

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

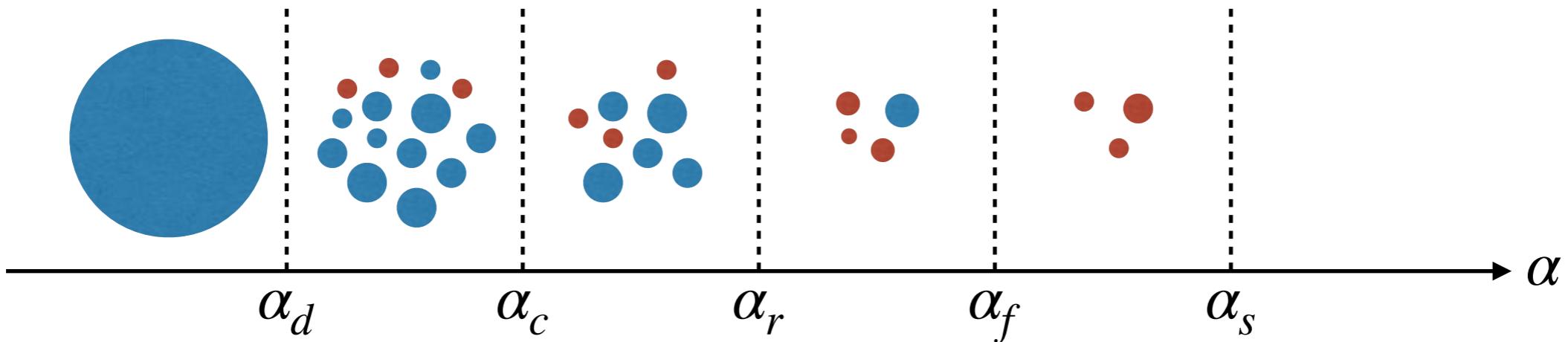
- Introduction to inference problems on sparse random graphs and trees (talk by Guilhem, thanks!)
- Motivations for running long numerical experiments on given random inference problems
- Results (published and unpublished)
 - Belief Propagation (BP) vs. Population Dynamics
 - BP for planted random hypergraph bicoloring
 - BP for asymmetric stochastic block model
 - Monte Carlo based algorithms for planted random coloring
 - Simulated Annealing
 - Replicated Simulated Annealing
 - Parallel Tempering

Outline

- Introduction to inference problems on sparse random graphs and trees (talk by Guilhem, thanks!)
- Motivations for running long numerical experiments on given random inference problems
- Results (published and unpublished)
 - Belief Propagation (BP) vs. Population Dynamics
 - 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

Motivations

- Understanding the behaviour of (solving) algorithms, possibly in terms of phase transitions
...but not easy connection, since most algorithms are out of equilibrium stochastic processes
- Constraint Satisfaction Problems (CSP)
on sparse random graphs with finite degree
→ many different phase transitions



...but no clear connection to algorithmic thresholds

Motivations

- In random CSP many states form a complex hierarchy (Replica Symmetry Breaking, RSB)
- Bayesian inference on sparse random graphs is simpler:

Bayes optimal condition



Nishimori condition
in planted models



Replica Symmetric (RS) states dominate the measure
Belief Propagation (BP) can compute the right marginals
(...but dynamical RSB is often present -> hard phase)

Motivations

- In random CSP many states form a complex hierarchy (Replica Symmetry Breaking, RSB)
- Bayesian inference on sparse random graphs is simpler:

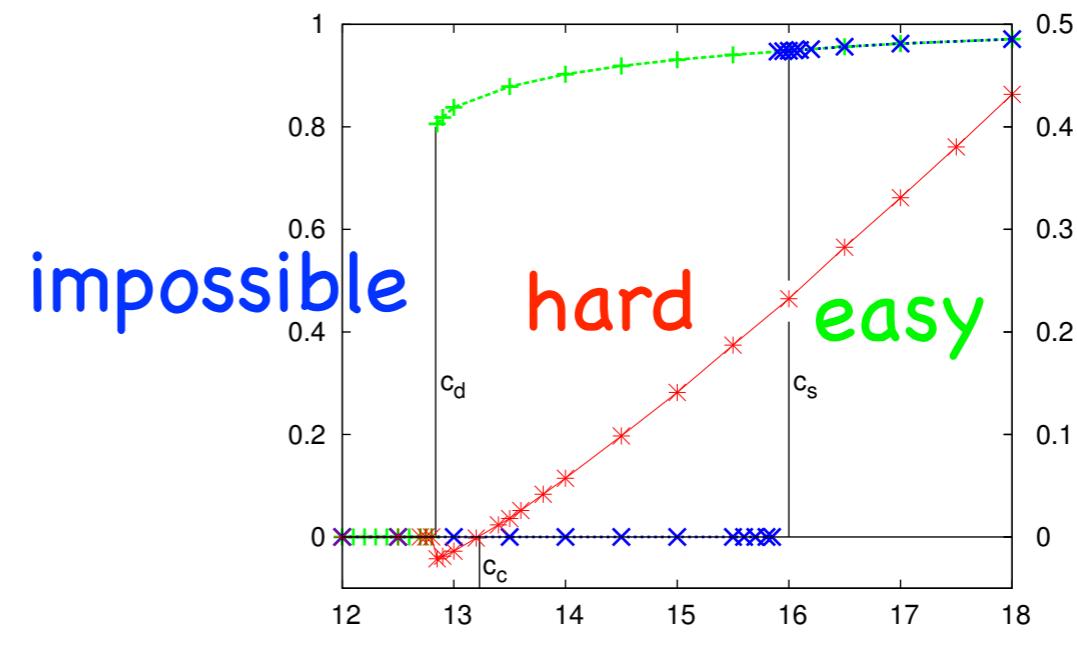
Bayes optimal condition



Nishimori condition
in planted models



Replica Symmetric (RS) states dominate the measure
Belief Propagation (BP) can compute the right marginals
(...but dynamical RSB is often present -> hard phase)



Two classes of algorithms (and more motivations)

- 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

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:
 - BP is run for a time much larger than the girth of the graph, i.e. BP messages run along cycles many many times before reaching convergence
 - BP has no information at all about the planted configuration and has to amplify the initial noise

Planted random hypergraph bicoloring

Application 3 in Guilhem's talk

- 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

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

damping factor
 $\gamma = 0.5$

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

Planted random hypergraph bicoloring

- BP initialization
 - with prob. $q_0 \rightarrow \hat{\eta}_{a \rightarrow i}^{(0)} = \eta_{i \rightarrow a}^{(0)} = \mathbb{I}[\sigma_i^* = 1]$
 - with prob. $1 - q_0 \rightarrow \hat{\eta}_{a \rightarrow i}^{(0)} = 0.5, \eta_{i \rightarrow a}^{(0)} \in [0.45, 0.55]$
- 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 $\left| \eta_{i \rightarrow a}^{(t+1)} - \eta_{i \rightarrow a}^{(t)} \right| < 10^{-8}$

Planted random hypergraph bicoloring

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

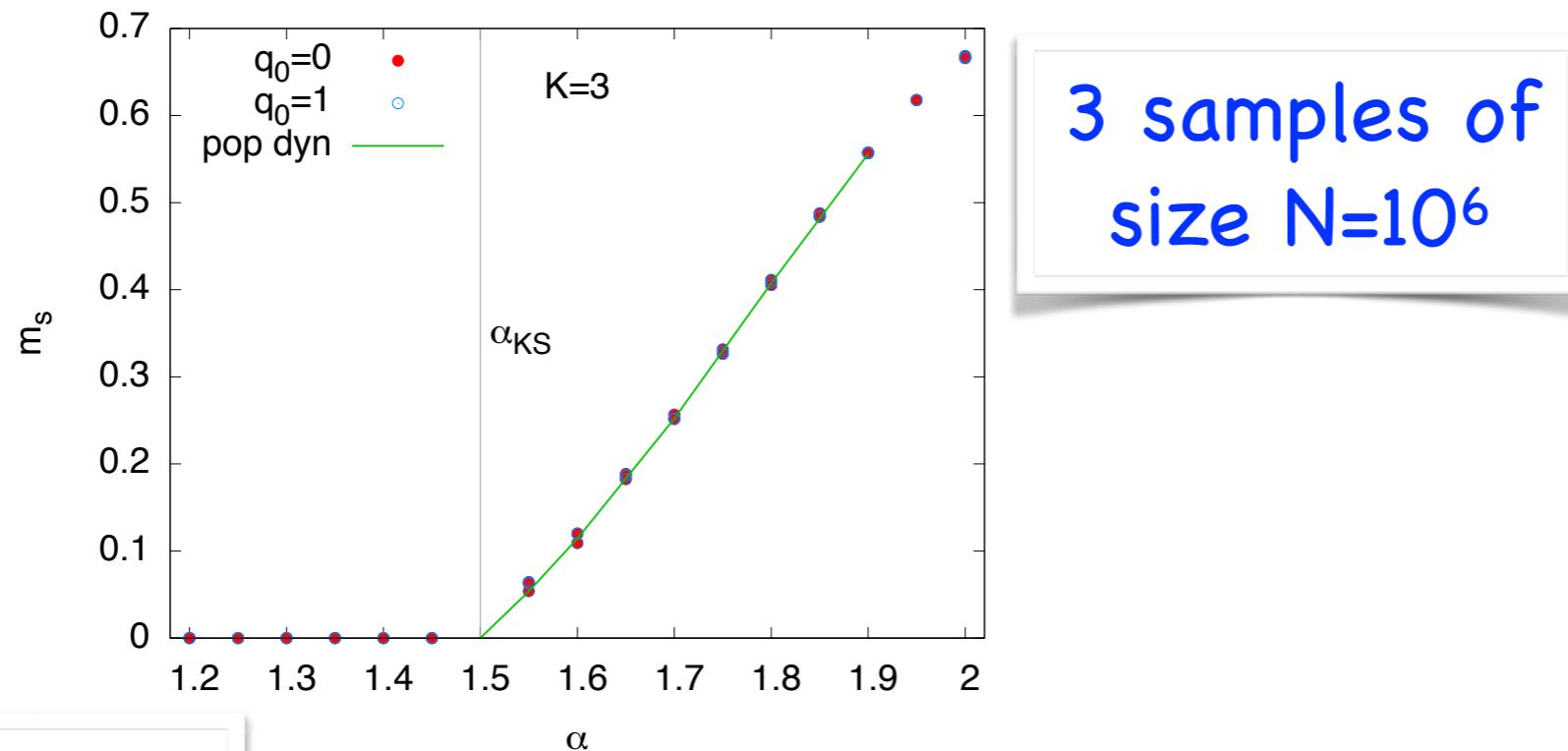
$$m_i^* = \frac{\prod_{a \in \partial i} \hat{\eta}_{a \rightarrow i}^* - \prod_{a \in \partial i} (1 - \hat{\eta}_{a \rightarrow i}^*)}{\prod_{a \in \partial i} \hat{\eta}_{a \rightarrow i}^* + \prod_{a \in \partial i} (1 - \hat{\eta}_{a \rightarrow i}^*)}$$

