Belief Propagation and Monte Carlo based algorithms to solve inference problems on sparse random graphs

Federico Ricci-Tersenghi
(La Sapienza, Roma)

in collaboration with
G. Semerjian and L. Zdeborova (arxiv:1806.11013)
and M. C. Angelini, A. Rampini

Outline

- Introduction to inference problems on sparse random graphs and trees
- Motivations for running long numerical experiments on given random inference problems
- New phases in sparse inference problems
- Numerical results (published and unpublished)
  - BP for planted random hypergraph bicoloring
  - BP for asymmetric stochastic block model
  - Monte Carlo based algorithms for planted random coloring
    - Simulated Annealing
    - Replicated Simulated Annealing
    - Parallel Tempering

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

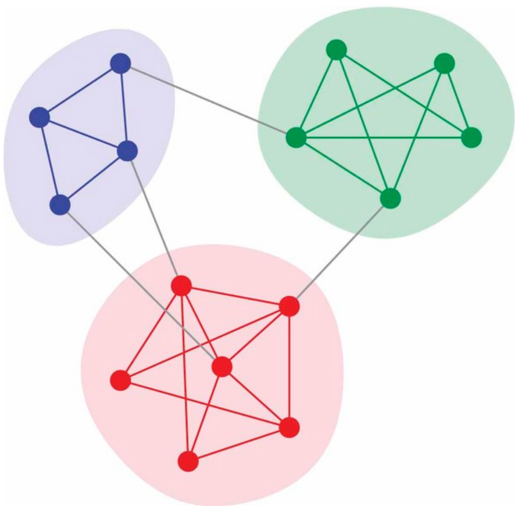
- Computing marginal probabilities is as hard as computing the partition function
- In random sparse inference problems -> estimate log(Z) via the Bethe approximation

Hidden partition model or stochastic block model (SBM)

- Random benchmark for the community detection problem
- Generate a partition of n nodes: e.g. q groups of size n/q
- Add independently edges between any pair of nodes according to the following probability

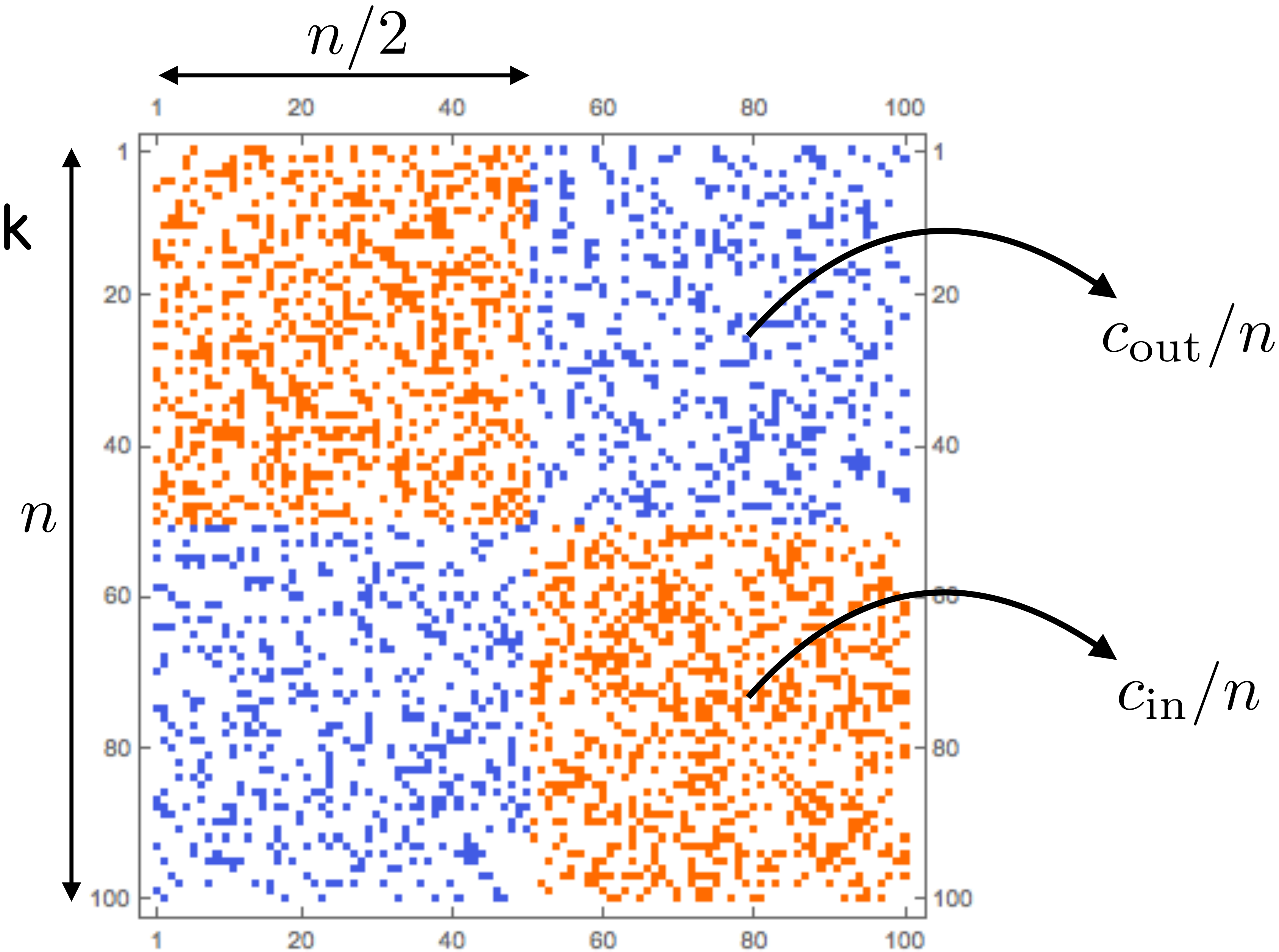
$$\mathbb{P}[(ij) \in E] = \begin{cases} c_{\text{in}}/n & \text{same group} \\ c_{\text{out}}/n & \text{different groups} \end{cases}$$

- Assortative model $c_{\text{in}} > c_{\text{out}}$
  Disassortative model $c_{\text{in}} < c_{\text{out}}$



The hidden partition model

Stochastic block model (SBM) with $q = 2$

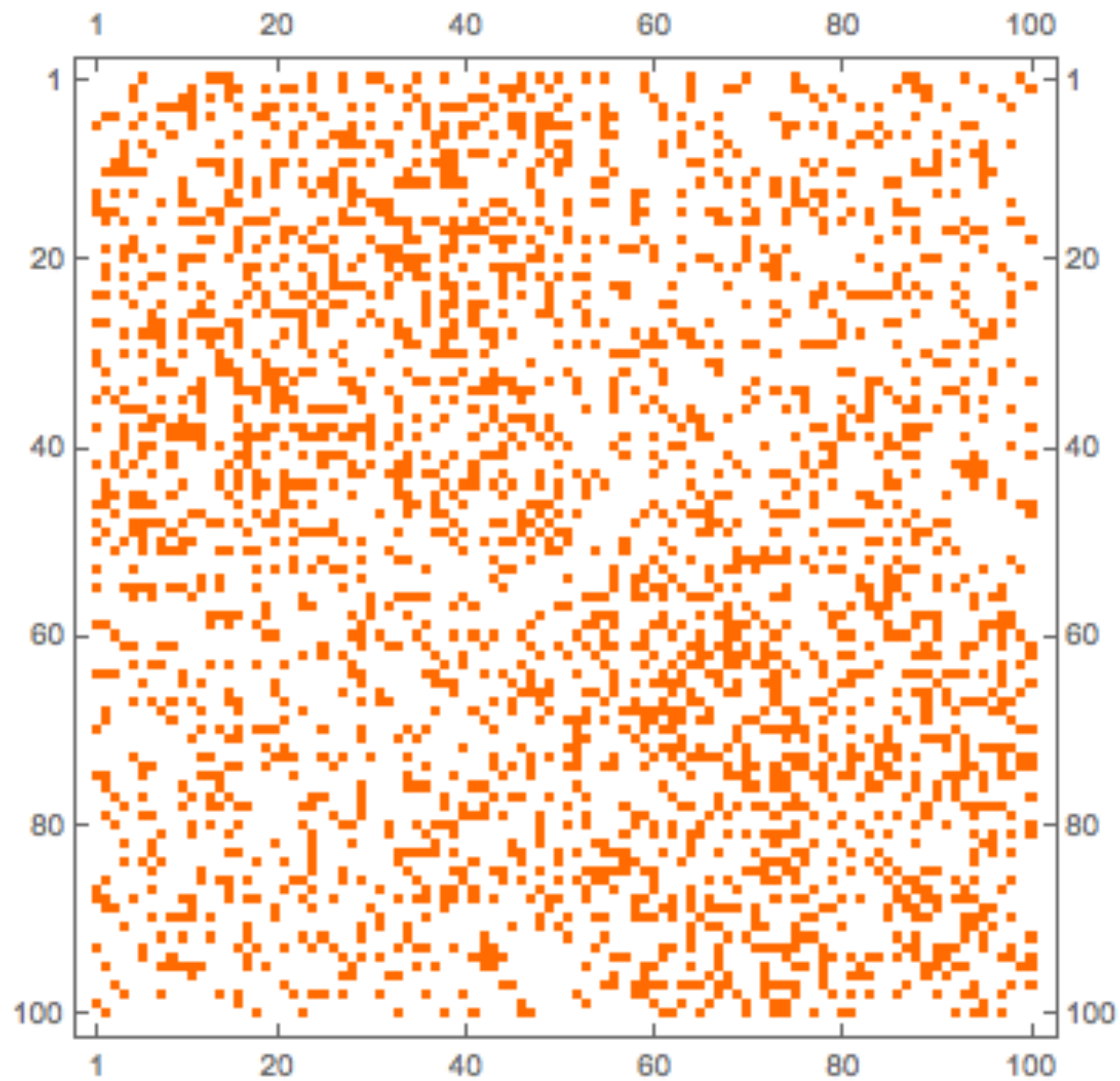


$$\mathbb{P}[(ij) \in E] = \begin{cases} c_{\mathrm{in}}/n & \text{same group} \\ c_{\mathrm{out}}/n & \text{different groups} \end{cases}$$

Assortative model:

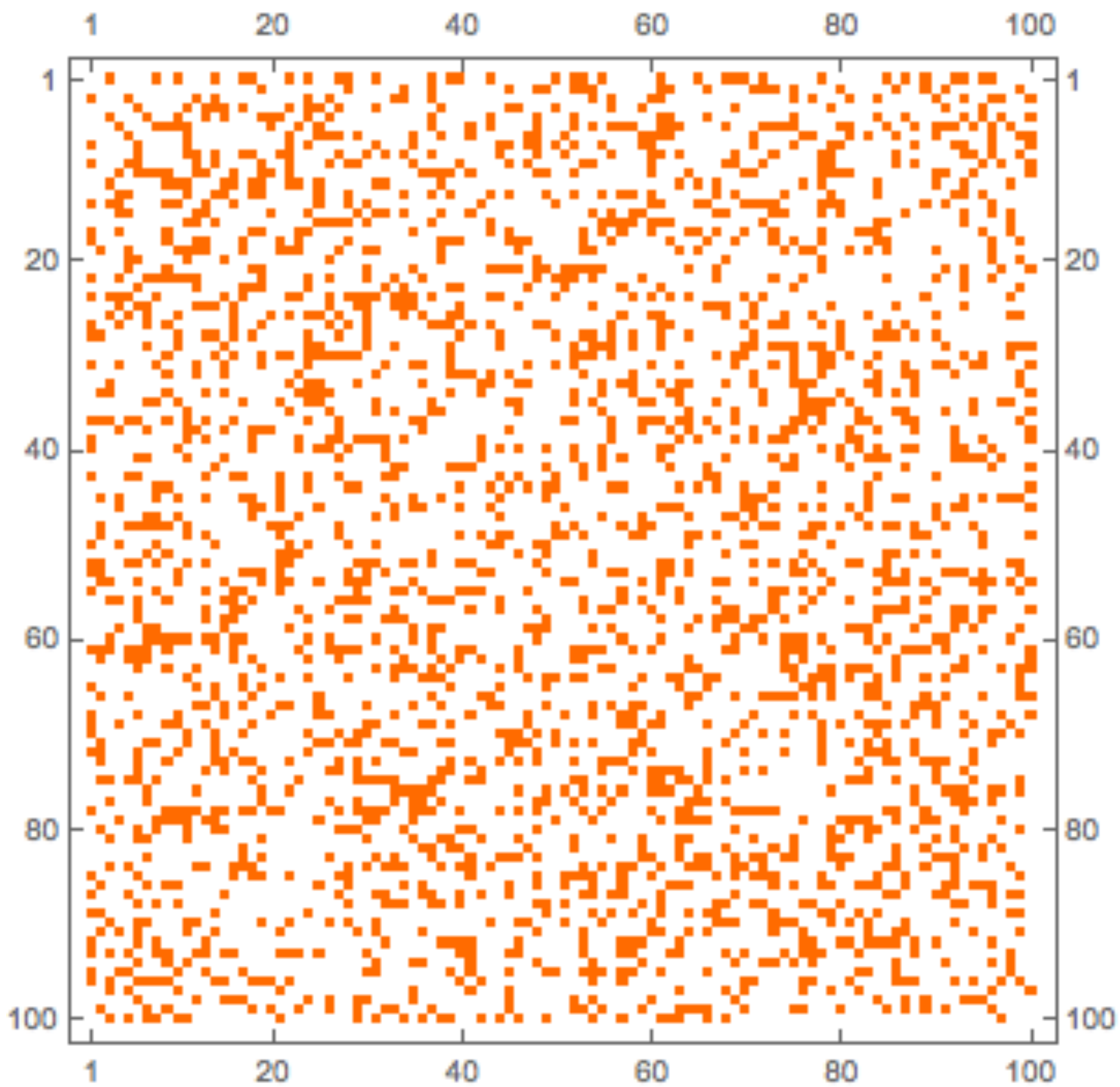
$c_{\mathrm{in}} > c_{\mathrm{out}}$

The hidden partition model



Colors are not provided !

The hidden partition model

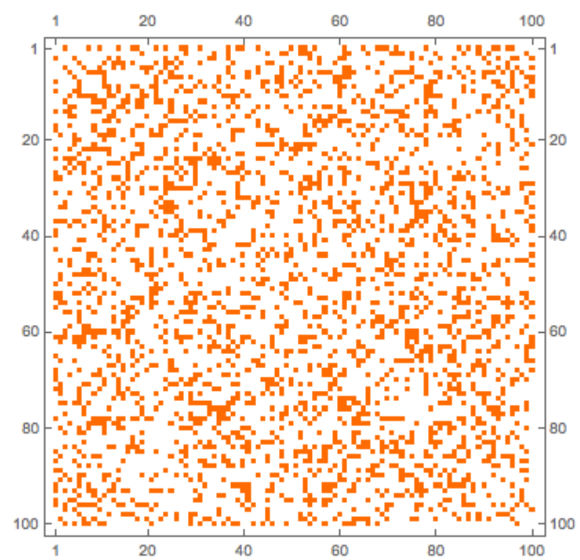


The right ordering neither !!

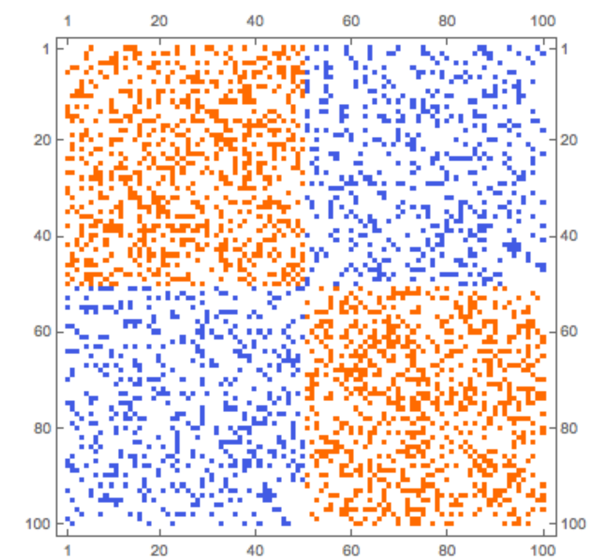
The hidden partition model

Given only the adjacency matrix

$A_{ij} = A_{ji} = \mathbb{I}[(ij) \in E]$



Infer the hidden partition



Hidden (true) partition -> x^*

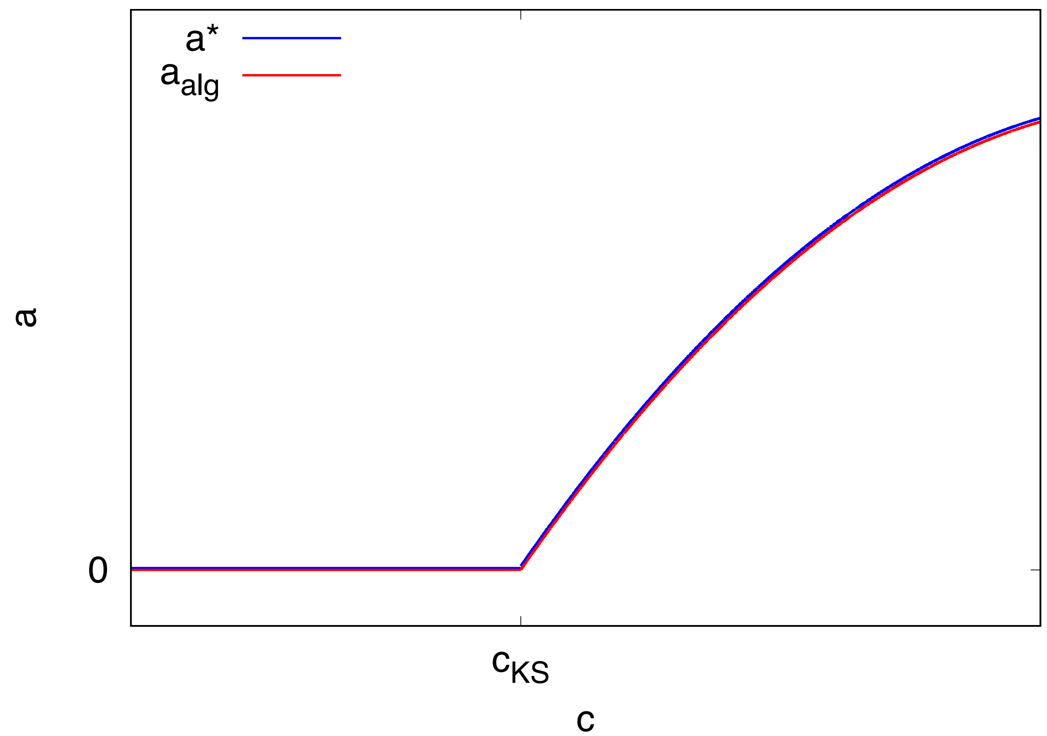
Estimated partition -> $\widehat{x}(A)$

Accuracy of inference is related to similarity -> $\langle \widehat{x}, x^* \rangle$

Conjectured phases in the symmetric SBM

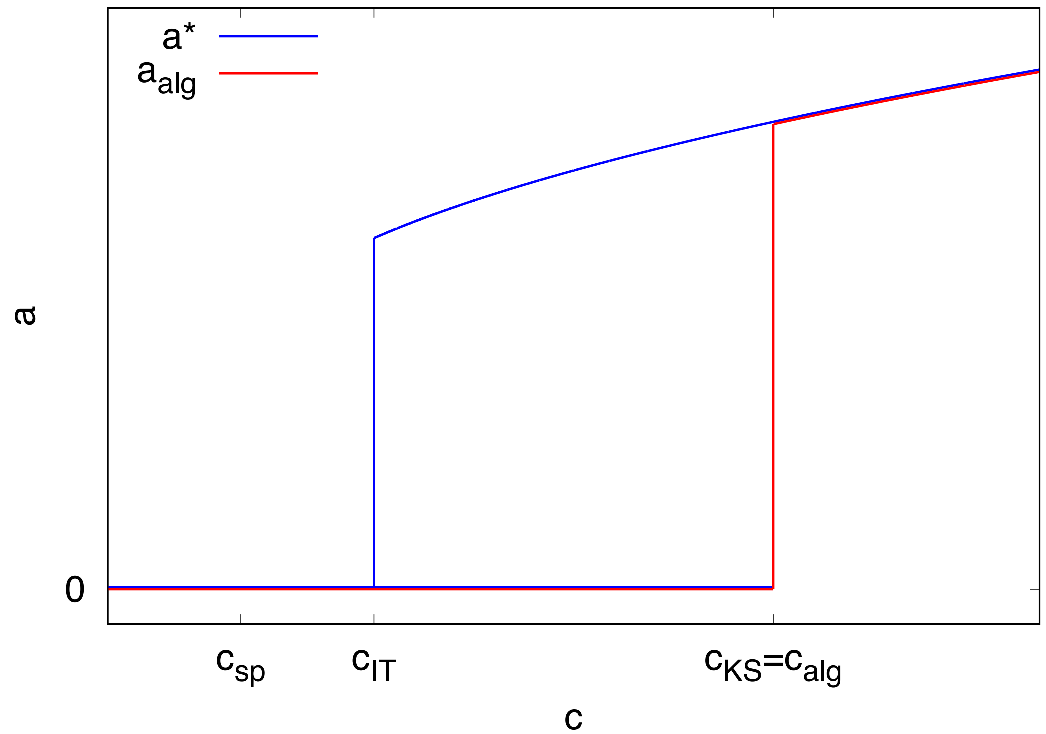
[Decelle, Krzakala, Moore, Zdeborova 11]

$q = 2, 3$



impossible / easy

$q \geq 5$



impossible / hard / easy

optimal (Bayes)   easily achievable (BP)

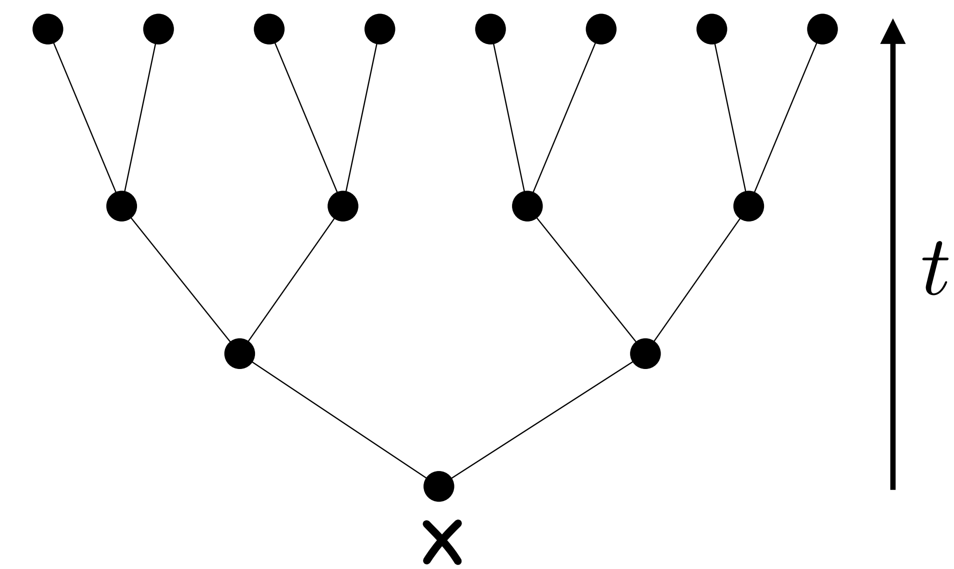
Kesten-Stigum (KS) and Information Theoretic (IT) transitions

A lucky situation!

