Statistical mechanics of random optimization & inference problems: from phase transitions to algorithmic behavior

Federico Ricci-Tersenghi
(La Sapienza, Roma)

in collaboration with M.C. Angelini (La Sapienza, Roma),
L. Budzynski, G. Semerjian (ENS, Paris)

talk given at the Workshop on Science of Data Science, ICTP, October 3rd, 2019

Motivations

- Better understanding limits of (robust) algorithms in performing fundamental tasks in optimization and inference
- Minimizing/maximizing a complex function:
  - solving any optimization problem
  - training a neural network
  - inferring via the likelihood
- Sampling a complex probability distribution
  - inferring via the posterior
  - out of equilibrium glassy physics

Ensembles of random problems

- No worst case analysis
- Typical case analysis -> statistical mechanics
- Ensemble of randomly generated problems depending on few key parameters (e.g. mean degree of a random graph)
- Tuning parameters one can
  - vary the hardness of random problems
  - undergo phase transitions
- Learn from hardest random problems

- Look for a connection between algorithmic complexity and phase transitions

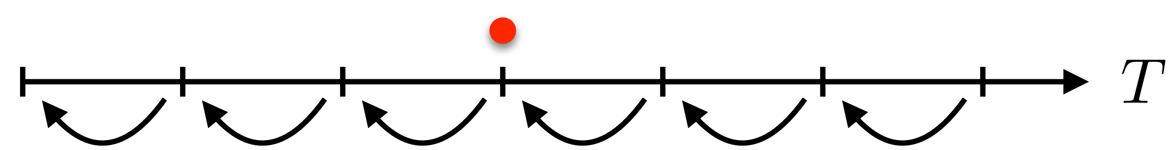
Which algorithms?

- Large and sparse random models (no AMP et simila)
- Two classes of algorithms
  - Belief Propagation run on a specific graph:
    - exact in some phases
    - Bayes optimal (if properly initialized)
    - not very robust on non-random graphs
  - Monte Carlo based algorithms:
    - (Replicated) Simulated Annealing, Parallel Tempering, …
    - very robust

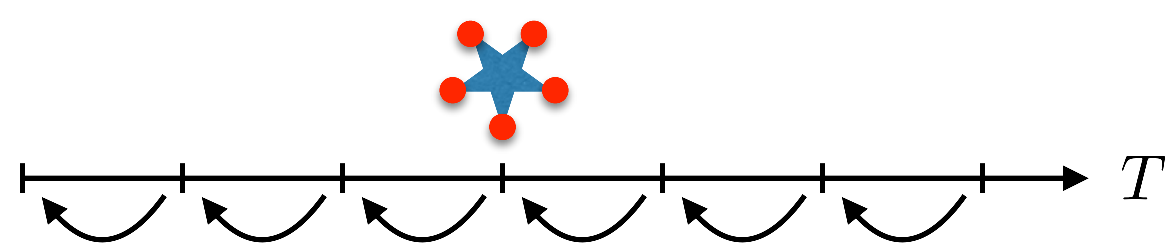
  Their limits are mostly unknown (working in the regime of times scaling linearly in the system size)

Monte Carlo based algorithms

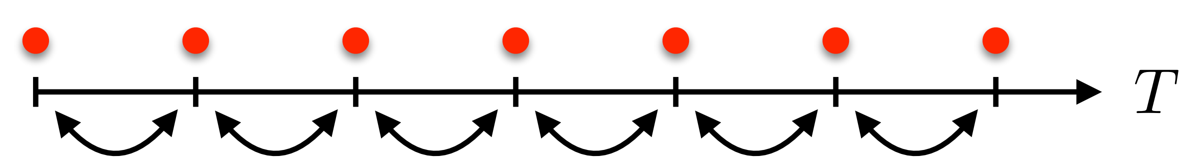
- Aim: minimize $E(\underline{\sigma})$
- Run MCMC sampling from $P_T(\underline{\sigma}) \propto \exp[-E(\underline{\sigma})/T]$
- Send $T \to 0$ and try to find the state of lowest energy
- Simulated Annealing (SA)



- Replicated Simulated Annealing (RSA)

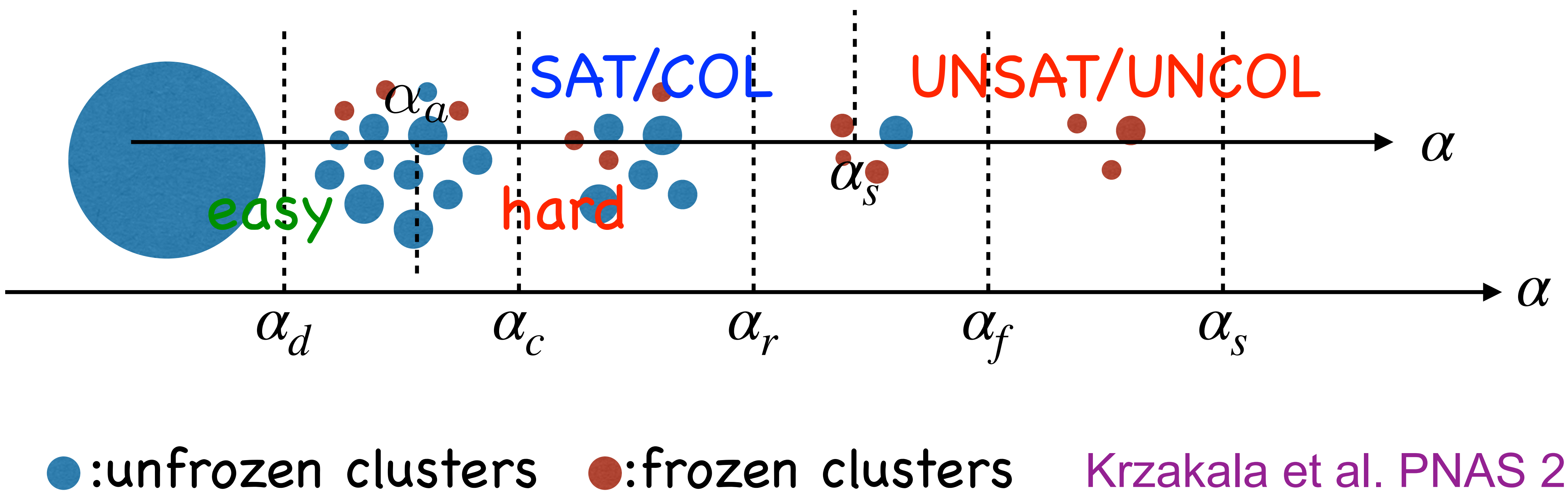


- Parallel Tempering (PT)



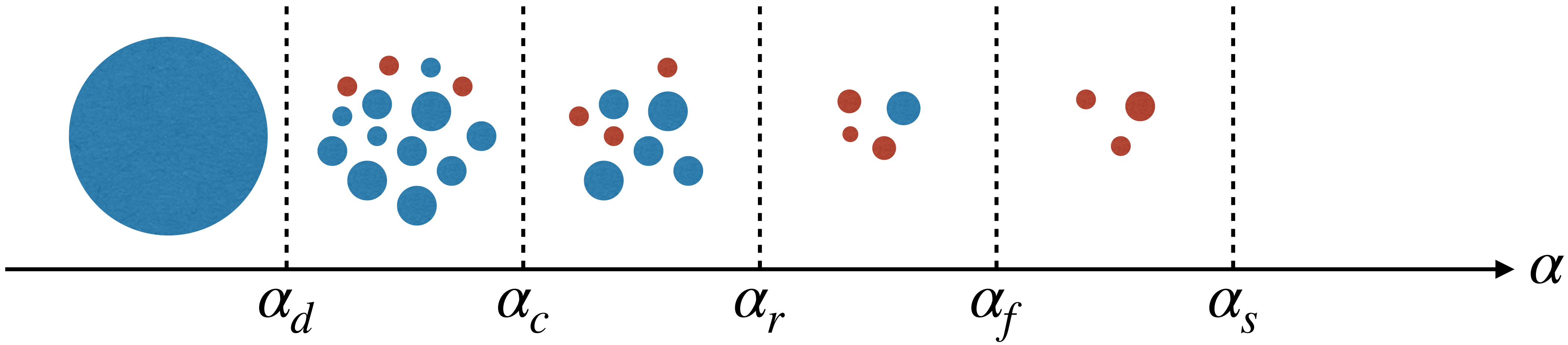
Random Constraint Satisfaction Pb.

