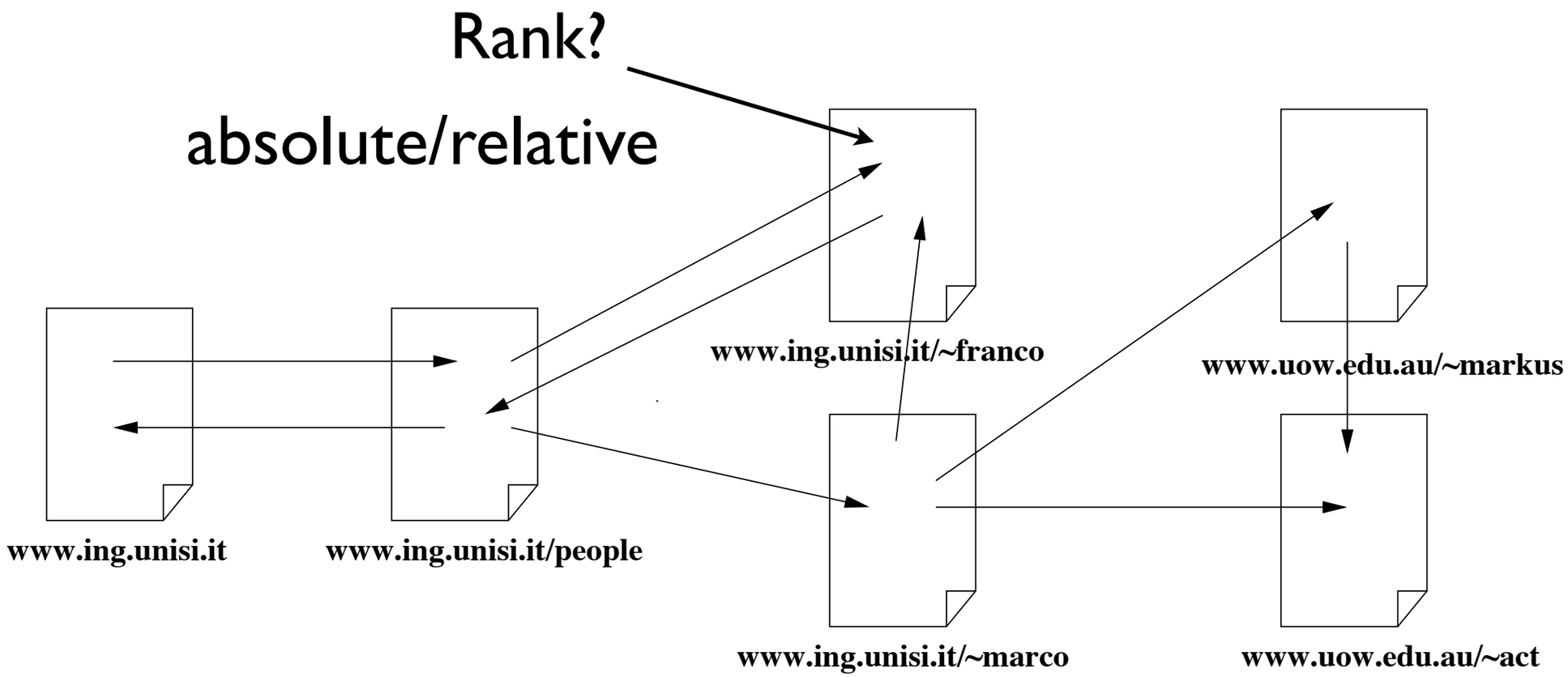
Supervision: which parts compose the castle? (back nodes)



Segmentation and recognition:
perhaps ... it's better not to separate in two steps ...

DaMn Bertinoro 2005

Ranking Web Pages



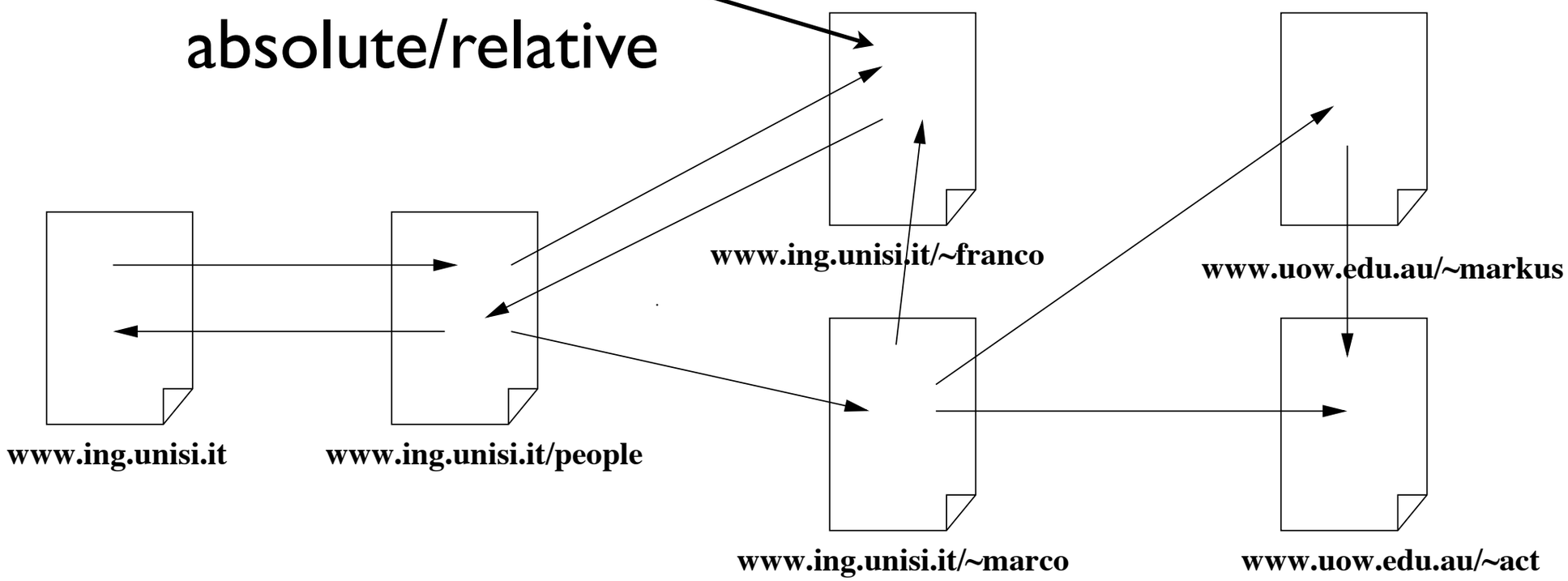
set a *supervision value* (topology & content)