- Detection of the planted configuration is signalled by

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

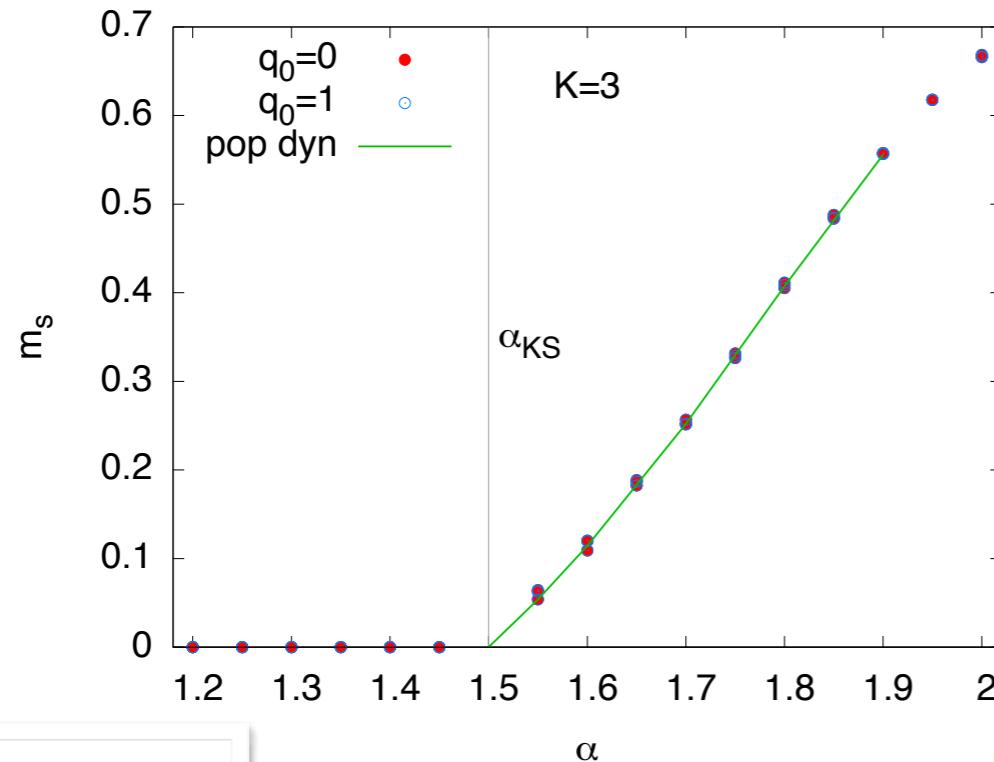
- The uninformative fixed point has $m_i^* = m_s = m_2 = q = 0$

Planted random hypergraph bicoloring

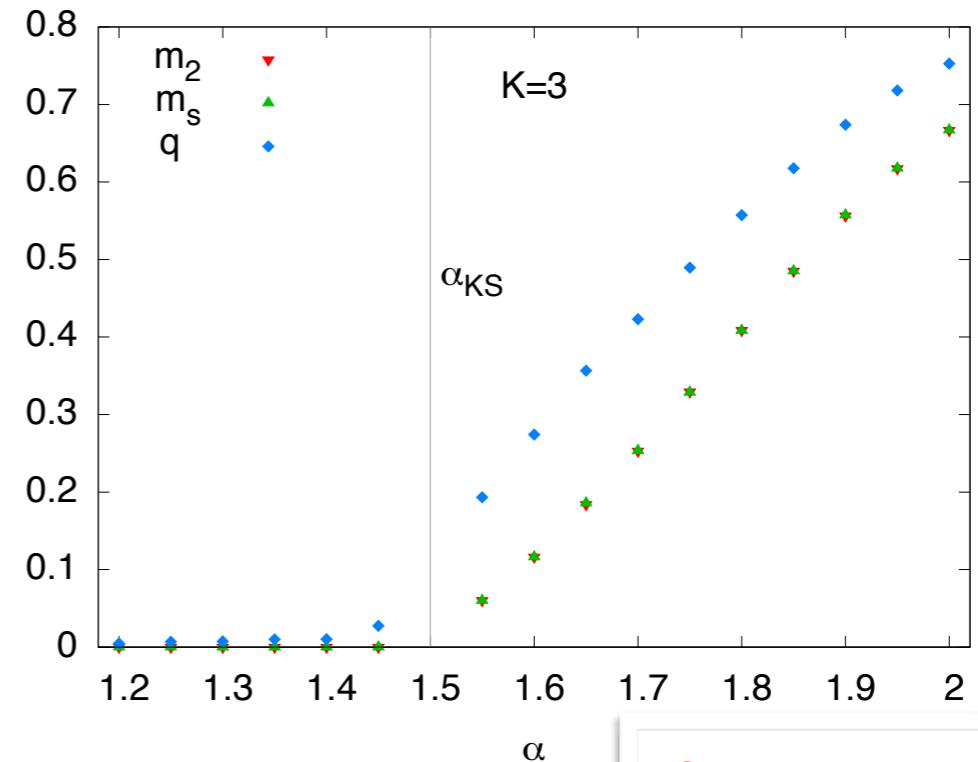


$m_s = 2a$
Guilhem's
notation

Planted random hypergraph bicoloring

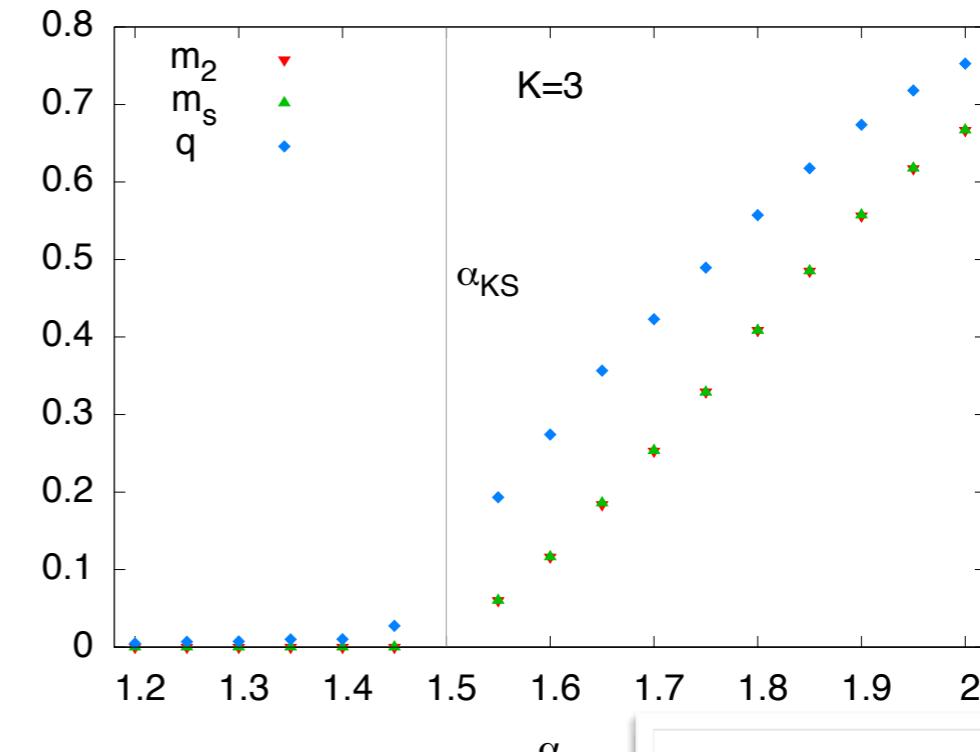
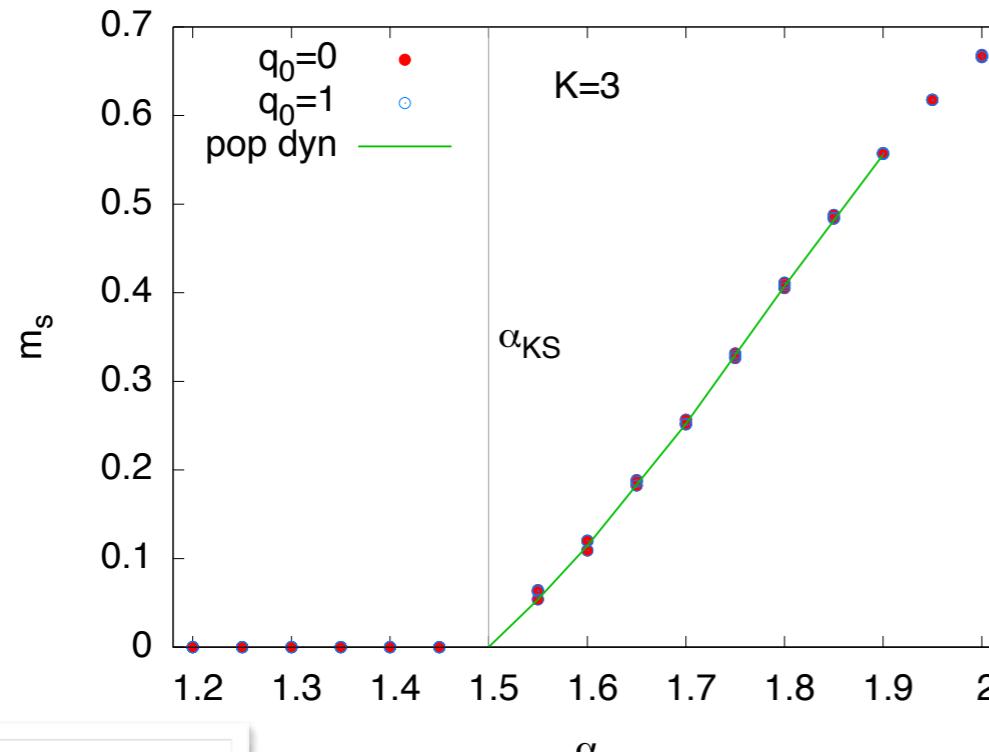