- Find a configuration of N variables that satisfies $M = \alpha N$ constraints, randomly generated
- Examples: random k-SAT, random graph q-coloring, ...
- key parameter: the mean degree of the random graph

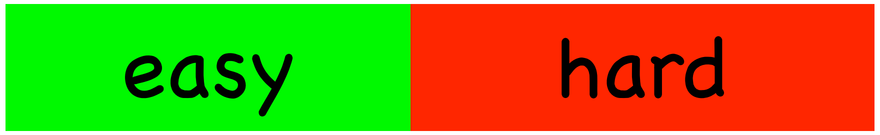


Krzakala et al. PNAS 2007

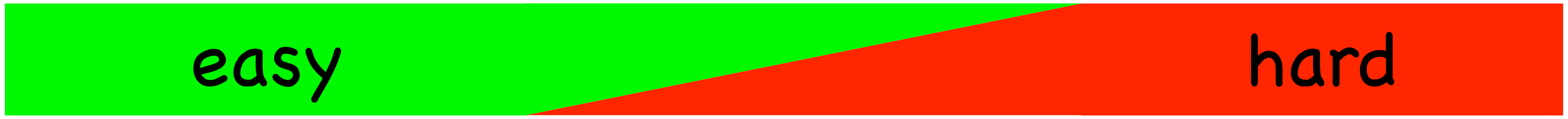
phase transitions & algorithms



BP-based alg. Semerjian, RT JSTAT 2009

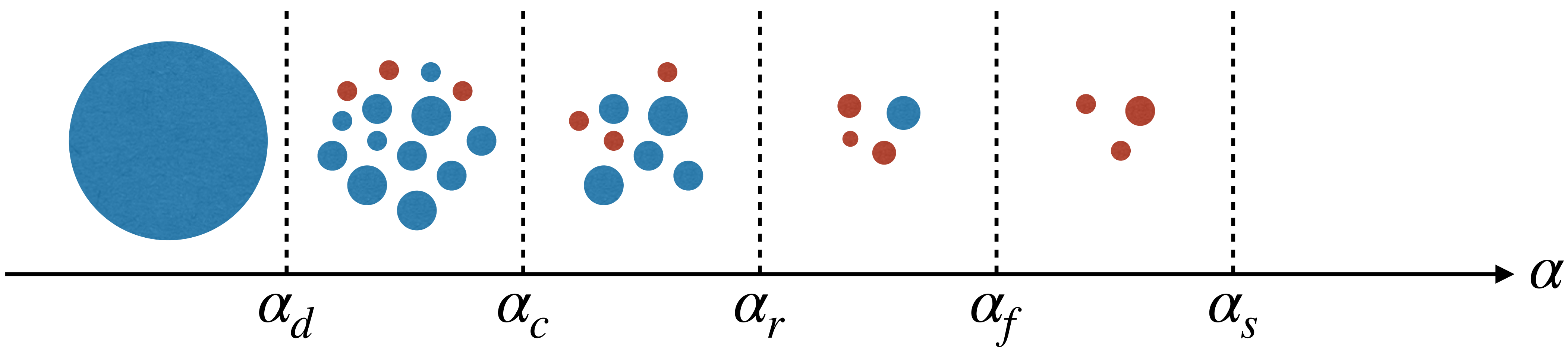


sampling uniformly solutions by MC



heuristic smart algorithms: message passing alg. (SID, BSP), Monte Carlo based, biased random walks (FMS, ASAT)

phase transitions & algorithms

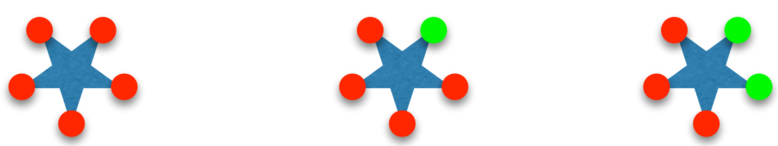


- Key observation: smart algorithms do not sample solutions uniformly (they never find frozen solutions)
- Conjectured ultimate algorithmic threshold is α_f
- Uniform measure over solutions not very useful to understand algorithms -> better biasing the measure (see Robust ensemble in Zecchina's talk)

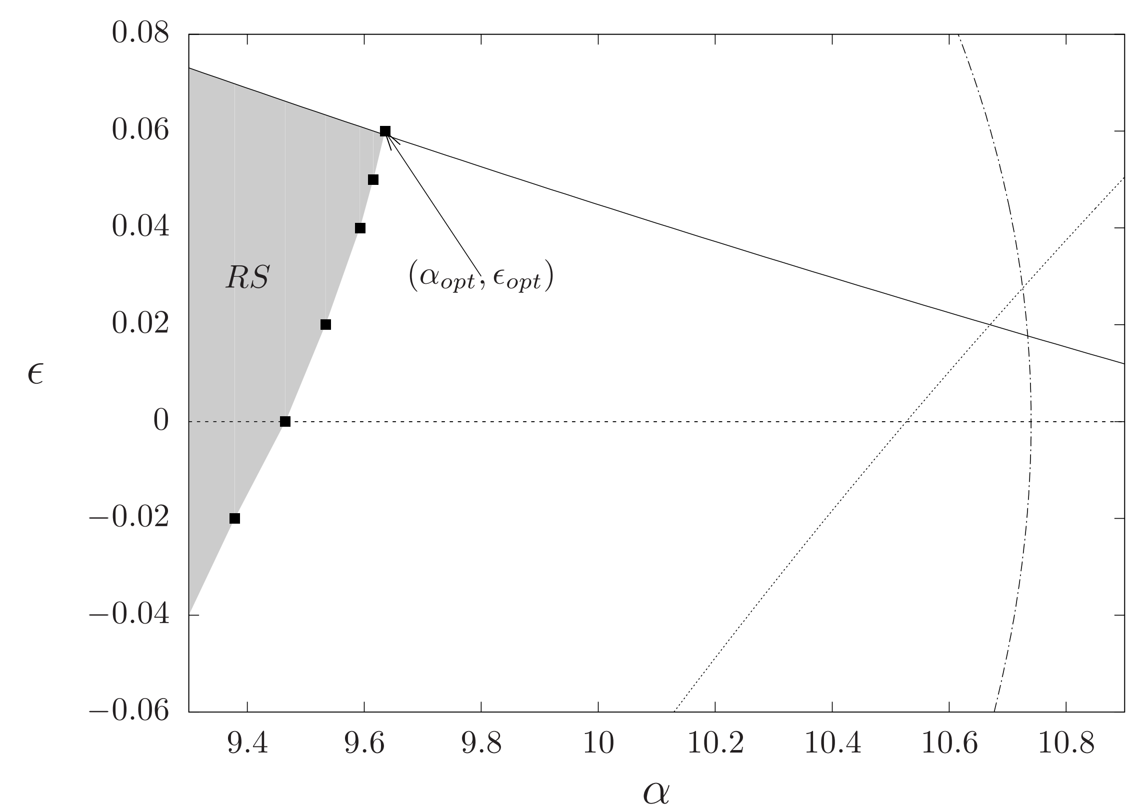
Biasing the measure: a first attempt

Budzynski et al. JSTAT 2019

- Random hypergraph bicoloring / NAE-k-SAT (k=5)
- Uniform measure: $\alpha_d = 9.465,\ \alpha_s = 10.46$ Gabrié et al. JPA 2017
- Simulated Annealing works until $\alpha_{alg} \approx 9.6$
- Bias the measure to reduce potentially freezing clauses:



$0 \qquad (1-\epsilon) \qquad 1$



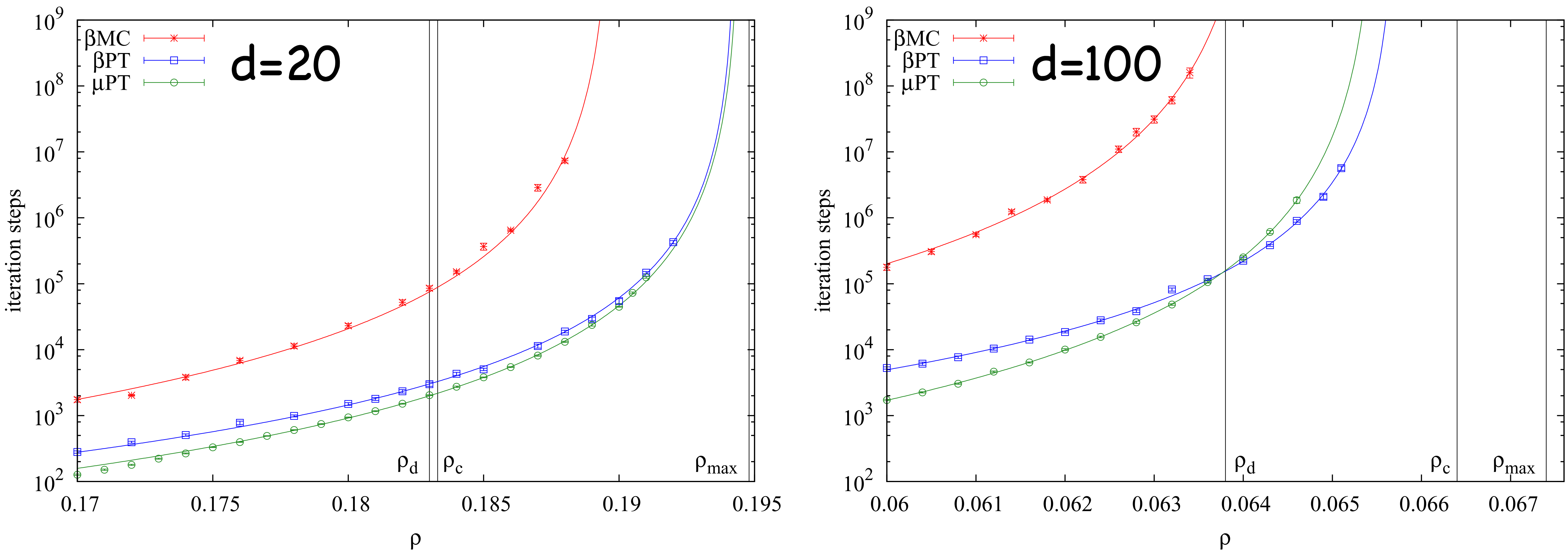
Algorithms & random CSP

- Many phase transitions found
- Several hints for the origin of the computational complexity: dynamical phase transition, long range correlations, frozen variables, glassy metastable states...
- Exact connection between phase transitions and algorithmic threshold is lacking (but in few cases)
- Smart algorithms find more easily solutions which are more attractive/accessible (one should count basins of attractions!)
- Biased measured can be a good solution
- ...but the story is still long...