we put supervision on "some nodes" and ask inference on the rest ...
if I can set relationships ... the same problem holds everywhere
The distinction learning set/test **has a different meaning!**

Web Spam?

is this spam?

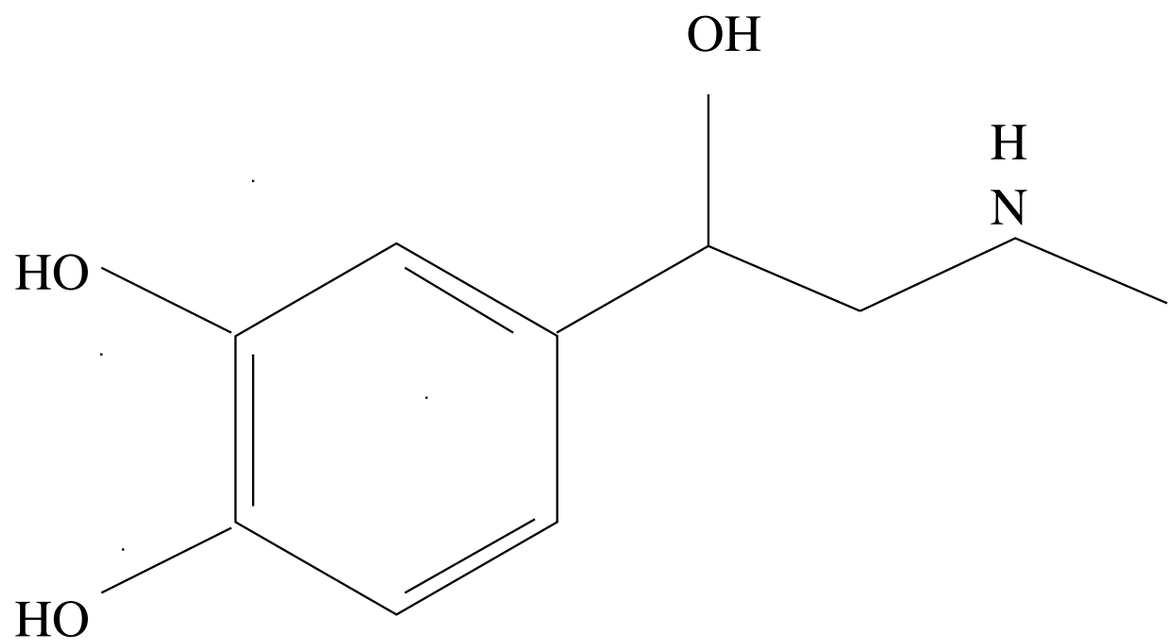
absolute/relative



Again: set a *supervision value* (topology & content)

Quantitative structure activity relationship (QSAR)

Unlike previous two examples ... a global property



DaMn Bertinoro 2005

Formal Statement

$\mathcal{L} = (\boldsymbol{G}, \boldsymbol{T})$ where $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{E})$ is graph

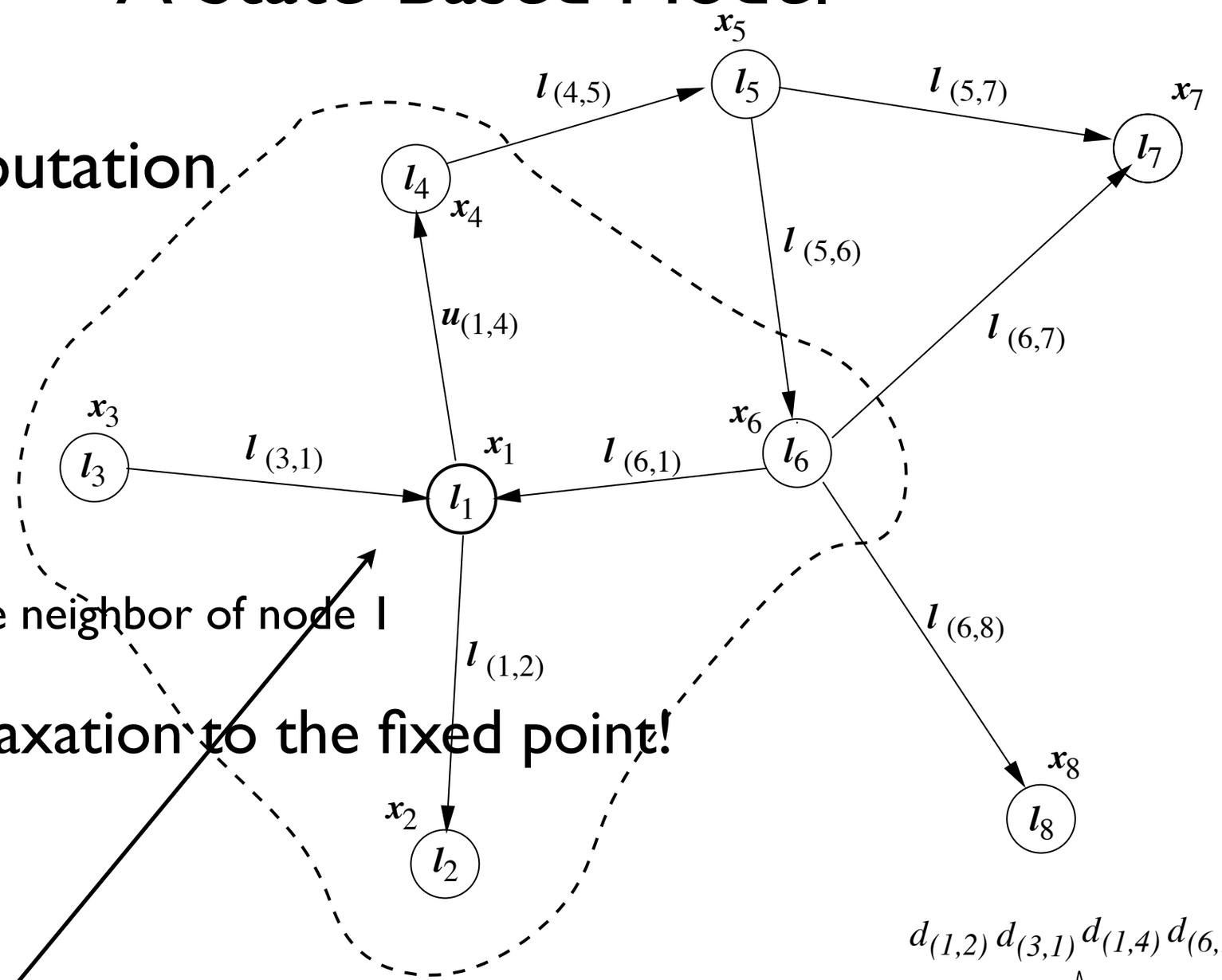
$\boldsymbol{T}$ a is set of pairs $\{(n_i, \boldsymbol{t}_i) | \, n_i \in N, \, \boldsymbol{t}_i \in \mathrm{R}^m, 1 \leq q\}$