$m_s = 2a$
Guilhem's
notation

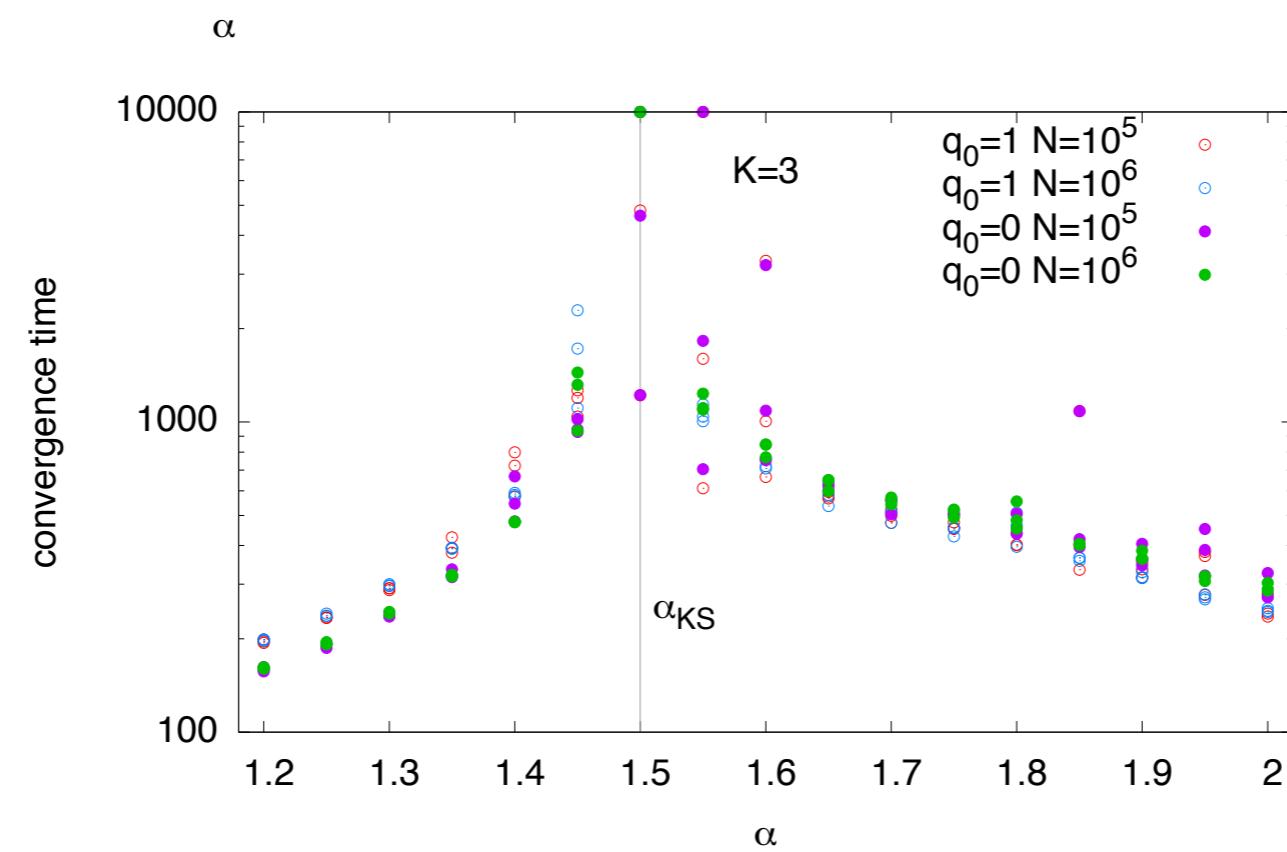


Bayes condition
 $m_s = m_2$

Planted random hypergraph bicoloring

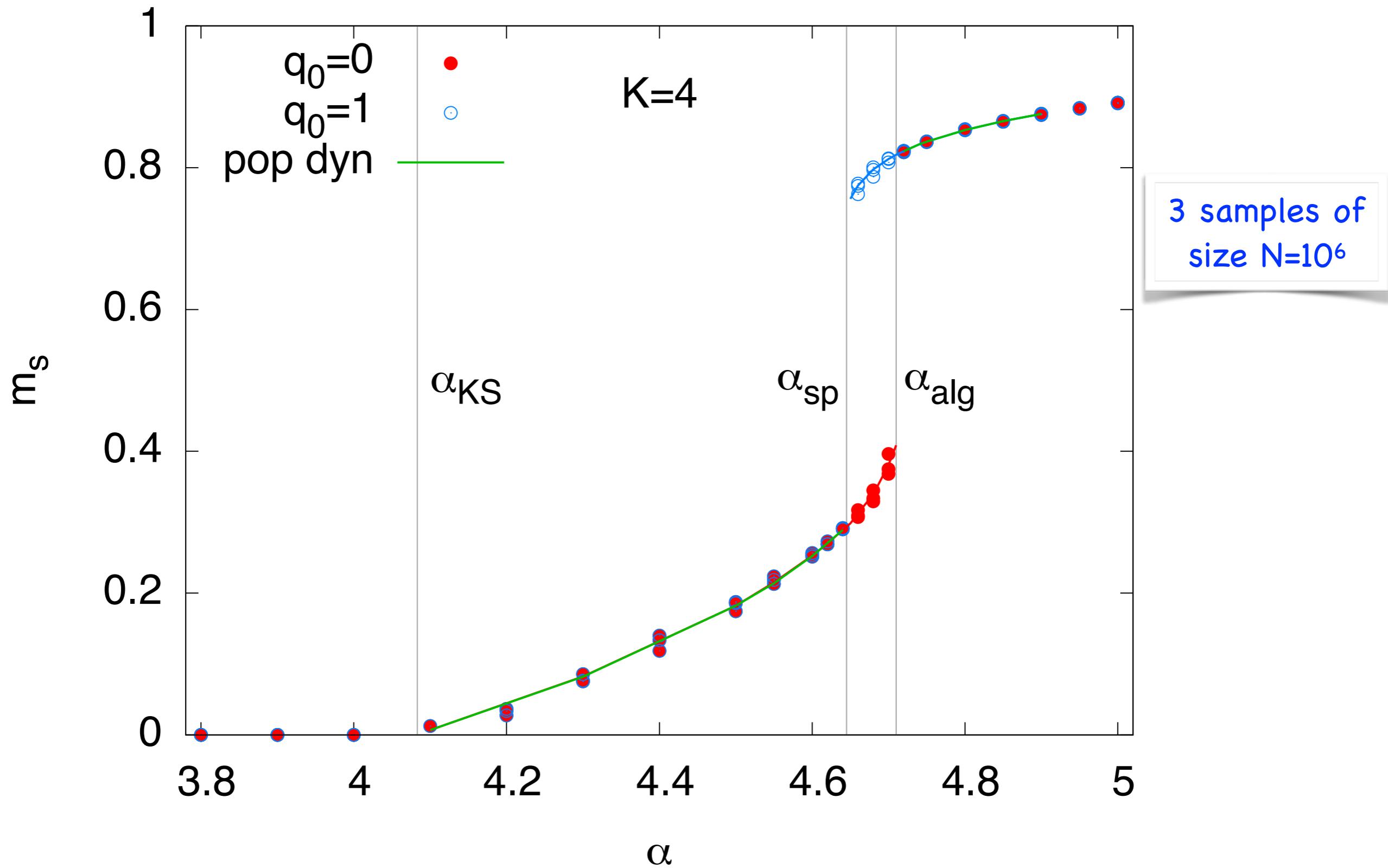


$m_s = 2a$
Guilhem's notation

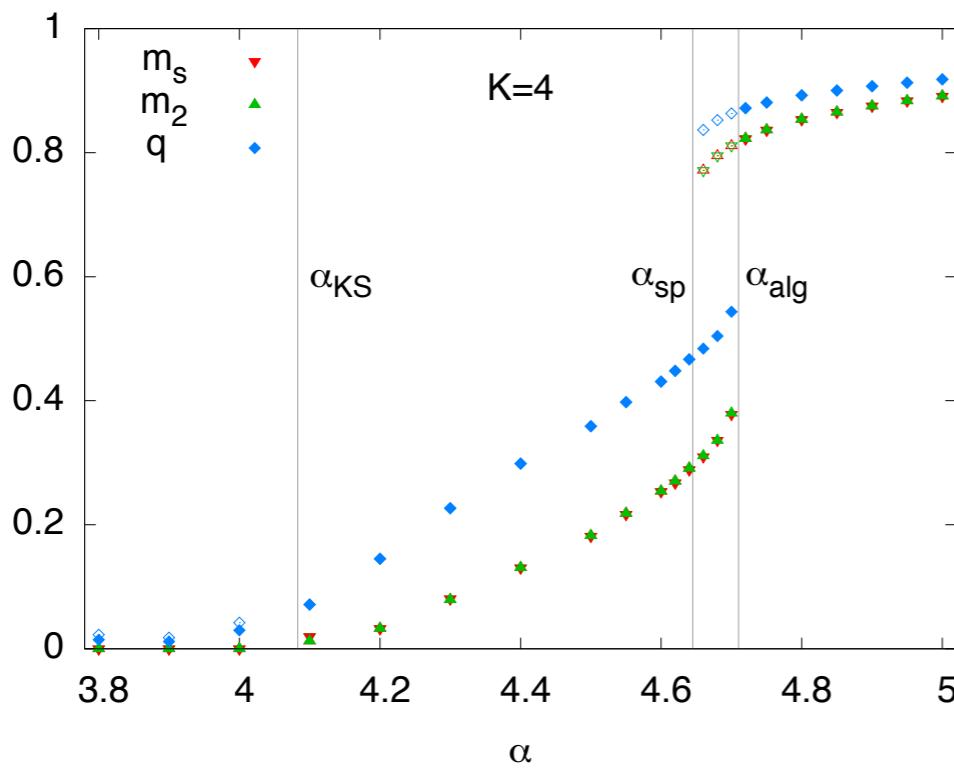
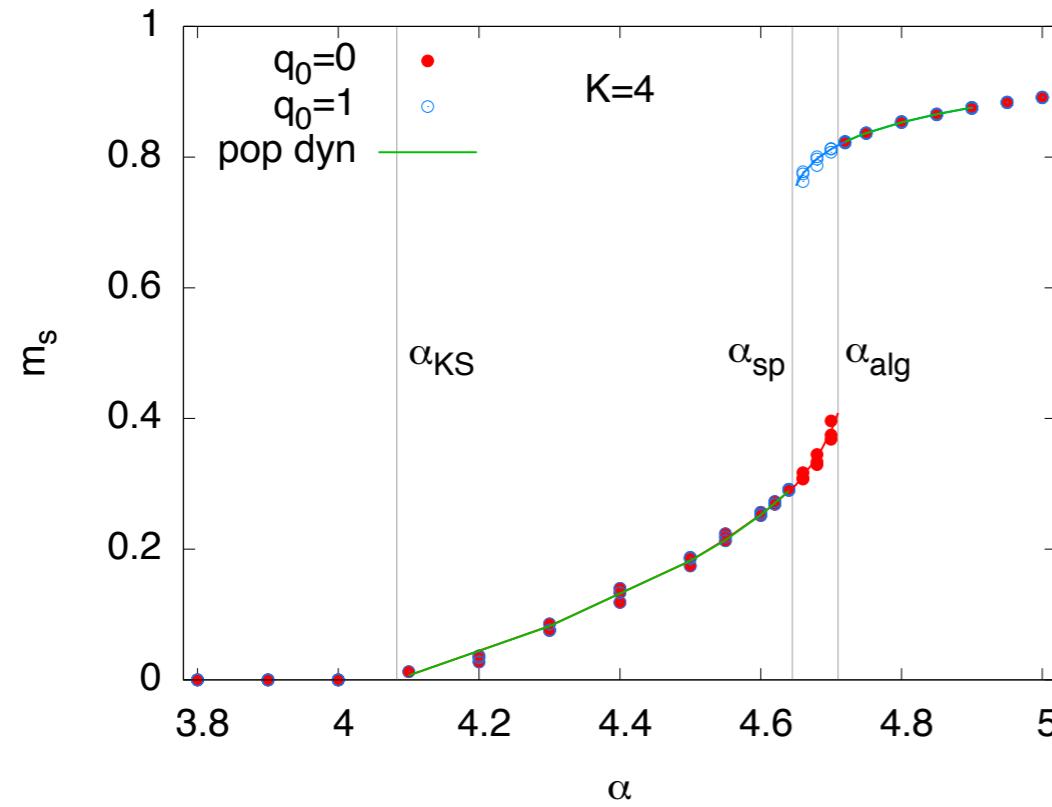


