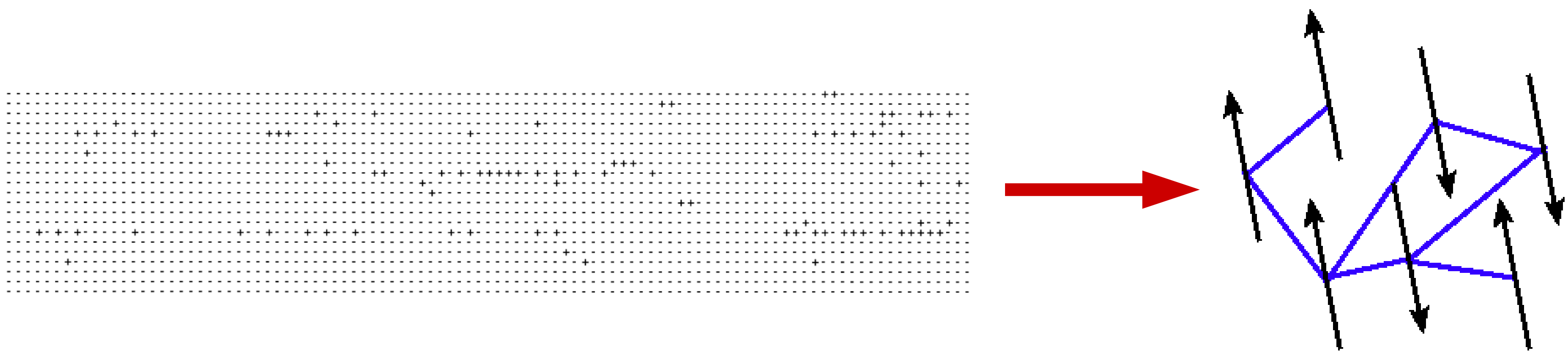
# Ising inverse problem :
# —
# Recovering the topology of the network !



**<u>Aurélien Decelle</u>** (LRI-TAU – Université Paris Sud)
Federico Ricci-tersenghi (Università di Roma – La Sapienza)

A.D., F. Ricci-Tersenghi — PRL 2014
A.D., F. Ricci-Tersenghi — … 2017

# LRI-TAU — Presentation

**Research in :**

- Developping Machine Learning methods:
  - Deep learning
  - Statistical physics and generative models
  - Reinforcement learning
  - Causality

- Applying ML to interdisciplinary thematics :
  - Solar physics
  - Social science
  - Particule physics (Higgs challenge)

**For further details, see http://tao.lri.fr**

# Outlines

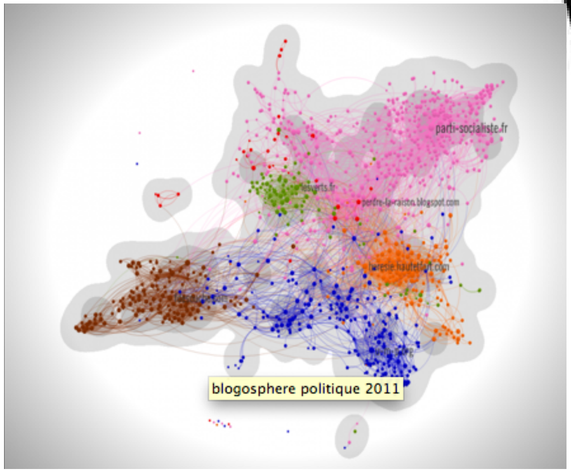
- **Motivations**
- **Setting**
- **Pseudo-Likelihood + Decimation**
- **Inferring many-body interactions**

# Motivations

**Why (Ising) inverse problems ?**

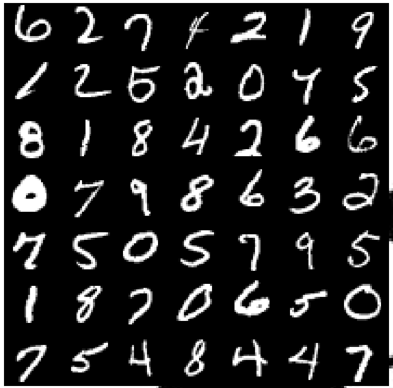
→ inferring parameters from observed configurations (this is what physicists do)

→ in social science: infer latent features of the system
(community detection (using potts model), …)

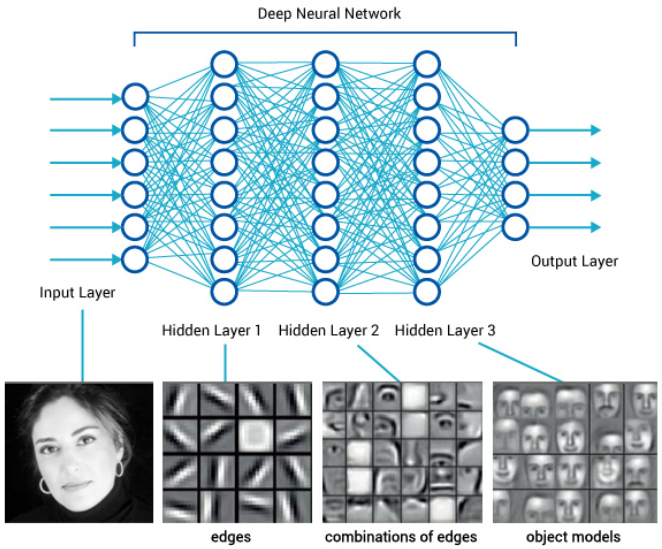



→ in neuroscience: infer the structure between neurons

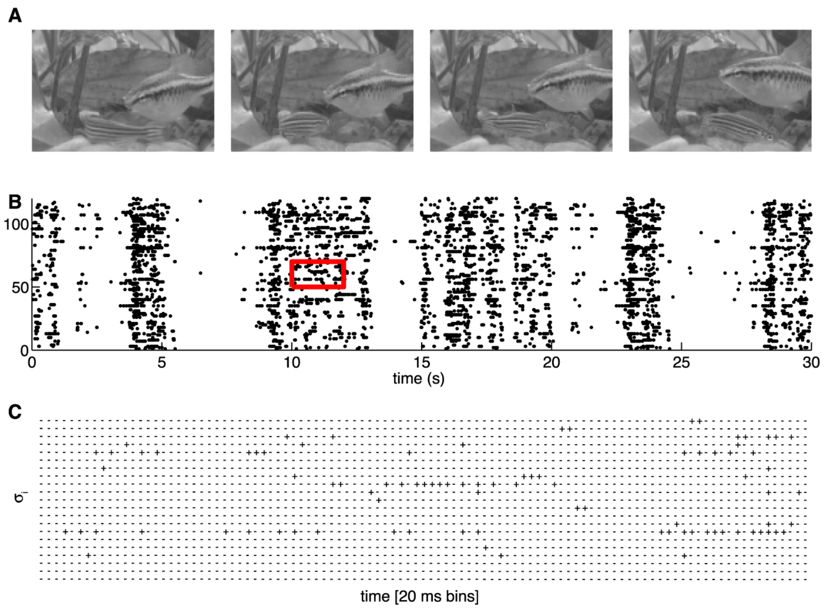
→ in Machine Learning : generative model of neural network
(typically Restricted Boltzmann Machines)



# Many applications

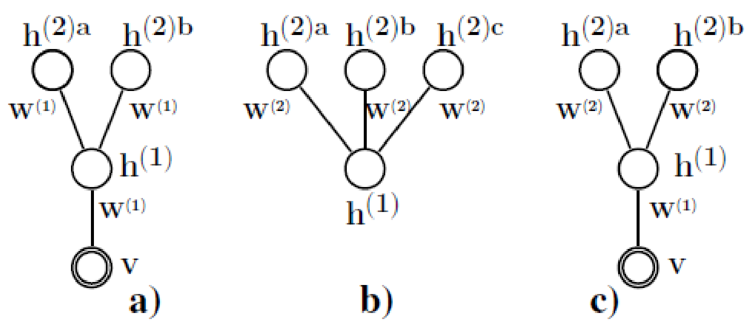


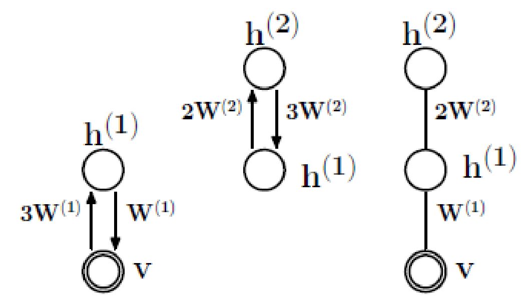

Machine Learning
(Lee et al.)