A unique graph!
A target on some nodes (arcs)

DaMn Bertinoro 2005

A State-Based Model

Local computation



the neighbor of node 1

We force relaxation to the fixed point!

$$x_1 = f_w(\underbrace{l_1, x_2, x_3, x_4, x_6}_{x_{ne[1]}}, \underbrace{l_{(1,2)}, l_{(3,1)}, l_{(1,4)}, l_{(6,1)}}_{l_{co[1]}}, \underbrace{l_2, l_3, l_4, l_6}_{l_{ne[n]}}, \overbrace{0, 1, 0, 1}^{d_{(1,2)}\, d_{(3,1)}\, d_{(1,4)}\, d_{(6,1)}})$$

DaMn Bertinoro 2005

What Does Learning Mean?

Banach fixed-point Theorem

target

output

$$e_{\boldsymbol{w}} = \sum_{i=1}^{q} (\boldsymbol{t}_i - \varphi_{\boldsymbol{w}}(\boldsymbol{G}, n_i))^2 + \beta L\left(\left\|\frac{\partial F_{\boldsymbol{w}}}{\partial \boldsymbol{x}}\right\|\right)$$

$L(y)$ is $(y-\mu)^2$ if $y > \mu$ and 0, otherwise, and the parameter $\mu \in (0,1)$