- Bayes optimality (student knows prior and model)
  -> noise in the data = noise assumed in doing inference
  -> Nishimori condition in statistical physics
  -> replica symmetry (or dynamical 1RSB)
  -> Belief Propagation (BP) returns the right marginals!
  (if properly initialized)

- BP algorithmic threshold seems to coincide with the KS bound (analytically known)
  -> algorithm behavior is predictable!

The (robust) tree reconstruction problem



- Draw the root label x according to the prior
- Broadcast it (with noise) along the tree
- Observe the labels on the leaves
- Is it possible to infer x in the $t \to \infty$ limit ?

- Robust version:
  - observe a fraction ε of labels on the leaves
  - infer root label in the limit $t \to \infty$ first and then $\varepsilon \to 0$

Graph-tree connection

- The analytical solution is derived on an infinite tree
  - Population Dynamics solves equations on an infinite tree
- The inference problem is defined on a given random graph of large but finite size (loops and non-edges)
- The 2 problems are locally equivalent
  ...but global graph-tree connection is delicate:
  - loopy BP is run for a time much larger than the girth of the graph, i.e. loopy BP messages run along cycles many many times before reaching convergence
  - BP on a given graph has no information at all about the planted configuration and has to amplify the initial noise

Two classes of algorithms

(and motivations for studying it)

- Belief Propagation run on a specific graph:
  - Bayes optimal (if properly initialized), but not very robust on non-random graphs coming from real world applications
  - Is fully predictable from the analytic solution derived on a tree and solved via the Population Dynamics algorithm ?
- 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)
  - Previous conjectures probably need to be improved

Planted random hypergraph bicoloring

- N binary variables $\sigma_i = \pm 1$ (black/white)
- Planted config. $\sigma_i^* = 1$ for $i < N/2$ and $\sigma_i^* = -1$ for $i \geq N/2$
- $M = \alpha N$ constraints involving K variables each
- Solution = all constraints are non-monochromatic
- Hypergraph built s.t. the planted config. is solution
- Given the hypergraph, write the posterior probability and use BP to compute marginals

damping factor
$\gamma = 0.5$

$$\eta_{i\to a}^{(t)} = \frac{\prod_{b\in\partial i\setminus a} \hat{\eta}_{b\to i}^{(t-1)}}{\prod_{b\in\partial i\setminus a} \hat{\eta}_{b\to i}^{(t-1)} + \prod_{b\in\partial i\setminus a}(1-\hat{\eta}_{b\to i}^{(t-1)})} ,$$

$$\hat{\eta}_{a\to i}^{(t)} = \gamma\, \hat{\eta}_{a\to i}^{(t-1)} + (1-\gamma)\frac{1-\prod_{j\in\partial a\setminus i} \eta_{j\to a}^{(t)}}{2-\prod_{j\in\partial a\setminus i} \eta_{j\to a}^{(t)} - \prod_{j\in\partial a\setminus i}(1-\eta_{j\to a}^{(t)})}$$

Planted random hypergraph bicoloring

- BP initialization
  - with prob. q_0 -> $\hat{\eta}_{a\to i}^{(0)} = \eta_{i\to a}^{(0)} = \mathbb{I}[\sigma_i^* = 1]$
  - with prob. $1 - q_0$ -> $\hat{\eta}_{a\to i}^{(0)} = 0.5 \ , \quad \eta_{i\to a}^{(0)} \in [0.45, 0.55]$
- Remind $\hat{\eta} = \eta = 0.5$ is always a solution to BP equations
- q_0 is the initial overlap with the planted configuration
- To solve the inference problem with no information on the planted configuration BP must be run with $q_0 = 0$
  The initial configuration is a small perturbation around the uninformative fixed point ($\hat{\eta} = \eta = 0.5$)
  Beyond KS, the perturbation grows
- Convergence if for every message $\left|\eta_{i\to a}^{(t+1)} - \eta_{i\to a}^{(t)}\right| < 10^{-8}$

Planted random hypergraph bicoloring

- At the fixed point $\{\eta^\star_{i\to a}, \hat{\eta}^\star_{a\to i}\}$ we compute the local magnetizations (marginal probabilities) as

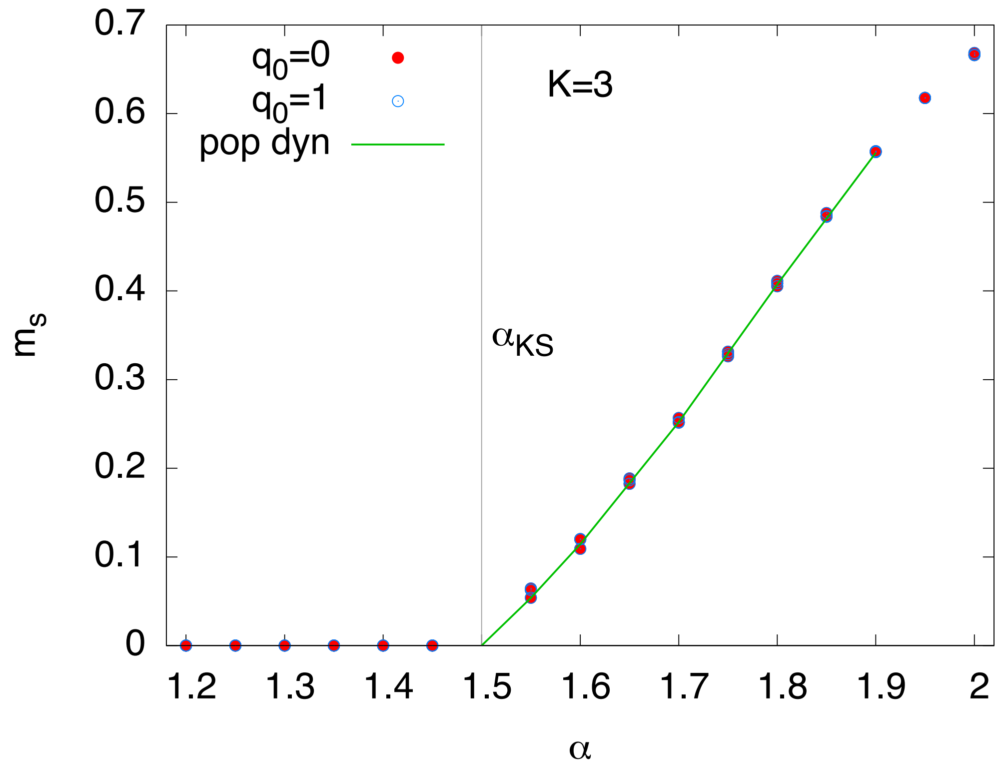
$$m_i^\star = \frac{\prod_{a\in\partial i}\hat{\eta}^\star_{a\to i} - \prod_{a\in\partial i}(1-\hat{\eta}^\star_{a\to i})}{\prod_{a\in\partial i}\hat{\eta}^\star_{a\to i} + \prod_{a\in\partial i}(1-\hat{\eta}^\star_{a\to i})}$$

- Detection of the planted configuration is signalled by

$$m_s = \frac{1}{N}\left|\sum_i m_i^\star s_i^*\right|, \qquad m_2 = \frac{1}{N}\sum_i (m_i^\star)^2, \qquad q = \frac{1}{N}\left|\sum_i \mathrm{sign}(m_i^\star) s_i^*\right|$$

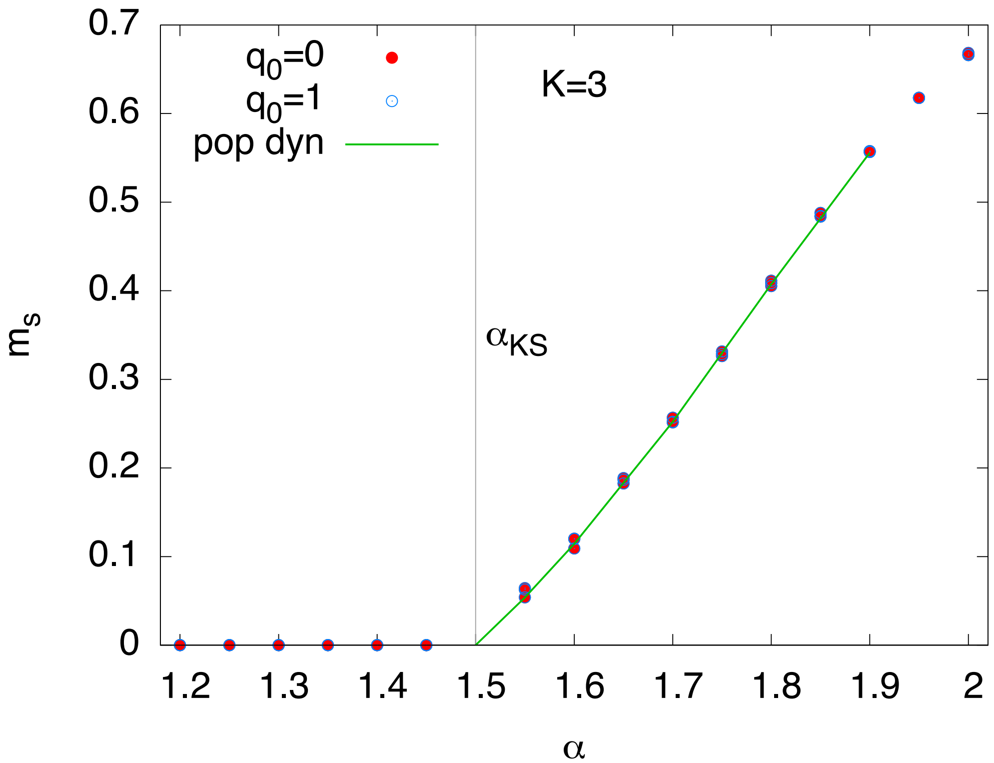
- The uninformative fixed point has $m_i^\star = m_s = m_2 = q = 0$