Bayes condition
 $m_s = m_2$

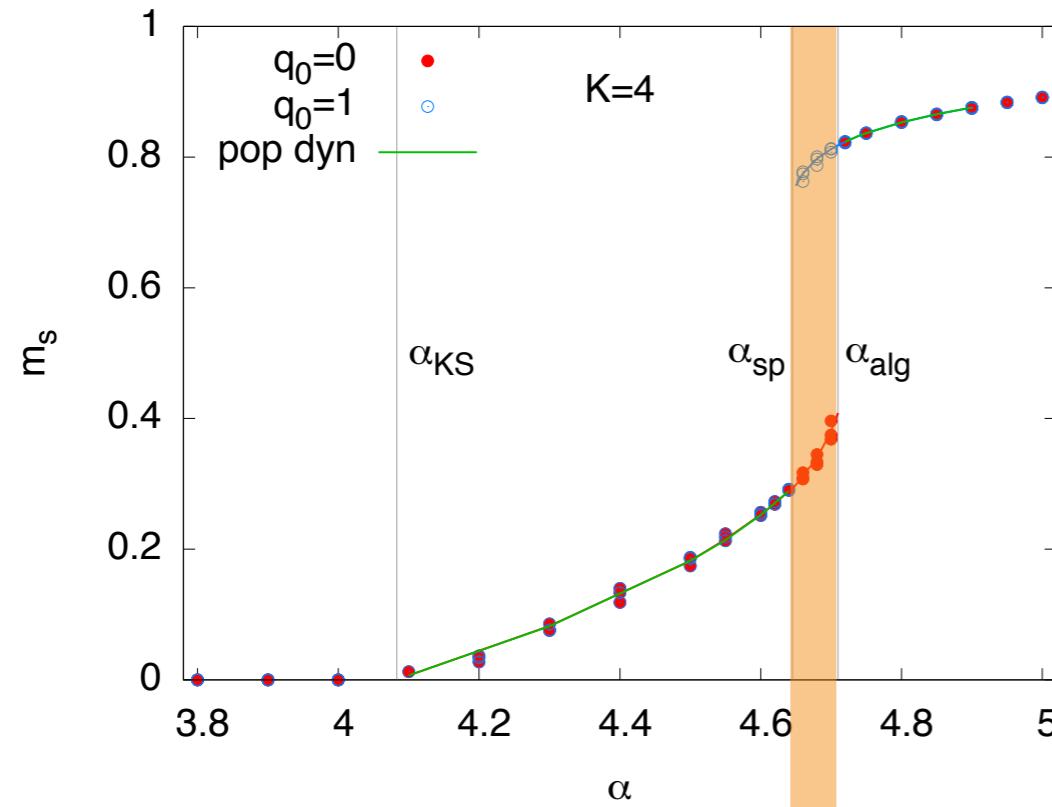
Planted random hypergraph bicoloring



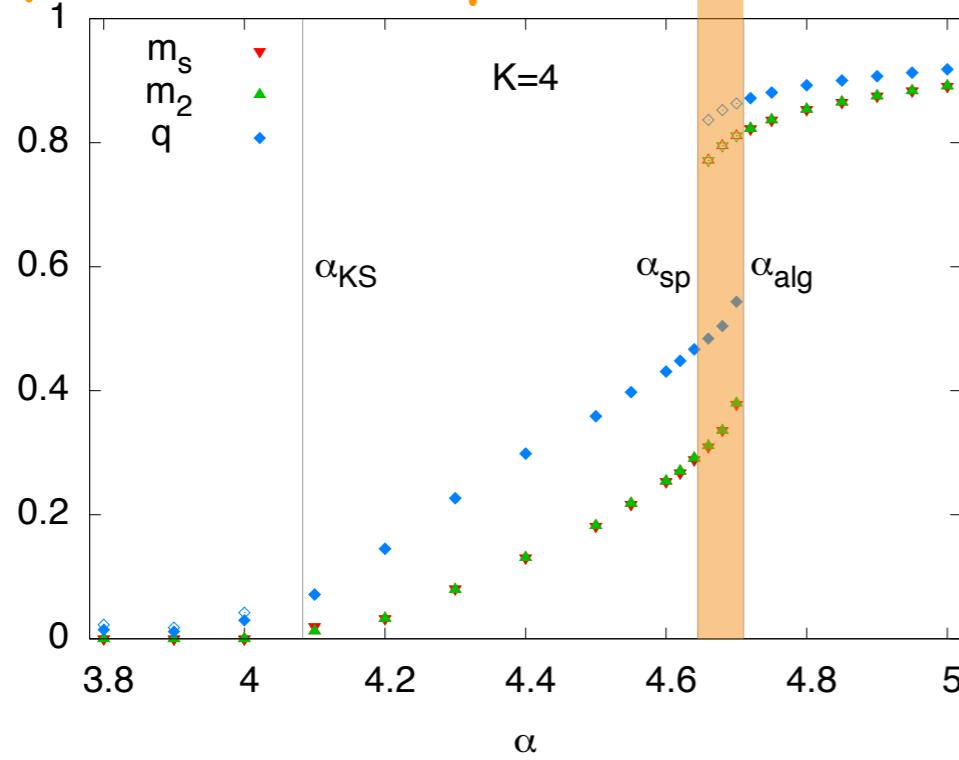
Planted random hypergraph bicoloring



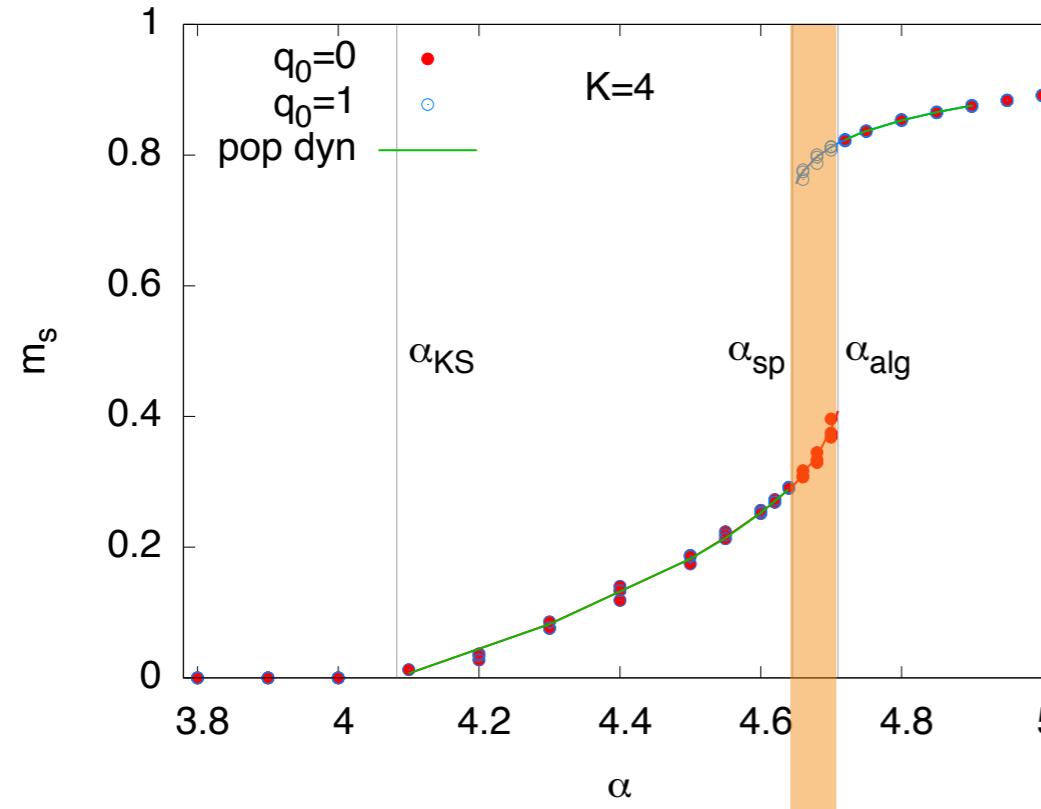
Planted random hypergraph bicoloring