DaMn Bertinoro 2005

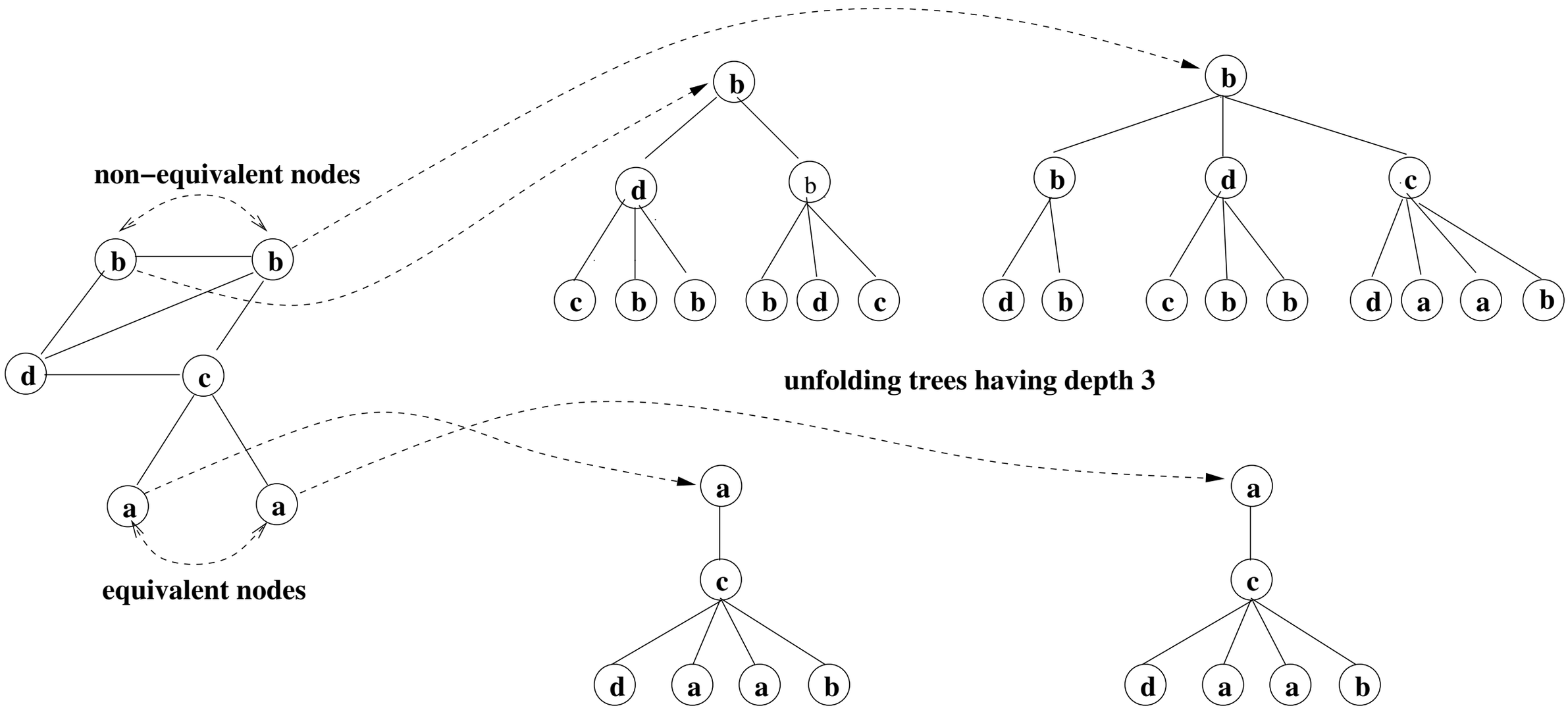
Computational Capabilities

Crucial question

The system dynamics is imposed by the “contraction map hyp”. Does it limit somehow the computational capabilities?

DaMn Bertinoro 2005

Unfolding Equivance



DaMn Bertinoro 2005

Preserving the Unfolding-Equivalence

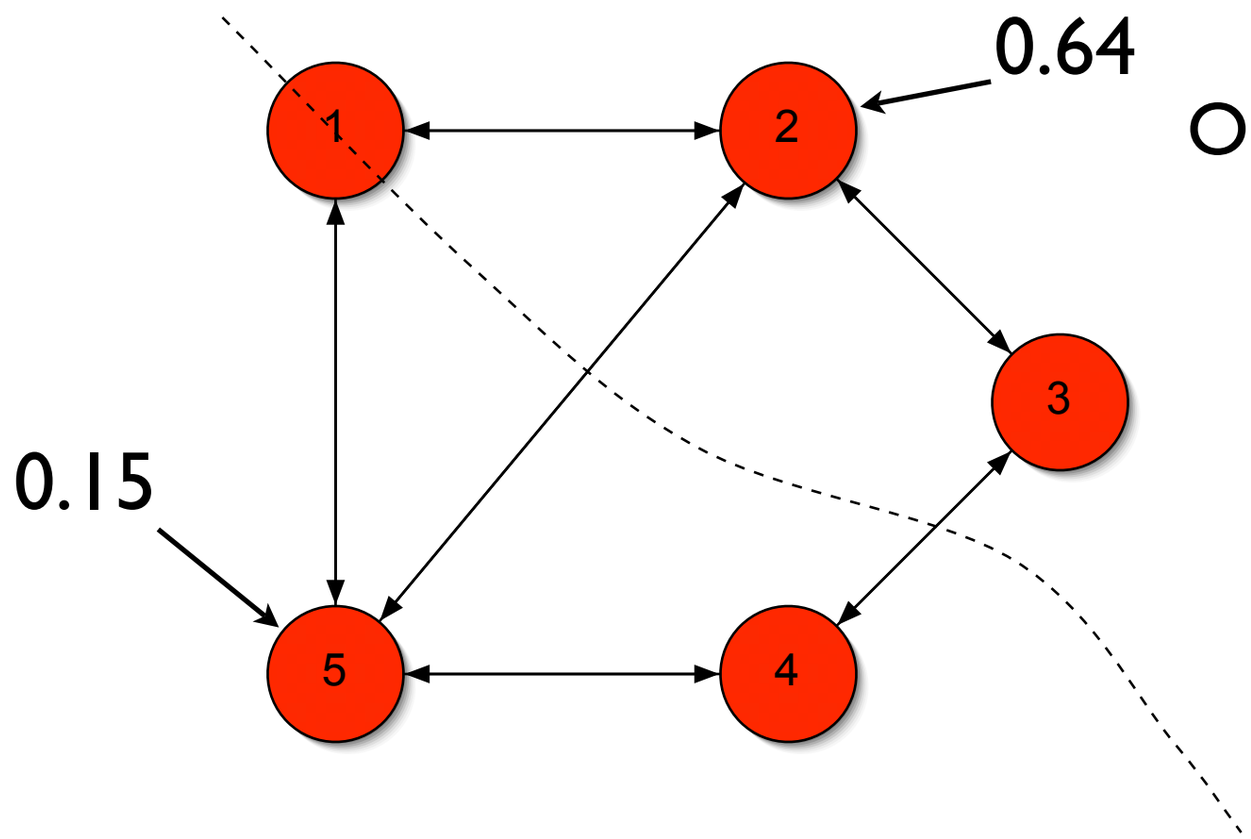
Definition 3 *A function* $l : \mathcal{G} \times \mathcal{N} \rightarrow \mathrm{R}^m$ *is said to preserve the unfolding equivalence on* $\mathcal{G}$*, if* $n \sim u$ *implies* $l(\boldsymbol{G}, n) = l(\boldsymbol{G}, u)$*, for any nodes* n, u *of* $n\mathcal{G}$*. The set of functions that preserve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

"Everything" preserving the unfolding-equivalence
can be calculated

DaMn Bertinoro 2005

What Cannot Be Calculated?