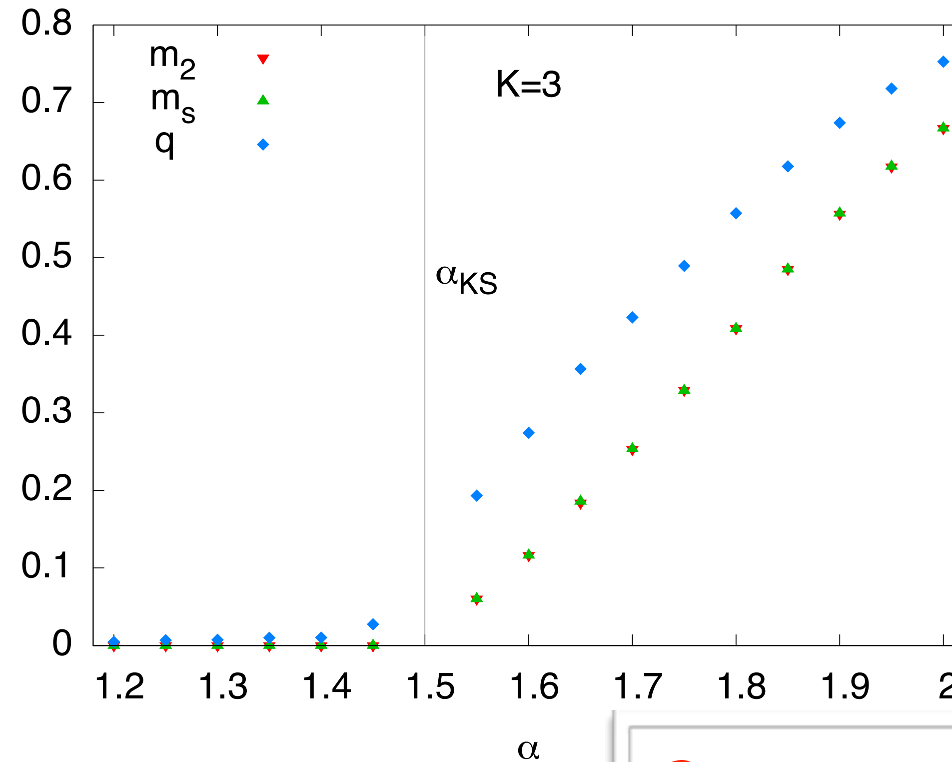
Planted random hypergraph bicoloring



3 samples of size N=10^6

Planted random hypergraph bicoloring

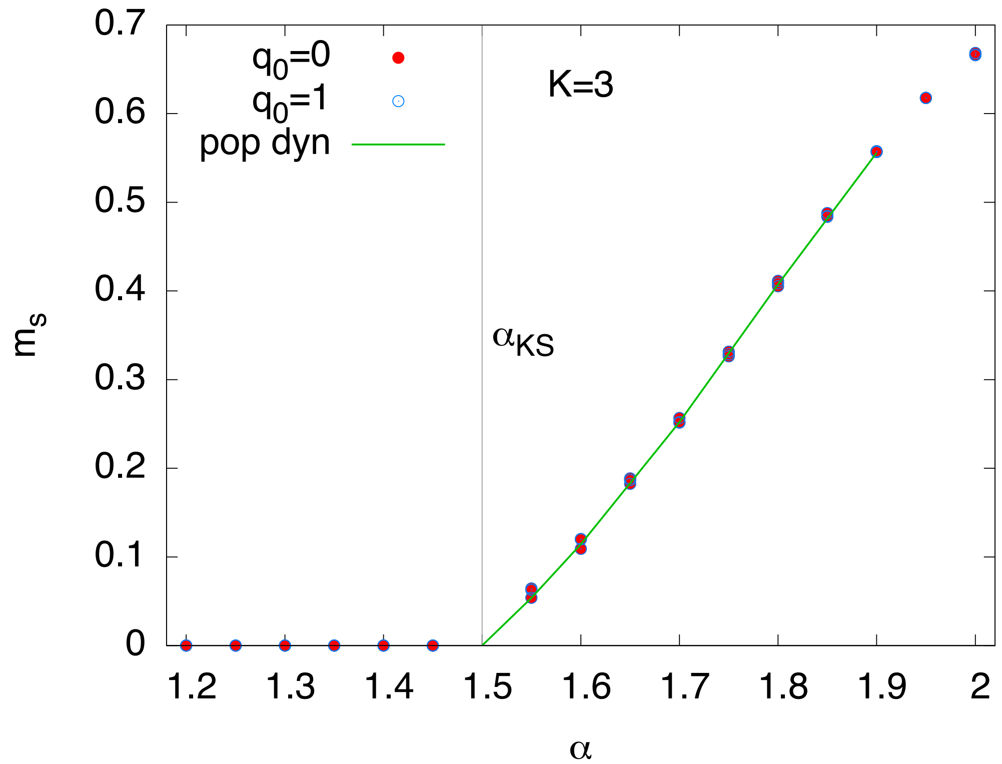


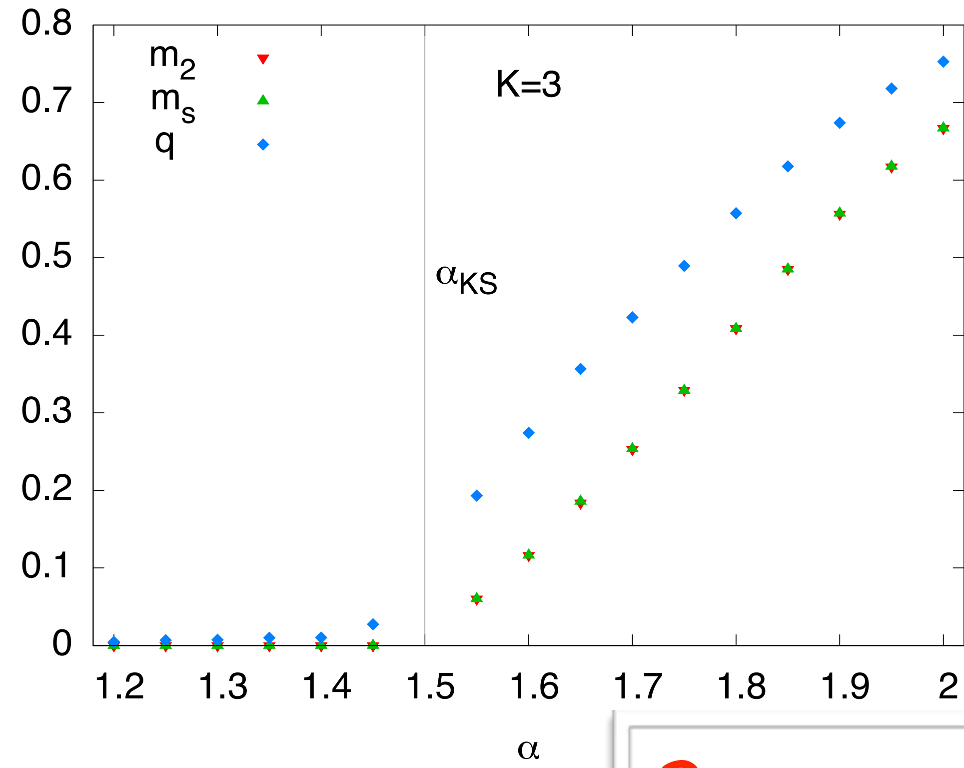


Bayes condition

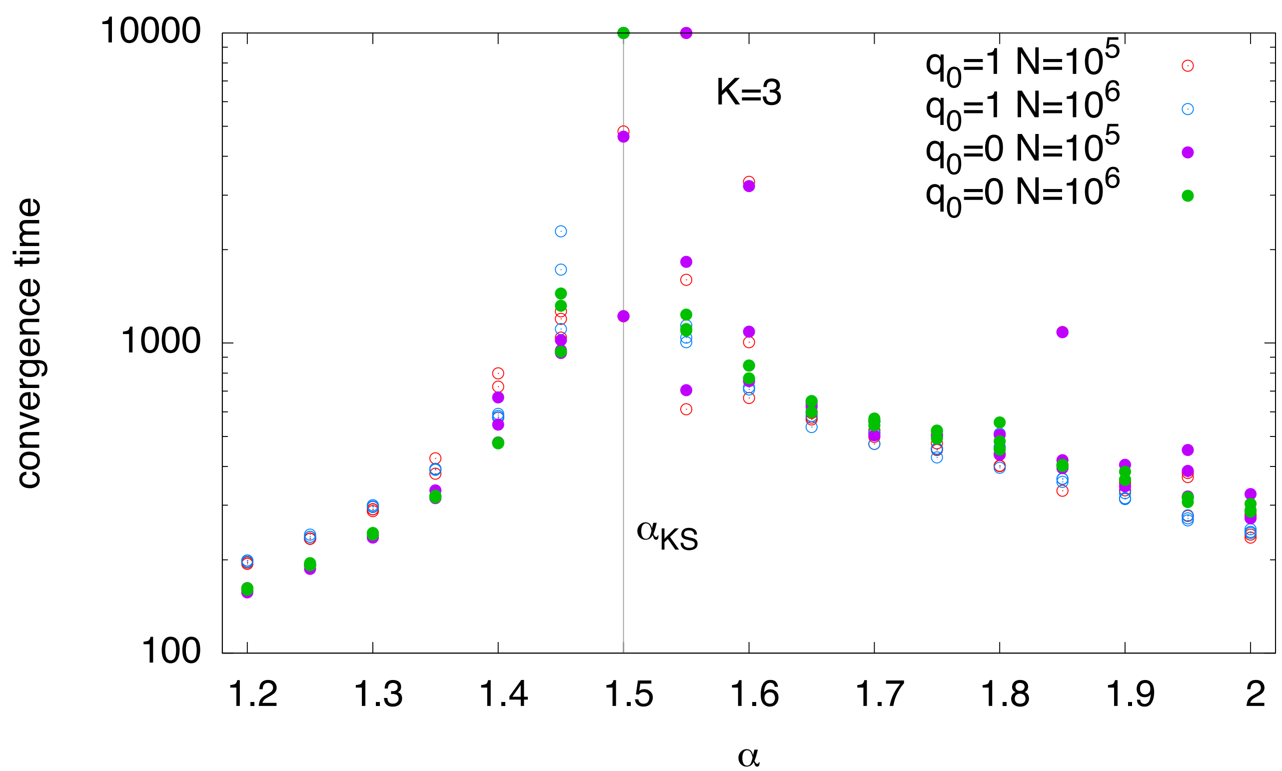
$$m_s = m_2$$

Planted random hypergraph bicoloring

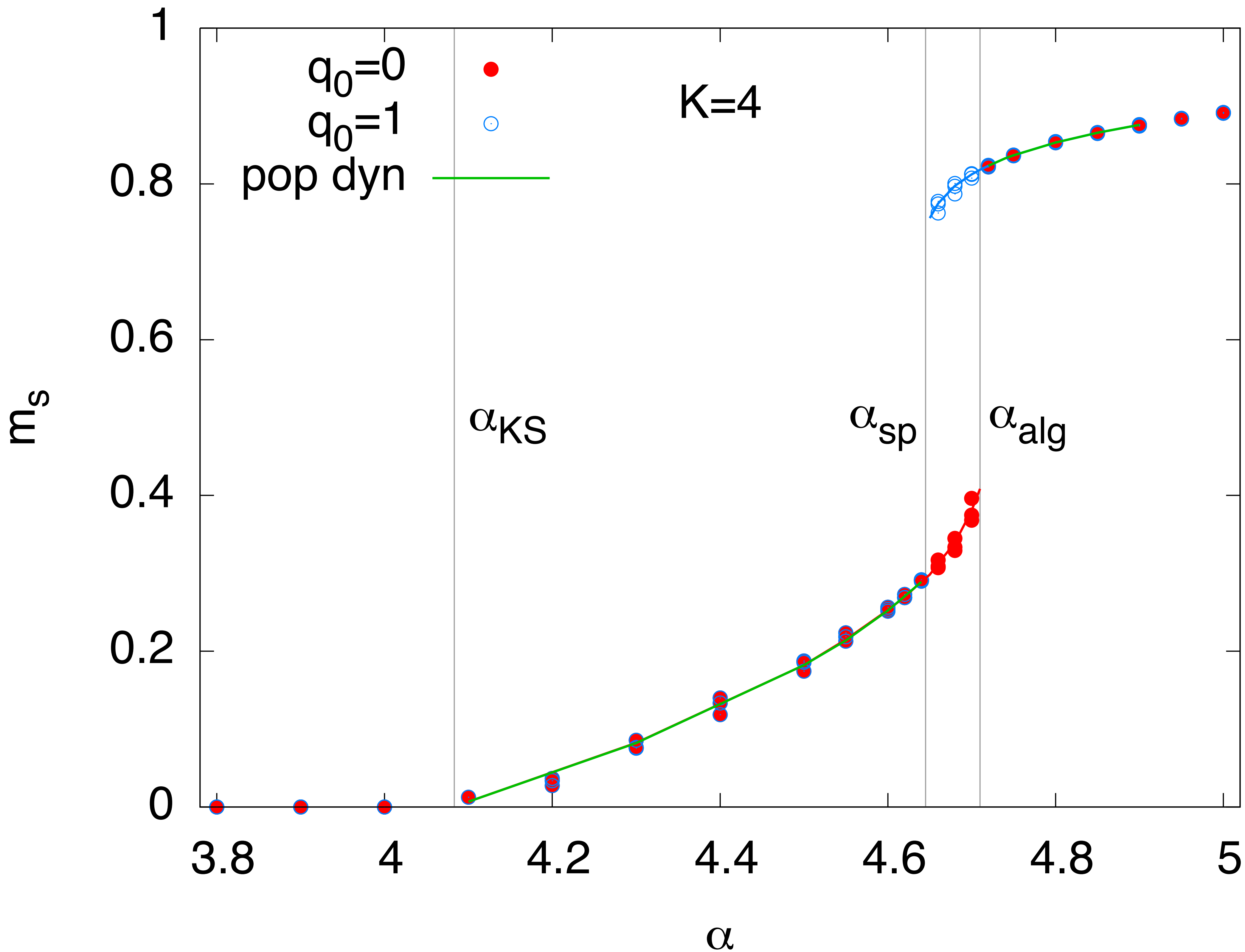




Bayes condition
$m_s = m_2$

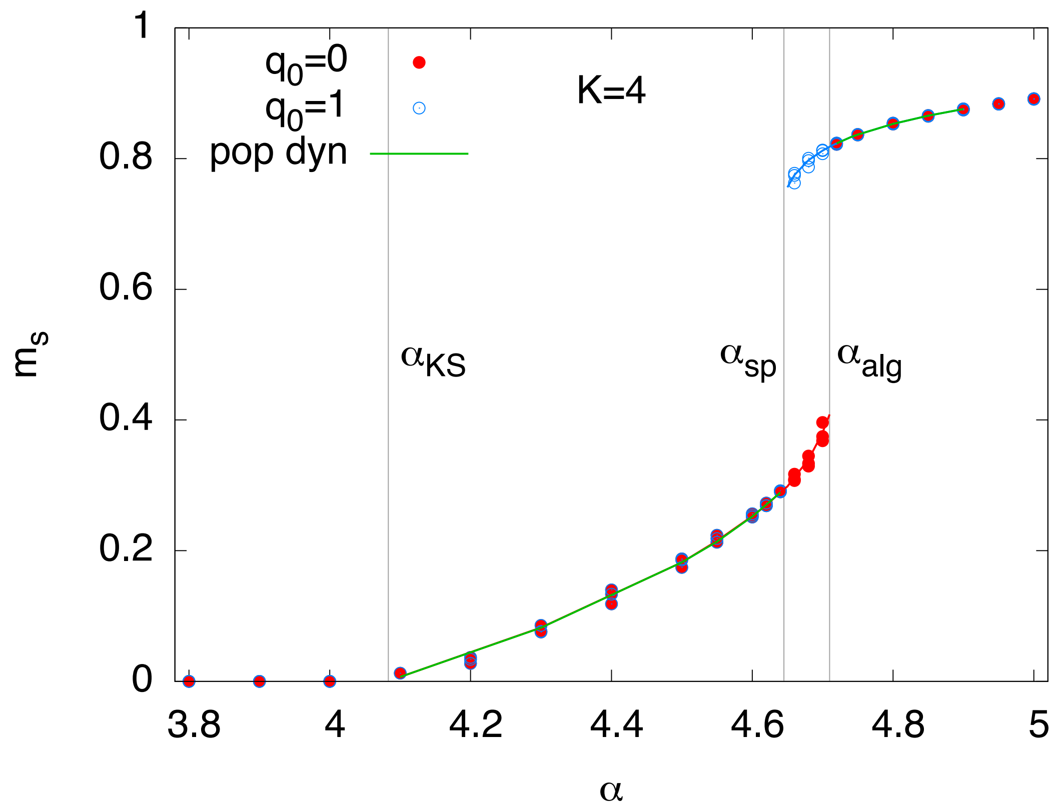


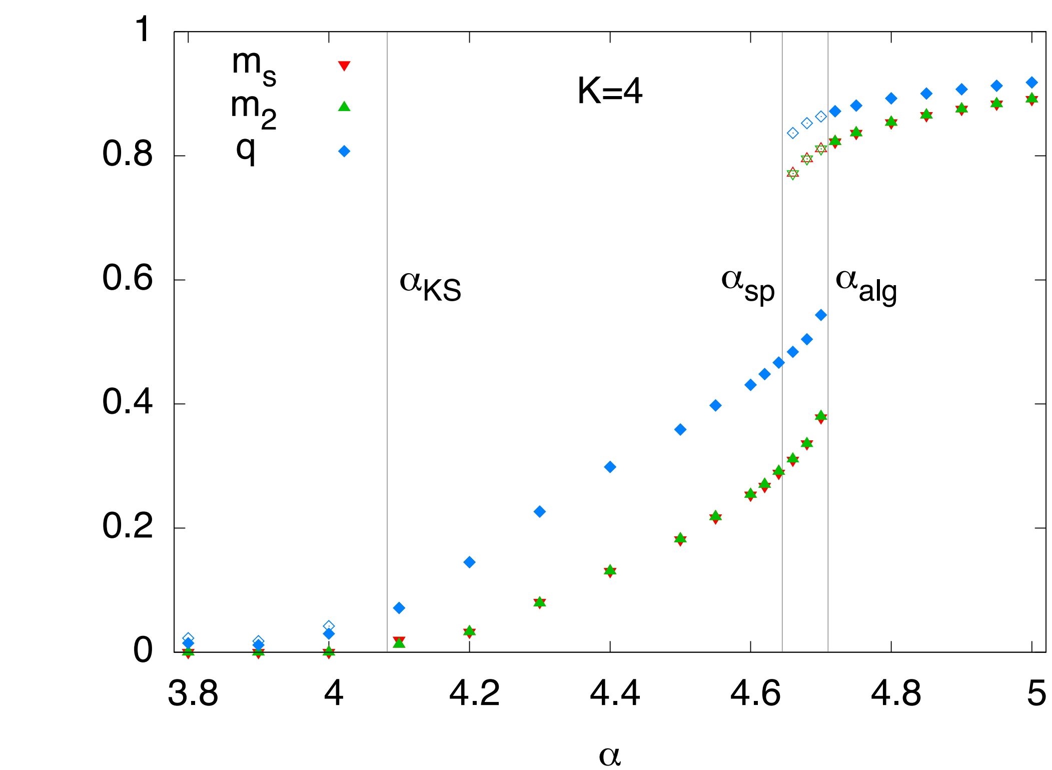
Planted random hypergraph bicoloring



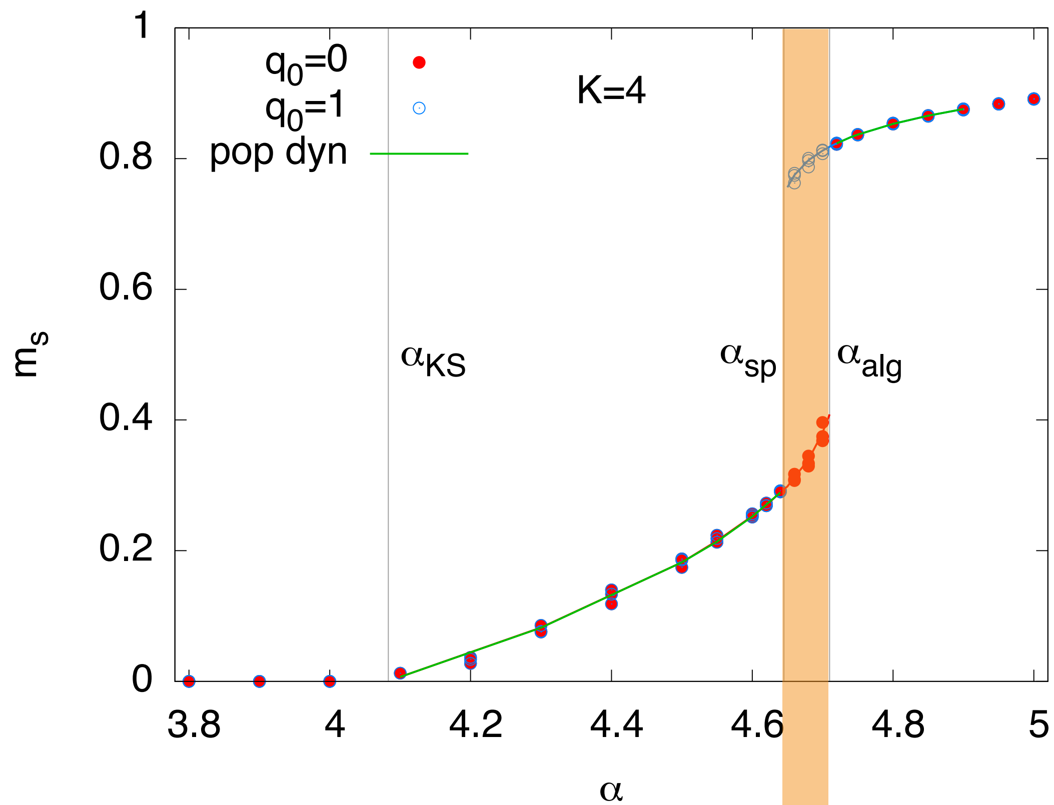
3 samples of size N=10^6

Planted random hypergraph bicoloring

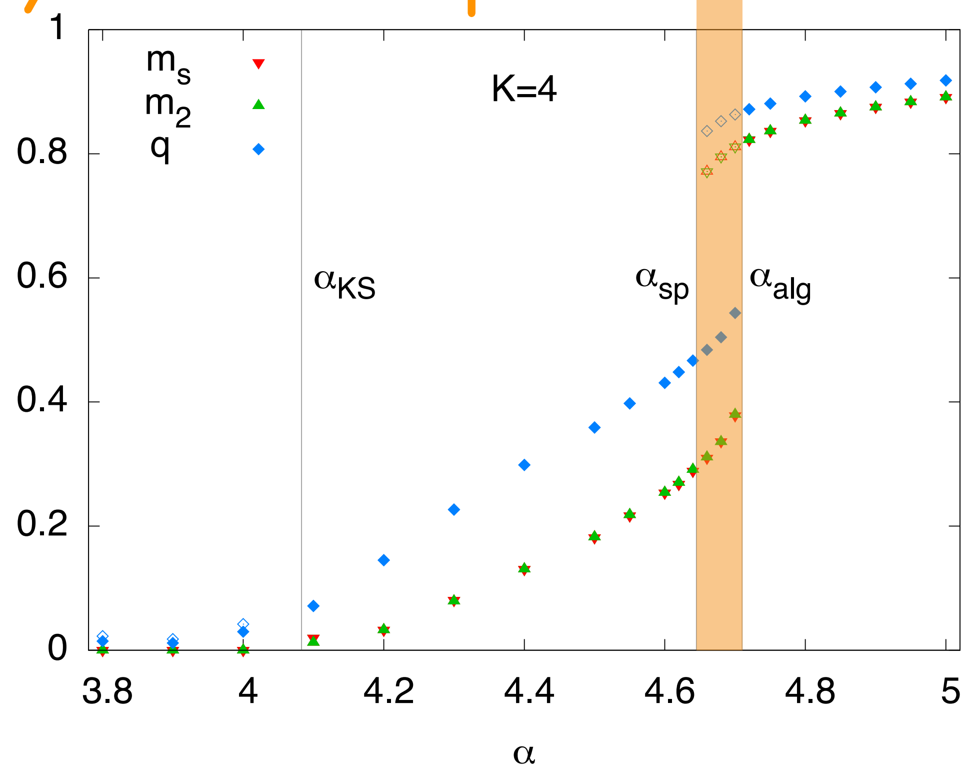




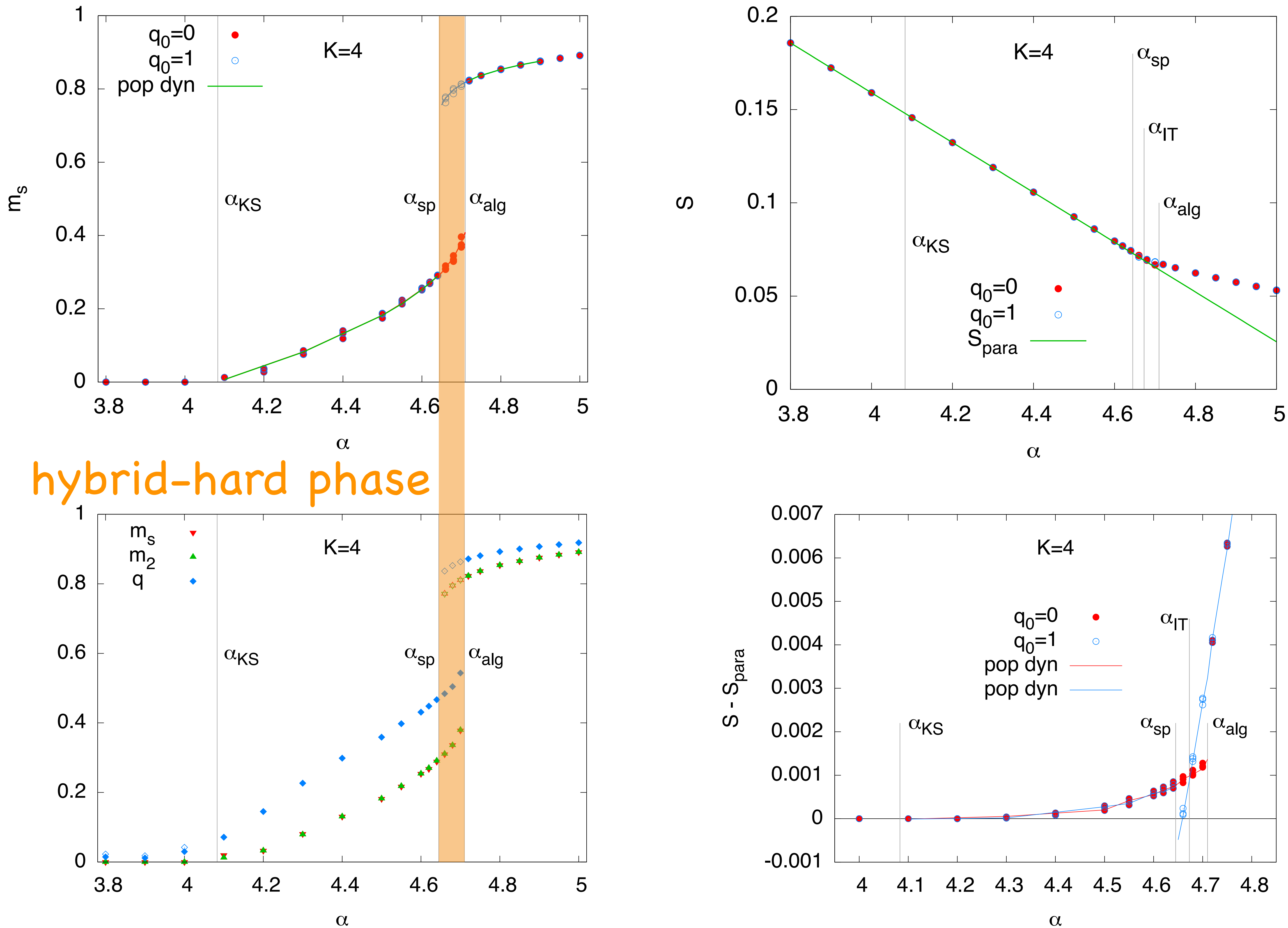
Planted random hypergraph bicoloring



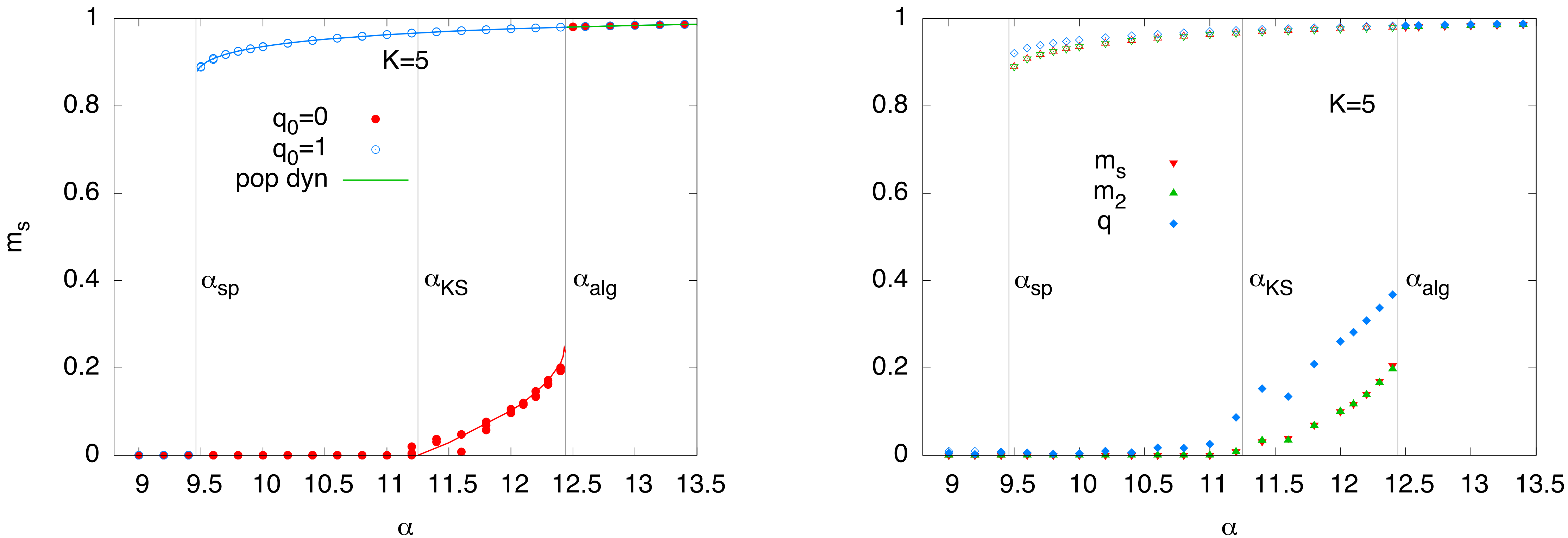
hybrid-hard phase



Planted random hypergraph bicoloring

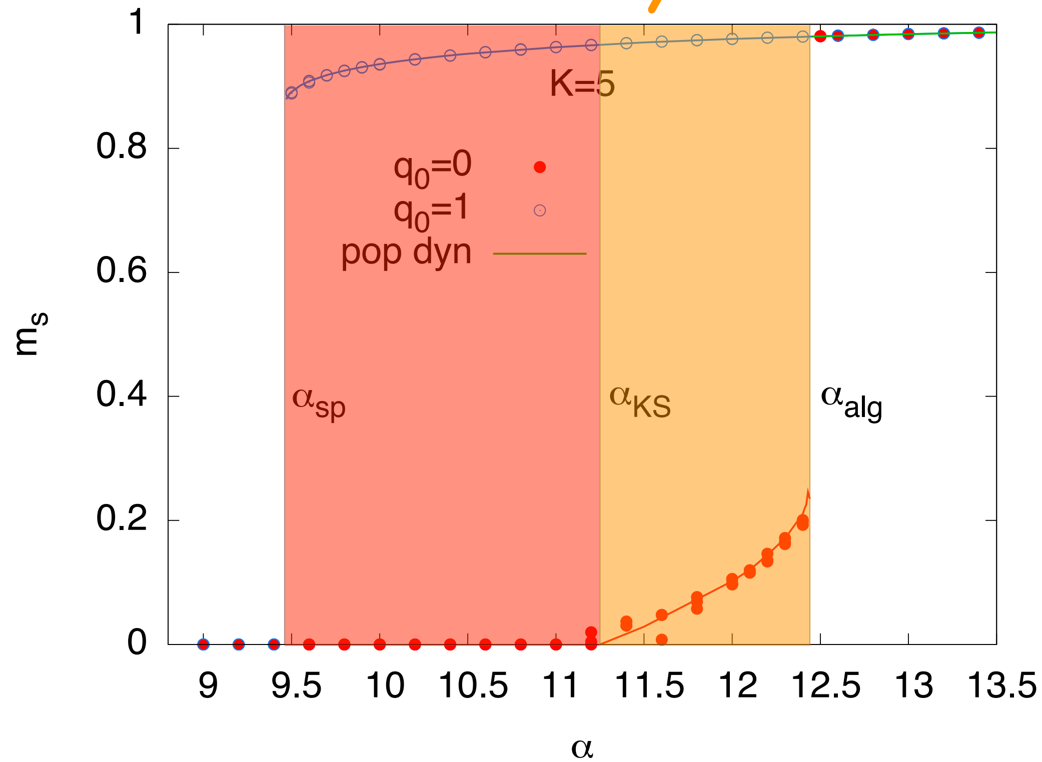


Planted random hypergraph bicoloring

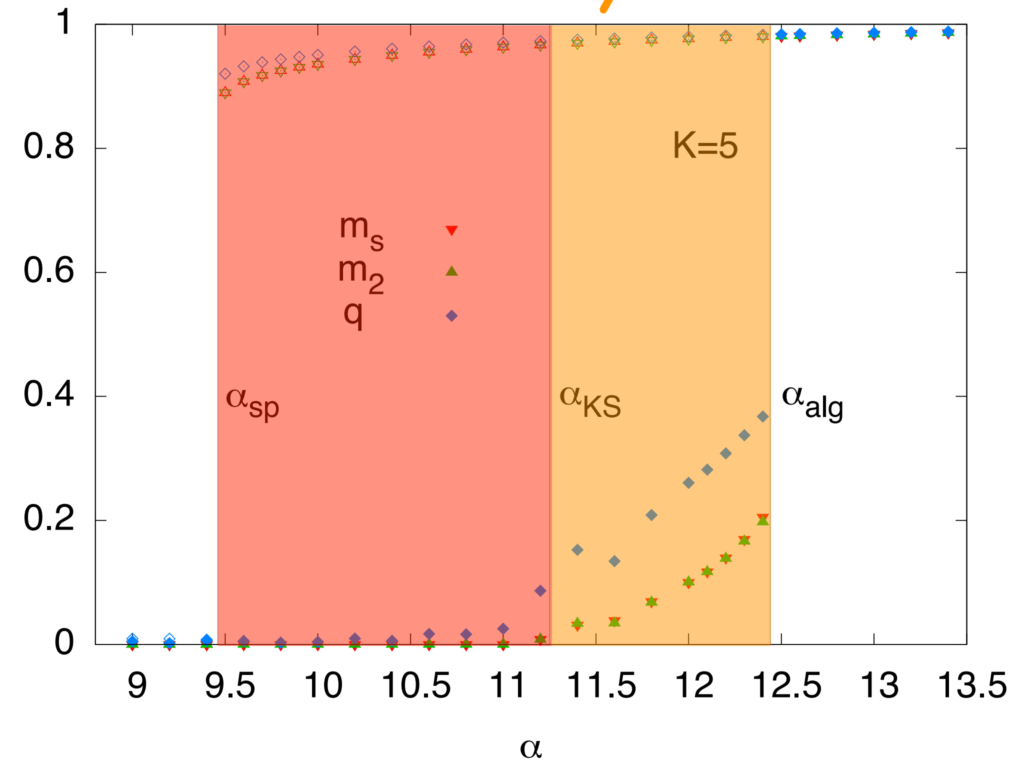


Planted random hypergraph bicoloring

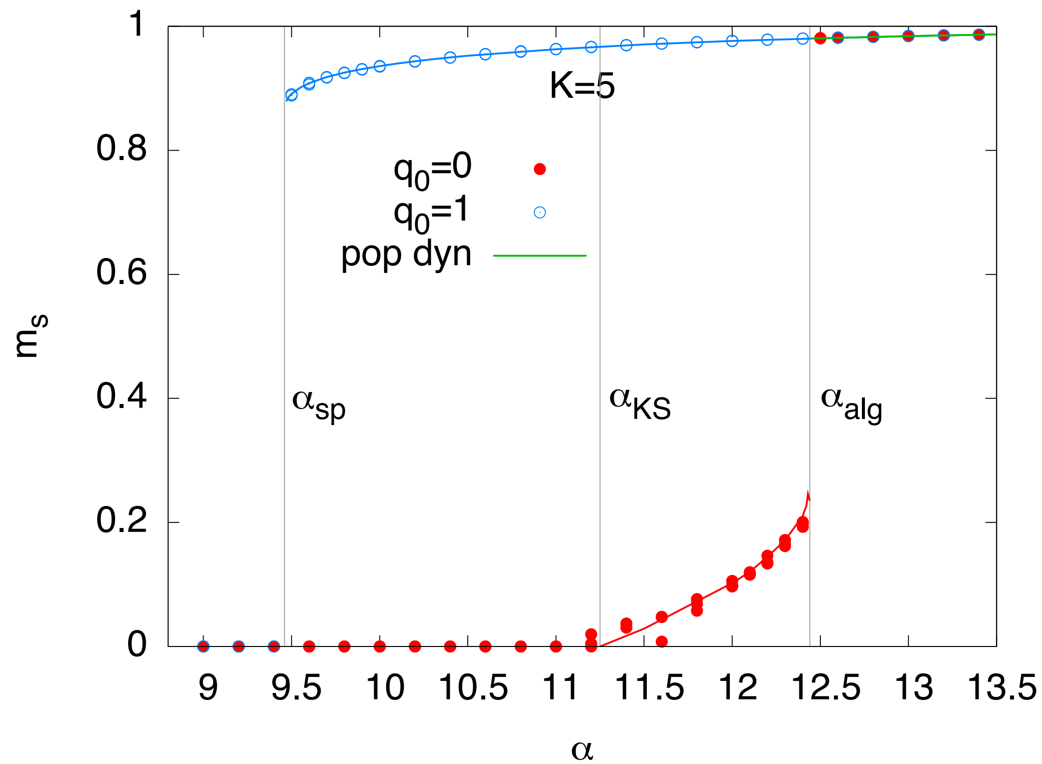
hard hybrid-hard

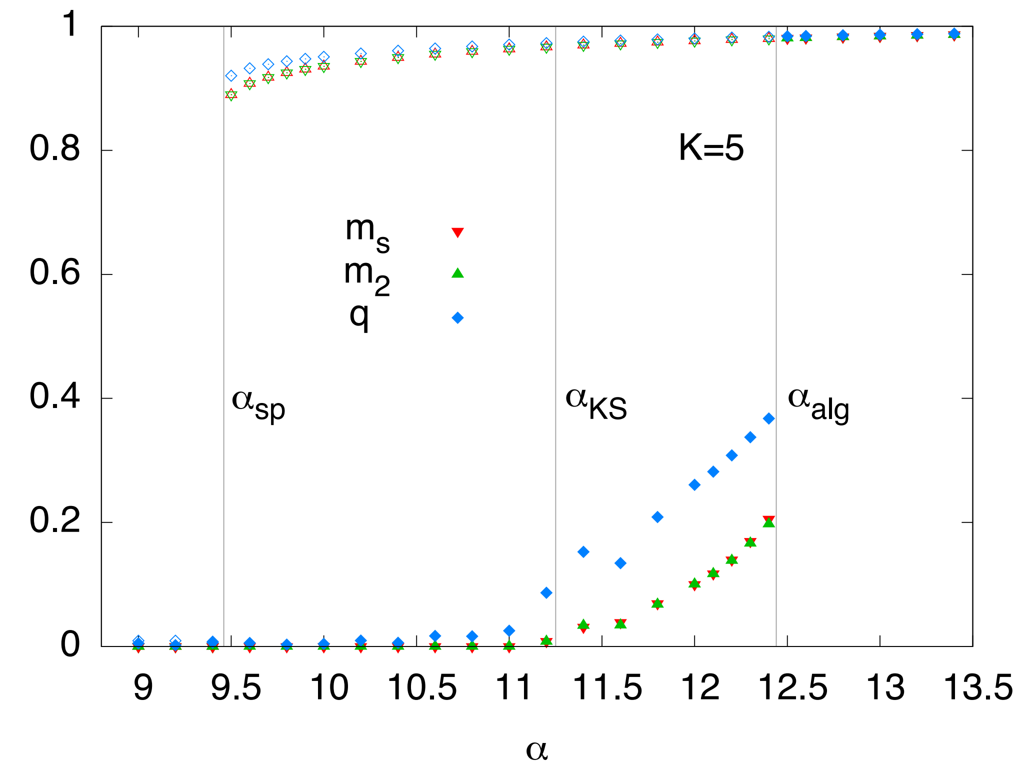


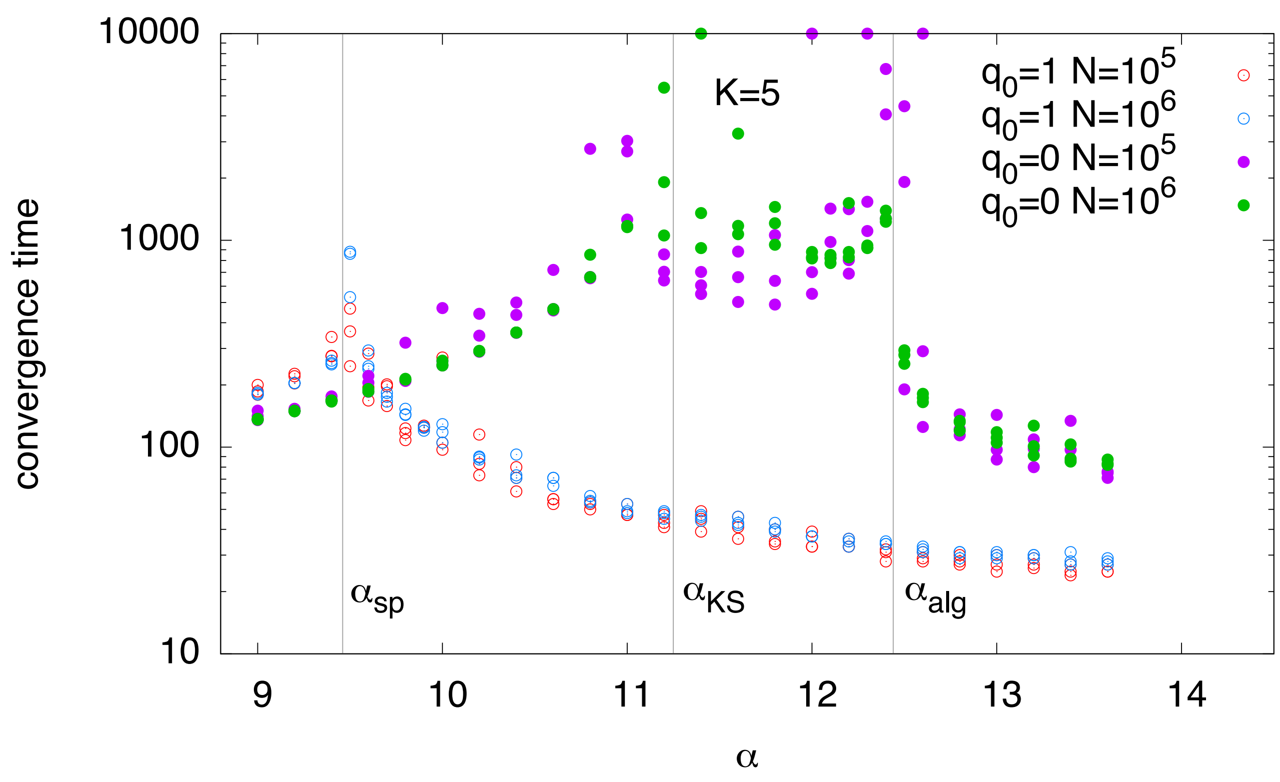
hard hybrid-hard



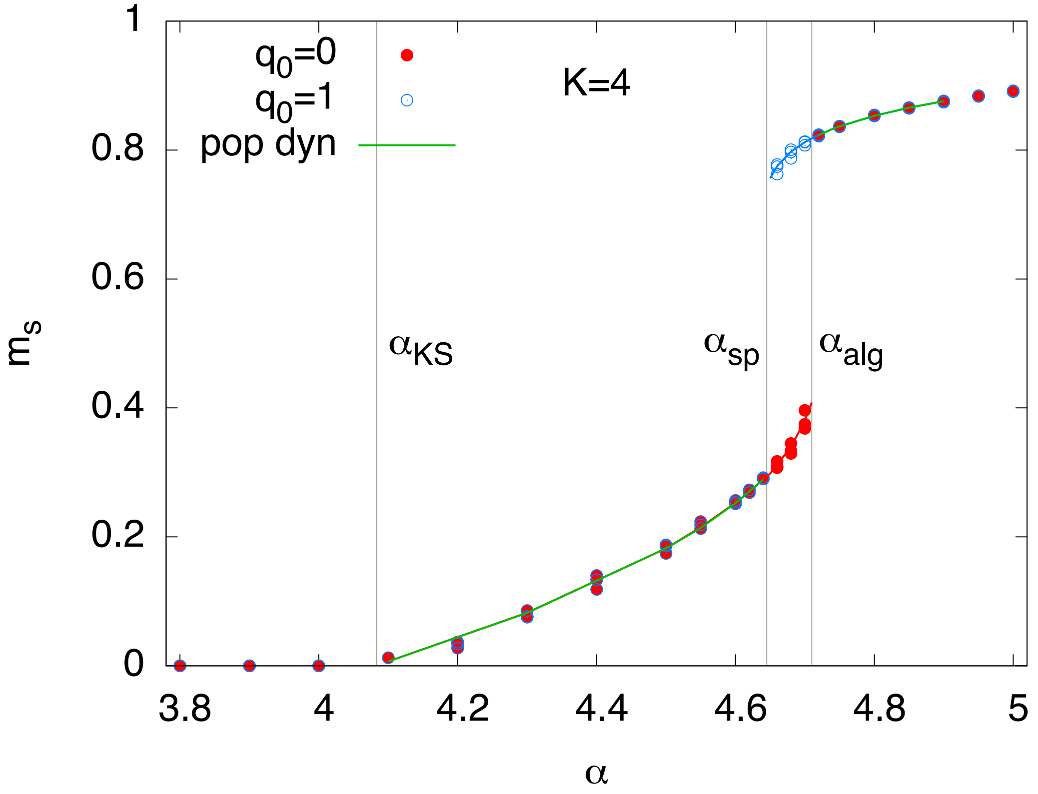
Planted random hypergraph bicoloring





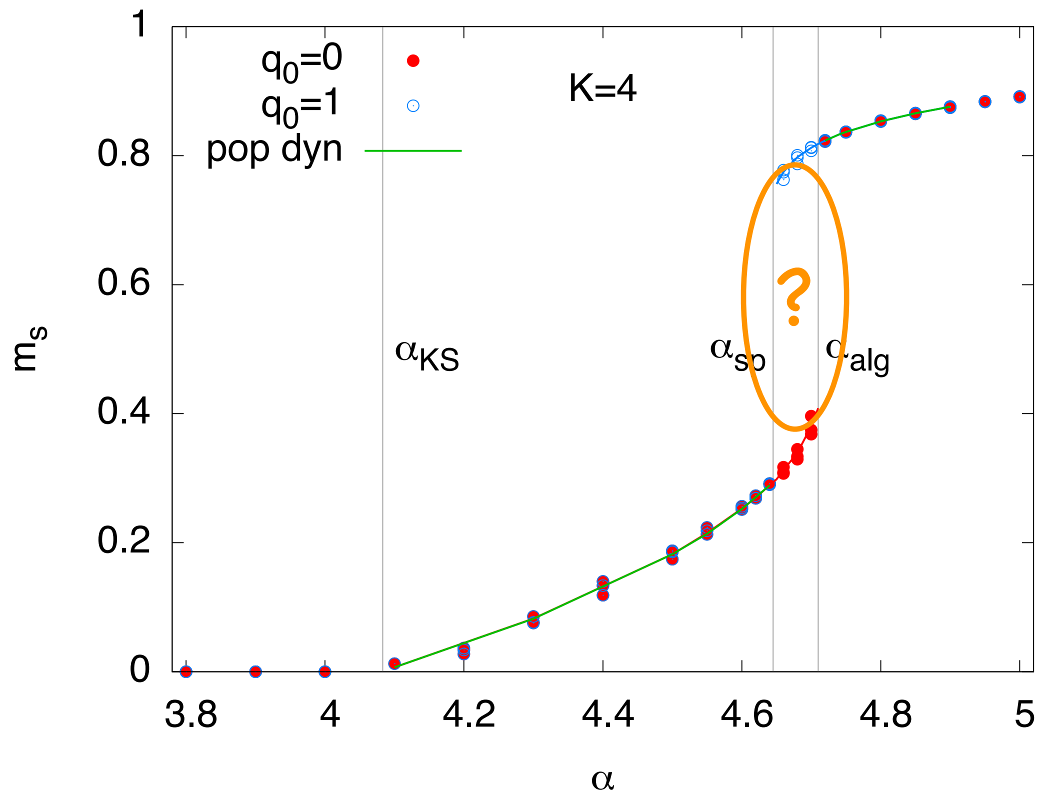


Finding the unstable fixed point via BP



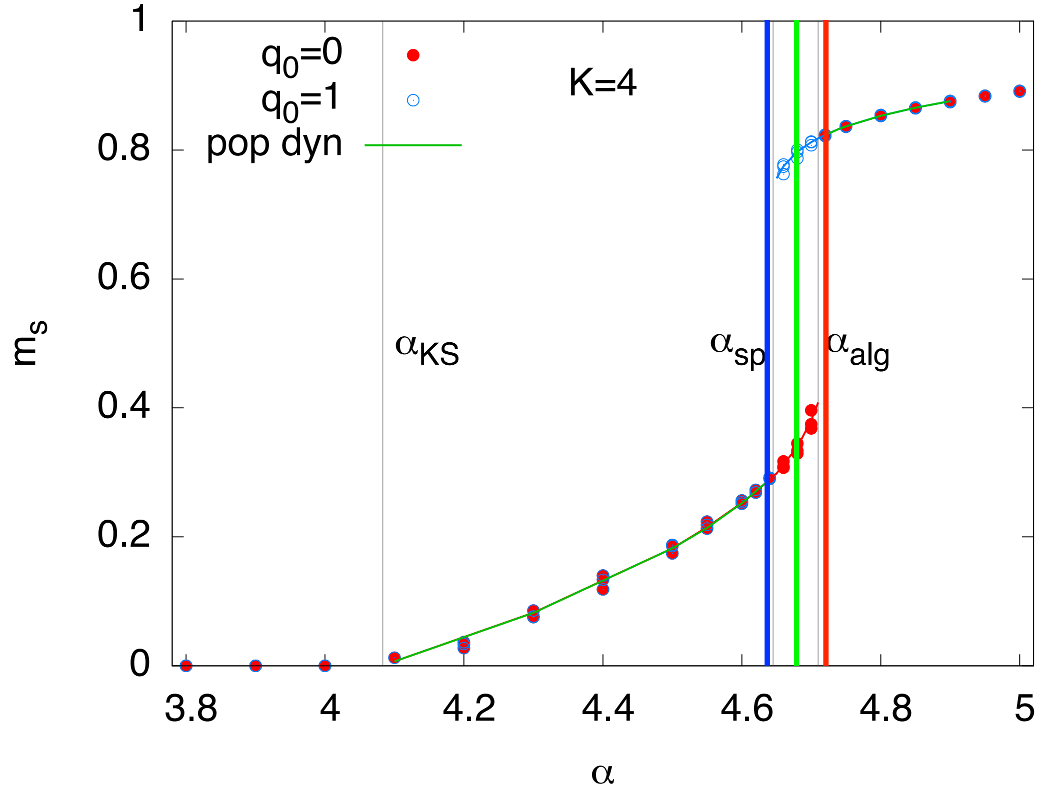
Hard (impossible?) to do via PopDyn

Finding the unstable fixed point via BP