…meanwhile use Parallel Tempering

Angelini, RT PRE 2019

- Most robust and general purpose optimization algorithm
- Largest Independent Set in d-regular random graph



Bayesian inference

- Teacher-student scenario
  - the teacher chooses a ground truth $\boldsymbol{x}^*$ from the prior $P_p(\boldsymbol{x})$ and a probabilistic model to generate the data $P_m(\boldsymbol{y}|\boldsymbol{x}^*)$
  - the teacher provides the student with the prior $P_p(\cdot)$, the model $P_m(\cdot|\cdot)$ and the data $\boldsymbol{y}$
  - the student uses Bayes formula to compute the posterior probability distribution

$$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{P_m(\boldsymbol{y}|\boldsymbol{x})P_p(\boldsymbol{x})}{Z(\boldsymbol{y})}$$

  - student problem is then sampling or maximizing the posterior probability distribution

Bayesian inference

- Statistical estimators are given in terms of marginal probabilities

$$\mu_i(x_i) = \sum_{\boldsymbol{x} \setminus x_i} P(\boldsymbol{x}|\boldsymbol{y})$$

  - $\widehat{x}_i = \sum_{x_i} x_i \mu_i(x_i)$ minimizes the MSE
  - $\widehat{x}_i = \mathrm{argmax}\ \mu_i(x_i)$ maximizes the MO

- Computing marginal probabilities is as hard as computing the partition function

Random sparse Bayesian inference

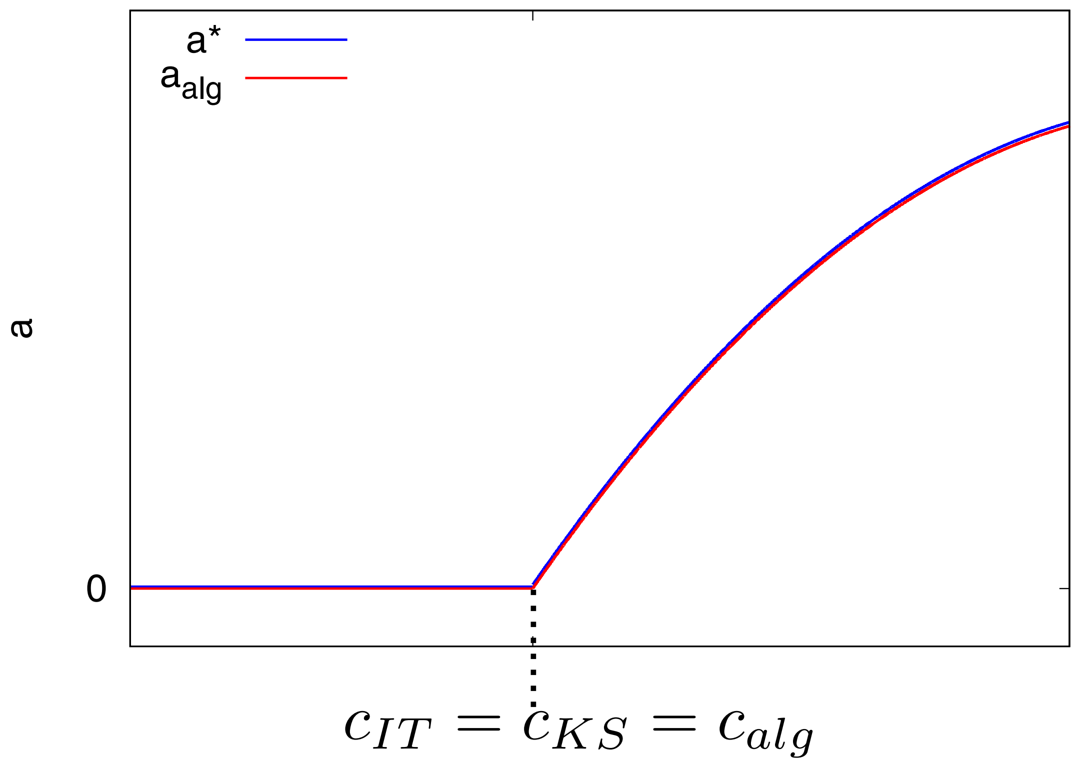
- Examples: SBM, planted random graph q-coloring, ...
  Given the random graph, infer hidden/planted structure

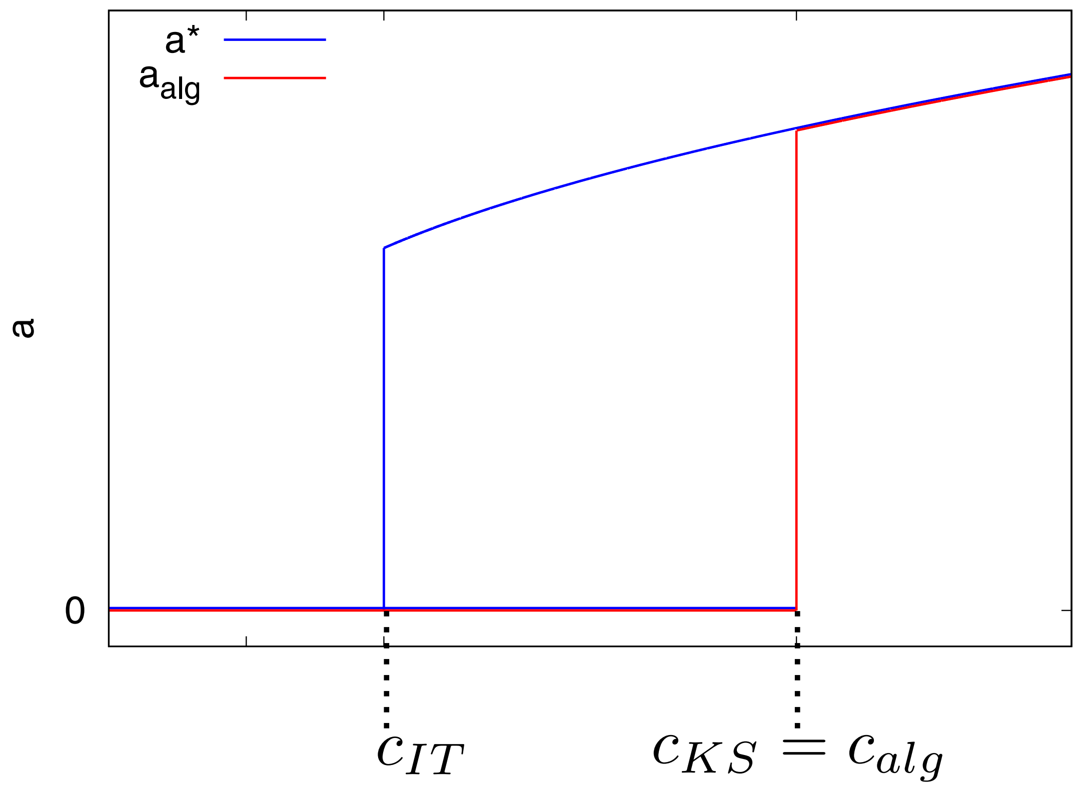
- Bayes optimality
  -> noise in the data = noise assumed in doing inference
  -> Nishimori condition in statistical physics
  -> replica symmetric free-energy
  -> Belief Propagation (BP) returns the right marginals!

Phase transitions & BP thresholds

Decelle et al. PRE 2011

optimal (Bayes)
easily achievable (BP)