Warning!
The symmetry constraints



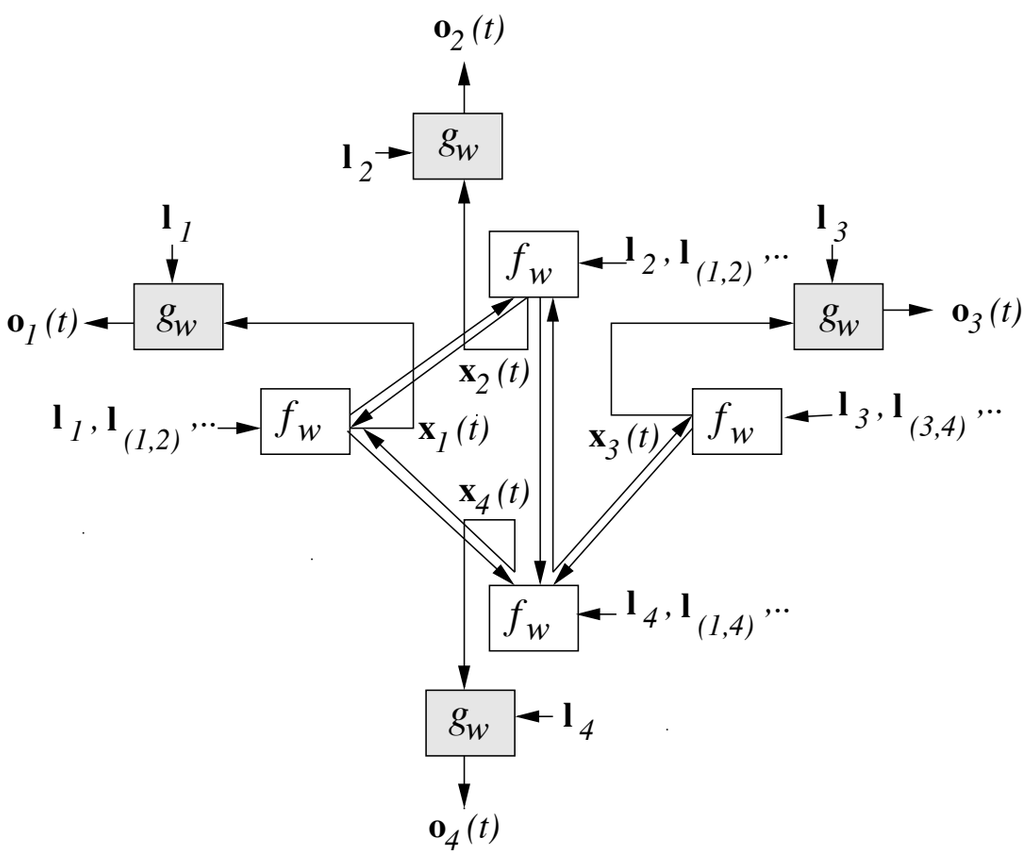
Only 3 different values ...

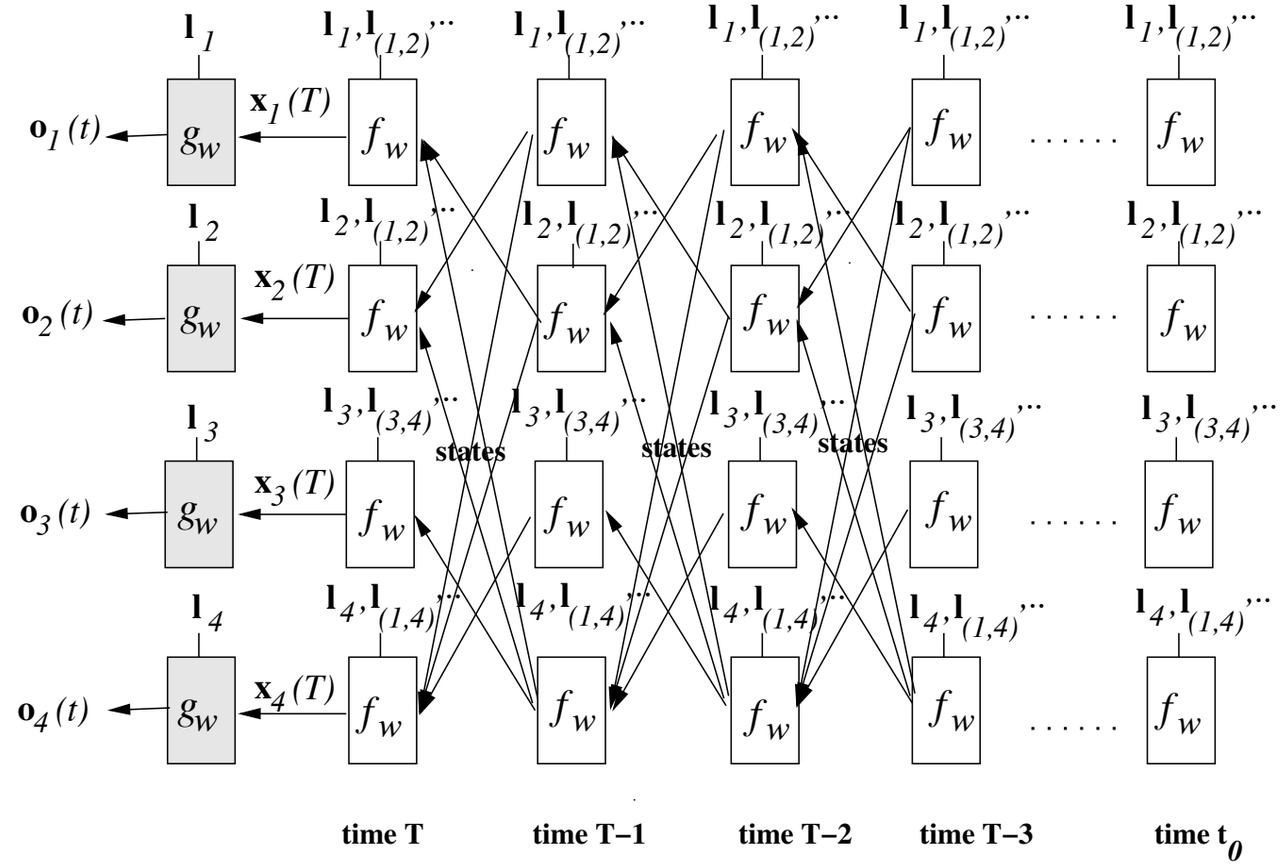
Impossible!
The function doesn't preserve unfolding equivalence

DaMn Bertinoro 2005

(Efficient) Gradient Computation

Classic Problem
Calculs of Variations, control literature
Almeida/Pineda (ANNs)