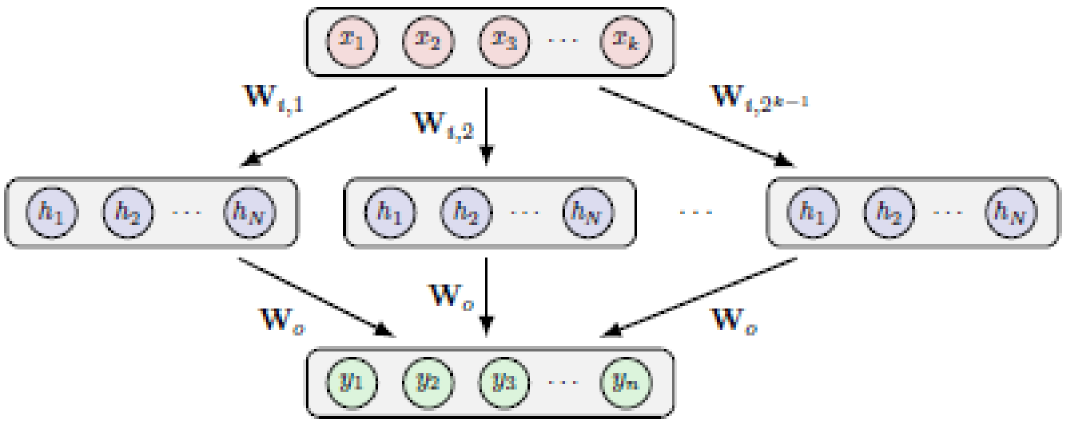


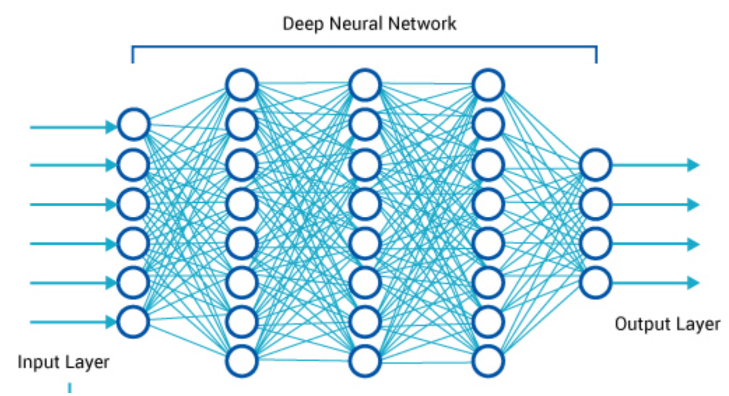

Neuron spiking (Tkacik et al.)

# Why the structure ?



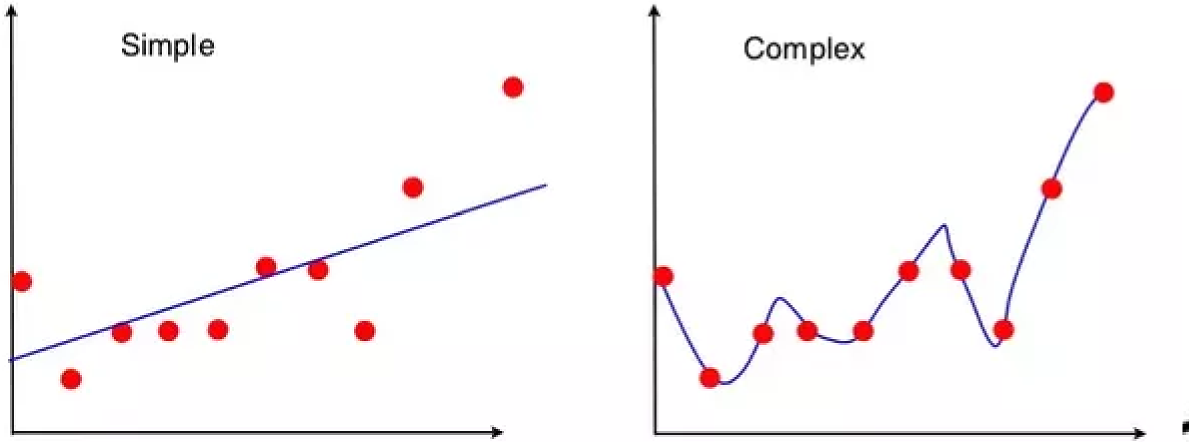



# Why the topology matters

In inverse problems, if you put all the possible parameters, you tend to **overfit** !

OverFIT !

- Lack of generalization
- No information on the structure/topology
- Fitting the noise !



# Can be a hard problem !

<u>Direct problems are already hard</u> : understanding equilibrium properties can be (very) challenging (e.g. spin glasses)

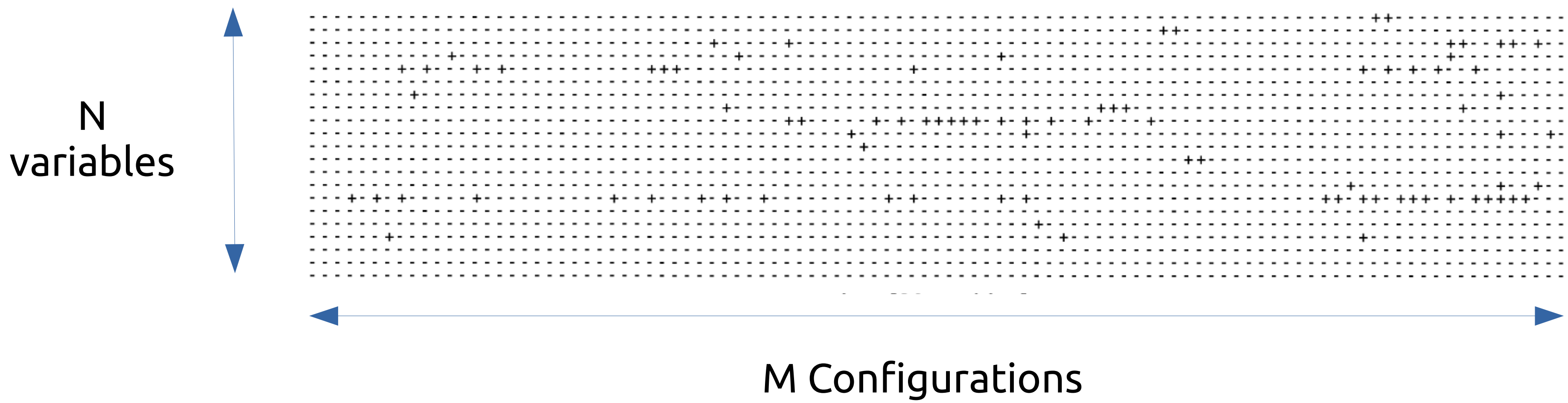
Inverse problems can be <u>**harder**</u> : maximizing the likelihood would involve to compute the partition function many times

You need to compute : $$< s_i s_j >= \sum_{\{s\}} \frac{s_i s_j \exp(-\beta H(\vec{s}))}{Z}$$

In particular, serious problems can appear because of
- Non-convex functions
- Slow convergence in the direct problem

# Setting

Set of configurations : $\{\sigma\}_{k=1..M}$ $\sigma_i^{(k)} = \pm 1$




- Define a model that can describe these data
- Find the parameters θ that match the data (according to the model)

# Setting

How can we find a good model that can explain the correlations and the biases !