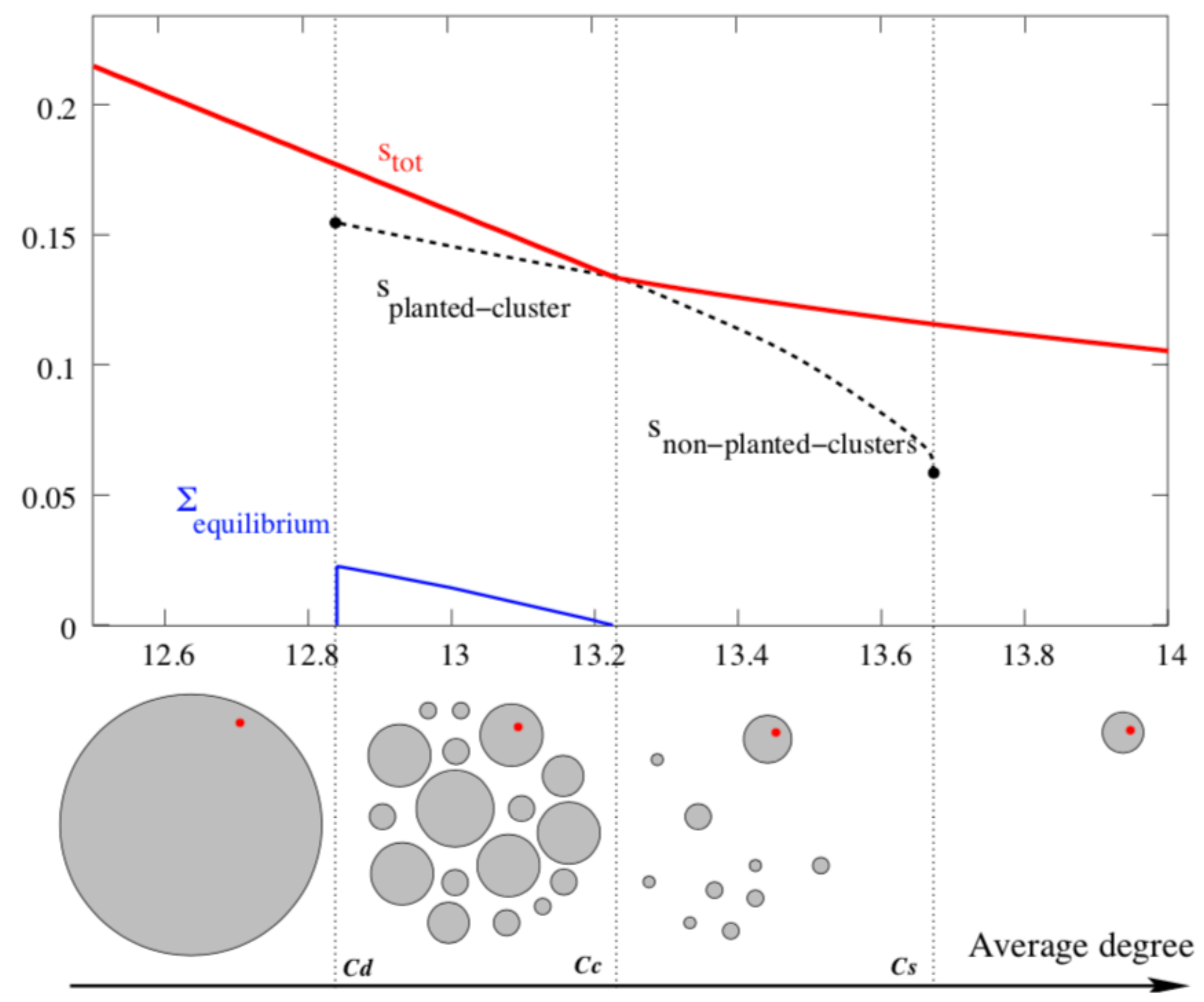


impossible / easy

impossible / hard / easy

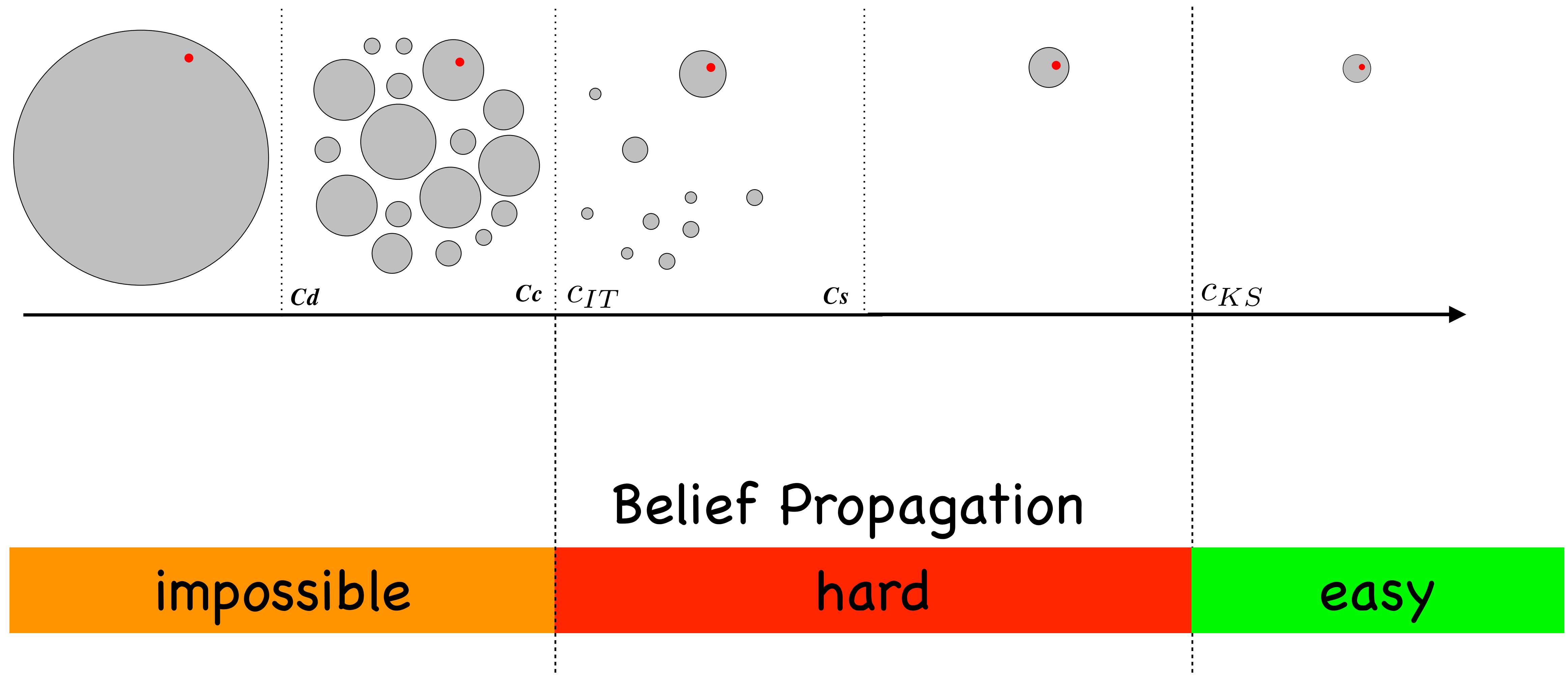
Quiet planting in random CSP

Krzakala Zdeborova PRL 2009



planted 5-coloring random graphs

phase diagram & algorithmic hardness



Glassy behavior in planted models

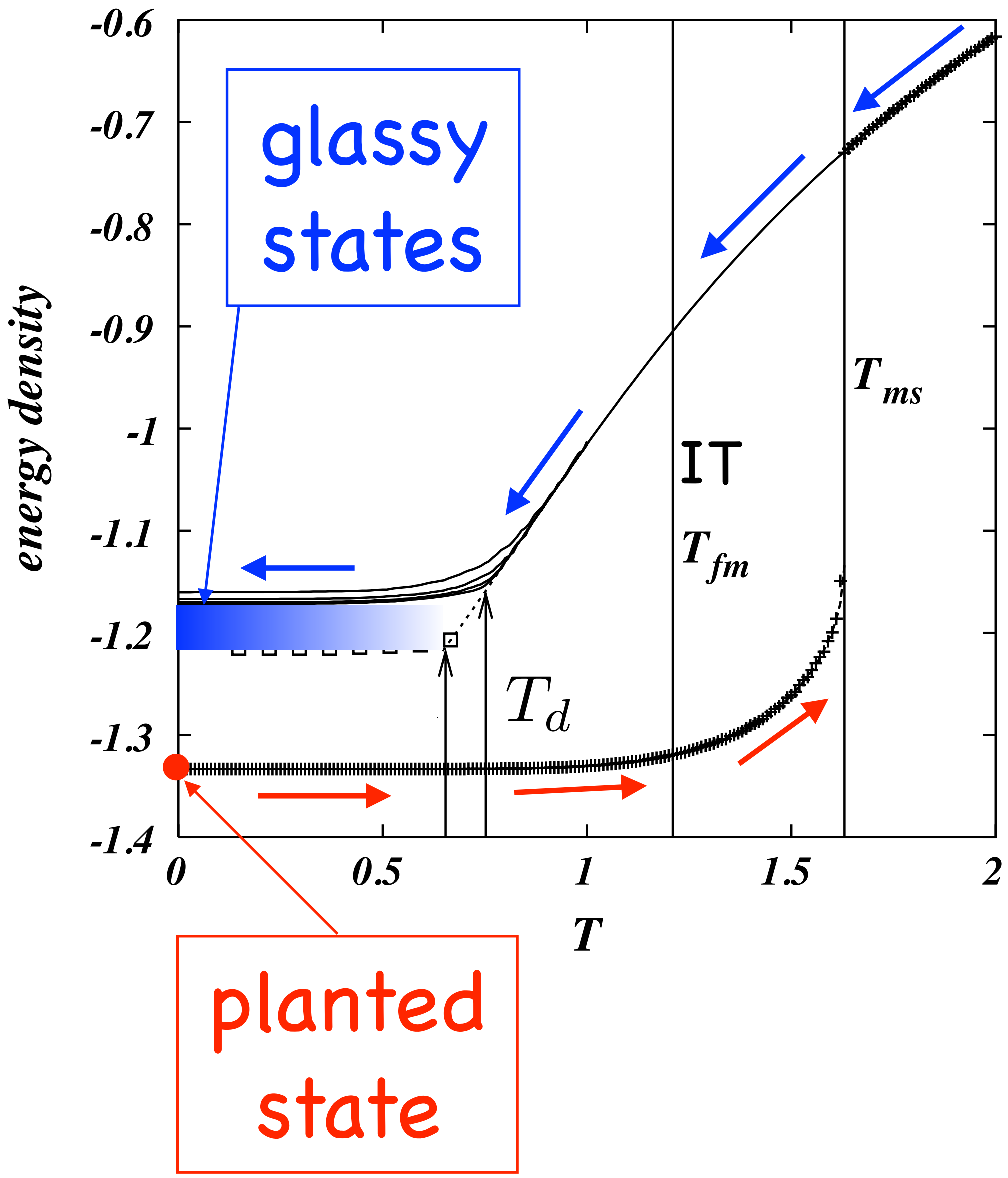