DaMn Bertinoro 2005

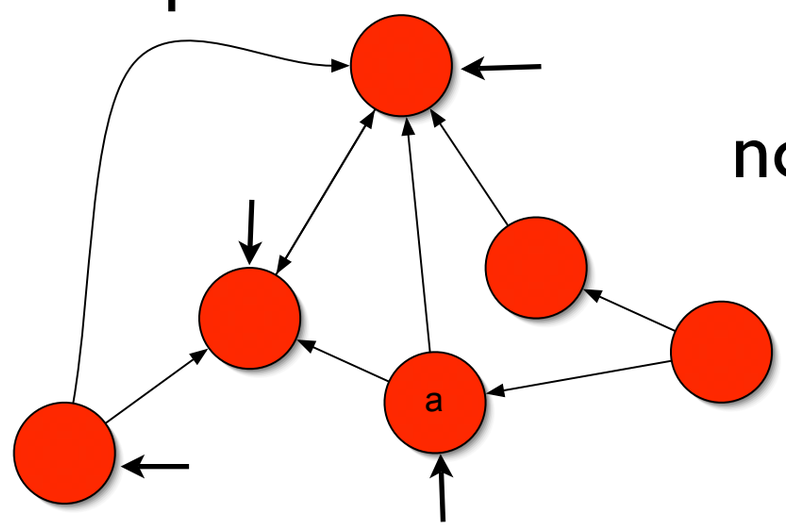
Experimental Results

Learning tasks in which the concept is based on:

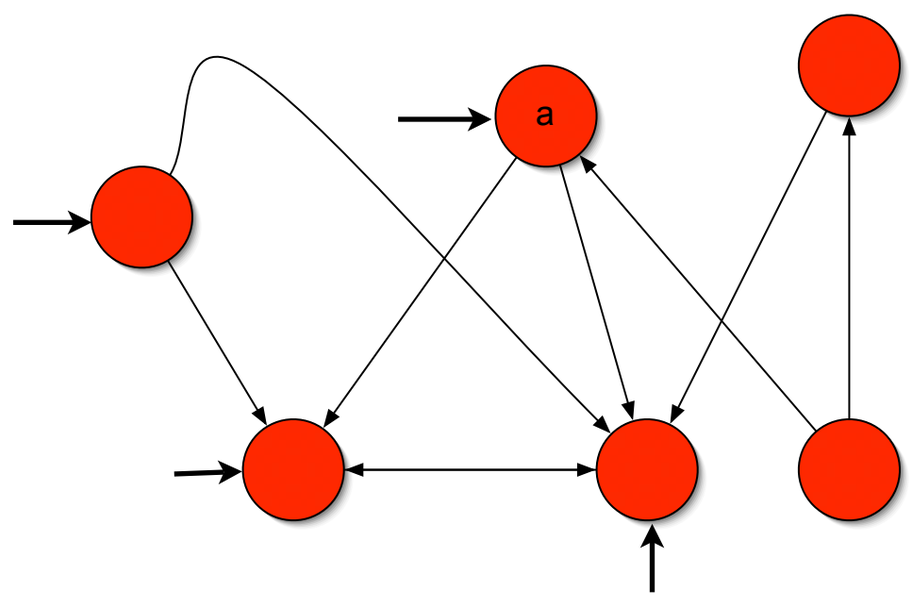
- the topologic properties only
- the content of the nodes only
- the joint dependence on both topology and content

The Clique Problem

"1" if the node is part of a 3-clique



no information in the nodes!



DaMn Bertinoro 2005

The Clique Problem

- 2000 random graphs, 20 nodes, 300 training set, 300 validation set
- 5-clique, 71.8% is "1"
- 2 states, 2-10 hidden units
- 84.6% on test set

Subgraph matching

Like for the "castle problem"

the whole picture

the "castle"

600 graphs (TS,VS,TS)
random labels, but S is in G
both label and links
s=5, h=5

around 10% more
G: 30 nodes, S: 5 nodes

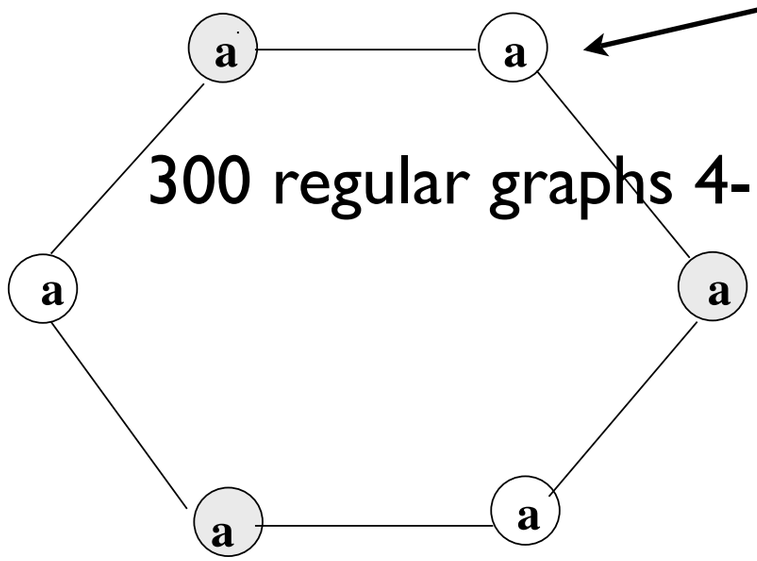
only label of the node

			Number of nodes in G				
			6	**10**	**14**	**18**	**Avg.**
Number of nodes in S	**3**	neural	92.4	90.0	90.0	84.3	89.1
		linear	93.3	84.5	86.7	84.7	87.3
		FNN	81.4	78.2	79.6	82.2	80.3
	5	neural	91.3	87.7	84.9	83.3	86.8
		linear	90.4	85.8	85.3	80.6	85.5
		FNN	85.2	73.2	65.2	75.5	74.8
	7	neural		89.8	84.6	79.9	84.8
		linear		91.3	84.4	79.2	85.0
		FNN		84.2	66.9	64.6	71.9
	9	neural		93.3	84.0	77.8	85.0
		linear		92.2	84.0	77.7	84.7
		FNN		91.6	73.7	67.0	77.4
	Avg.	neural	91.8	90.2	85.9	81.3	
		linear	91.9	88.5	85.1	80.6	
		FNN	83.3	81.8	71.3	72.3	
	Total avg.	neural			87.3		
		linear			86.5		
		FNN			77.2		