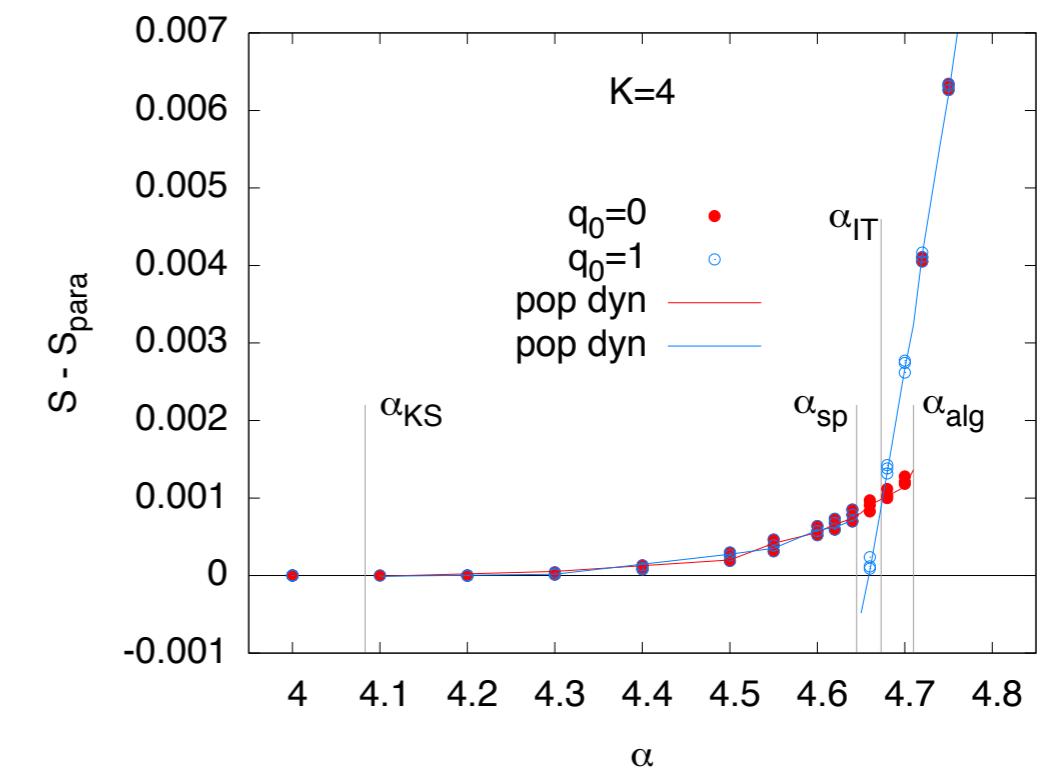
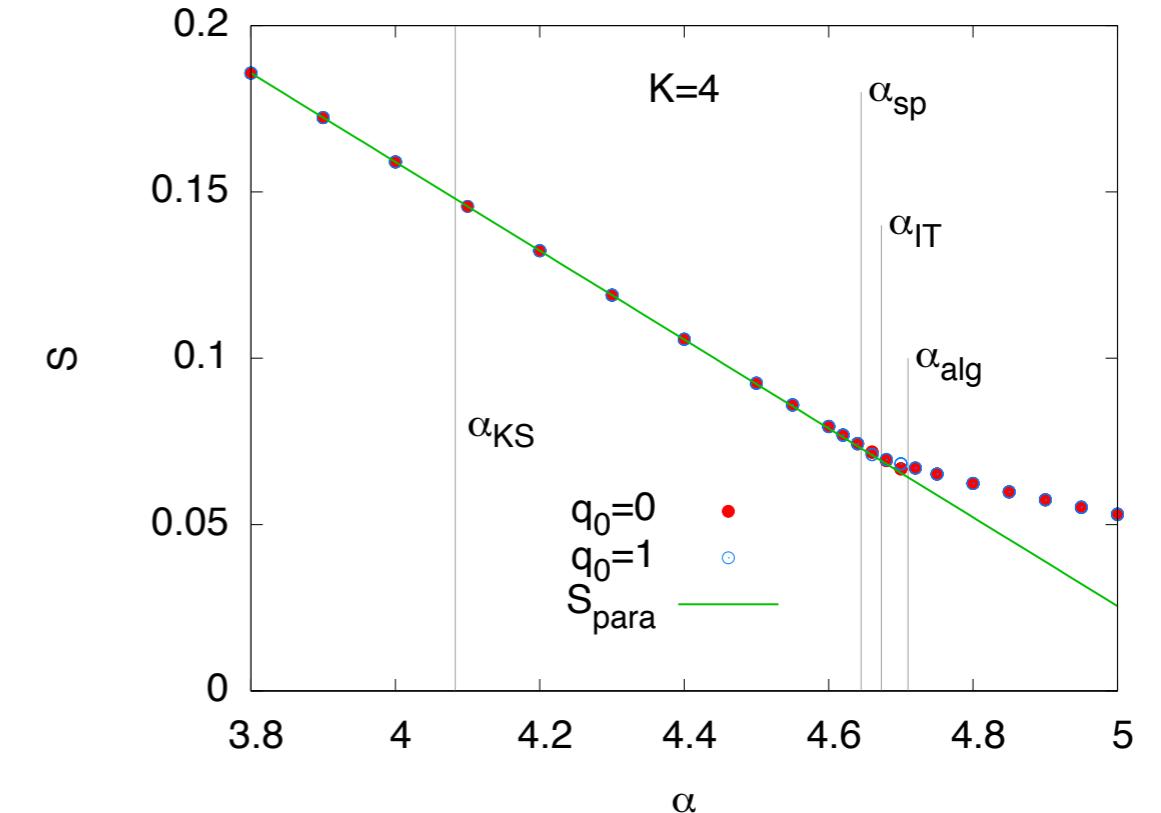
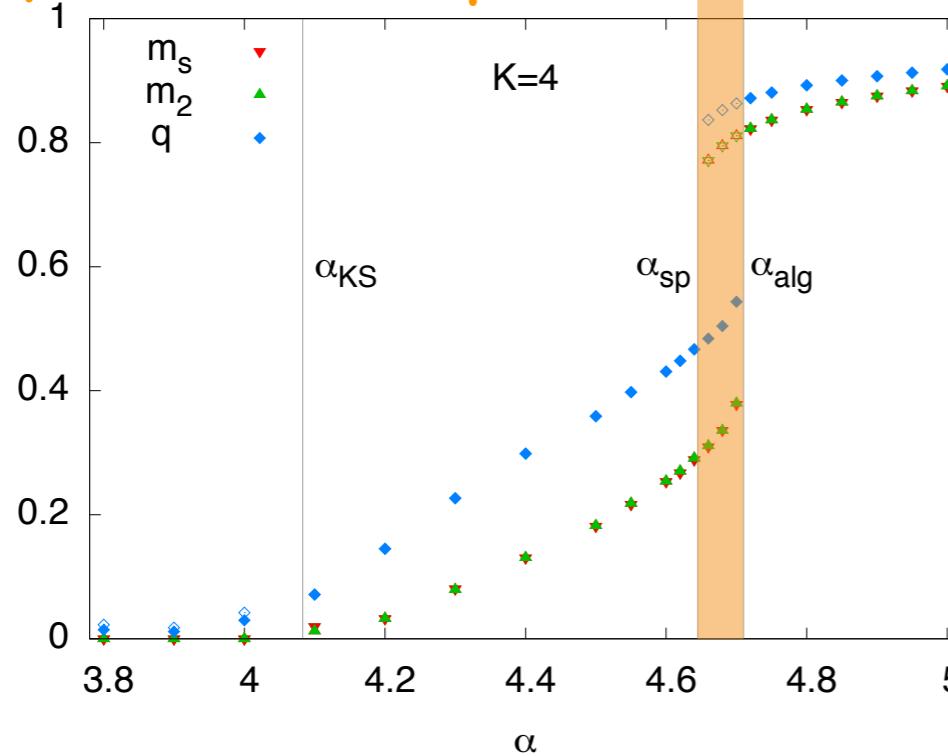
hybrid-hard phase



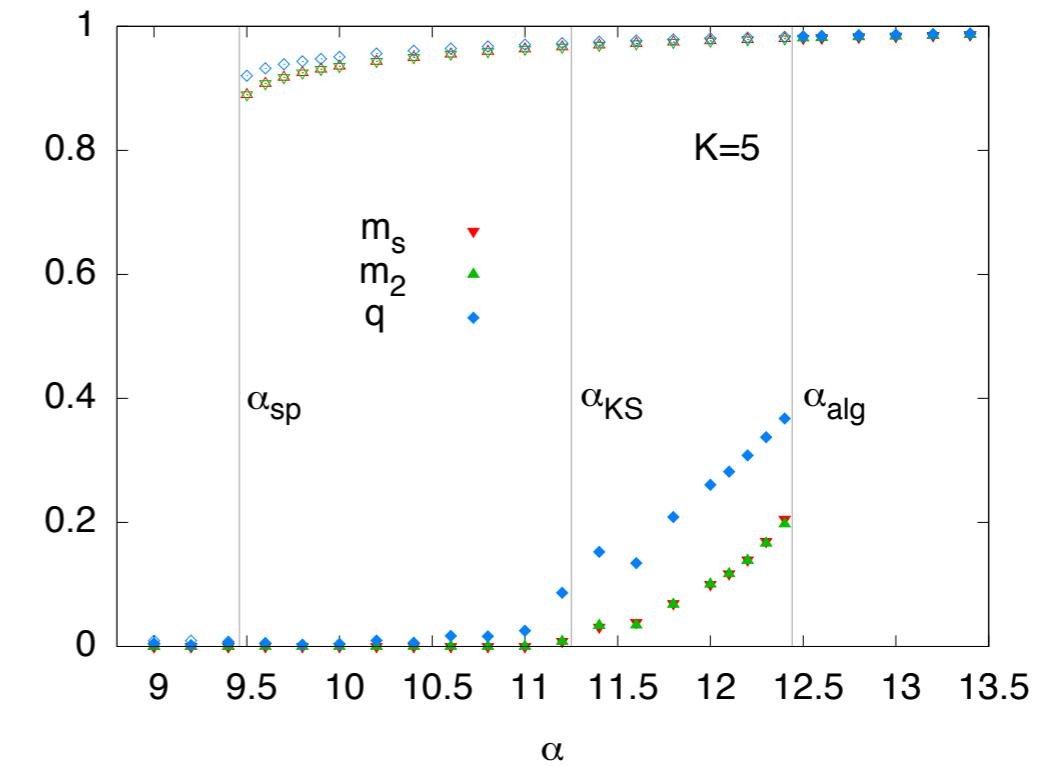
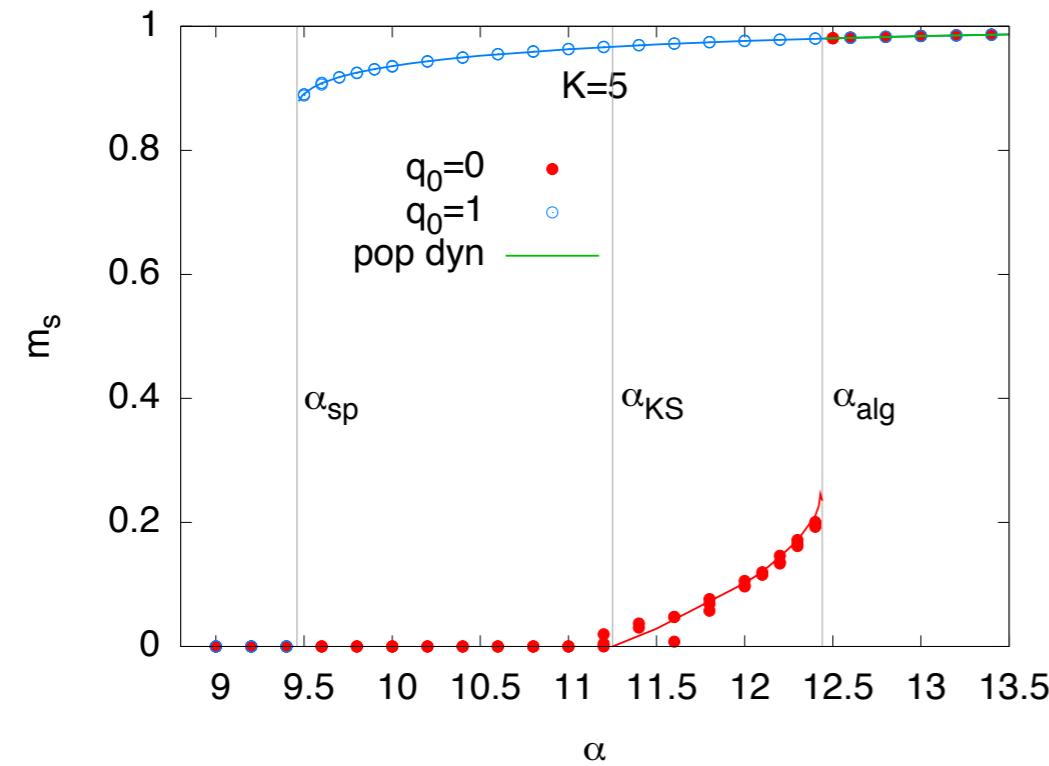
Planted random hypergraph bicoloring