Maximum entropy model :

$$S(p) = -\sum_{\{s\}} p(\{s\}) \log(p(\{s\}))$$

$$\text{argmax}\left( S(p) + \sum_{i<j} \lambda_{ij}(\langle s_i s_j \rangle_p - \langle s_i s_j \rangle_{data}) + \sum_{i} \lambda_i(\langle s_i \rangle_p - \langle s_i \rangle_{data}) \right)$$

# Setting

**Maximum entropy**
modelize any correlations

**The Ising model**

$$p(\sigma)=\frac{\exp\left(\sum_{i<j} J_{ij} s_i s_j + \sum_i h_i s_i\right)}{Z}$$

Static process : no time correlations (altough possible)

Maximizing the likelihood

$$p(\theta|\{\sigma\}) \propto p(\{\sigma\}|\theta)$$

$$p(\{\sigma\}|\theta) = \prod_k \frac{exp(\sum_{i<j} J_{ij} s_i^{(k)} s_j^{(k)} + \sum_i h_i s_i^{(k)})}{Z}$$

Reproduce the correlations and biases

$$\langle s_i s_j \rangle, \langle s_i \rangle$$

# Setting

Maximizing the likelihood

$$p(\theta|\{\sigma\}) \propto p(\{\sigma\}|\theta)$$

$$p(\{\sigma\}|\theta) = \prod_k \frac{exp(\sum_{i<j} J_{ij} s_i^{(k)} s_j^{(k)} + \sum_i s_i^{(k)})}{Z}$$

$$\mathcal{L} = \sum_k \sum_{i<j} J_{ij} s_i^{(k)} s_j^{(k)} + \sum_i h_i s_i^{(k)} - \log(Z)$$

**Gradient ascent :**

$$\Delta J_{ij} = \frac{\partial \mathcal{L}}{\partial J_{ij}} \propto (\langle s_i s_j \rangle_{\text{Data}} - \langle s_i s_j \rangle_p)$$

# Two directions

**Convex problem — but exponential complexity for log(Z)**

**Mean Field approach !**

Direct process : $J_{ij} = f(C_{ij}^{-1}, h_i)$

Polynomial in N !

The approximation can be improved :
1) naïve MF (independant spins)
2) TAP, correction or order √N$^{-1}$
3) Bethe Approx, tree like structure

- can't be used  with hidden variables !
- can't recover properly the topology !

**Maximizing likelihood !**

Exactly ? N=20 max

Approx to the likelihood : PseudoLikelihood
1) polynomial in N and M
2) can be improved

Useful to recover the graph
Can deal with many-bodies interactions

# Pseudo-Likelihood

**Goal:** find a function that can be maximized and would infer correctly the J's, h's

$$p(\vec{s}) = p(s_i|\vec{s}_{j\neq i})p(\vec{s}_{j\neq i})$$

we keep only this part !

$$p(s_i|\vec{s}_{j\neq i}) = \frac{e^{\beta s_i(\sum_{j\neq i} J_{ij} s_j + h_i)}}{2\cosh(\beta \sum_{j\neq i} J_{ij} + h_i)}$$

# Pseudo-Likelihood

$$p(s_i|\vec{s}_{j\neq i}) = \frac{e^{\beta s_i(\sum_{j\neq i} J_{ij}s_j + h_i)}}{2\cosh(\beta\sum_{j\neq i} J_{ij}s_j + h_i)}$$

Then we can maximize the following quantity :

$$\mathcal{PL} = \sum_{k=1}^{M}\sum_{i=1}^{N}\log(p(s_i^{(k)}|\vec{s}_{j\neq i}^{(k)}))$$

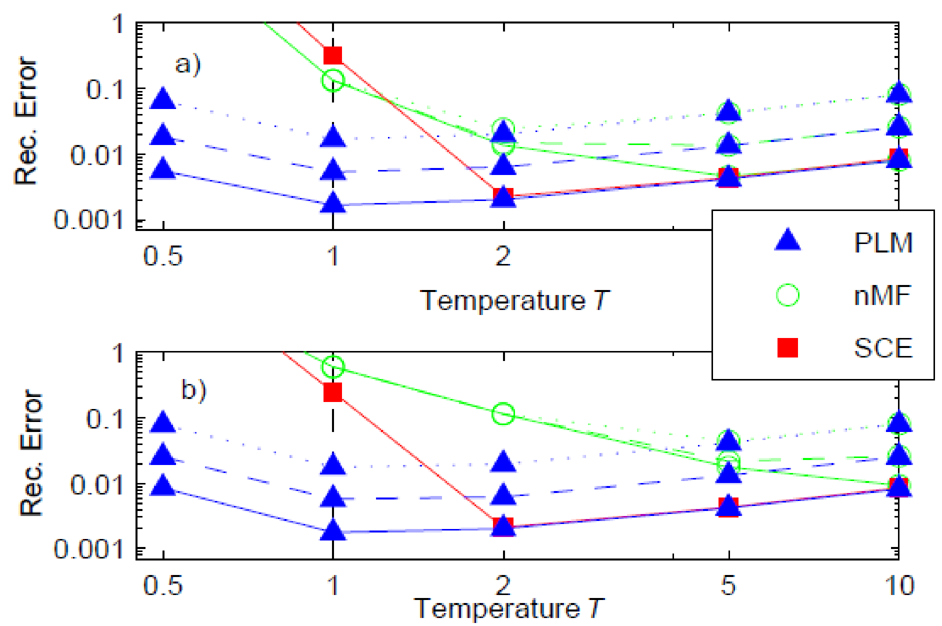
<u>Why should it work ?</u>

1) Maximizing the marginal of site i, ~ok
2) When data are following Gibbs, infer the true value for infinite sampling
3) Convex function, complexity goes as $O(N^3M)$

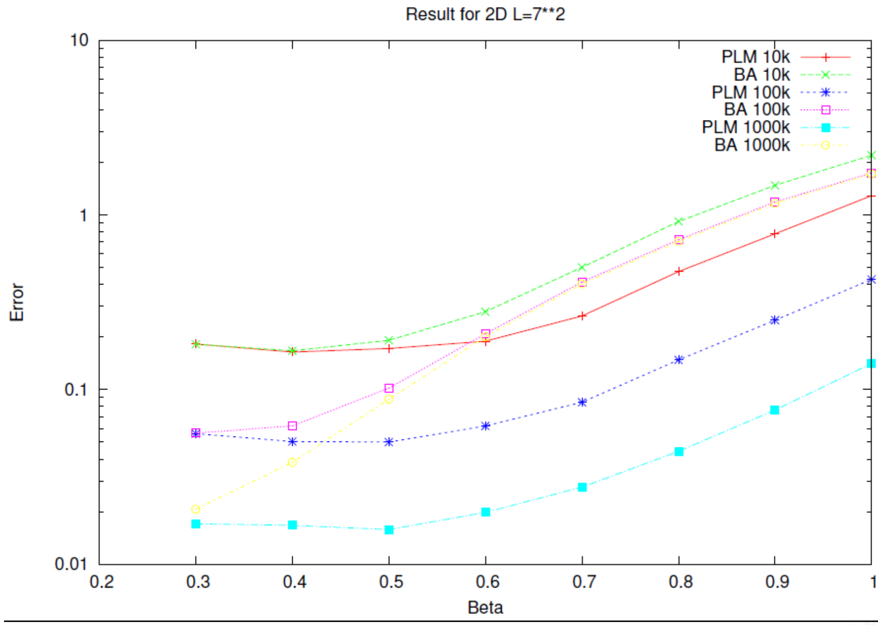
# How well it goes ?

<u>With reasonnable sampling you get good results !</u>

SK model, N=64, with M=$10^6$, $10^7$, $10^8$
b) with sparsity



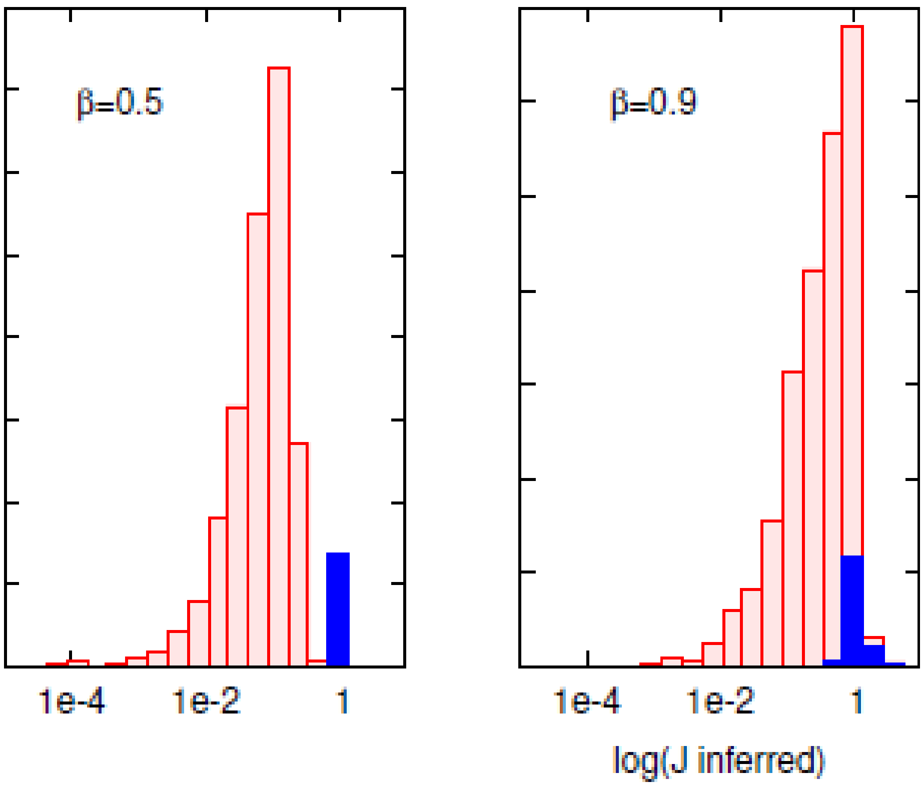

2D ferro model,
N=49, with M=$10^4$, $10^5$, $10^6$




E. Aurell and M. Ekeberg 2012

# What about the topology ?

Results for a 2D diluted ferromagnet (N=49)