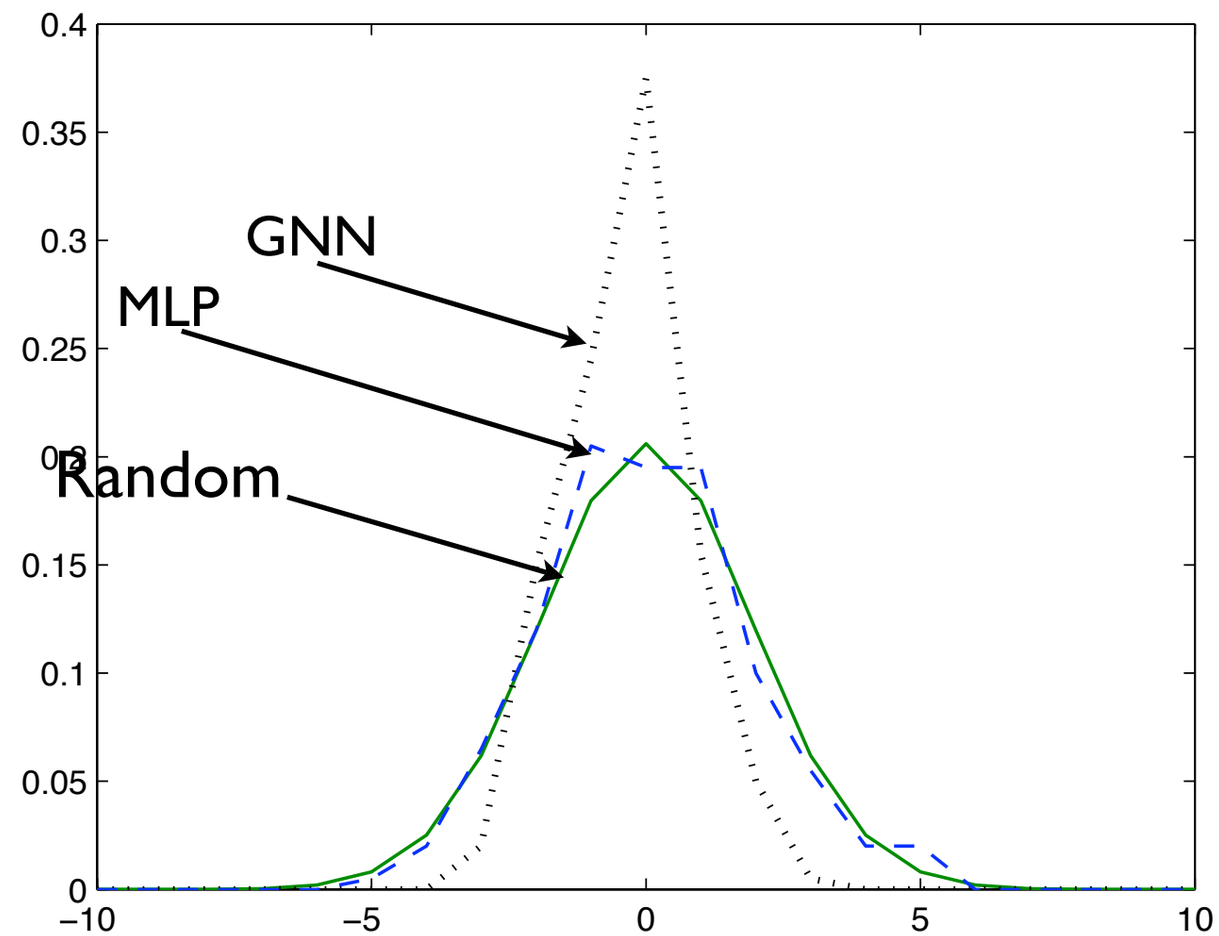
The Half-Hot Problem

$$\tau(\boldsymbol{G}, n) = 1 \text{ for a half of the nodes of } \boldsymbol{G} \text{ and } \tau(\boldsymbol{G}, n) = -1$$

5-dim label (integer 0-5 range)



300 regular graphs 4-10 nodes



it doesn't preserve unfolding equivalence!