EUROPHYSICS LETTERS 15 August 2001

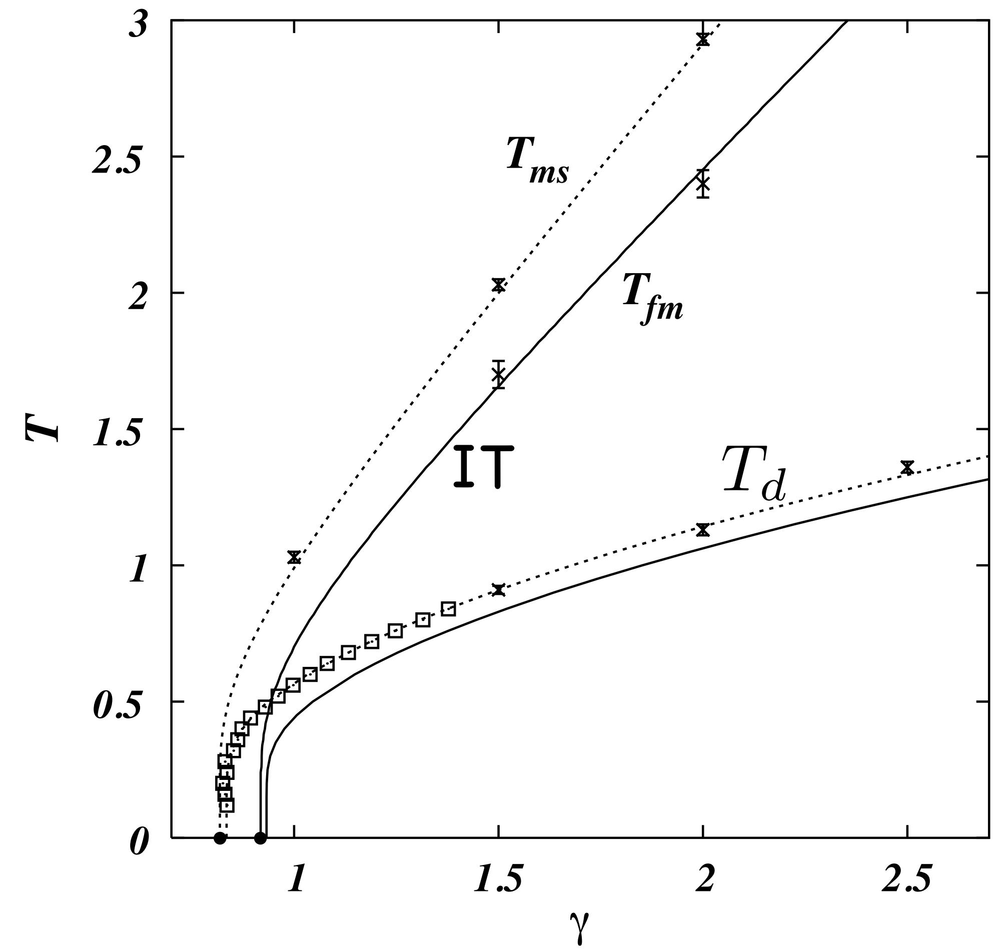
Europhys. Lett., **55** (4), pp. 465–471 (2001)

A ferromagnet with a glass transition

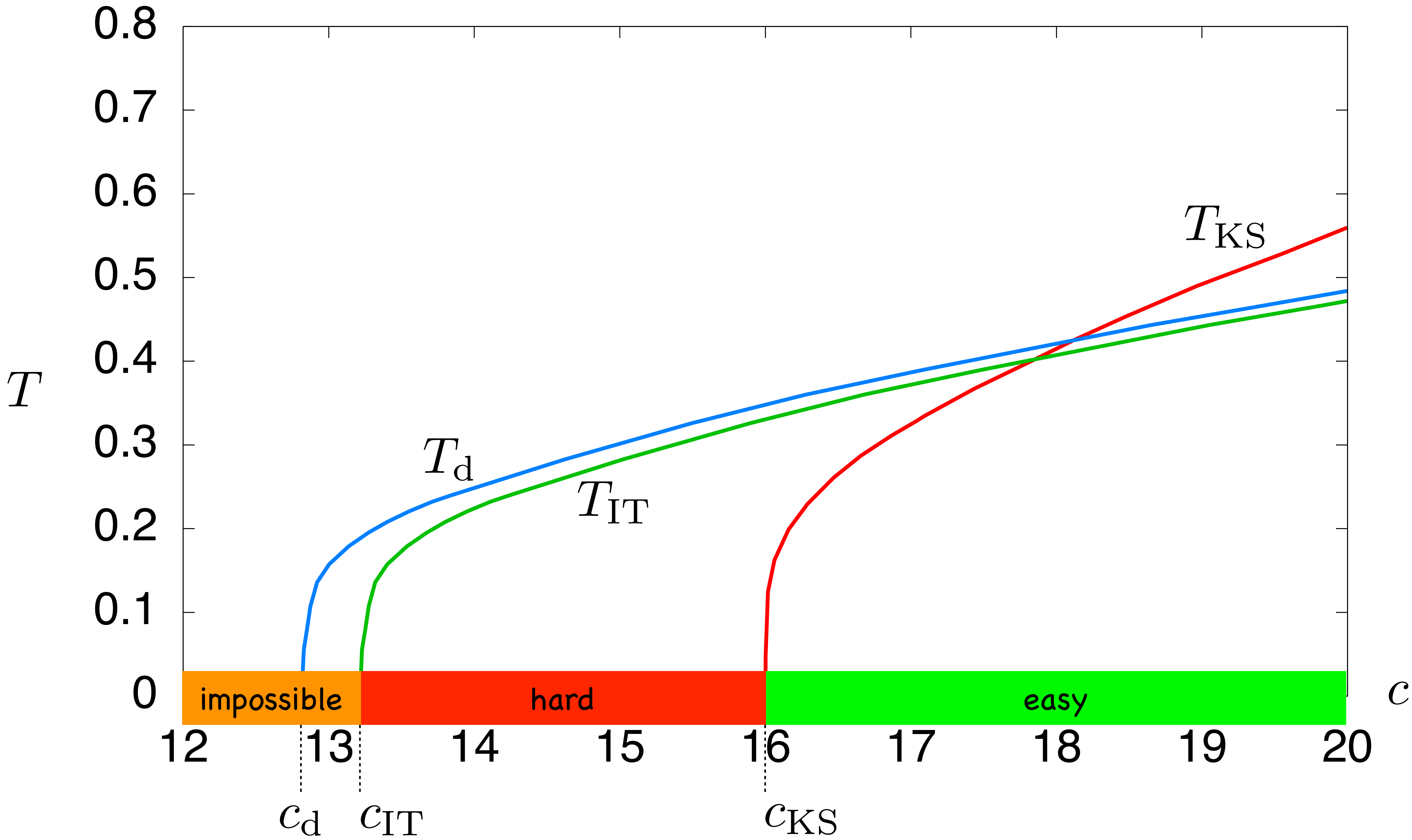
S. Franz[1](*), M. Mézard[2](**), F. Ricci-Tersenghi[1](***), M. Weigt[3](**) and R. Zecchina[1](***)