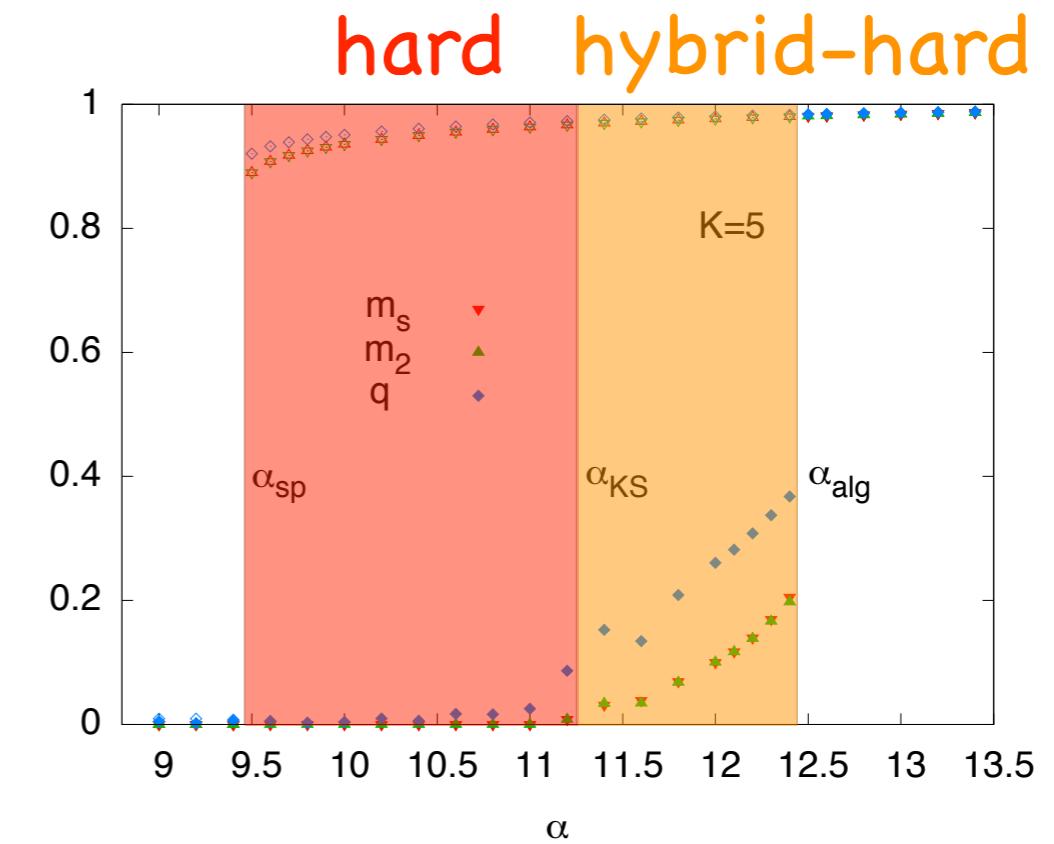
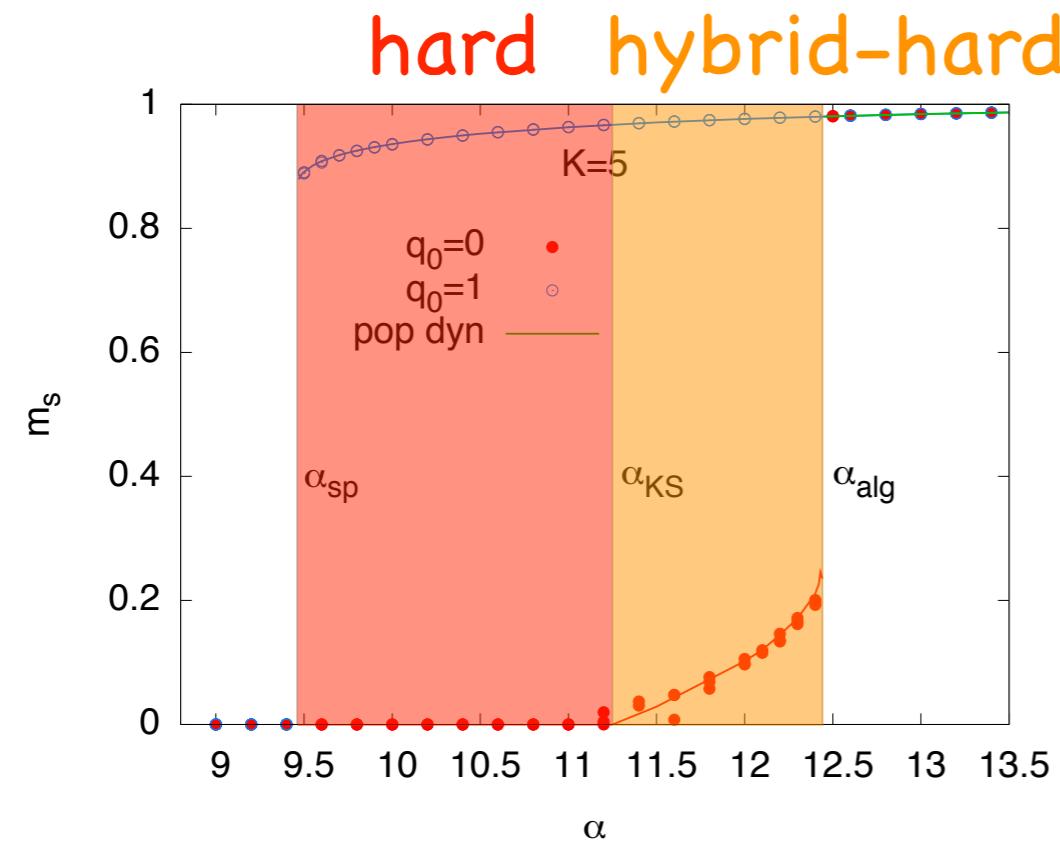
hybrid-hard phase