# Using prior distribution

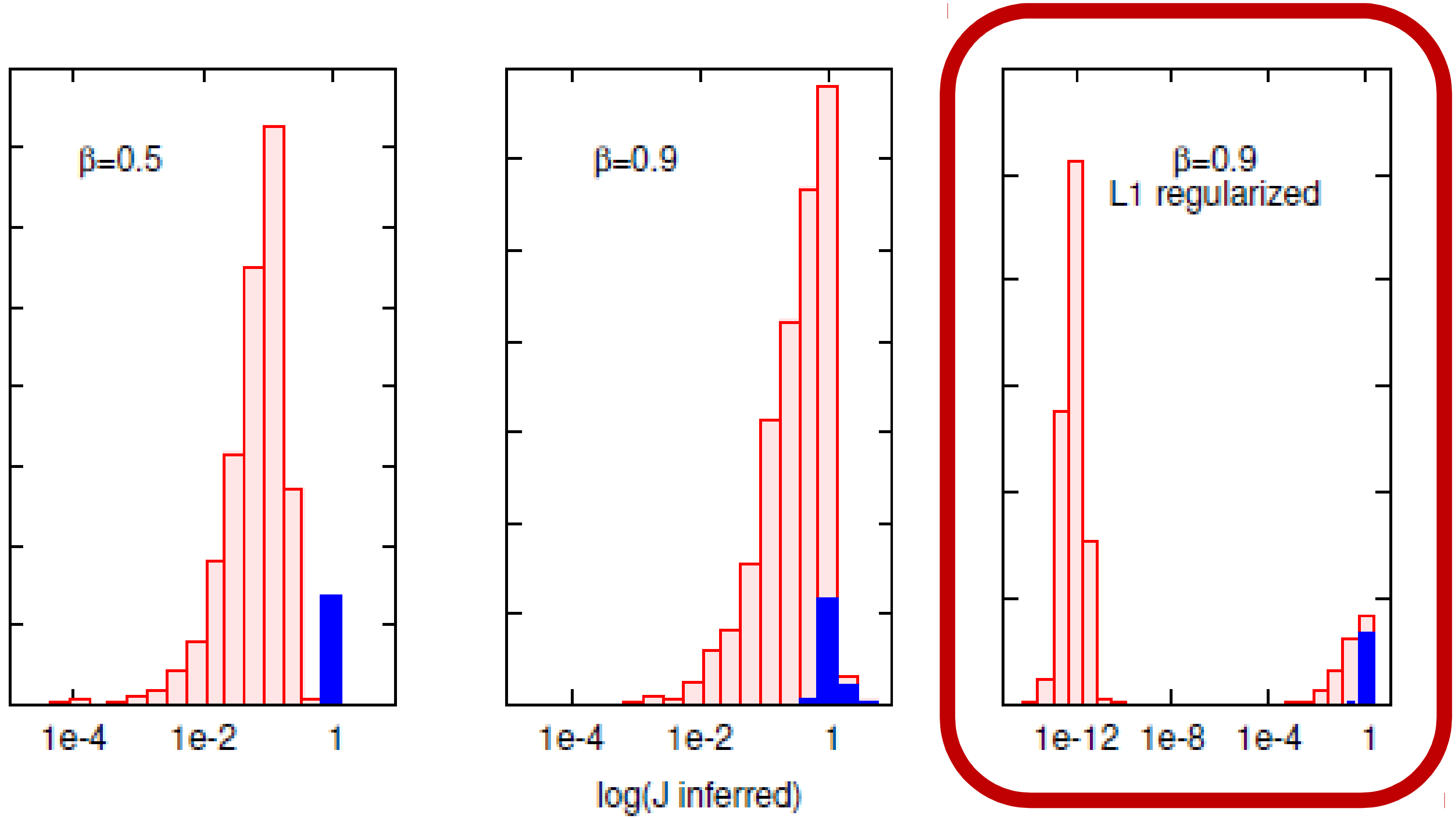
We know that a Laplace prior impose sparsity in the inference process !

$$\mathcal{PL}_i = \sum_{k=1}^{M} \log(1 + e^{-2\beta s_i^{(k)} (\sum_j J_{ij} s_j^{(k)} + h_i)})) - \lambda \sum_j |J_{ij}|$$

**But how do I fix λ ?**

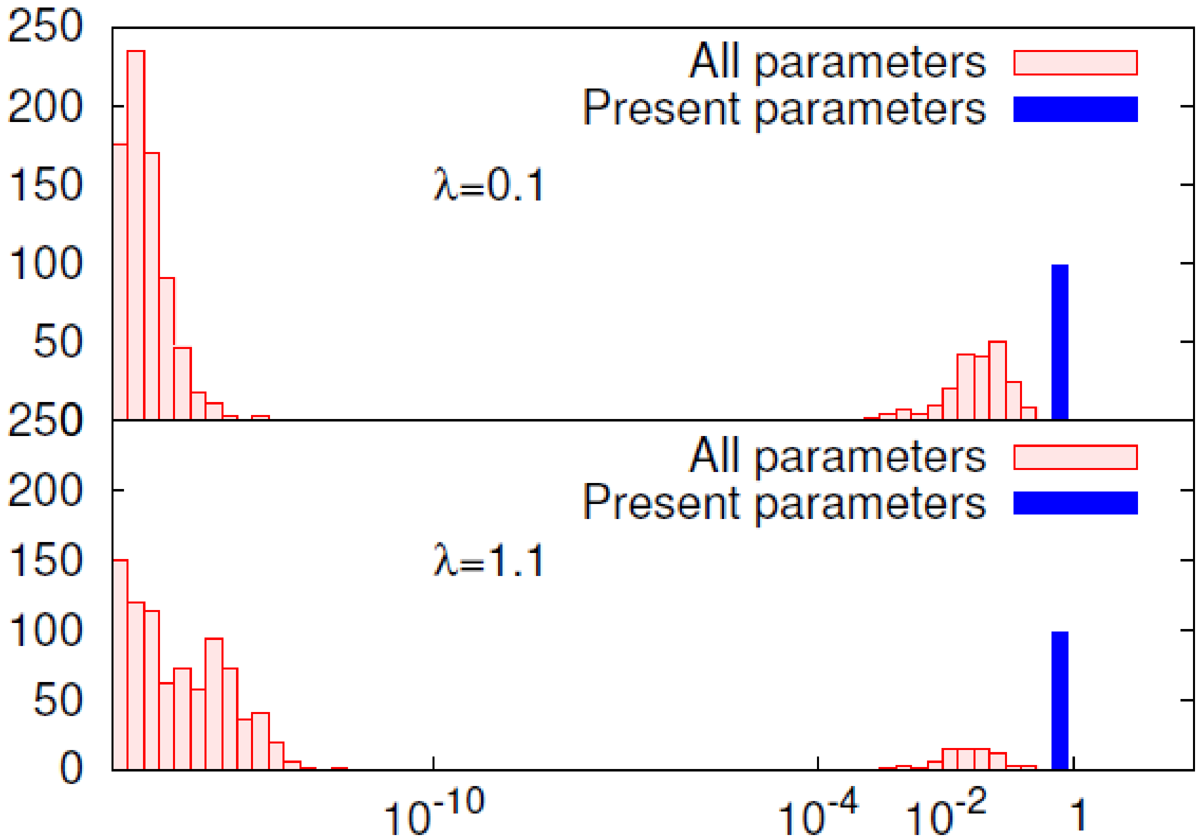
# What about the topology ?