Hard (impossible?) to do via PopDyn

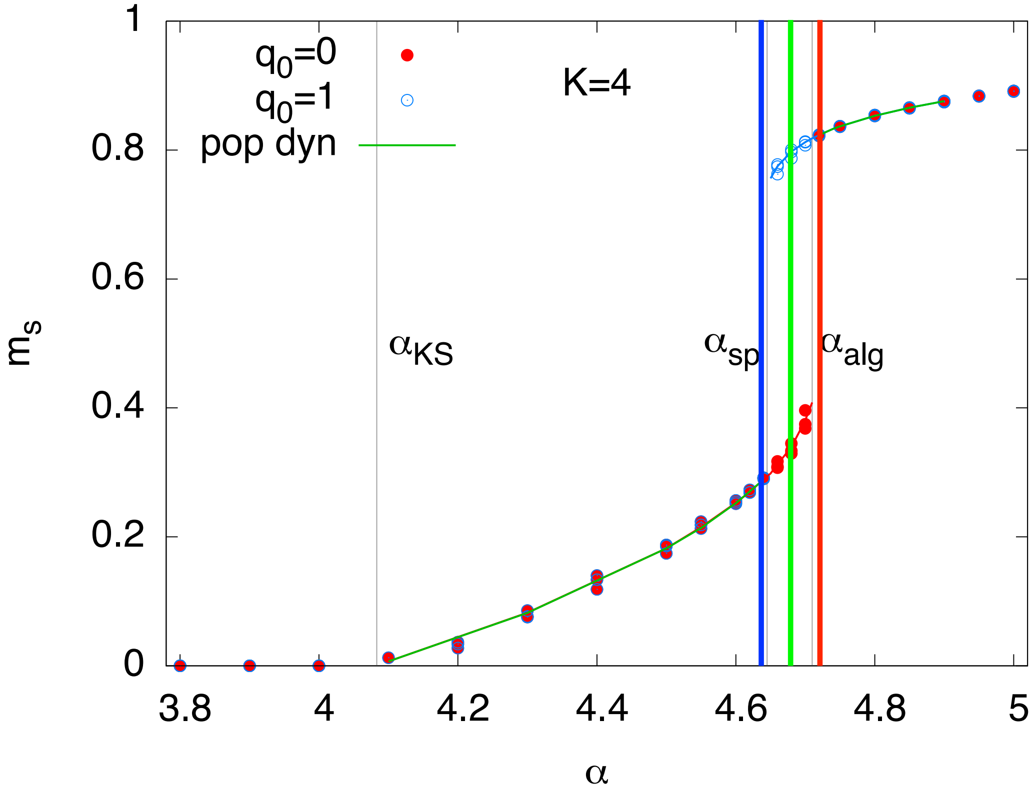
Finding the unstable fixed point via BP

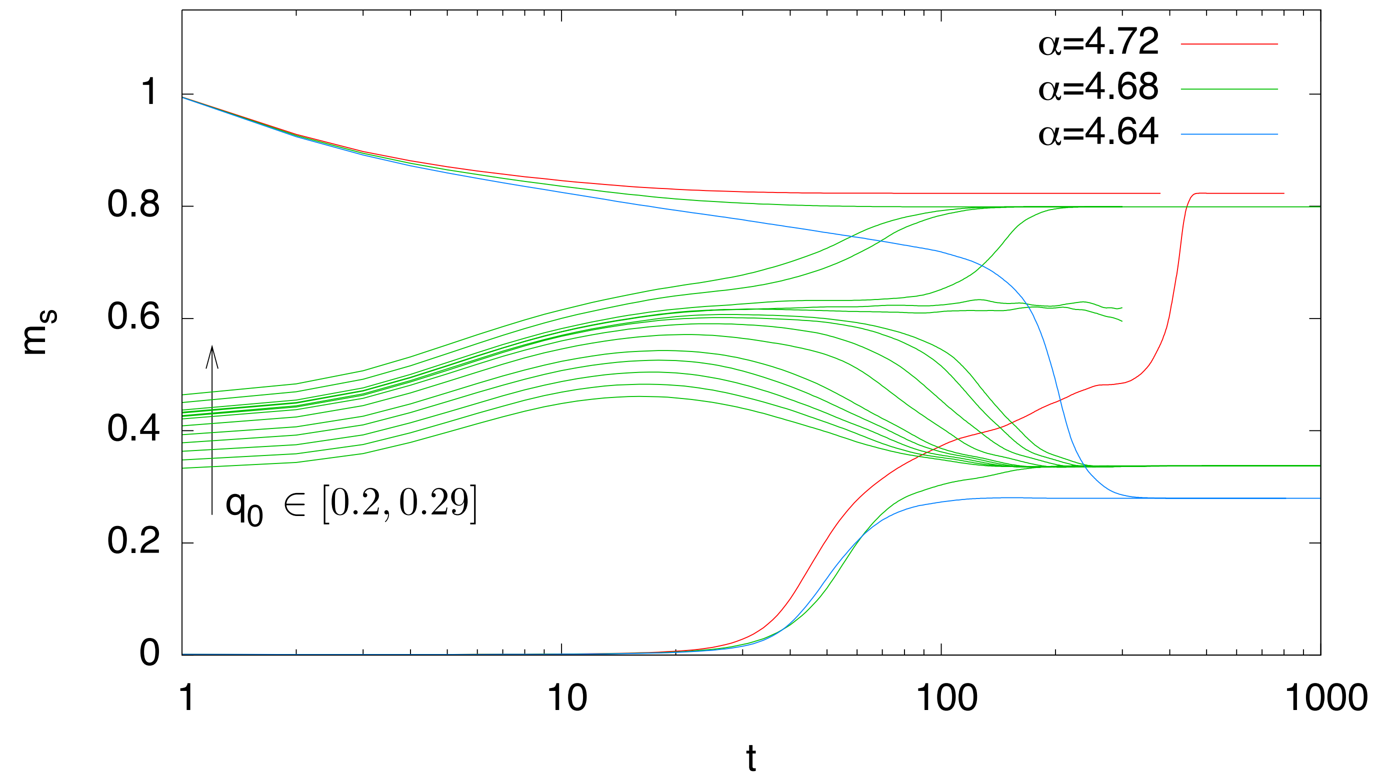


Hard (impossible?) to do via PopDyn

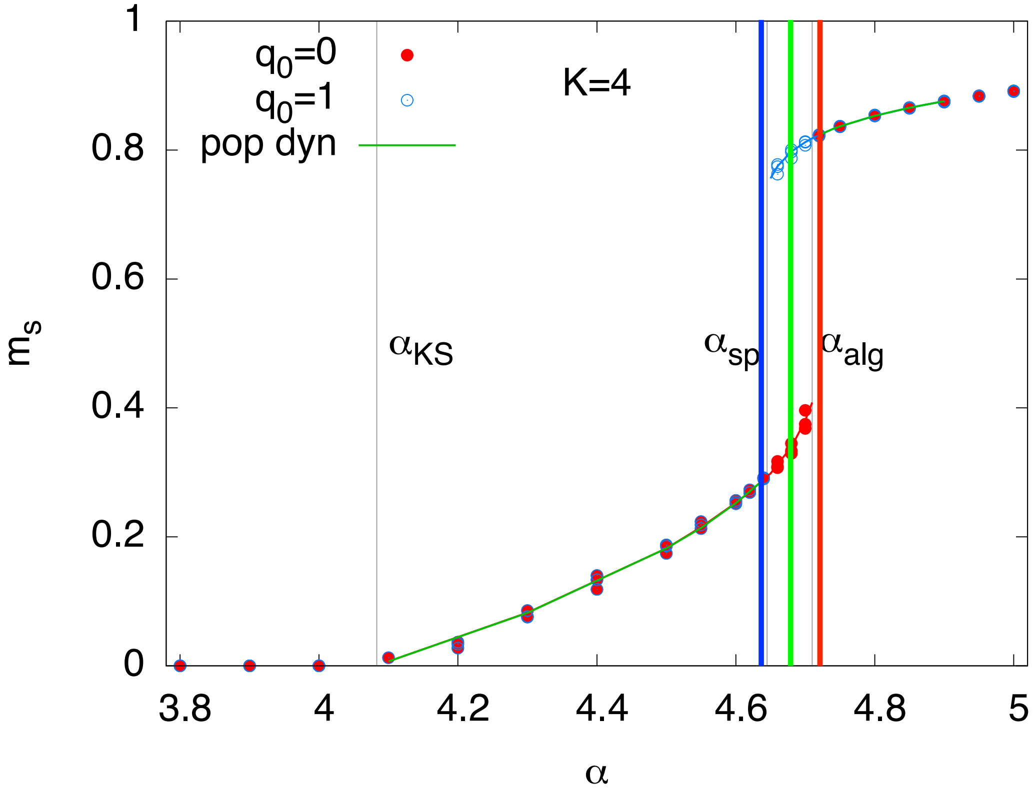
Finding the unstable fixed point via BP

Hard (impossible?) to do via PopDyn

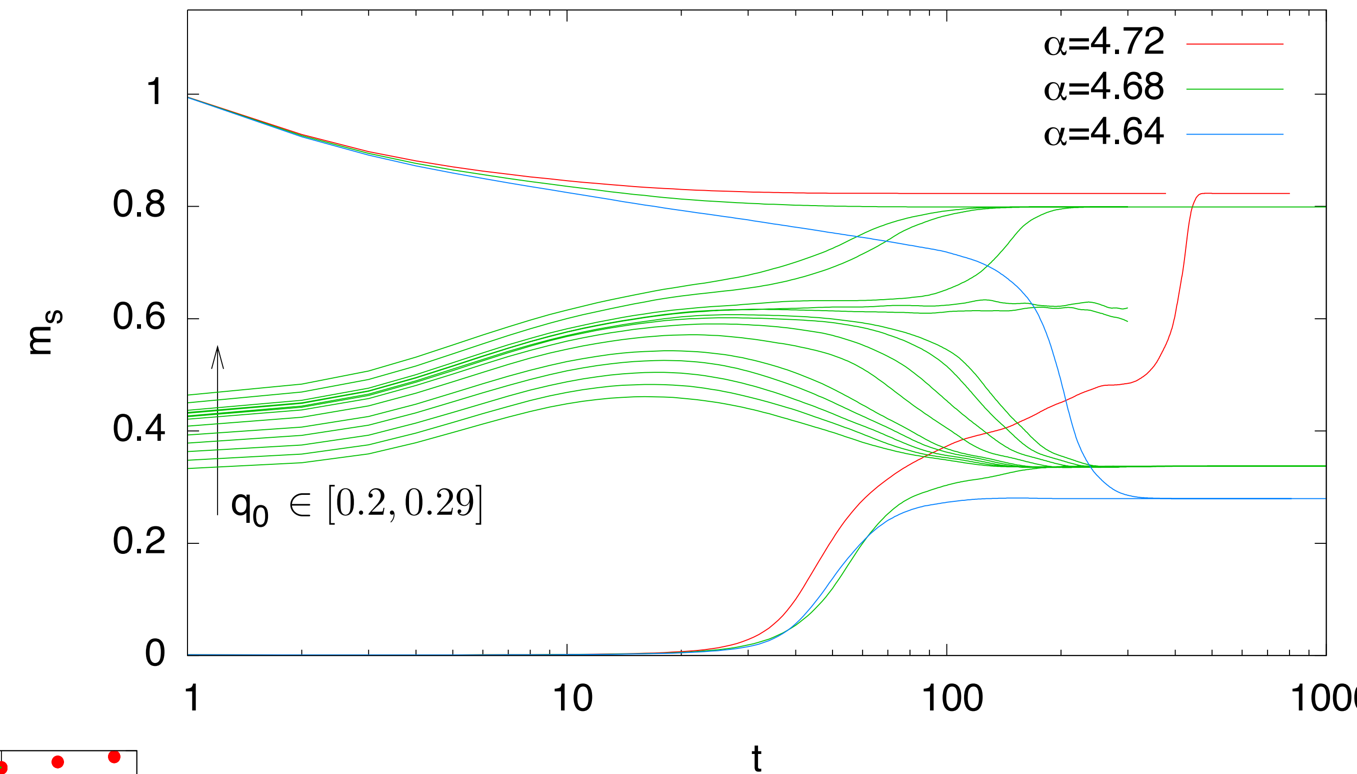


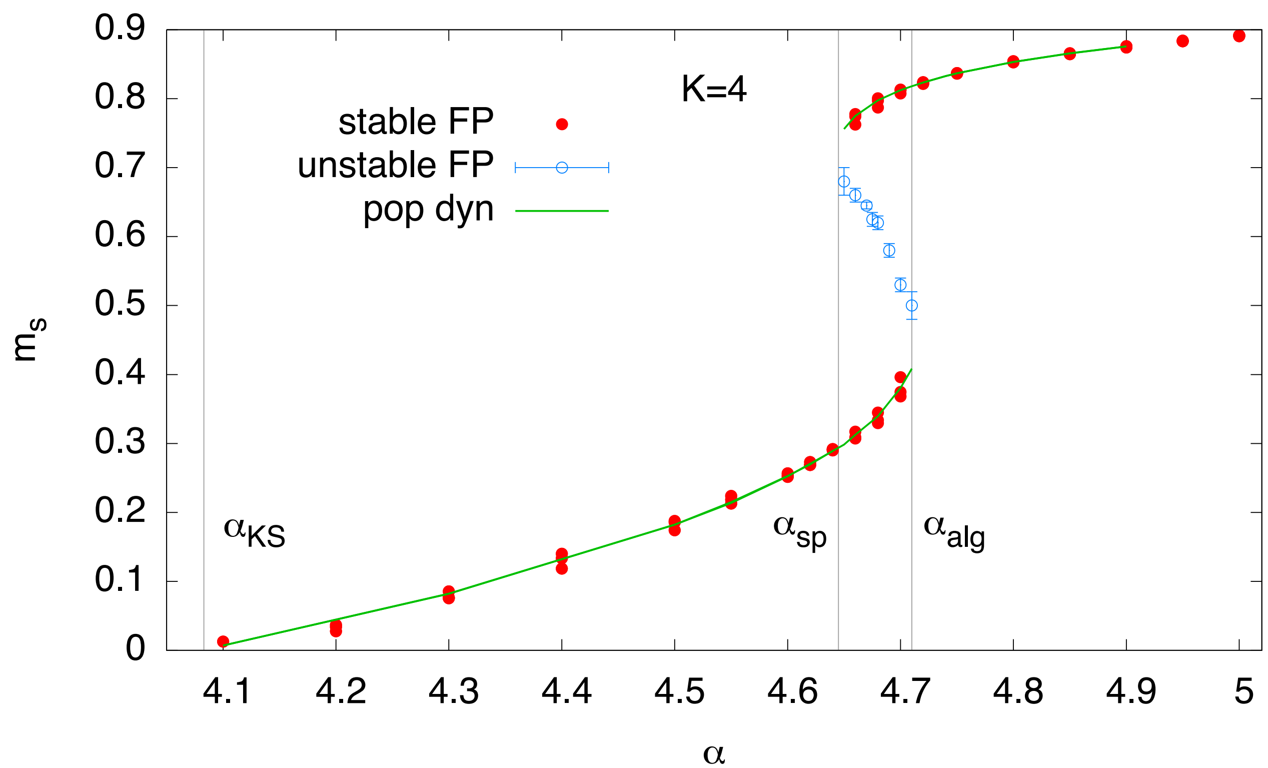


Finding the unstable fixed point via BP



Hard (impossible?) to do via PopDyn

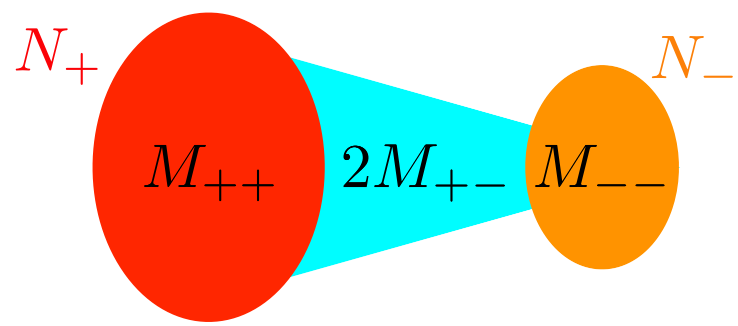




Asymmetric SBM

- 2 hidden communities ($\sigma_i = \pm 1$) of sizes $N_\pm = N\frac{1 \pm \overline{m}}{2}$
- Random graph with M_{++}, M_{--} edges in each group and $2M_{+-}$ between the groups

$$M_{\sigma\tau} = \frac{N_\sigma N_\tau}{2N} d \left(1 + \theta \frac{(\sigma - \overline{m})(\tau - \overline{m})}{1 - \overline{m}^2} \right)$$



- Each group has the same mean degree
- θ is the SNR:

  $\theta = 0$ single random graph $\mathcal{G}(N, M = dN/2)$

  $\theta = 1$ two disconnected random graphs

  $\mathcal{G}(N_+, N_+ d/2)$, $\mathcal{G}(N_-, N_- d/2)$

Asymmetric SBM

- Posterior distribution -> Ising model on a sparse graph

$$\mathbb{P}[\underline{\sigma}] \propto \exp\left[H \sum_i \sigma_i + J \sum_{(ij)\in E} \sigma_i \sigma_j - K \sum_i d_i \sigma_i - \frac{d\theta N}{2(1-\overline{m}^2)} \left(\frac{1}{N} \sum_i \sigma_i - \overline{m} \right)^2 \right]$$

prior   likelihood edges   likelihood non-edges