Ferromagnetic sparse 3-spin model

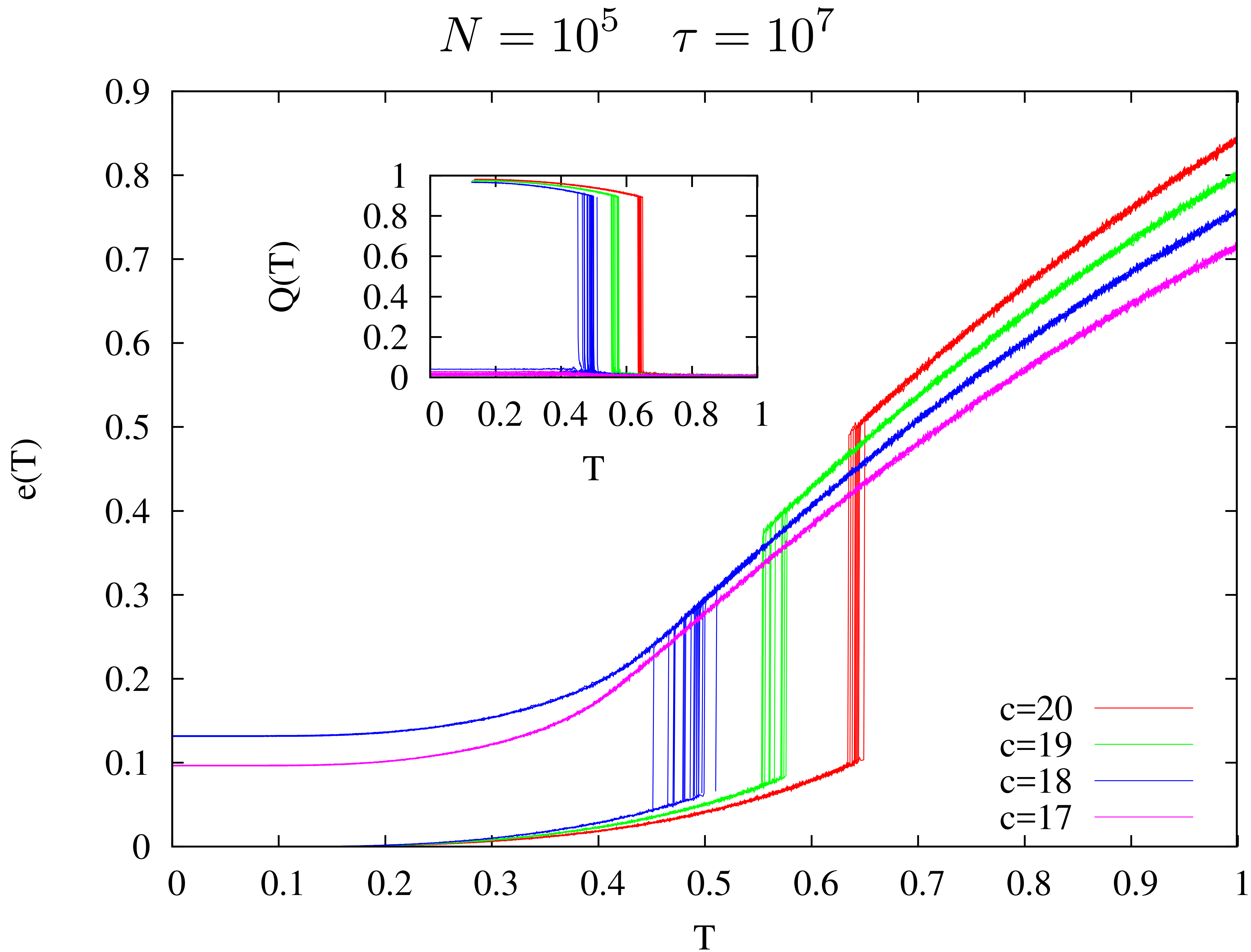




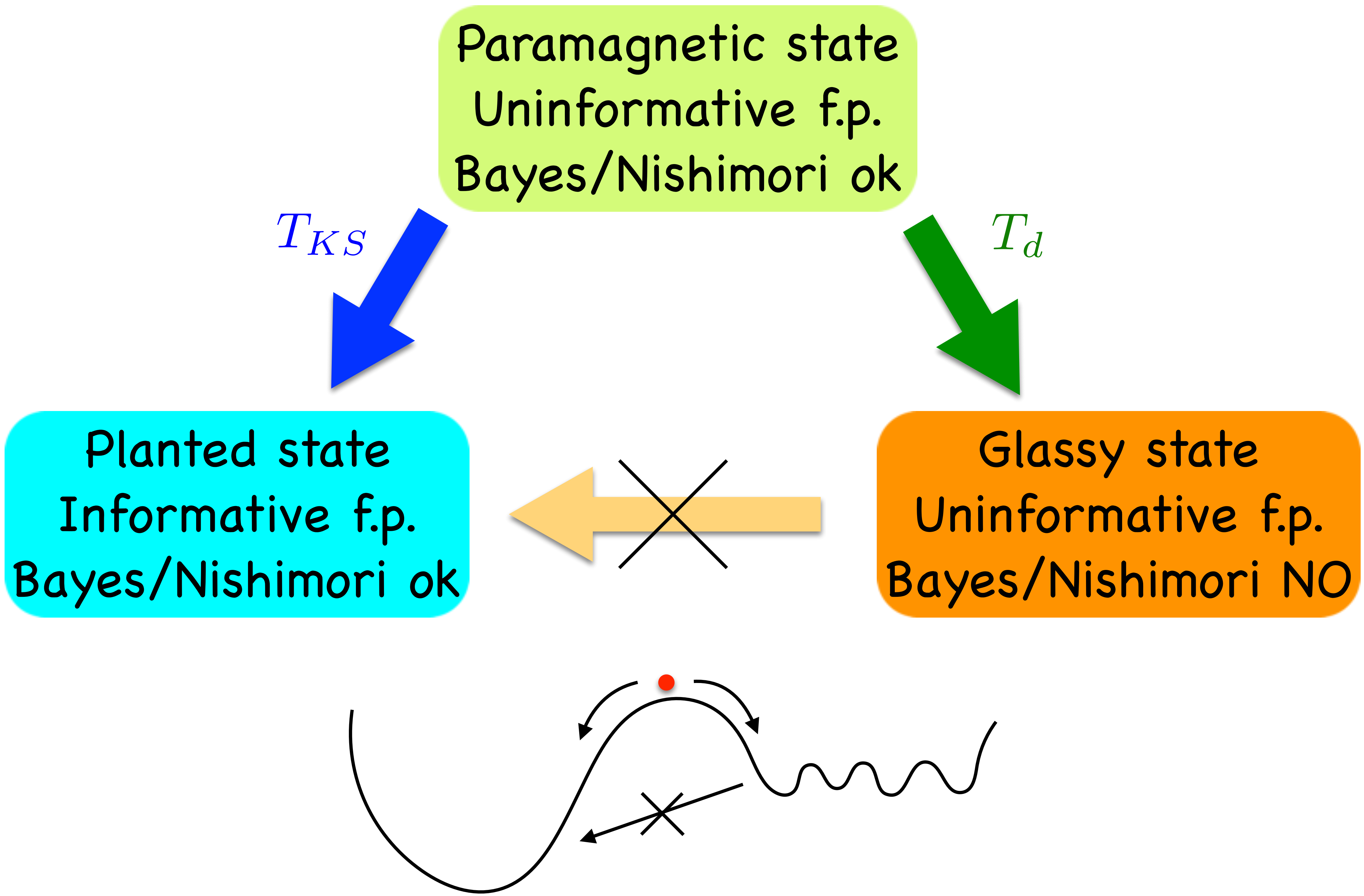
Phase diagram for predicting MC behavior



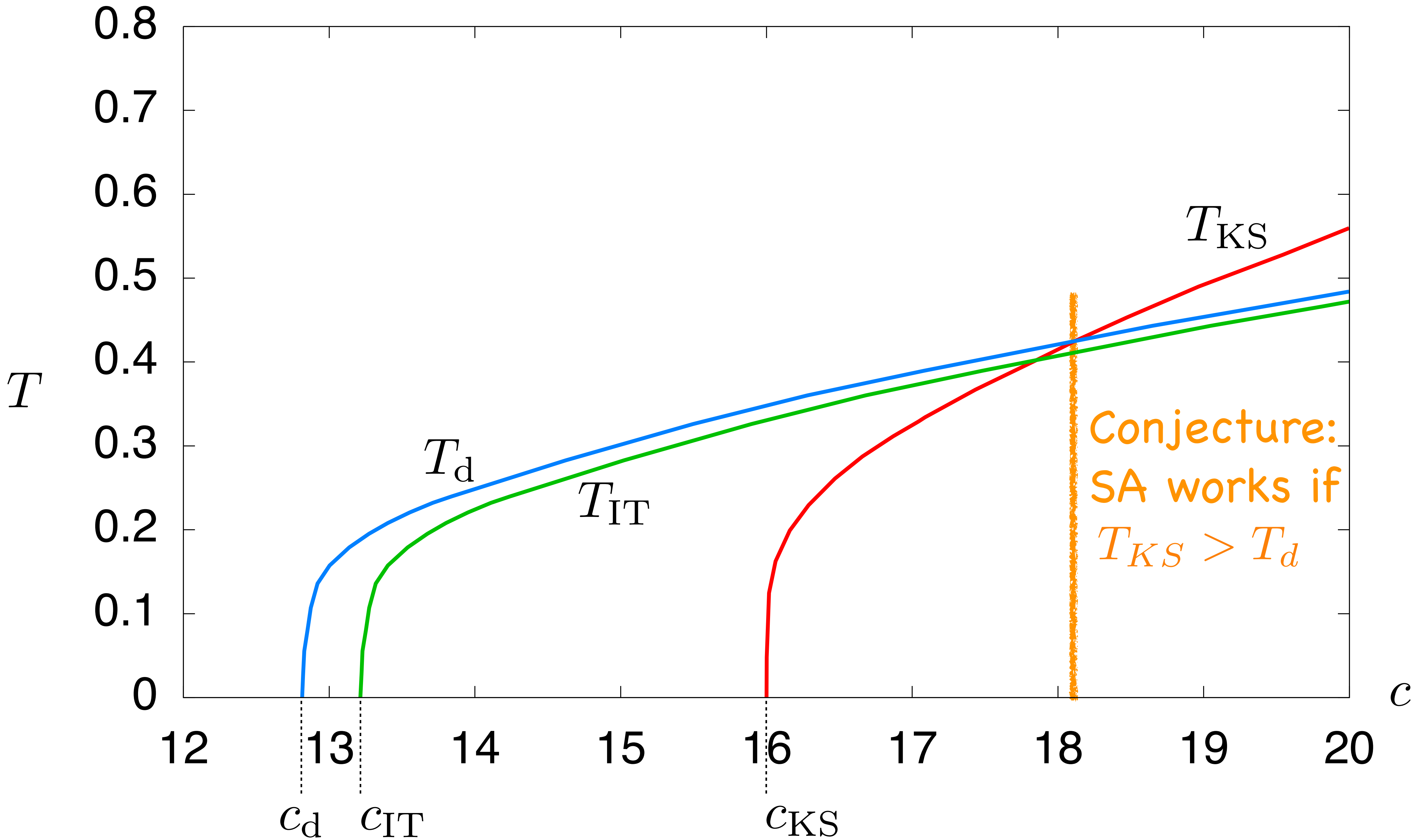
SA in planted models (q=5 coloring)