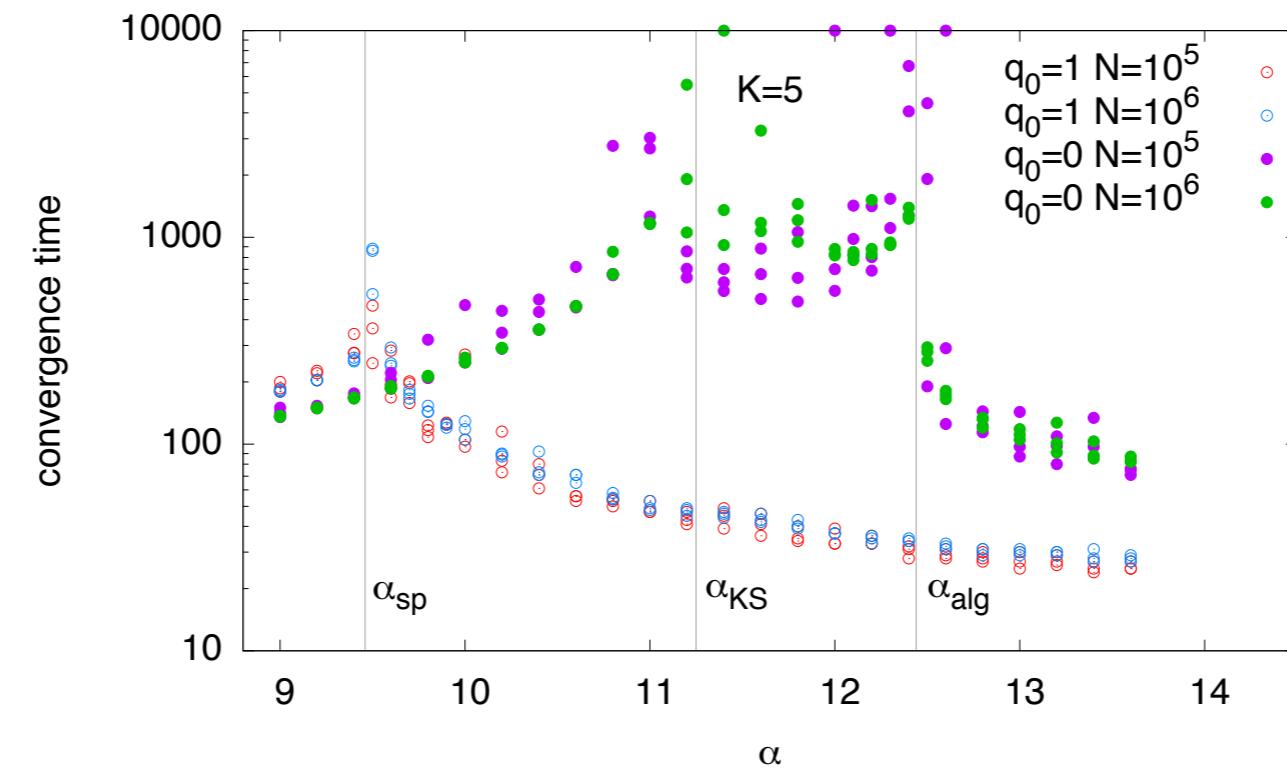
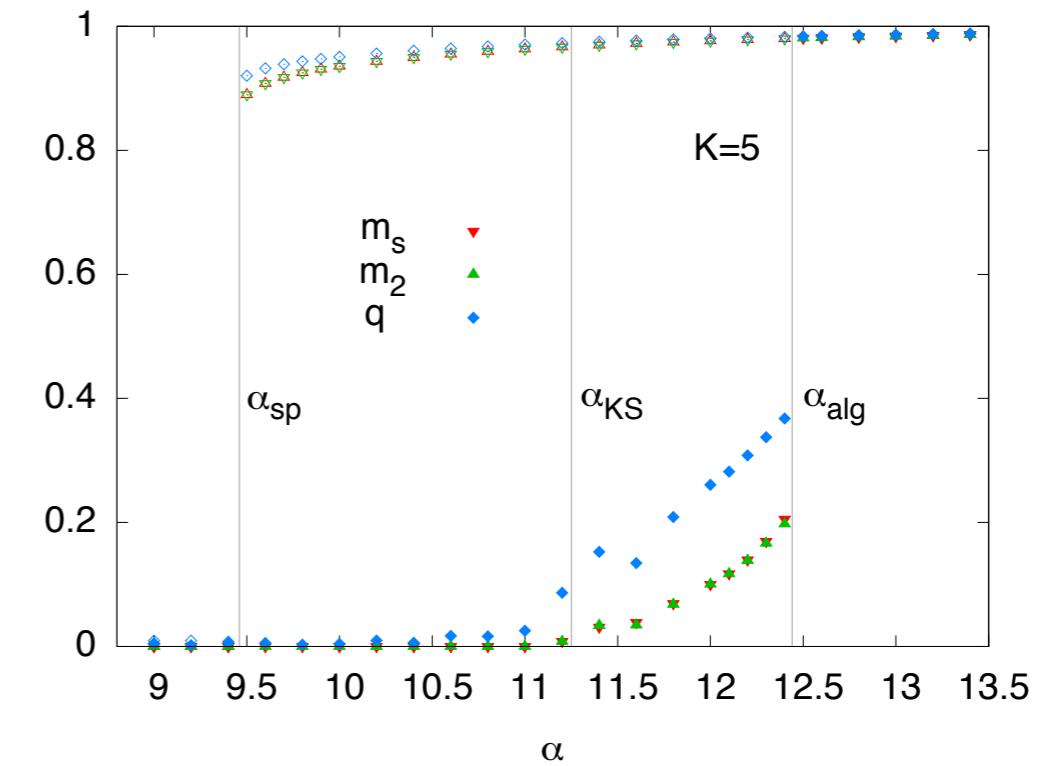
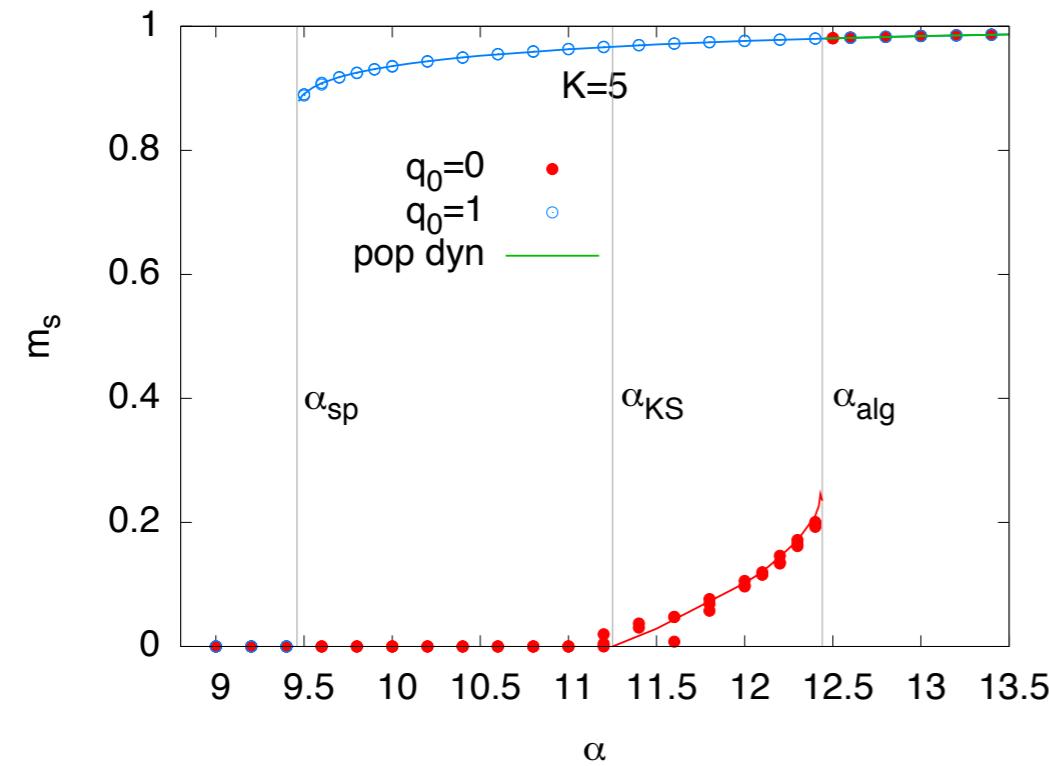
Planted random hypergraph bicoloring