$$H = \operatorname{atanh}(\overline{m})$$

$$J = \frac{1}{4} \log \left[\frac{(1+\theta)^2 - \overline{m}^2 (1-\theta)^2}{(1-\theta)^2 (1-\overline{m}^2)} \right]$$

$$K = \frac{1}{4} \log \left[\frac{1 - \overline{m}^2 + \theta(1+\overline{m})^2}{1 - \overline{m}^2 + \theta(1-\overline{m})^2} \right]$$

- BP equations would involve $O(N^2)$ messages due to the green global constraint, but they can be simplified to take into account only the $O(N)$ messages running on the sparse graph

Asymmetric SBM

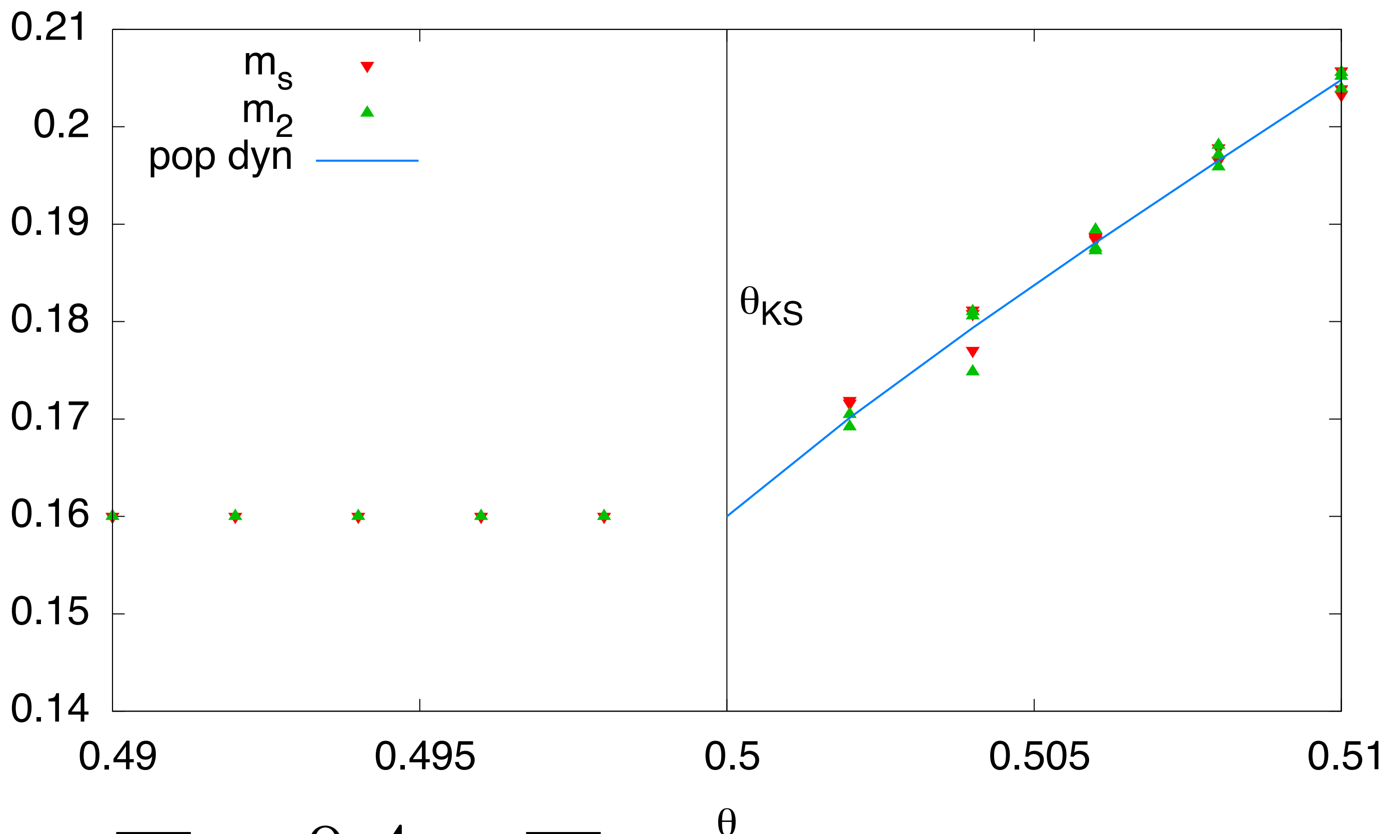
- Simplified BP equations

$$u_{i\to j}^{(t+1)} = \gamma u_{i\to j}^{(t)} + (1-\gamma)\text{atanh}\left[\tanh(J)\tanh\left(\widetilde{H}_i^{(t)} + \sum_{k\in\partial i\setminus j} u_{k\to i}^{(t)}\right)\right]$$

$$\widetilde{H}_i^{(t)} = H - d_i K + \frac{d\theta}{1-\overline{m}^2}\left(\overline{m} - \frac{1}{N}\sum_{\ell} m_\ell^{(t)}\right)$$

- Local magnetizations (marginals) $m_i^{(t)} = \tanh\left(\widetilde{H}_i^{(t)} + \sum_{k\in\partial i} u_{k\to i}^{(t)}\right)$
- Convergence criterion $\left|m_i^{(t+1)} - m_i^{(t)}\right| < 10^{-8} \quad \forall i$
- BP initialized close to a configuration $\underline{\tau}^0$ with
  $\mathbb{P}[\tau_i^0 = s] = q_0\,\delta_{s,\sigma_i^*} + (1-q_0)\dfrac{1+\overline{m}\,s}{2}$
- Uninformative fixed point has all messages $u_{i\to j}^{\star} = K$
  It becomes locally unstable at $\theta_{\text{KS}} = 1/\sqrt{d}$

Asymmetric SBM (d=4)