Results for a 2D diluted ferromagnet (N=49)

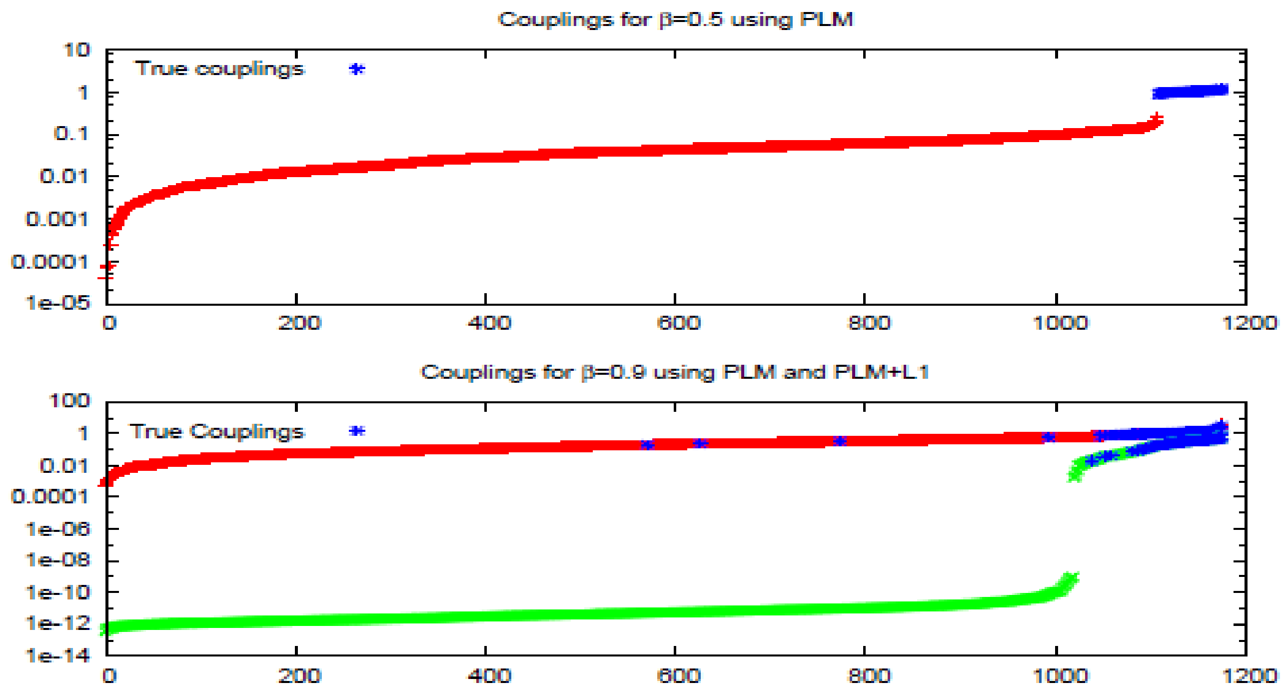


# What about the topology ?

Results for a 2D diluted ferromagnet (N=49)



# Decimating ?




**In RED : PLM**

**In BLUE : true couplings**

**In GREEN : PLM-L1**

Progressively decimating parameters with a small absolute values
Not NEW :

- In optimization problem using BP (Montanari et al.)
- Brain damage (Lecun)

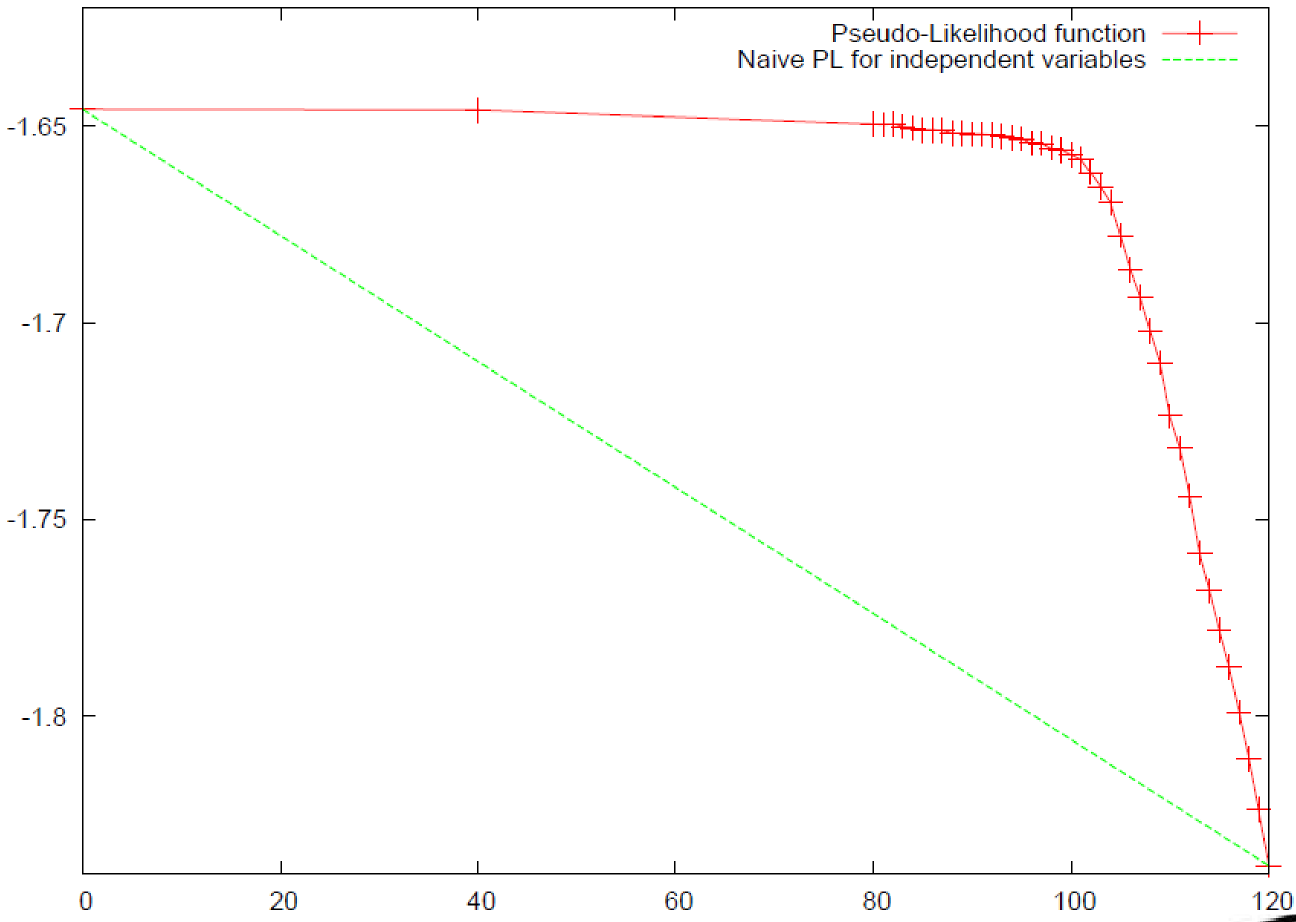
# Decimation algorithm

Given a set of equilibrium configurations and all unfixed parameters

1. Maximize the Pseudo-Likelihood function over all non-fixed variables
2. Decimate the smallest variables (in magnitude) and fixed them
3. If (criterion is reached)
   exit
4. Else
   goto 1.