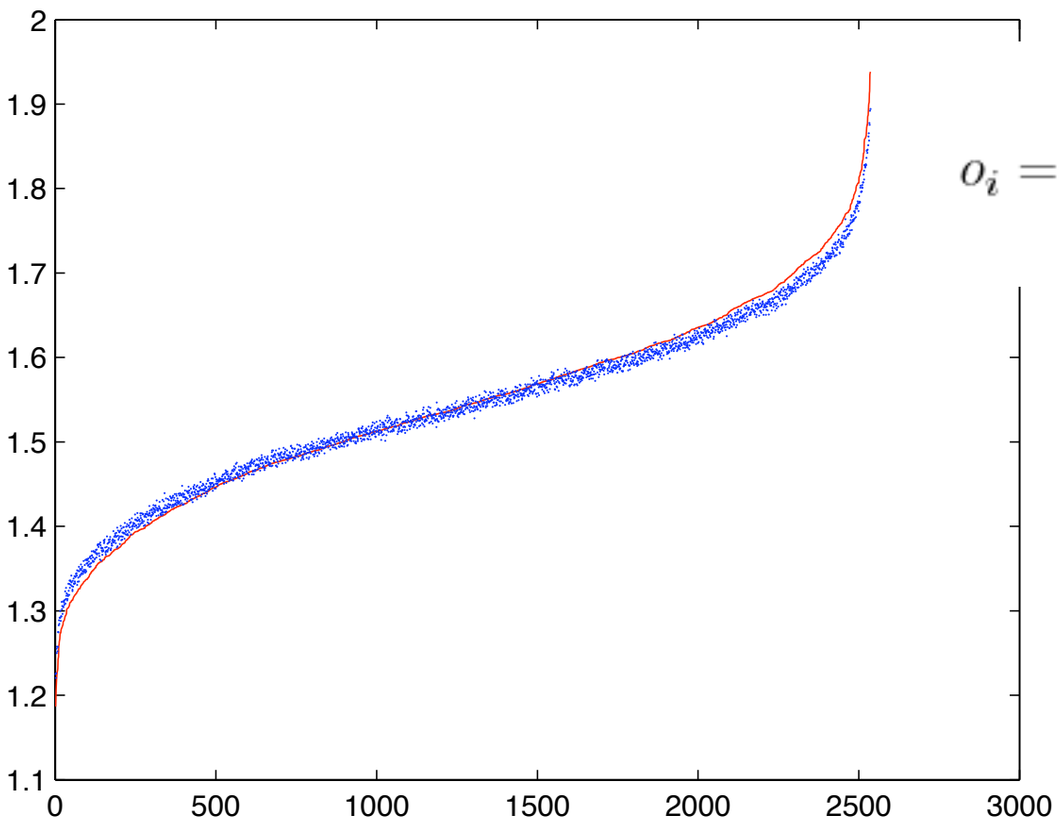
$$e_{\boldsymbol{w}} = \sum_{n,u \in N,\, n \neq u} (o_n - 1)(o_u - 1)$$

$$\delta_i = \frac{|\boldsymbol{G}_i|}{2} - b$$

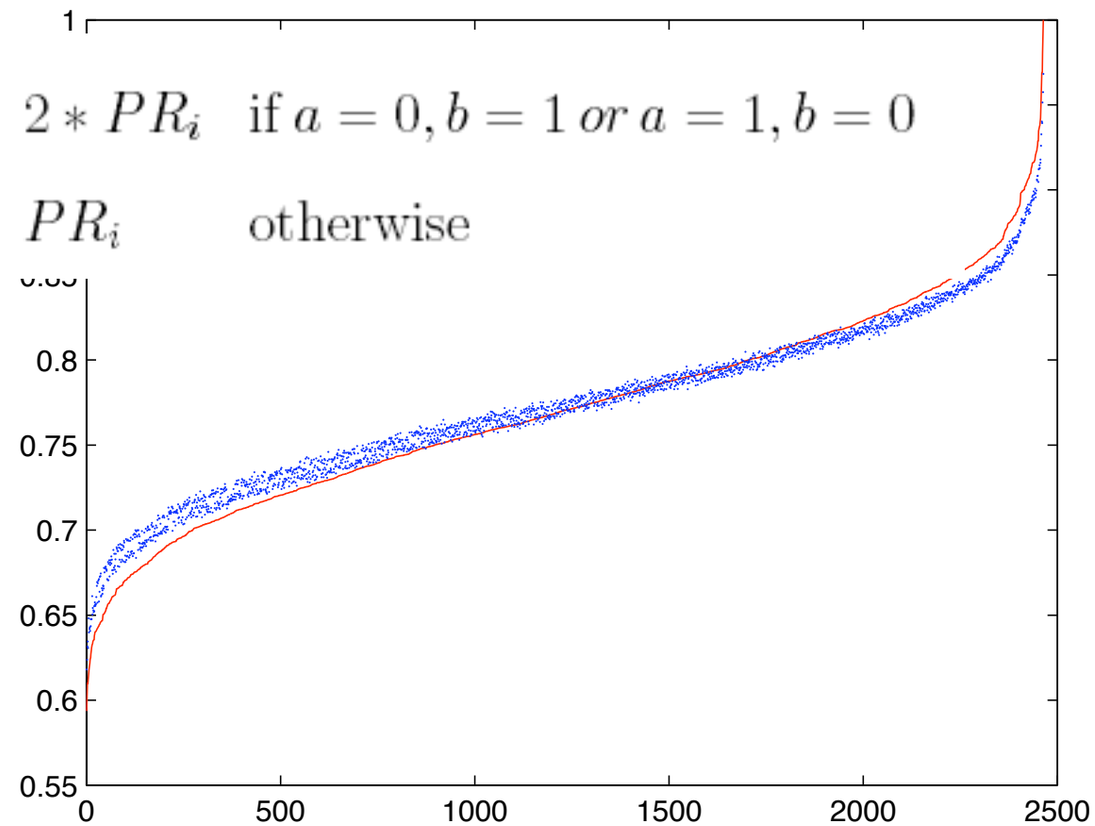
hot nodes

Learning PR-like Functions

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



$$o_i = \begin{cases} 2 * PR_i & \text{if } a = 0, b = 1 \; or \; a = 1, b = 0 \\ PR_i & \text{otherwise} \end{cases}$$



5,000 nodes
50 + 50 (Training & Validation Sets), 5 hidden units (linear model)

DaMn Bertinoro 2005

Experimental steps

- s/w based on MathLab 7.0.1
- It's evolving ... updating of feature (functions on arcs, non-stationarity, ...)
- download s/w at http://airgroup.dii.unisi.it

Conclusions

- Graphs in data ... graphs in functions (Compiling data & function on the same graph)
- Only preliminary experimental results ...
- We'd love trying with problems like Web spam, social networks (blogs)!
- Related approaches to be investigated (mainly with Kernel machines: Risi Kondor, Bernard Schoelkopf), but also with ILP + Statistics

Related Papers

download from http://nautilus.ing.unisi.it

- M. Gori, M. Maggini, and L. Sarti, "**Exact and Approximate Graph Matching Using Random Walks**", IEEE TPAMI (2005)
- A technical report (draft)

Comments very welcome!

Thanks for your attention!

DaMn Bertinoro 2005