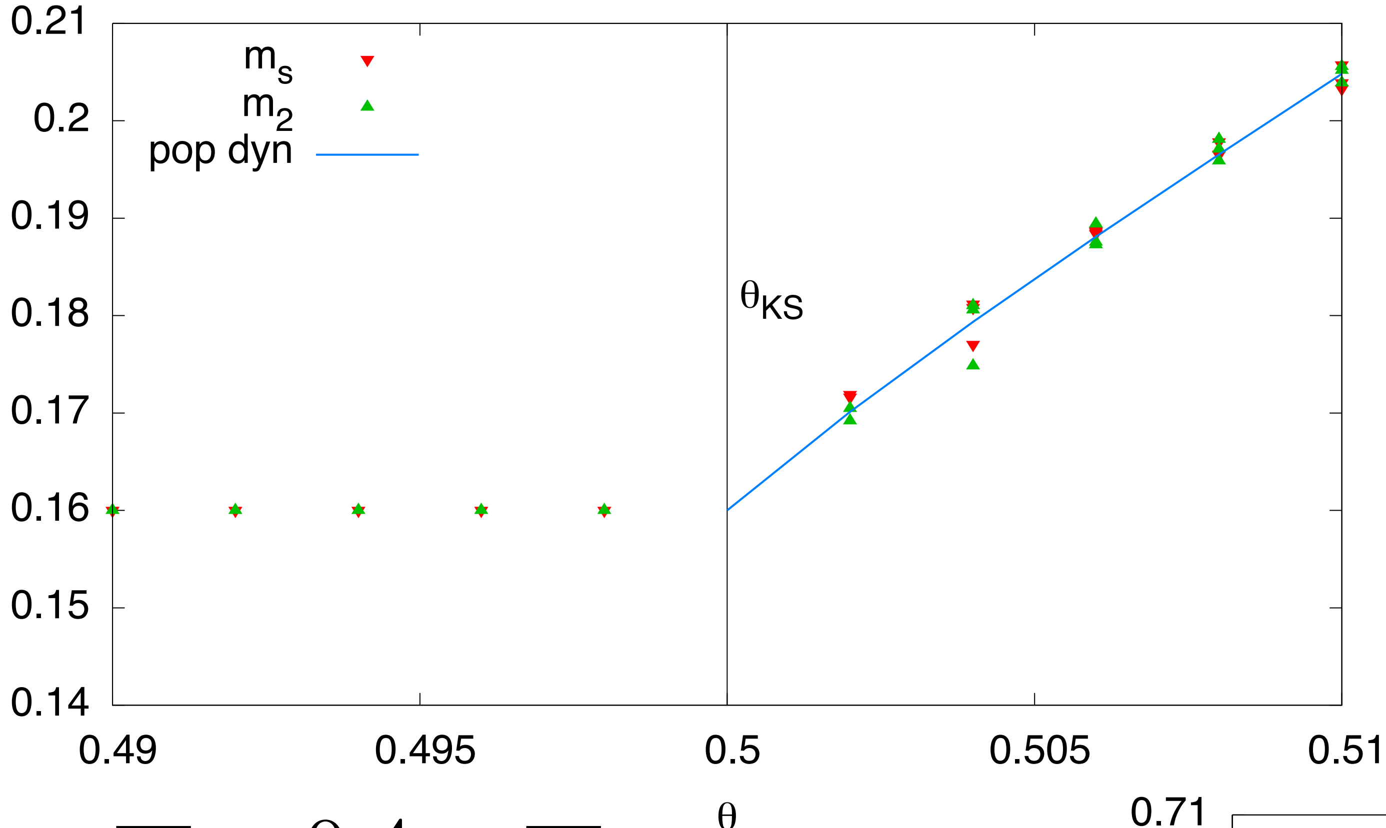


$\overline{m} = 0.4 < \overline{m}_c$

3 samples of size N=10^7

critical asymmetry
$\overline{m}_c = 1/\sqrt{3} \simeq 0.577$

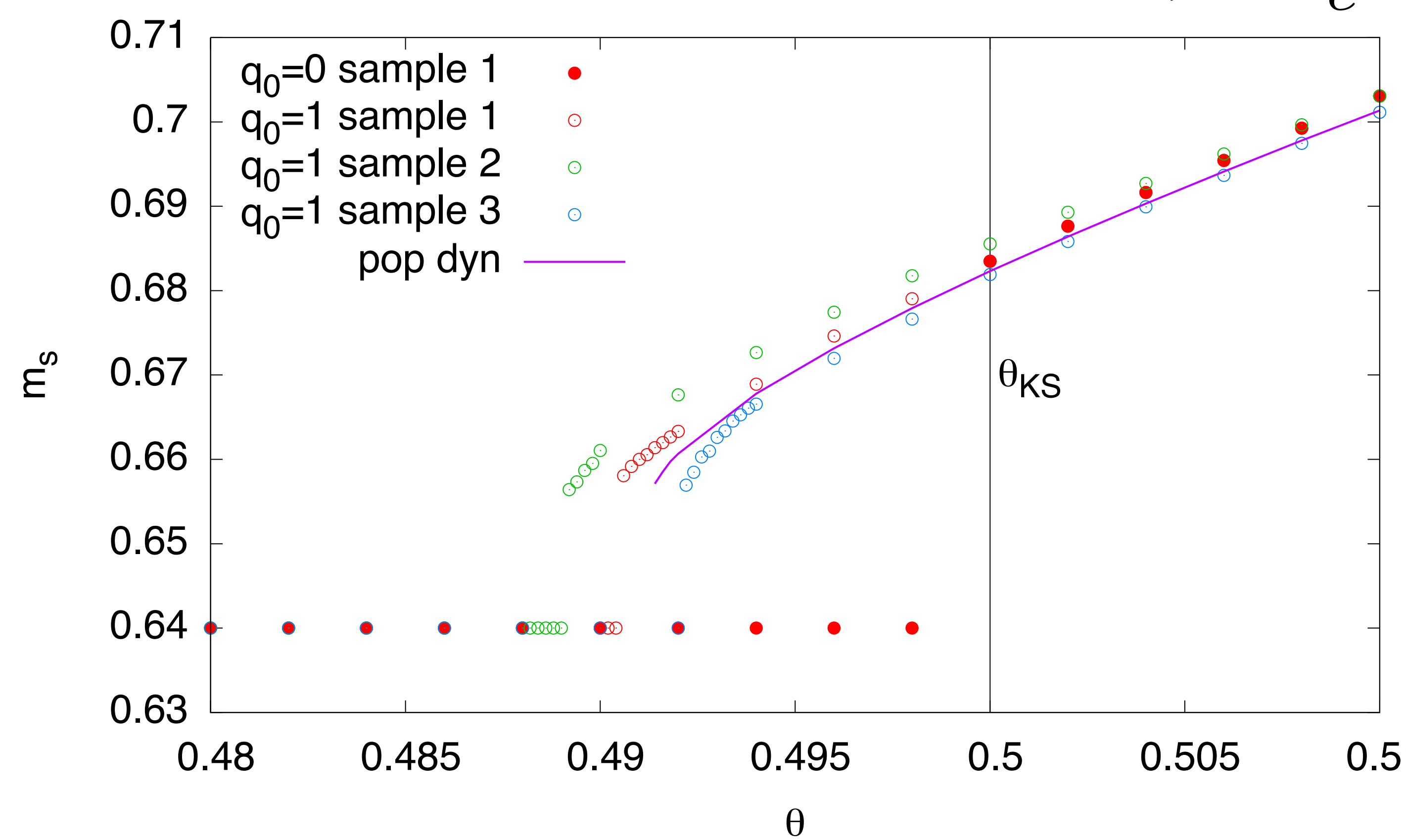
Asymmetric SBM (d=4)



3 samples of size N=10^7

$\overline{m} = 0.4 < \overline{m}_c$

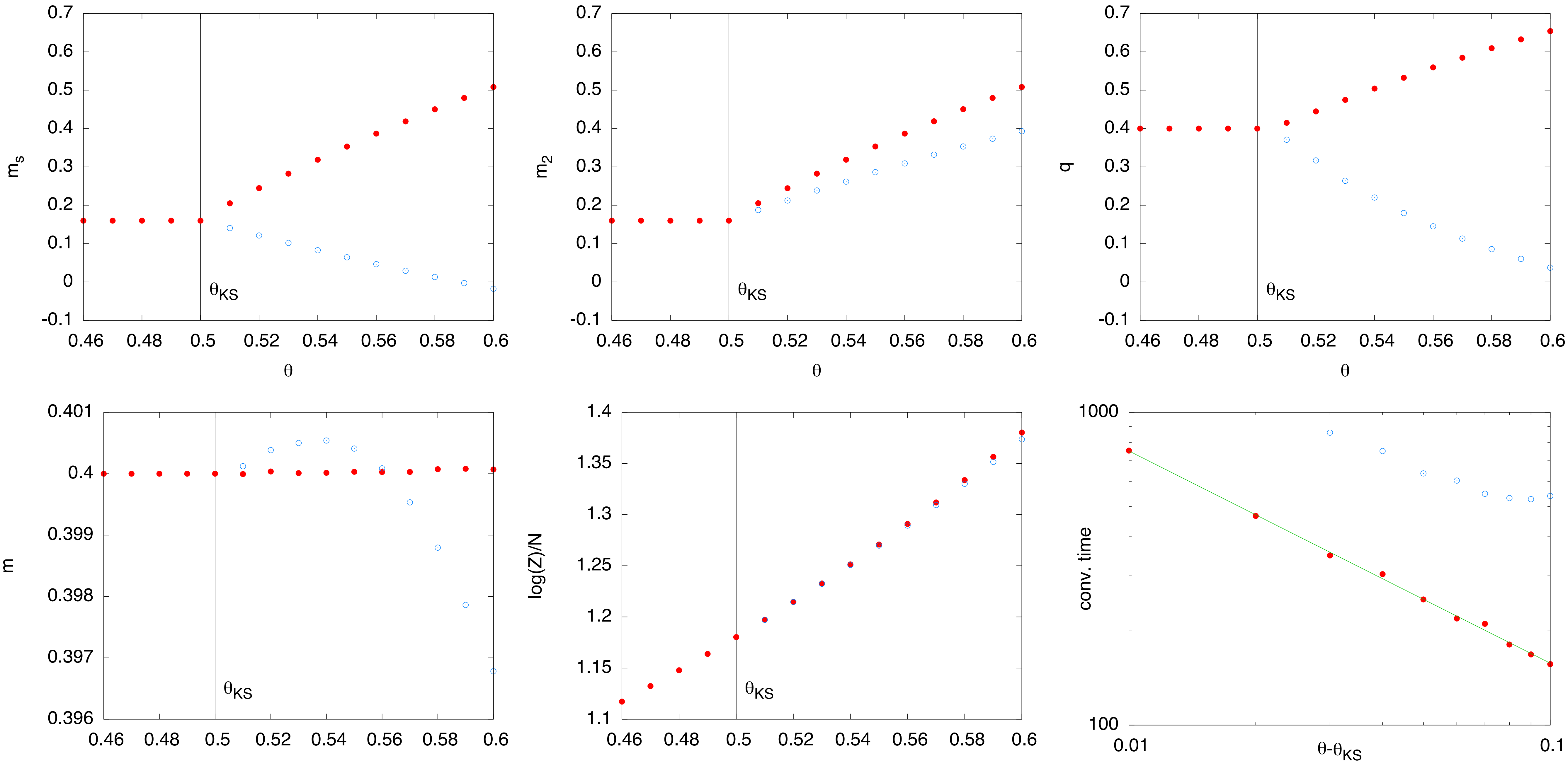
$\overline{m} = 0.8 > \overline{m}_c$



critical asymmetry
$\overline{m}_c = 1/\sqrt{3} \simeq 0.577$

Asymmetric SBM: more BP fixed points

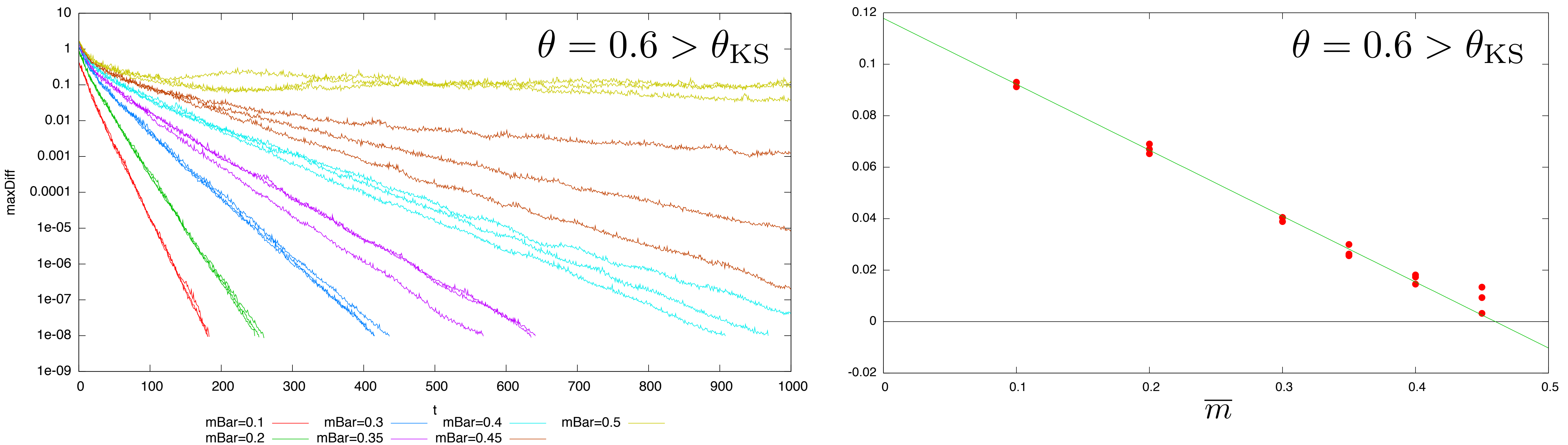
$$d = 4\ ,\ \overline{m} = 0.4\ ,\ N = 10^7$$



dominating fixed point   sub-dominating fixed point

Asymmetric SBM: RSB states?

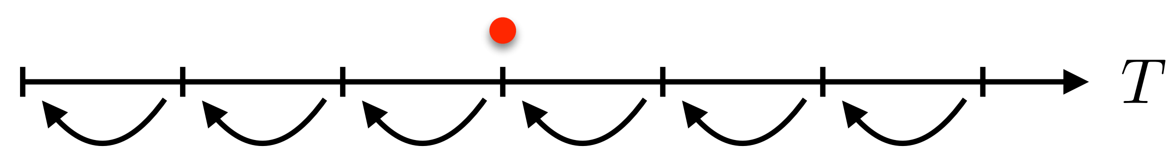
- The sub-dominating fixed point does not satisfy the Bayes/Nishimori condition
  -> may undergo a transition to a RSB phase
  -> BP does not converge anymore...



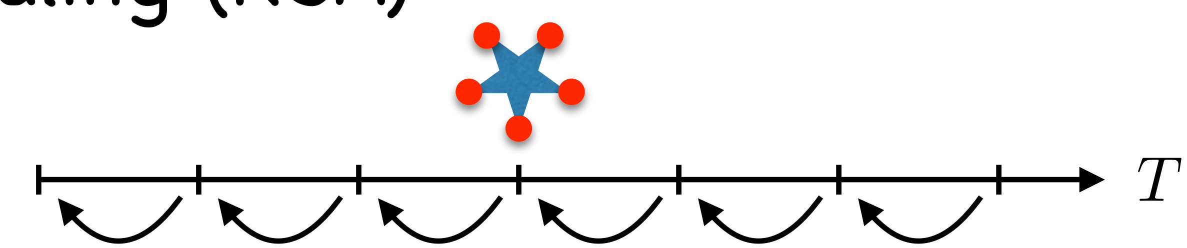
- From the algorithmic/dynamic point of view the asymmetric SBM seems more complex than expected

Monte Carlo based algorithms

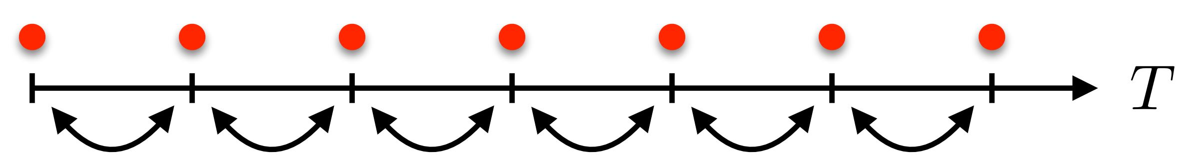
- Aim: minimize $E(\underline{\sigma})$ = number of violated constraints
- In planted models $E(\underline{\sigma}^*) = 0$
- Run MCMC sampling from $P_T(\underline{\sigma}) \propto \exp[-E(\underline{\sigma})/T]$
- Send $T \to 0$ and try to find the ground state
- Simulated Annealing (SA)



- Replicated Simulated Annealing (RSA)

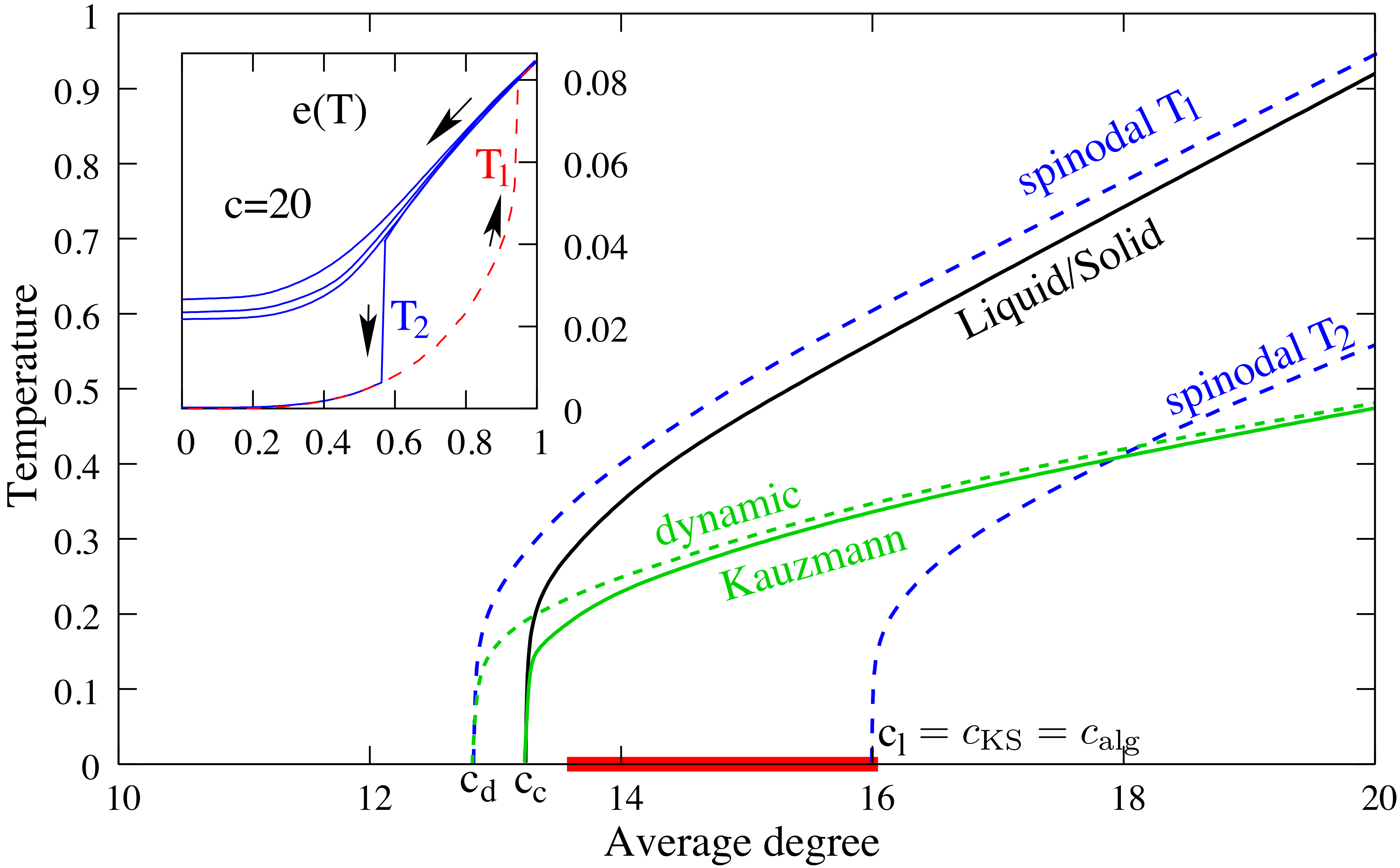


- Parallel Tempering (PT)



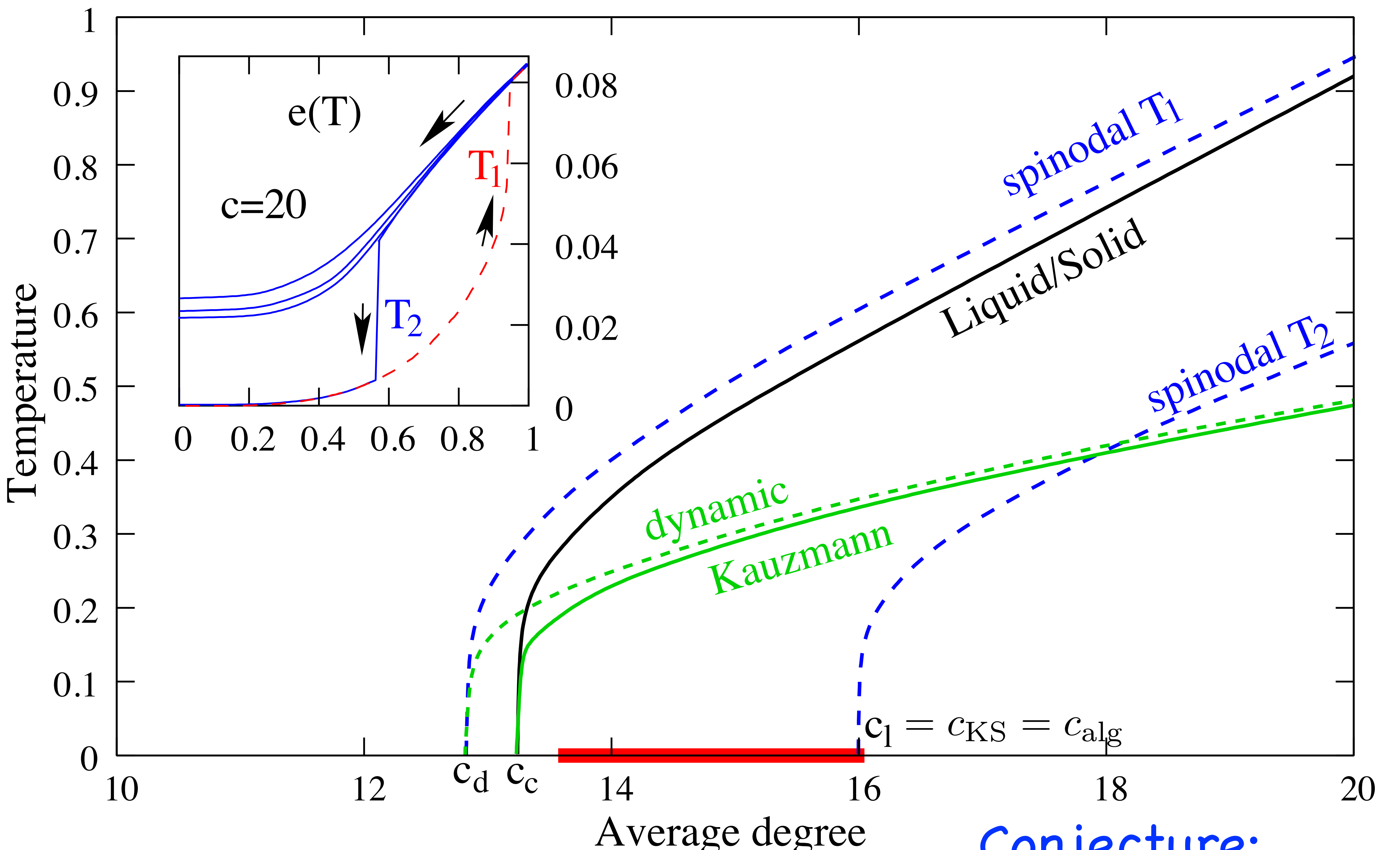
SA in planted models (q=5 coloring)