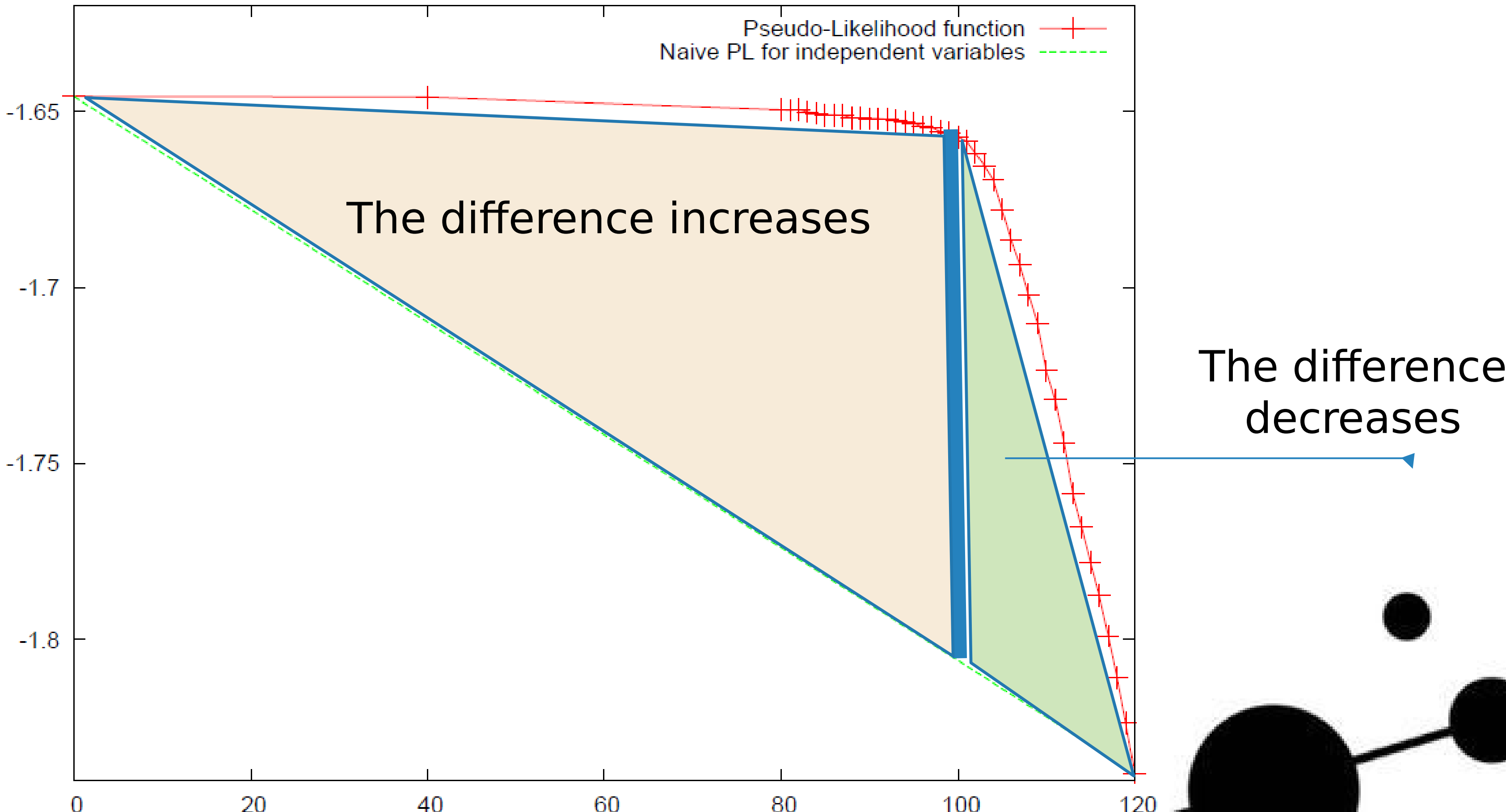
# Example of PL

Random graph with 16 nodes

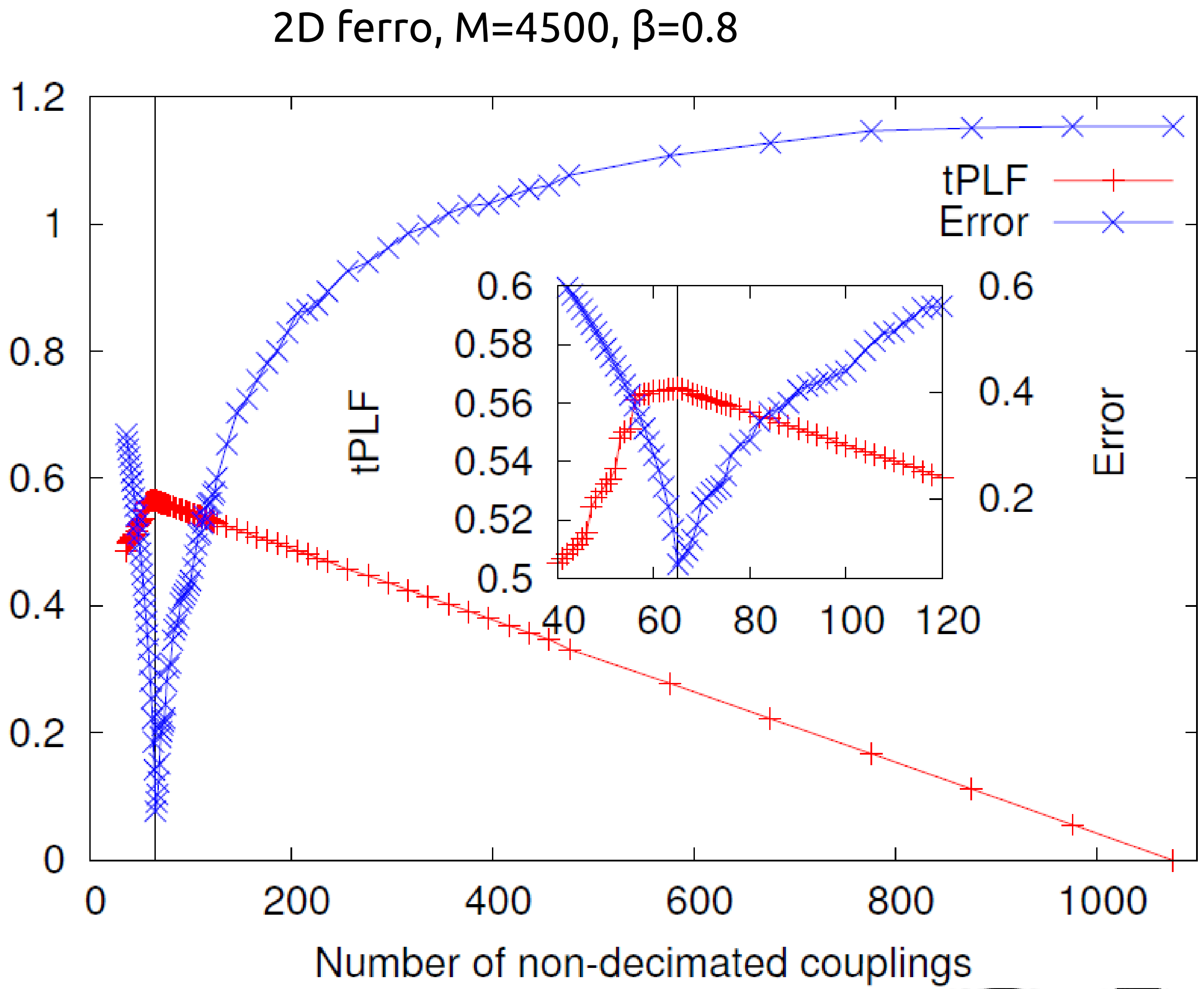


# Example of PL

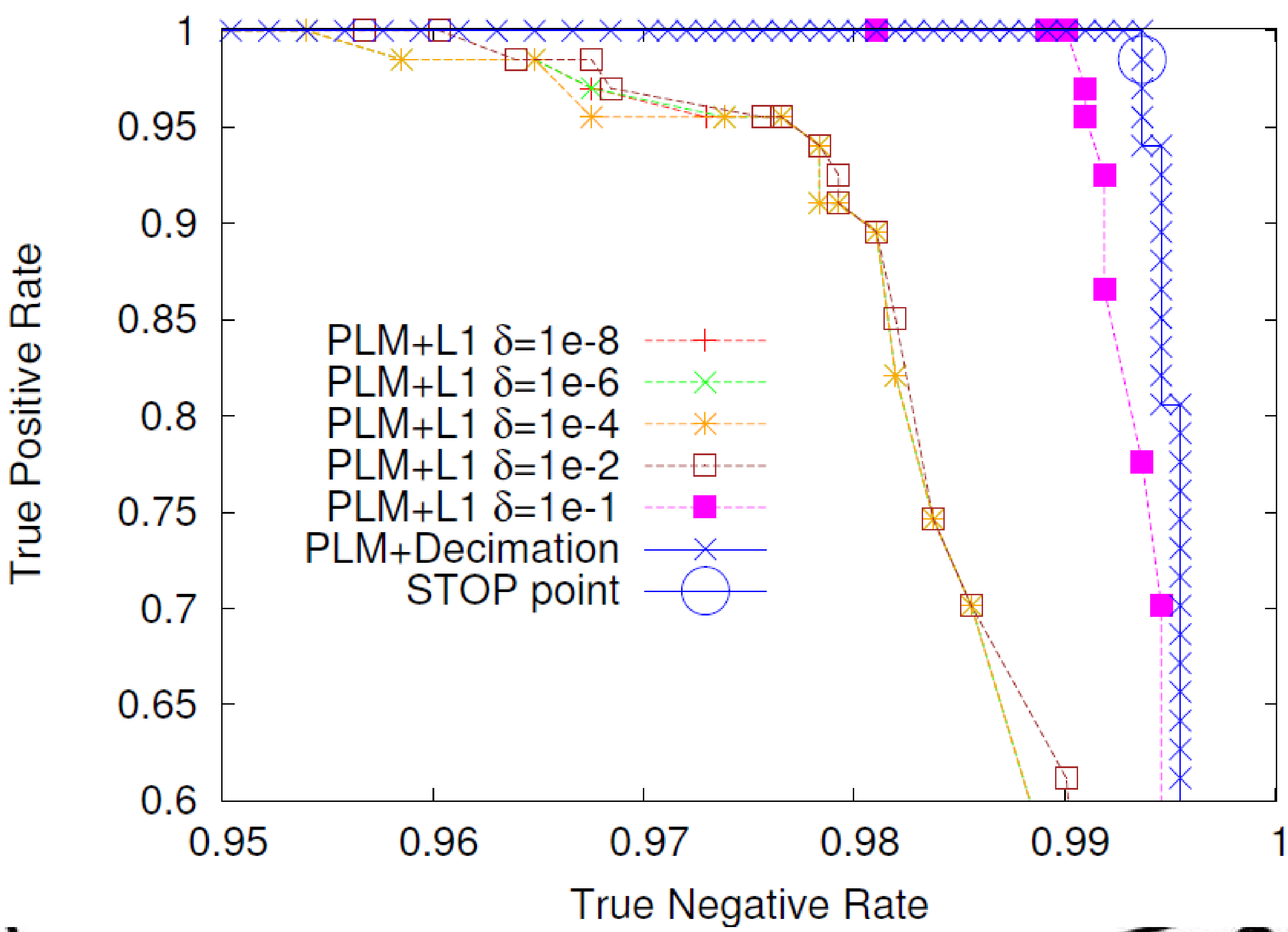
Random graph with 16 nodes



# What happened ?



# Roc comparison



# Many body interactions

Systems can sometimes have many-body interactions !

$$\mathcal{H} = \sum_{i<j} J_{ij} s_i s_j + \sum_{i<j<k} J_{ijk} s_i s_j s_k$$

<u>Easy generalization of the PseudoLikelihood :</u>

$$p(s_i | \vec{s}_{j \neq i}) \propto e^{\beta s_i (\sum_{i<j} J_{ij} s_i s_j \sum_{j<k} J_{ijk} s_j s_k)}$$