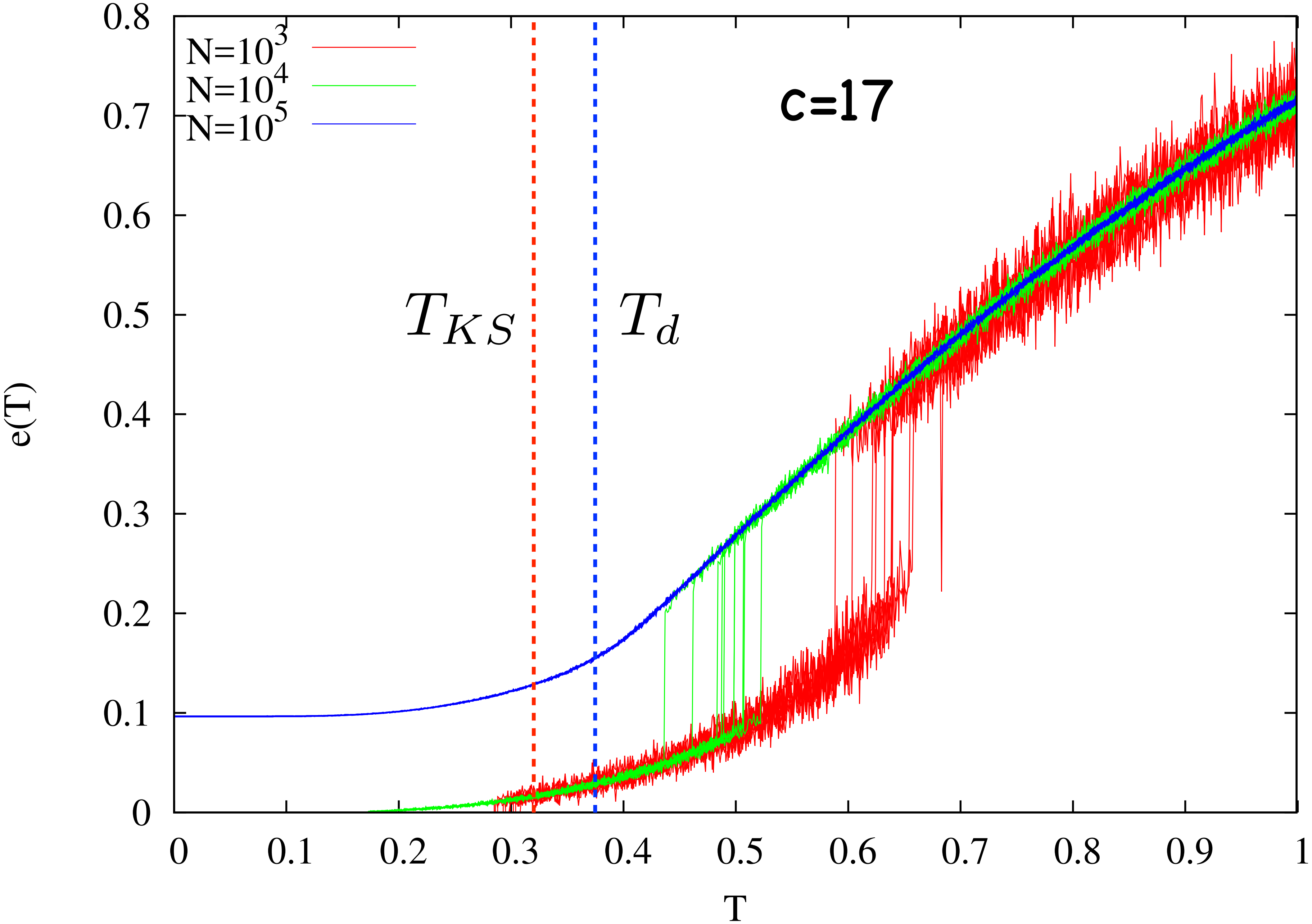
Picture for SA in planted models



Phase diagram for predicting MC behavior

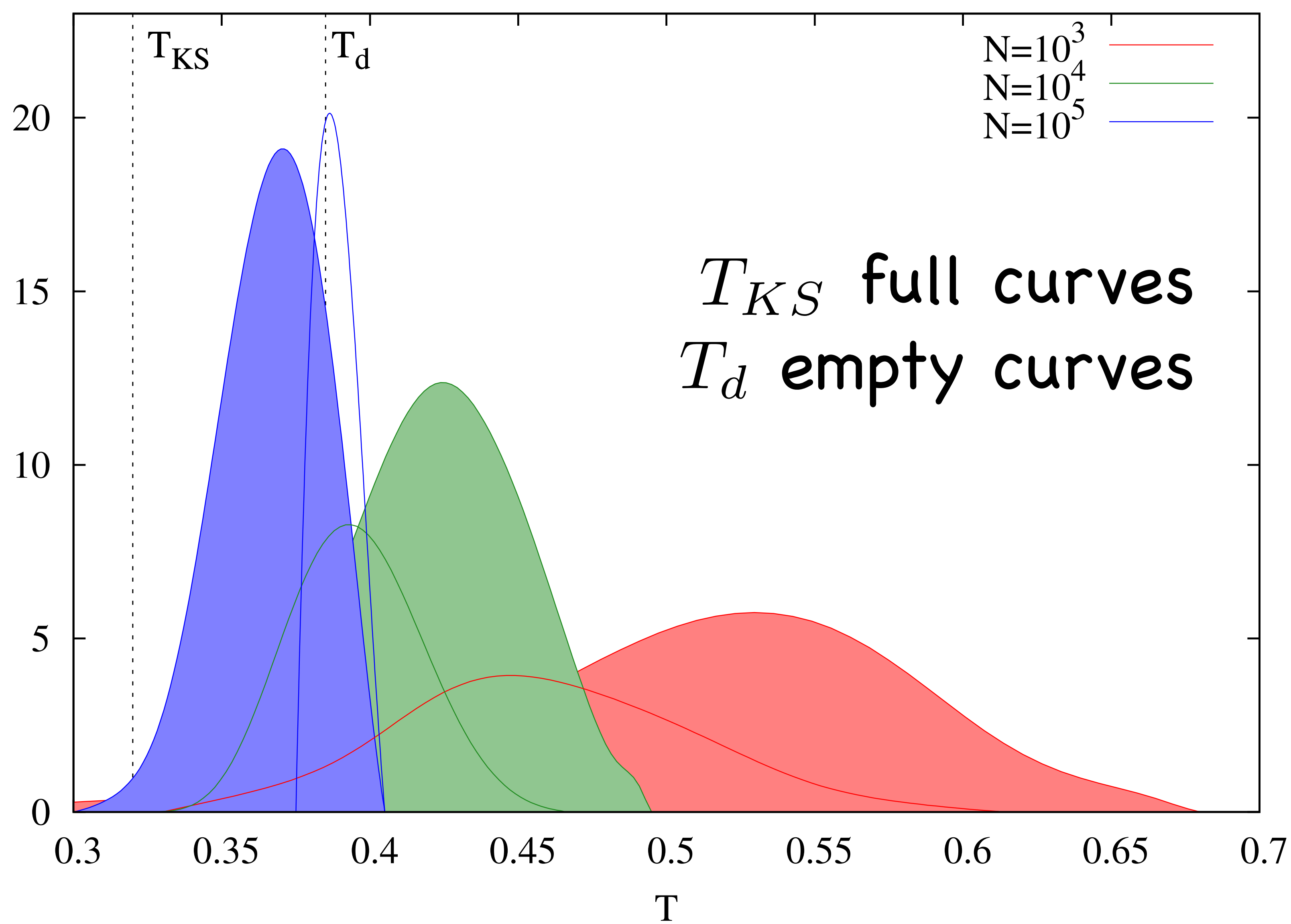


Strong finite size effects

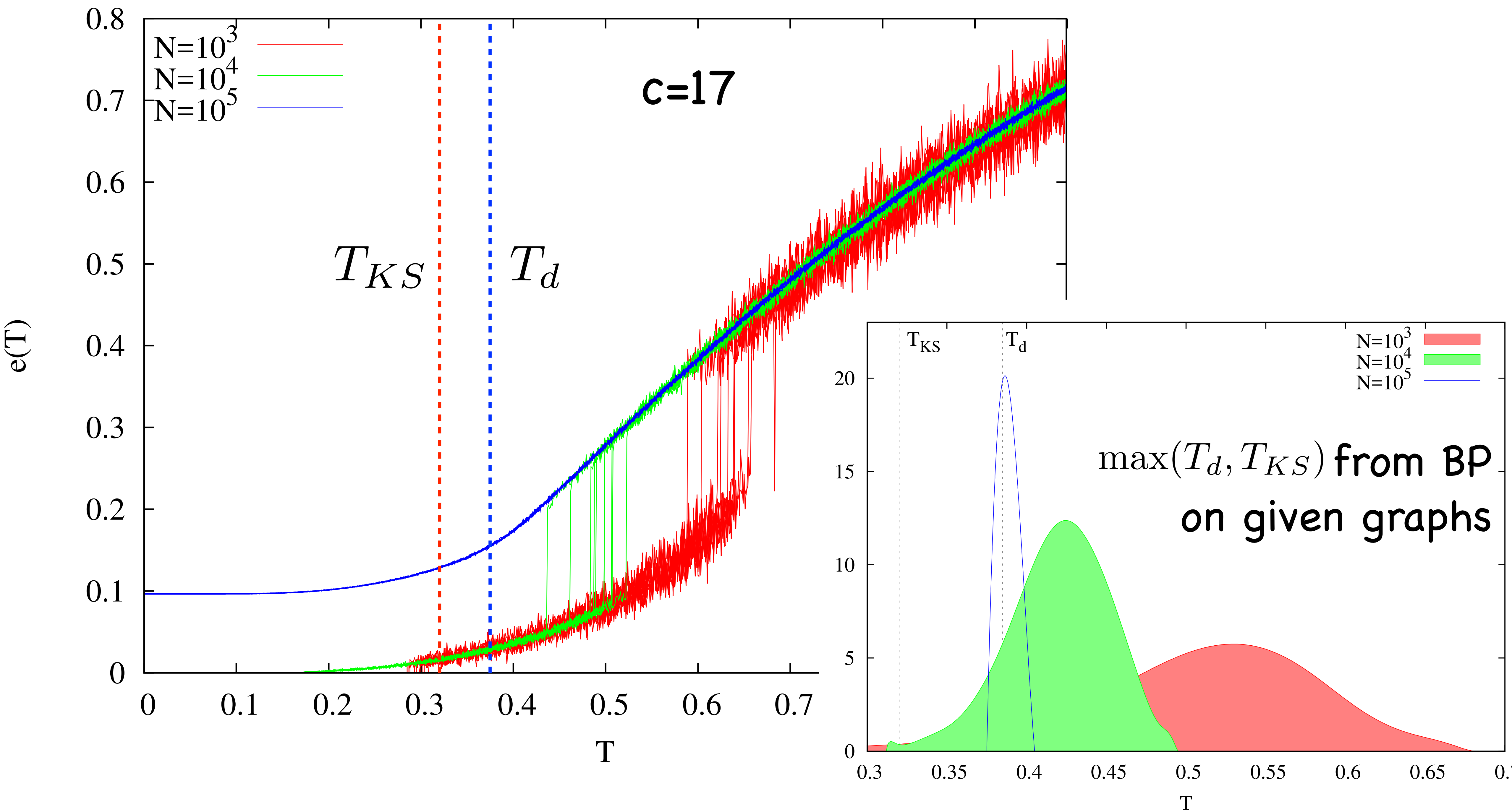


Analytical prediction of finite size effects

- Solution on the infinite tree -> thermodynamic limit
- BP on a given graph -> finite size effects (unexpected!)