Krzakala & Zdeborova, PRL 2009



SA in planted models (q=5 coloring)

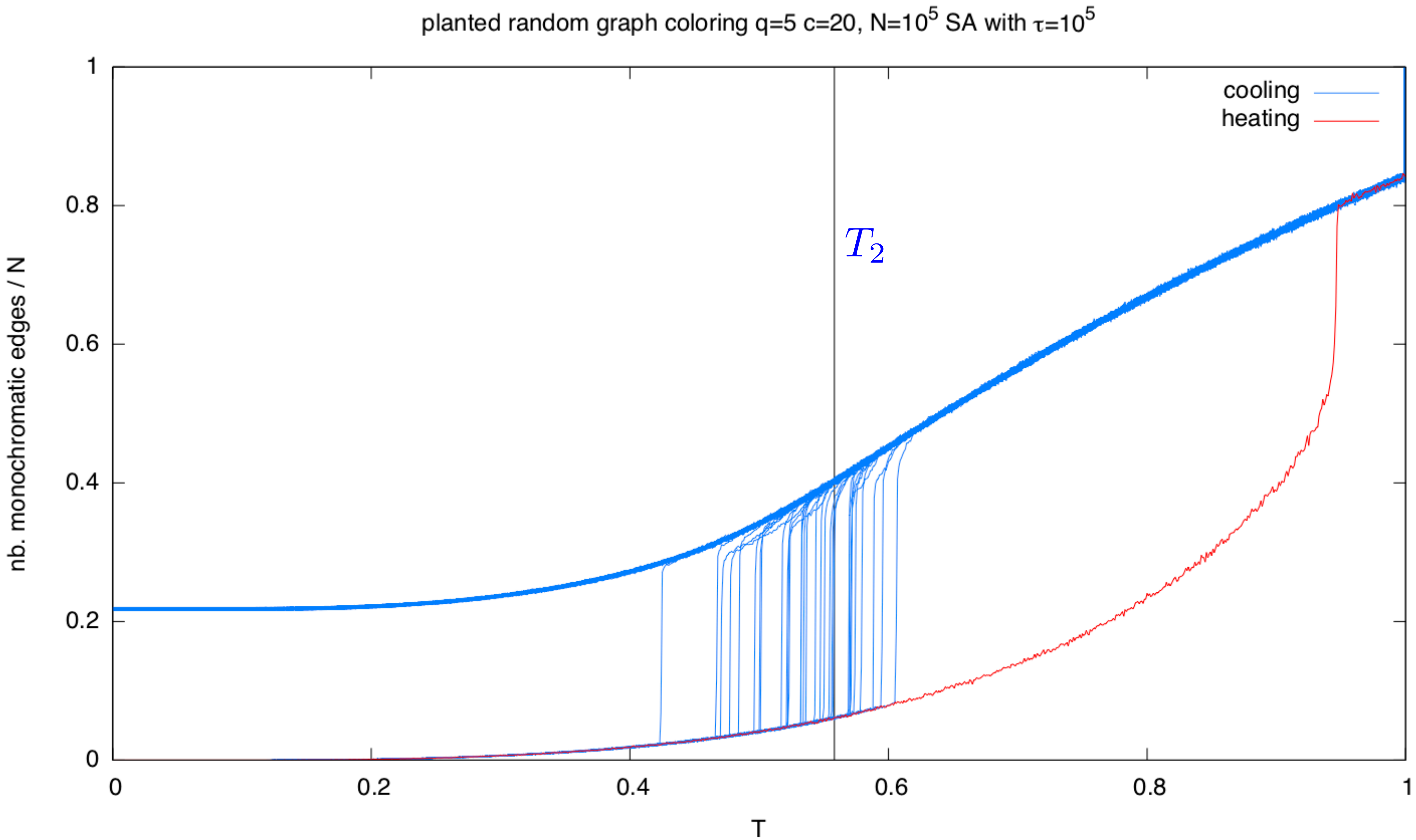
Krzakala & Zdeborova, PRL 2009



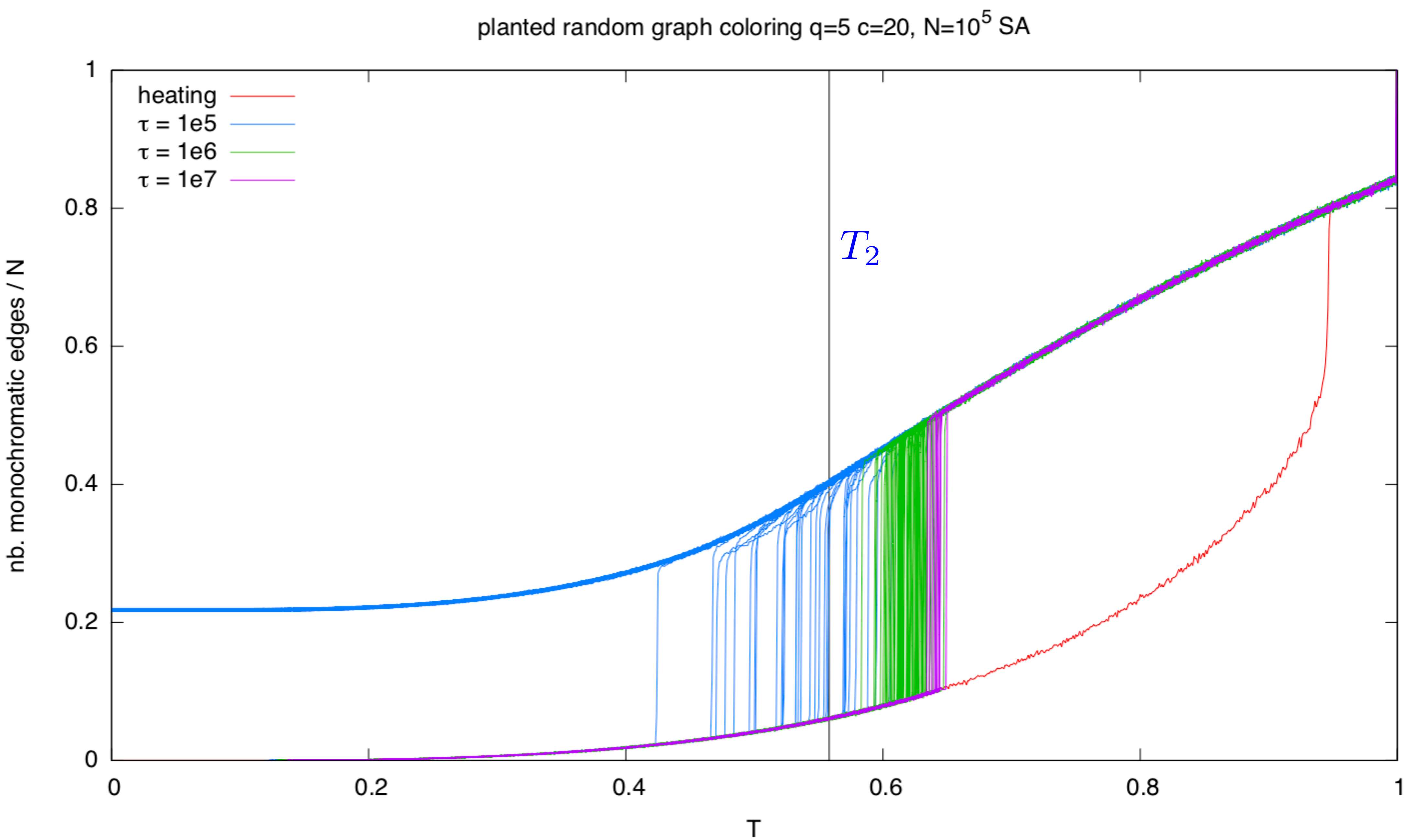
Conjecture:

SA works for $c > c_{KS}$

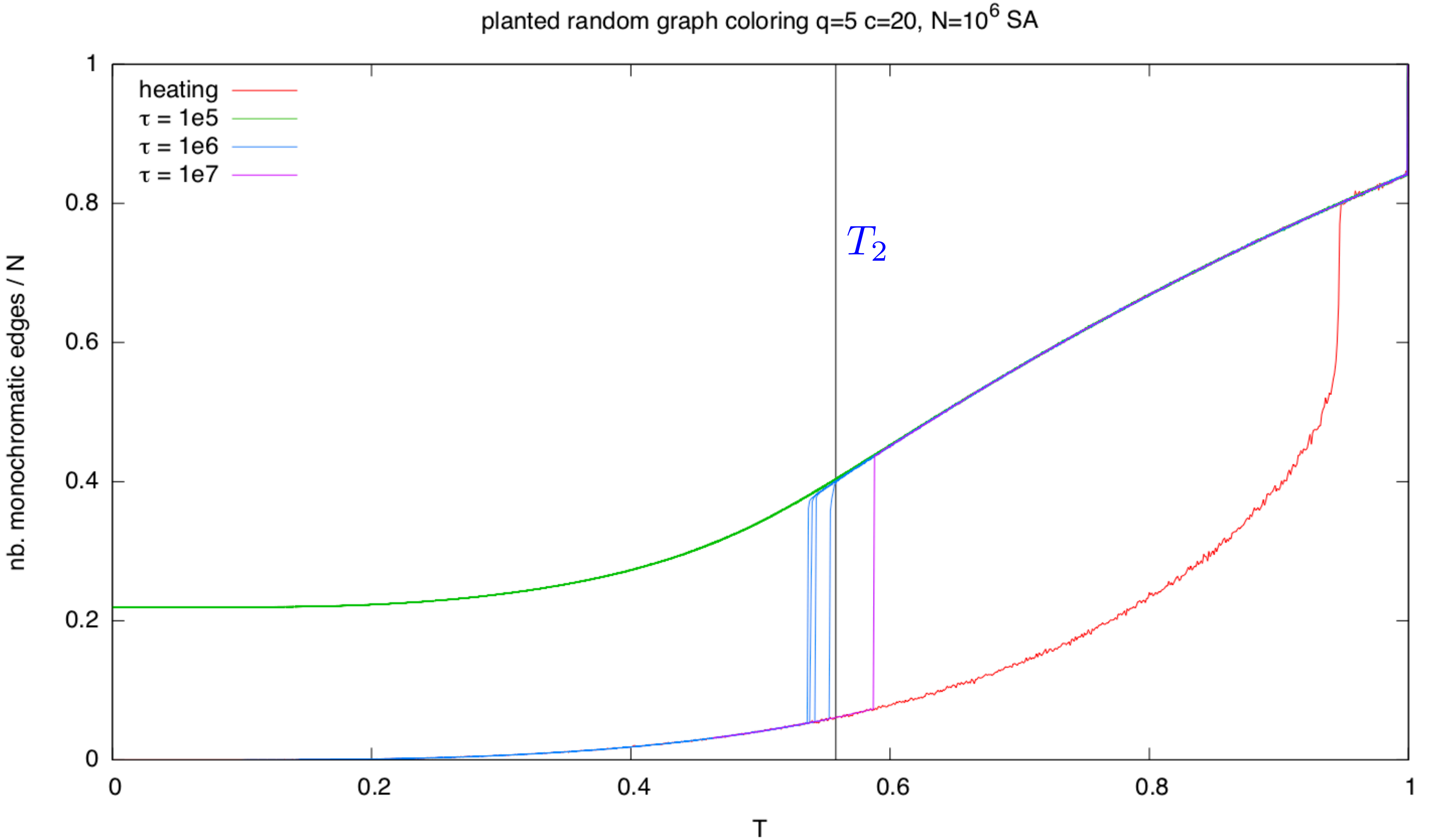
SA in planted models (q=5 coloring)



SA in planted models (q=5 coloring)

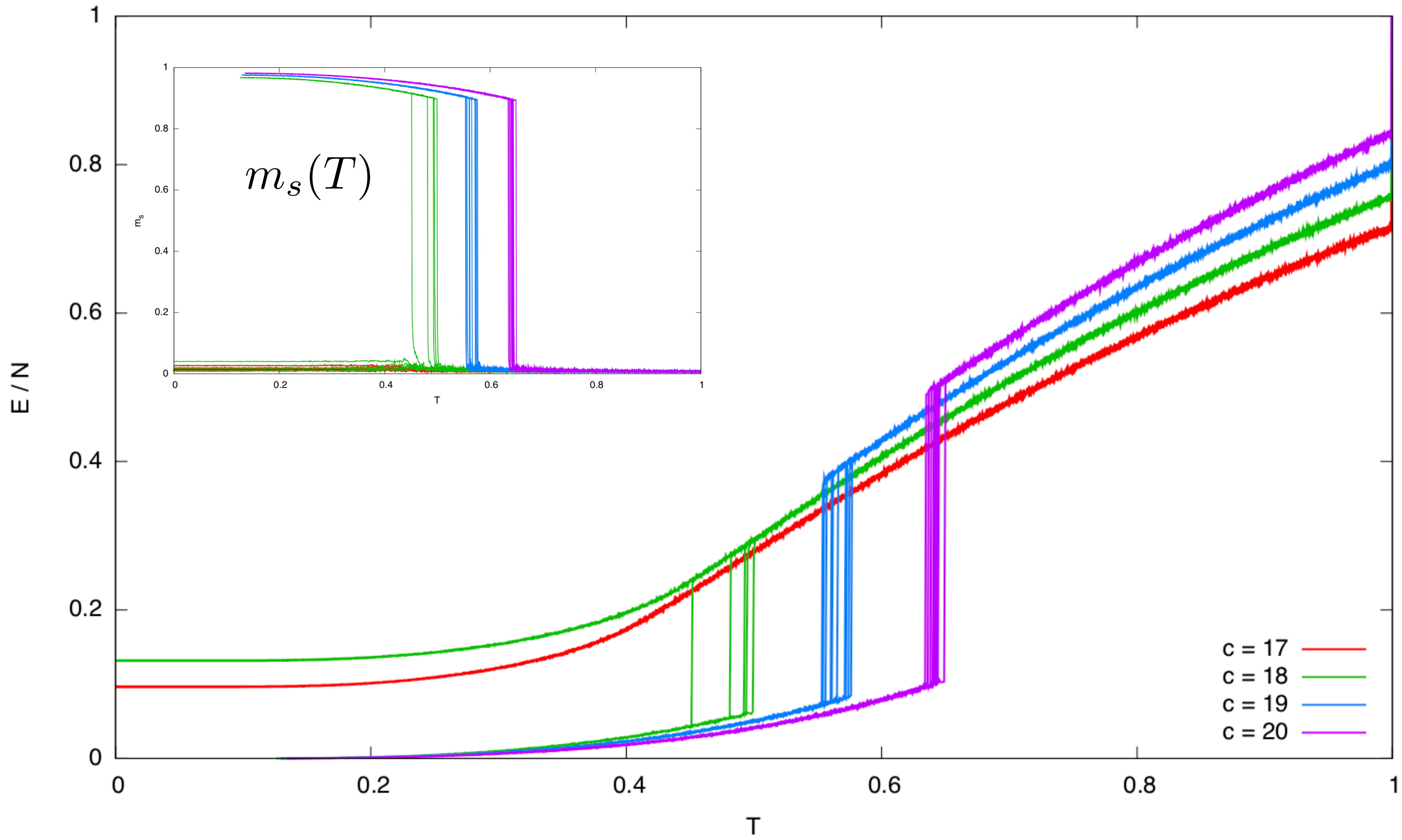


SA in planted models (q=5 coloring)

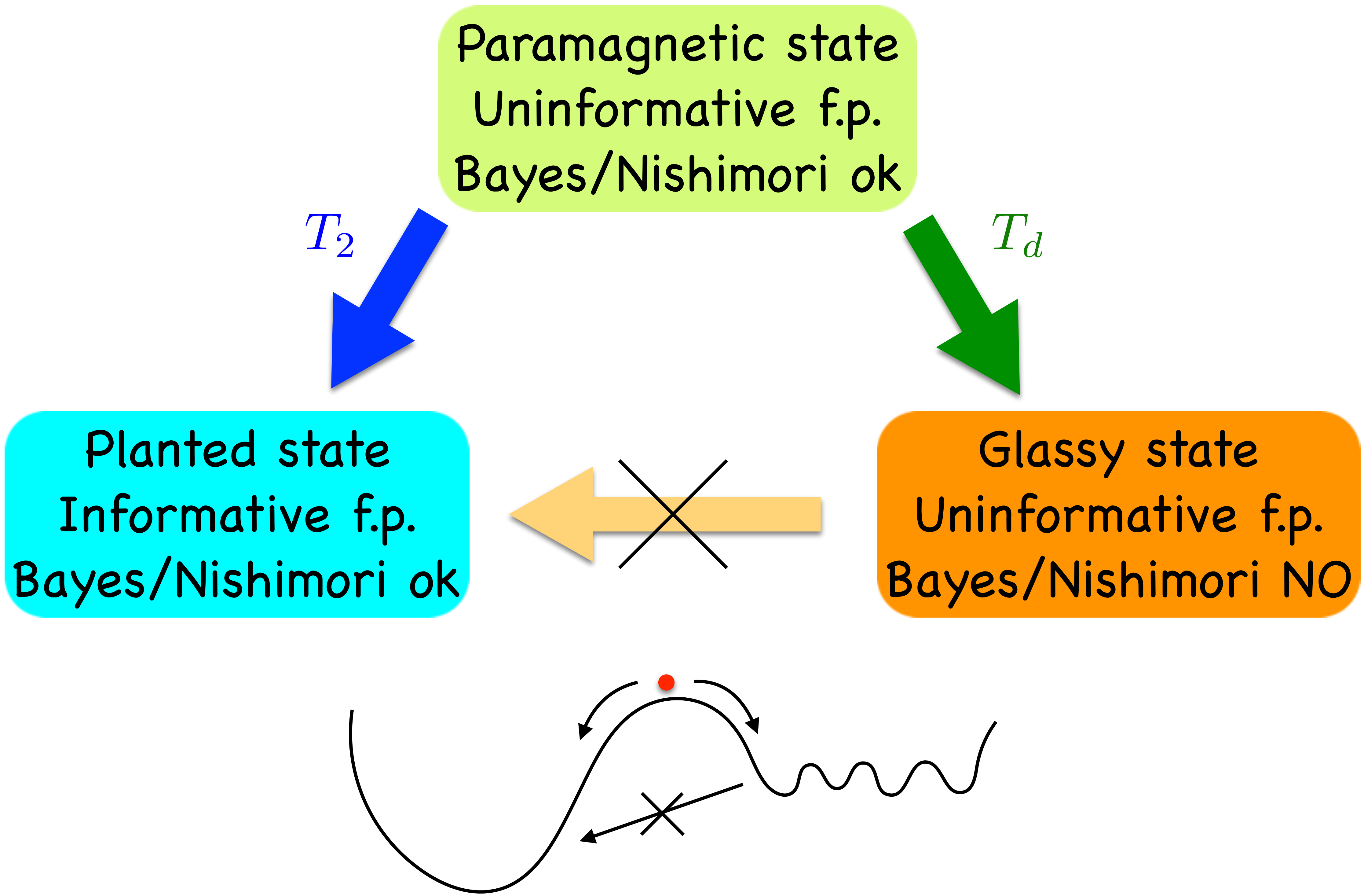


SA in planted models (q=5 coloring)