**<u>Problem</u>** : derivative w.r.t all parameters → complexity $O(N^4 M)$
Get worse and worse for interaction between many spins !
You don't want to add all possible parameters (meaningless)

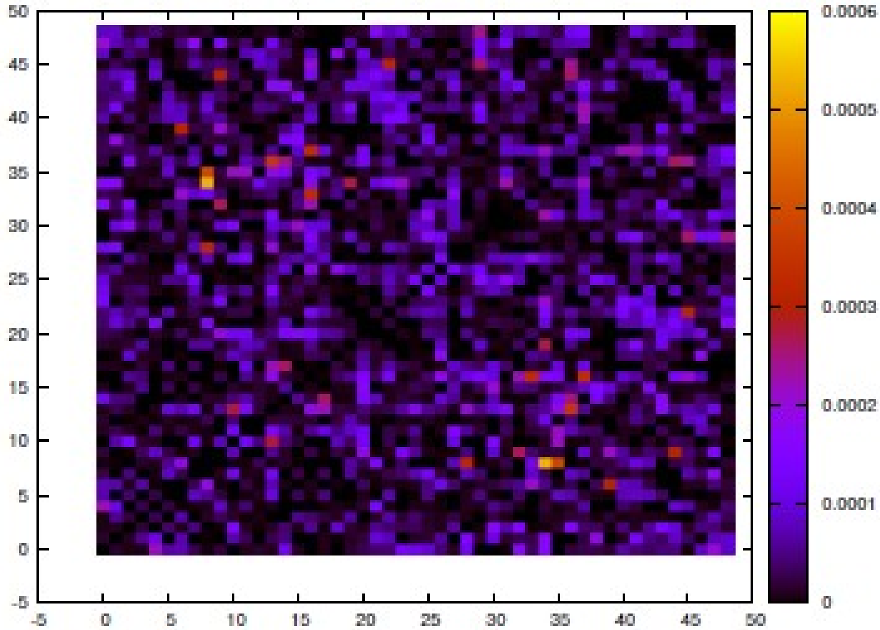
# Experiment

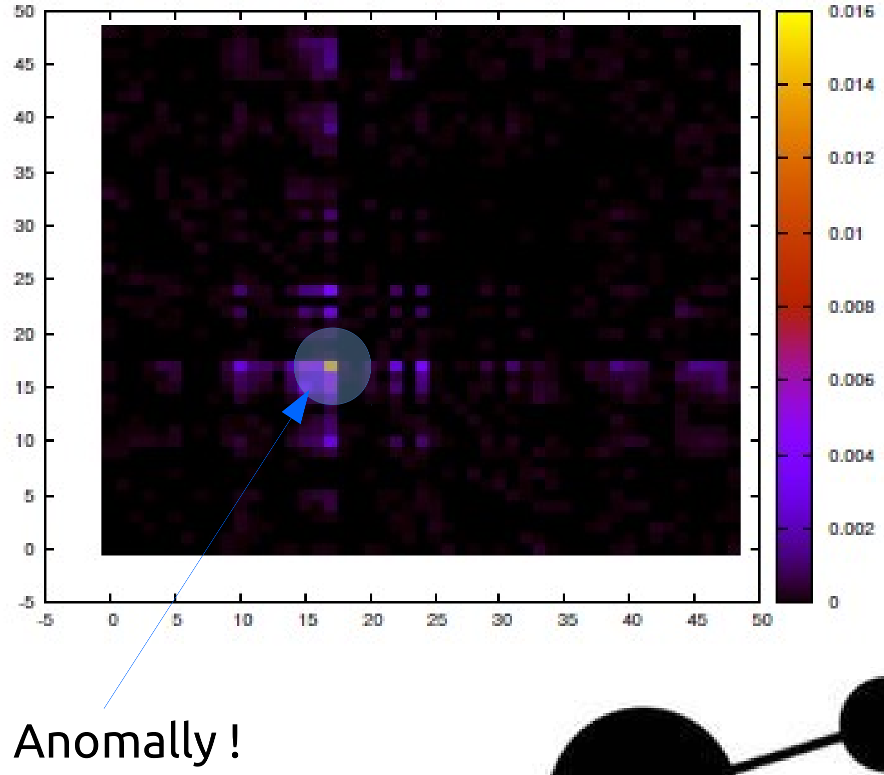
Let’s consider the following experience

- Take a system S1, 2D ferro without field
- Take a system S2, 2D ferro without field but with some 3-body interactions
- Make the inference on the two models with a pairwise model and a model with 3B interactions included

# Experiment

On the **left** : inference on S1 with the correct model
On the **right** : inference on S2 with only pairwise interactions