Strong finite size effects (understood)



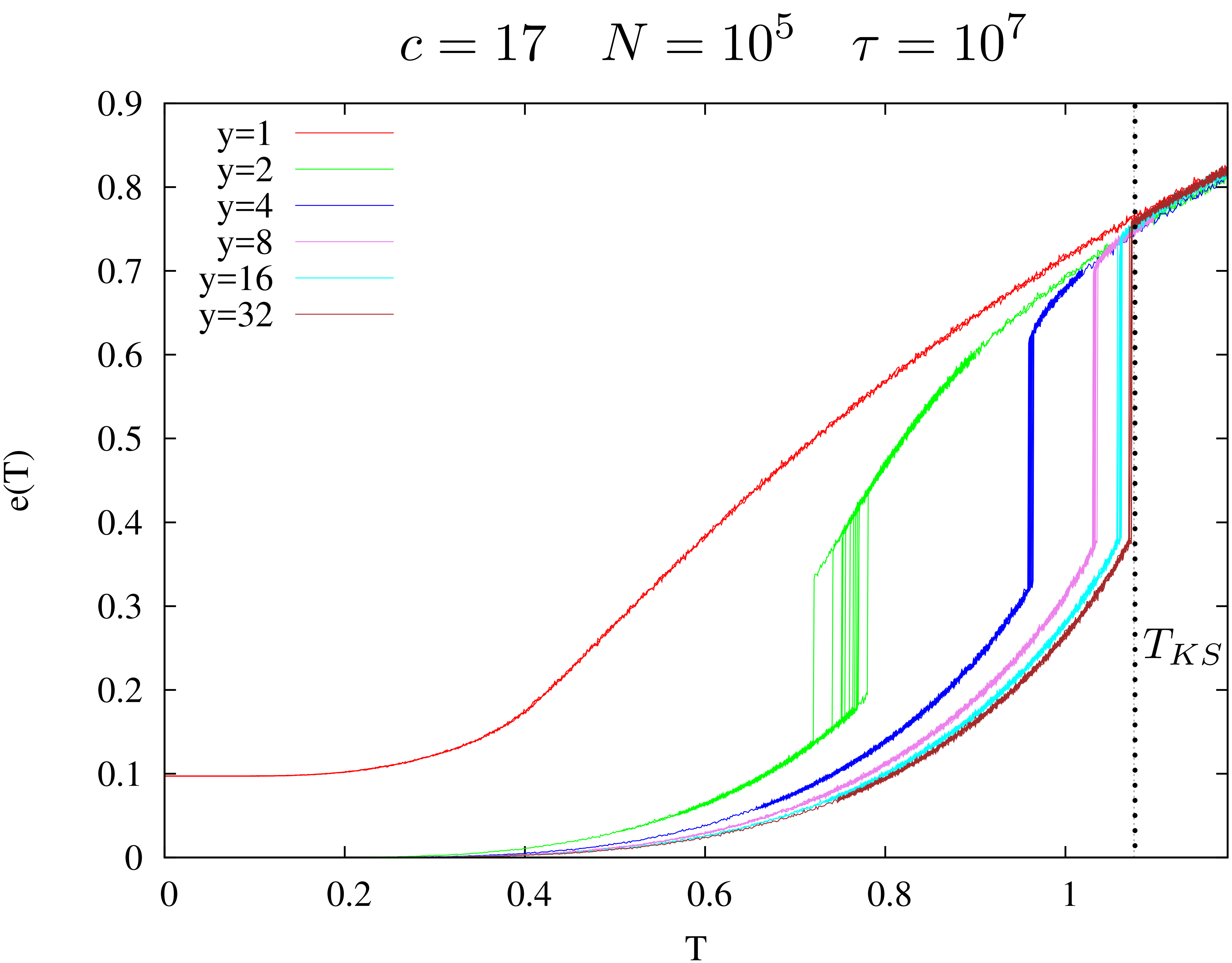
Replicated SA

- Proposed by Zecchina & co. to sample states of larger entropy with higher probability
- Very simple implementation
  - R replicas $\underline{\sigma}^a$ with $a = 1, \ldots, R$
  - energy function prefers replicas to be close by

$$\sum_a E(\underline{\sigma}^a) - \frac{\gamma}{R} \sum_{a<b} \sum_i \delta_{\sigma_i^a, \sigma_i^b}$$

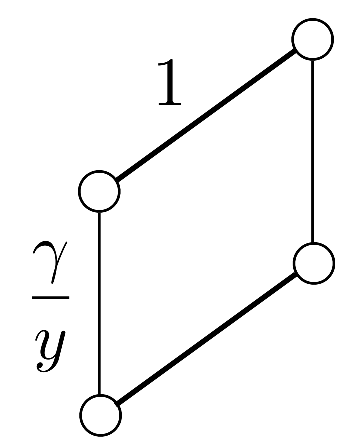
  - $\gamma = 1$ in all next plots

RSA in planted random coloring (q=5)



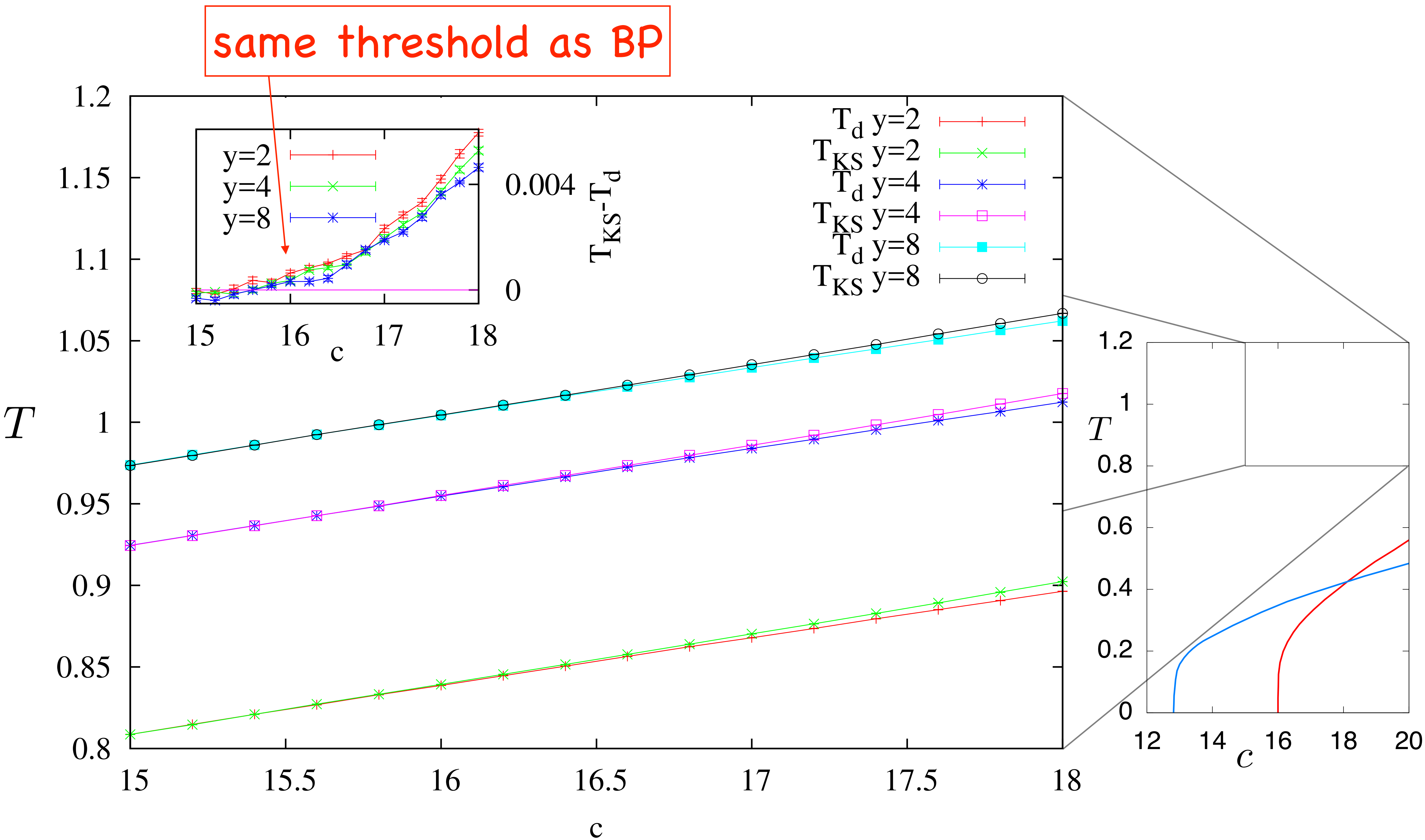
Replicated model: analytical results

- Exact solution is costly: supervariables with q^y values
- Run BP on the replicated graph (there are short loops!)



- For large y short loops become weaker -> the BP threshold coincides with RSA
- For y>1 the glass transition is continuous
- Computations on the infinite tree returns badly wrong thresholds (unless supervariables are used)

New critical lines for the replicated model



Conclusions and perspectives

- Analytic predictions for the algorithmic thresholds of SA and RSA (for y large)
- Mandatory to consider the dynamical transition towards glassy states (not seen by AMP or BP)
- Finite size effects can be studied via BP on finite graphs
- Glass transition in replicated model (y>1) is continuous

- Linear time MC algorithm work up to the boundary of the hard phase. Can super-linear MC algorithm work inside the hard phase?

Thanks!