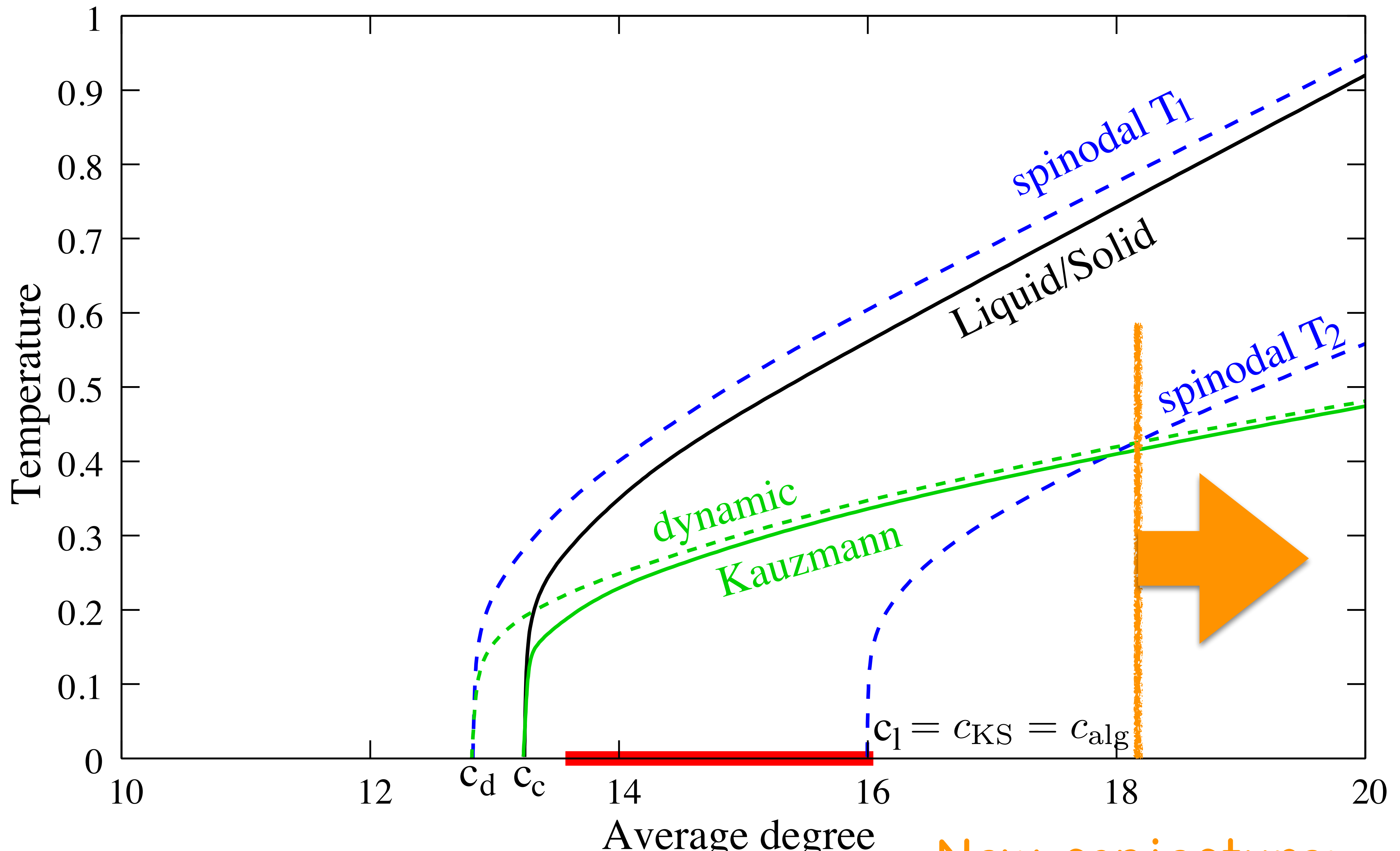
$$N = 10^5 \quad \tau = 10^7$$



Picture for SA in planted models



SA in planted models (q=5 coloring)



New conjecture:
SA works if $T_2 > T_d$

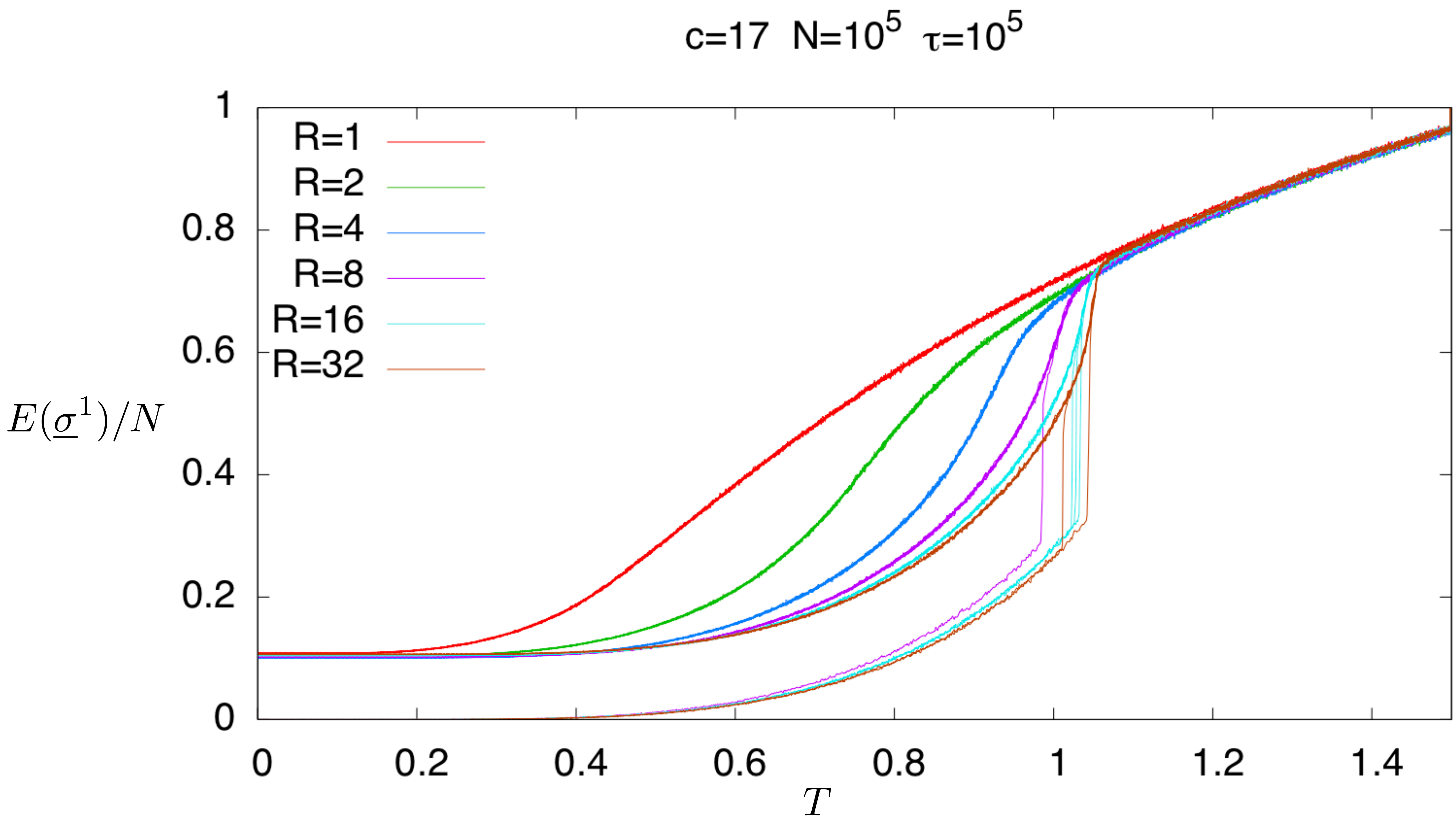
Replicated SA

- 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 favours close by replicas

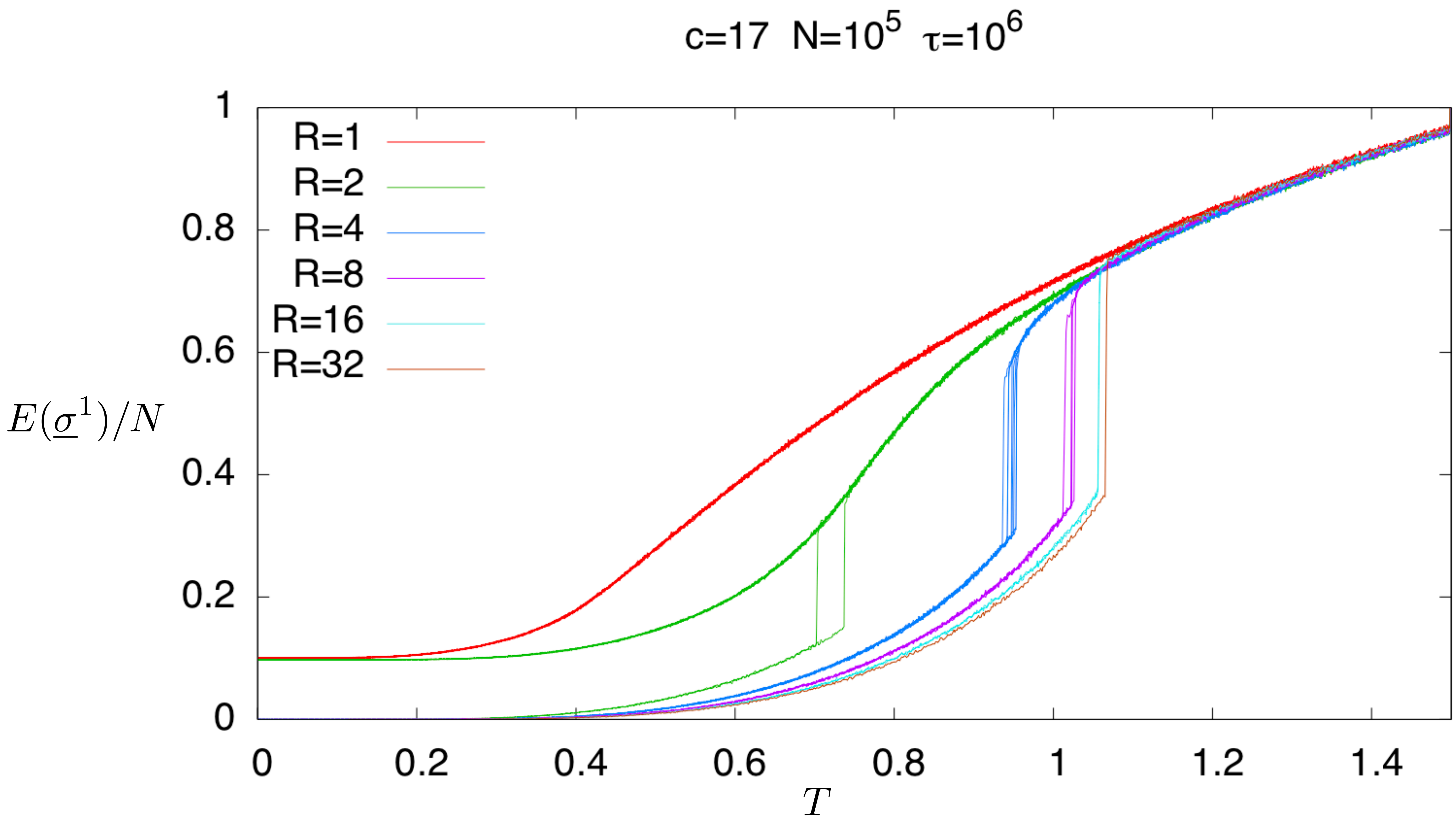
$$\sum_a E(\underline{\sigma}^a) - \frac{\gamma}{R} \sum_{a<b} \sum_i \delta_{\sigma_i^a, \sigma_i^b}$$

  - $\gamma = 1$ in all next plots

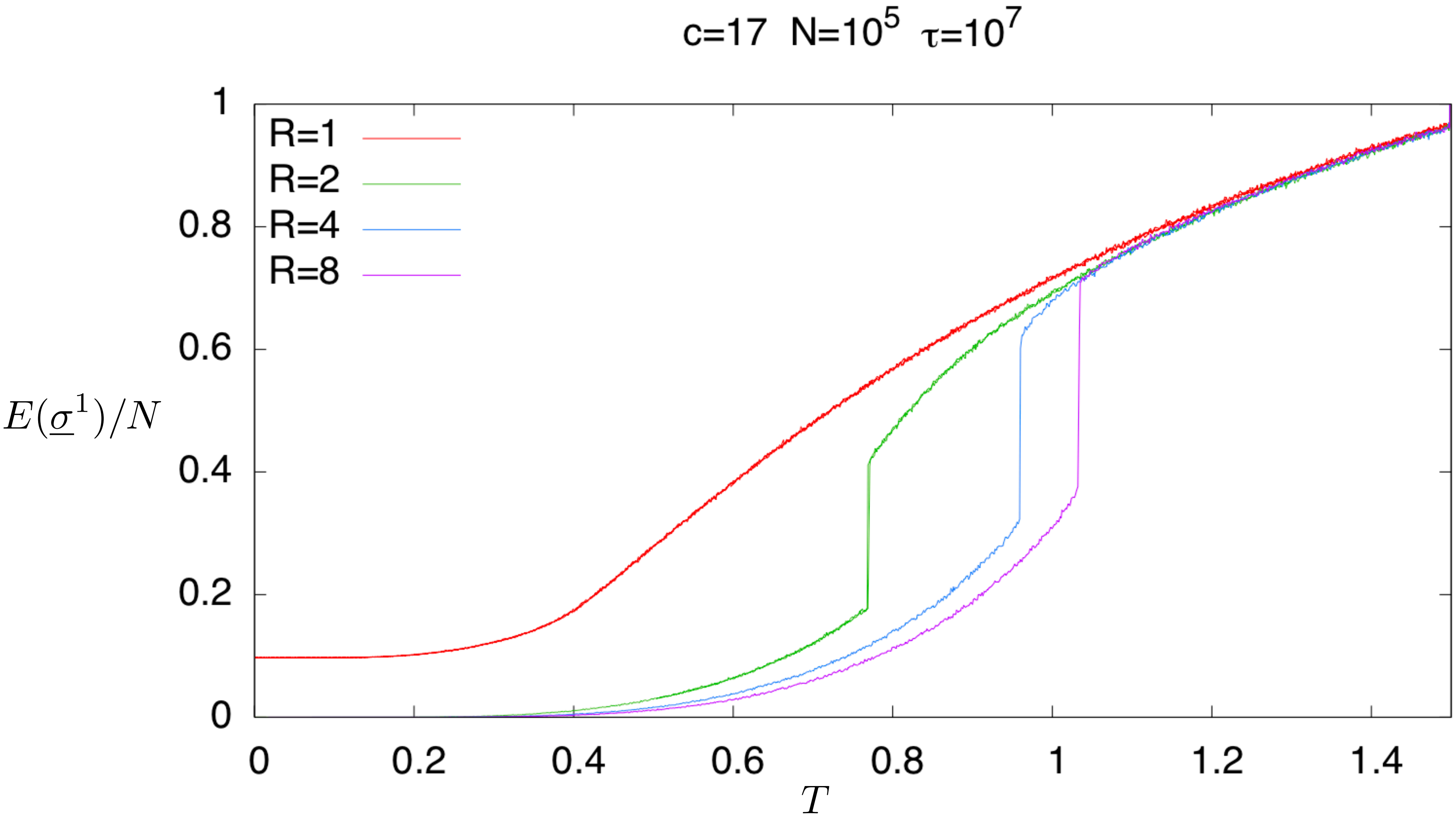
RSA in planted random coloring (q=5)



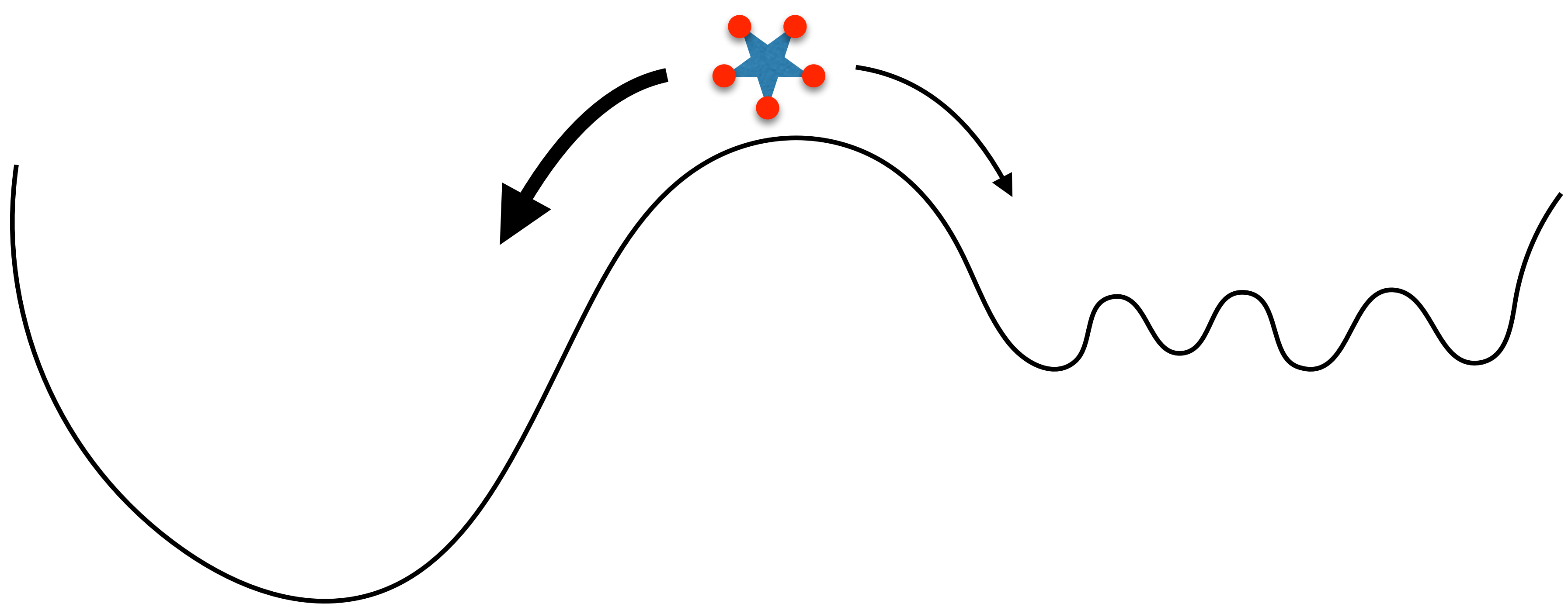
RSA in planted random coloring (q=5)



RSA in planted random coloring (q=5)



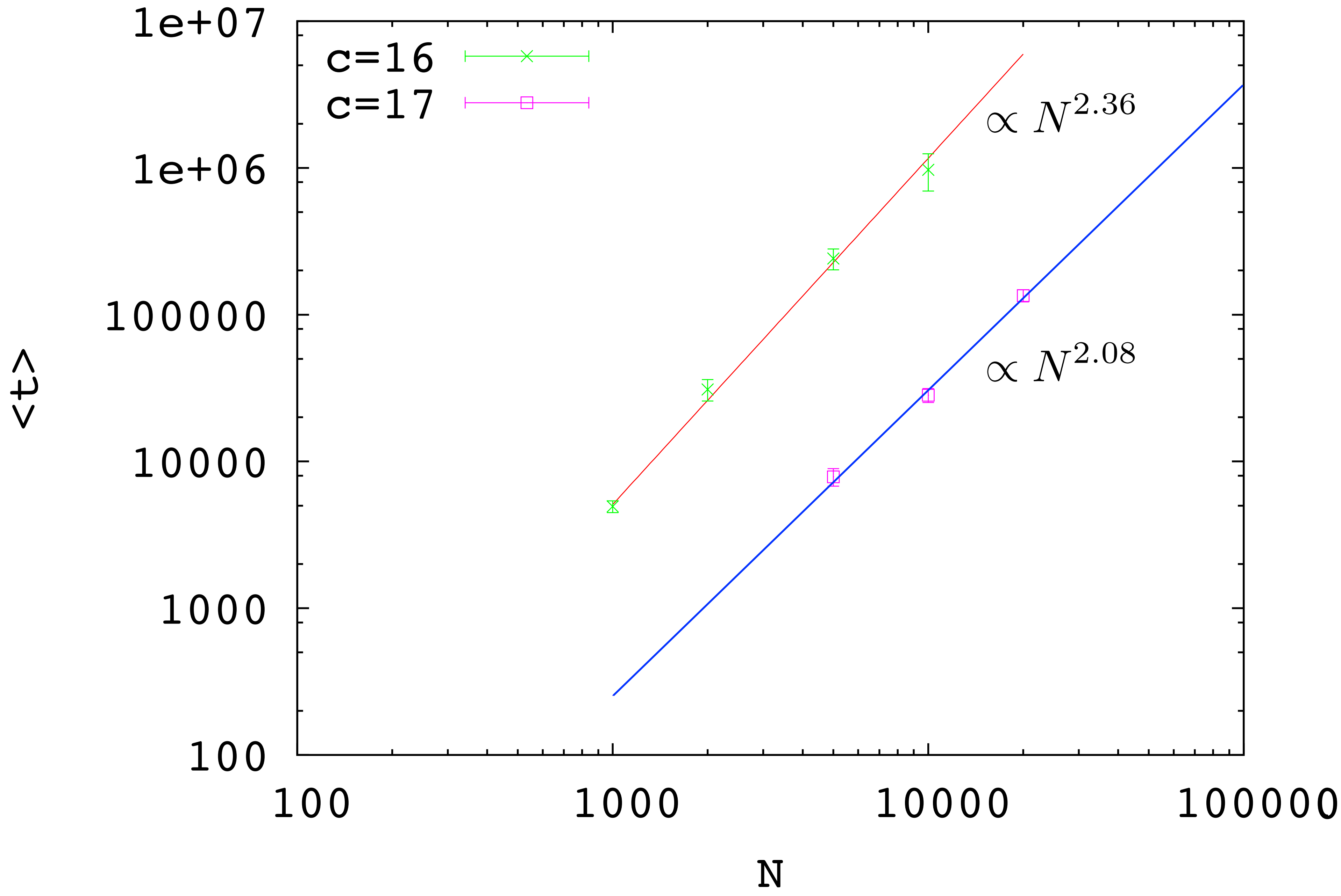
Picture for RSA in planted models



Parallel Tempering (PT)

- 30 temperatures equally spaced in the range $\beta = 1/T \in [0, 6]$
- very robust method
  - practically no optimization
  - the only relevant parameter is the lowest temperature
- total running time = 30 * <t> (shown in the next plot)

PT in planted random coloring (q=5)



Partial conclusions from preliminary data

- When Monte Carlo based algorithms can find the planted configuration?
  - Simulated Annealing -> strictly above $c_{\rm alg}$ when $T_2 > T_d$
  - Replicated Simulated Annealing -> starting from $c_{\rm alg}$
  - Parallel Tempering -> even below $c_{\rm alg}$? (finite N)

- The picture seems robust with respect to the model:
  - planted random graph coloring (q=5)
  - planted random hypergraph bicoloring (K=4 and K=5)

Thanks!