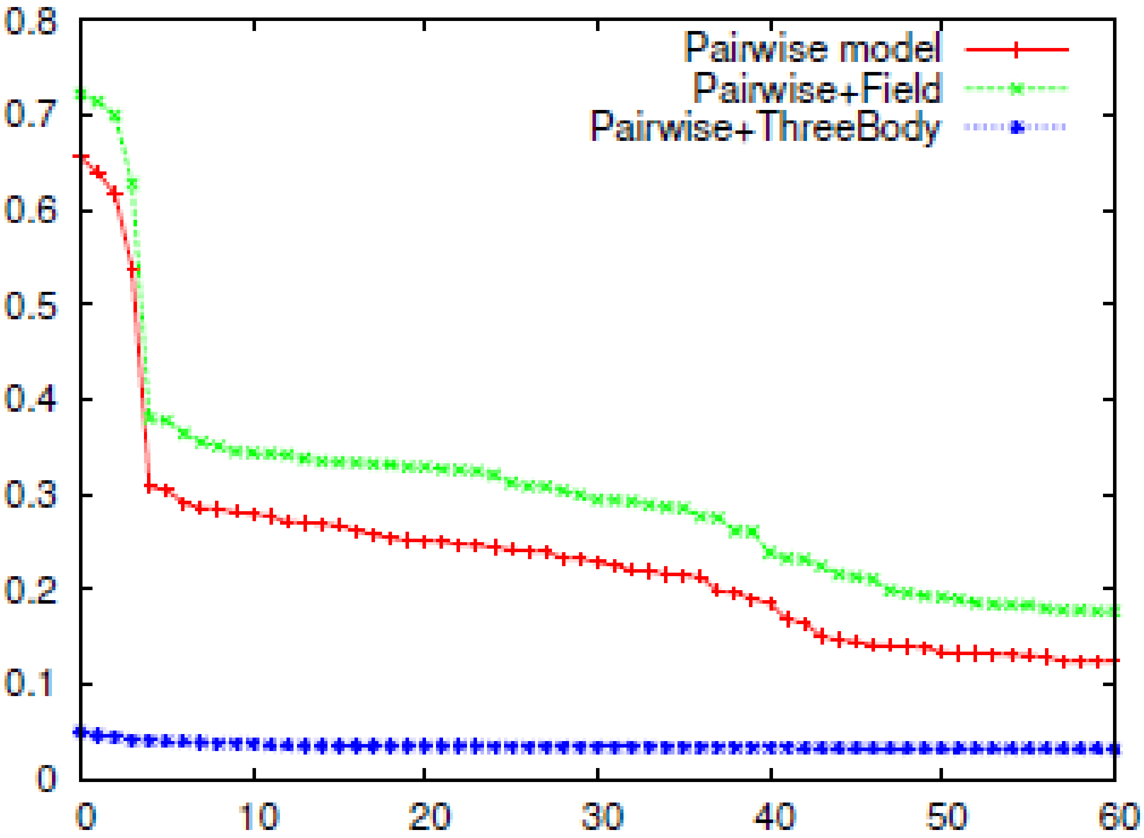




**<u>But:</u>** this can be corrected using a magnetic field !

# Experiment

Error on the three points correlations function

$$\langle s_i s_j s_k \rangle$$

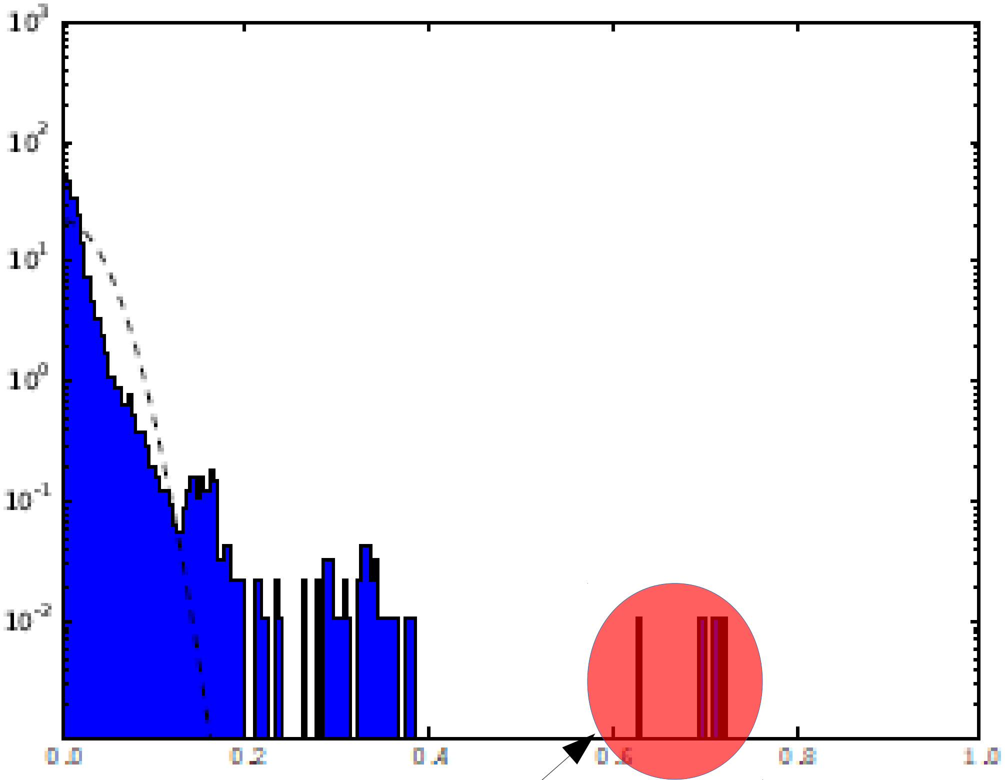



Take the error on the 3points correlation functions, plot them by decreasing order!

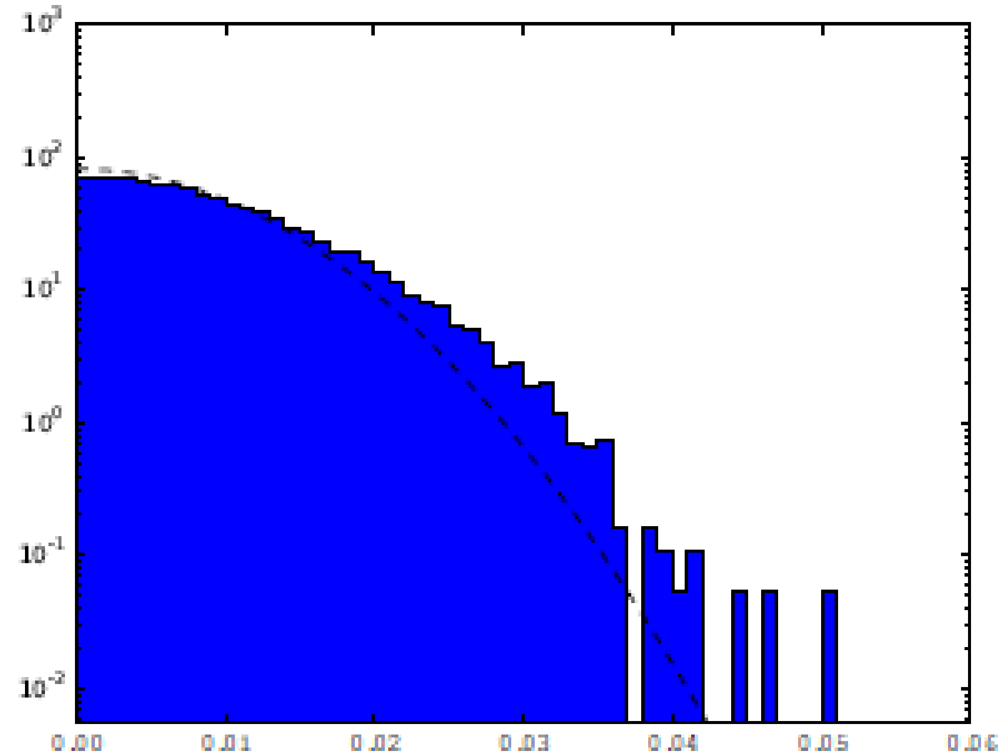
**Can you guess how many three-body interactions there are ?**

# Experiment

- Wrong model –
Histogram of the error on the 3p-corr




- Correct model –
Histogram of the error on the 3p-corr



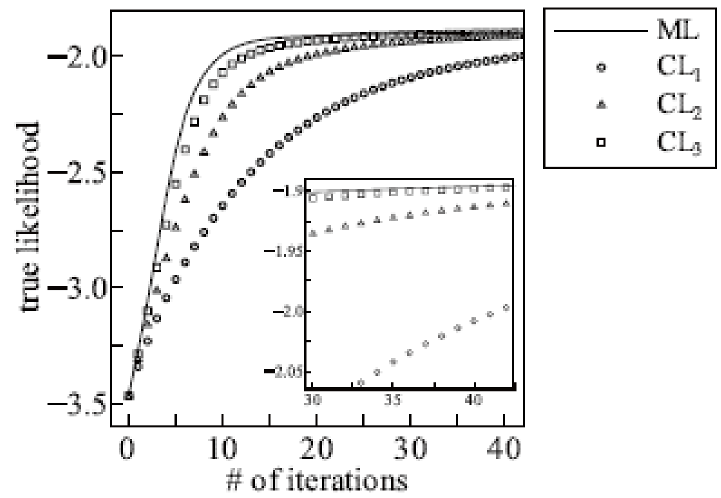

**<u>4 outliers</u>** → these are the ones that were added !

# Extension & Application

- **Dynamical case** : A.D. and P. Zhang (2015)
- **Cheating students** : S. Yamanaka, M. Ohzeki, A.D. (2014)
- **XY model** : P. Tyagi, L. Leuzzi
- **Non-linear wave and many-bodies** : P. Tyagi, L. Leuzzi

**Using higher order Likelihood ? (cf Yasuda et al.)**

$$p(s_i, s_j | \vec{s}_{k \neq i,j})$$




**Application to model with hidden variables ?
(Machine Learning)**