Planted random hypergraph bicoloring

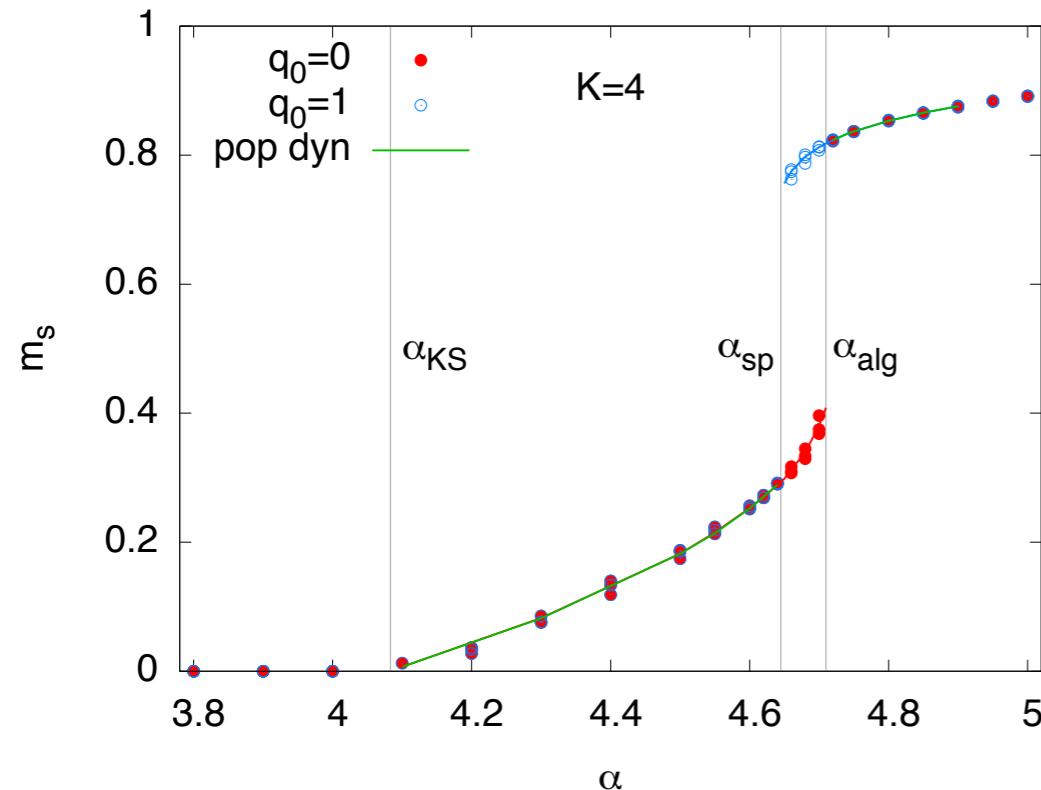


Planted random hypergraph bicoloring



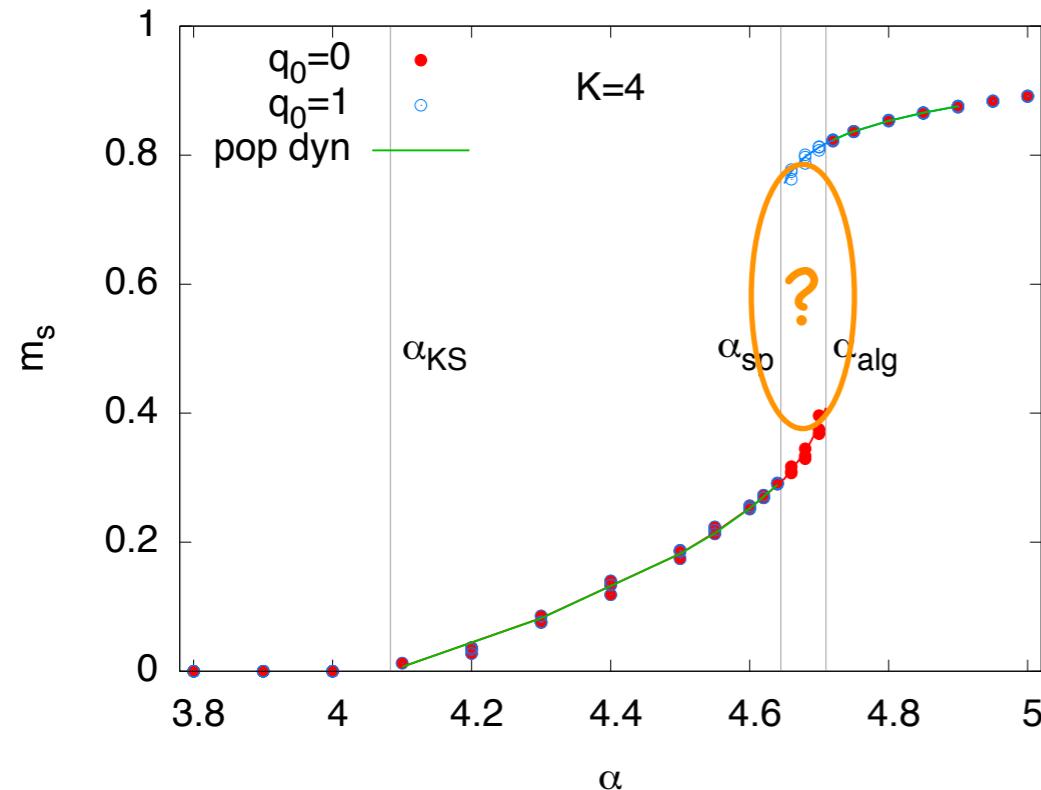
Finding the unstable fixed point via BP

Hard (impossible?) to do via PopDyn



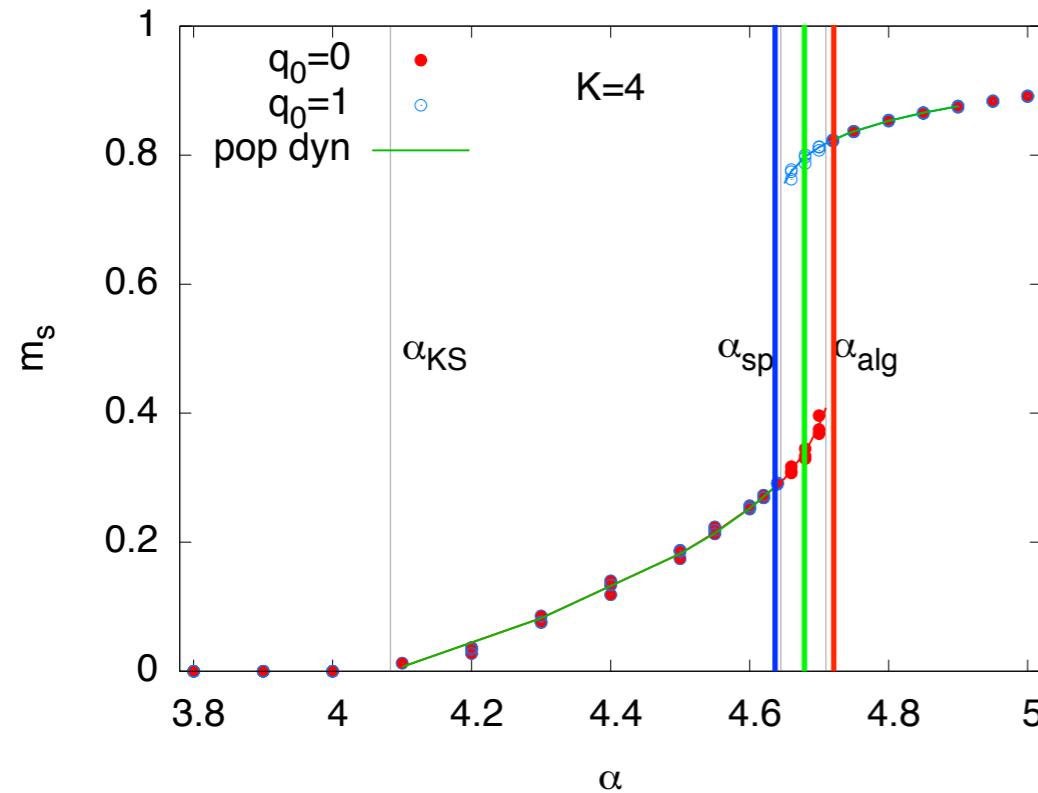
Finding the unstable fixed point via BP

Hard (impossible?) to do via PopDyn



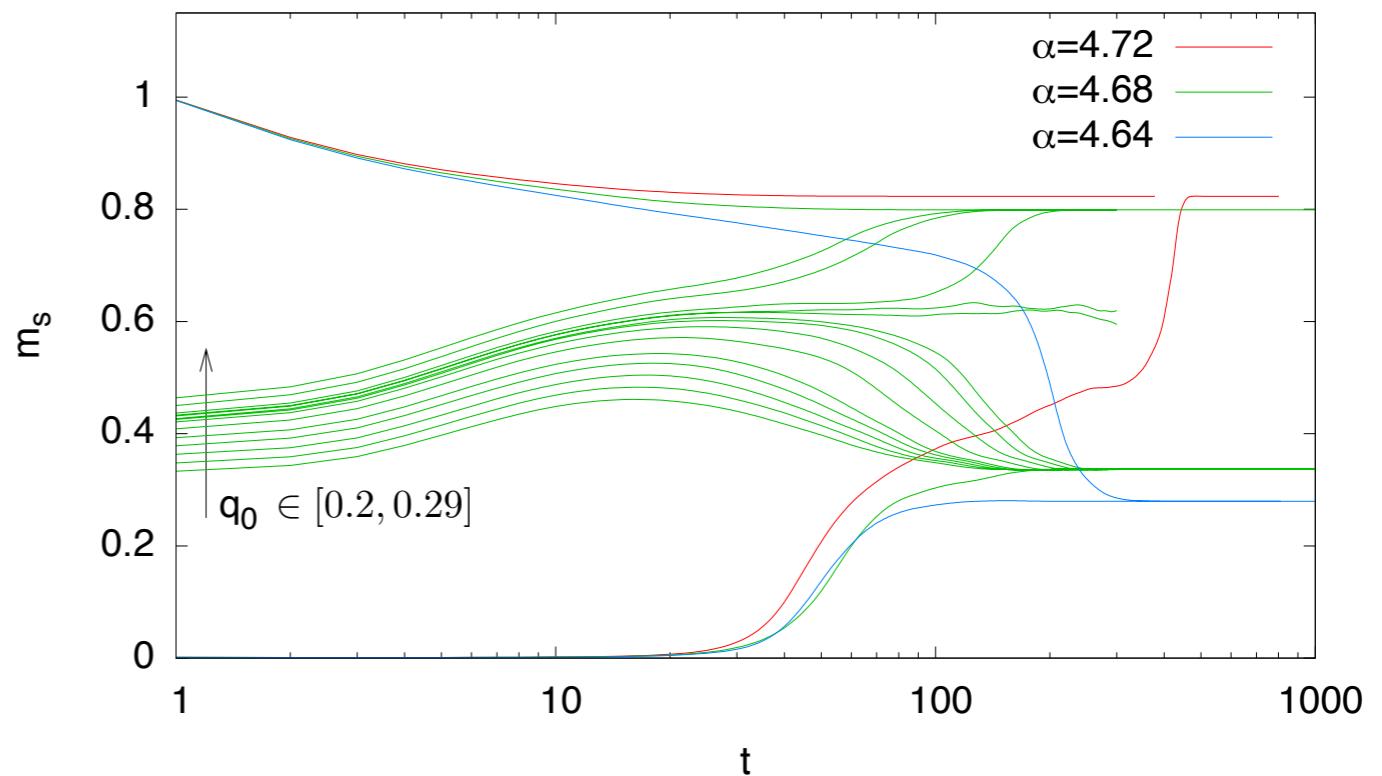
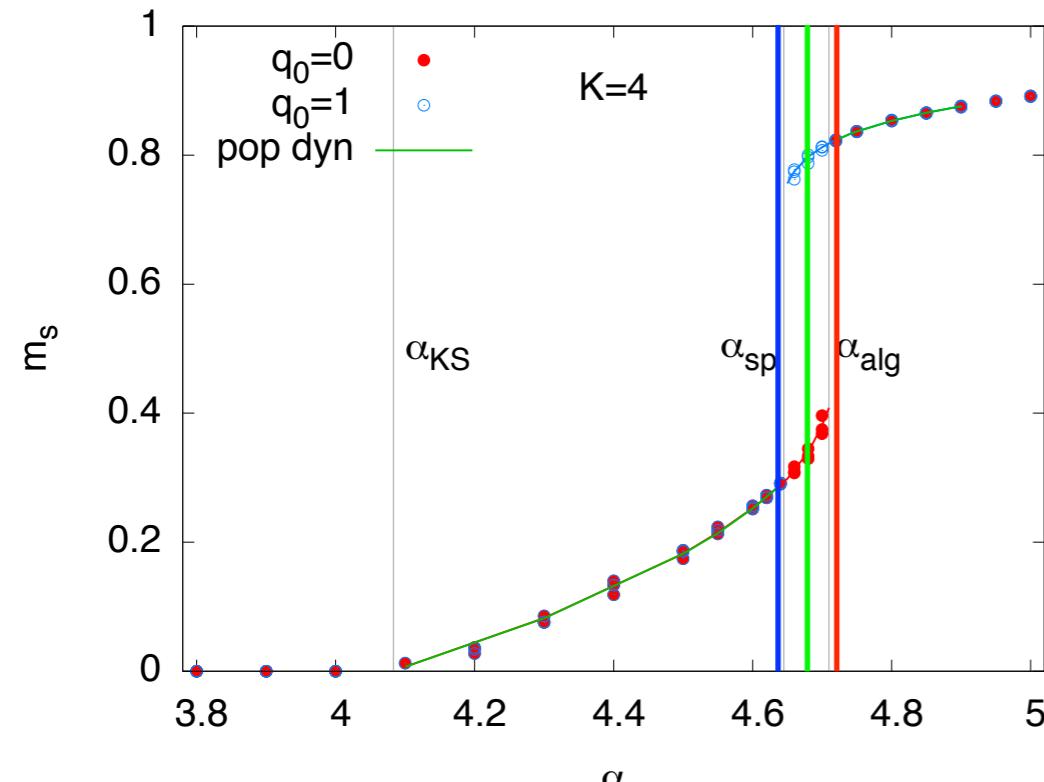
Finding the unstable fixed point via BP

Hard (impossible?) to do via PopDyn



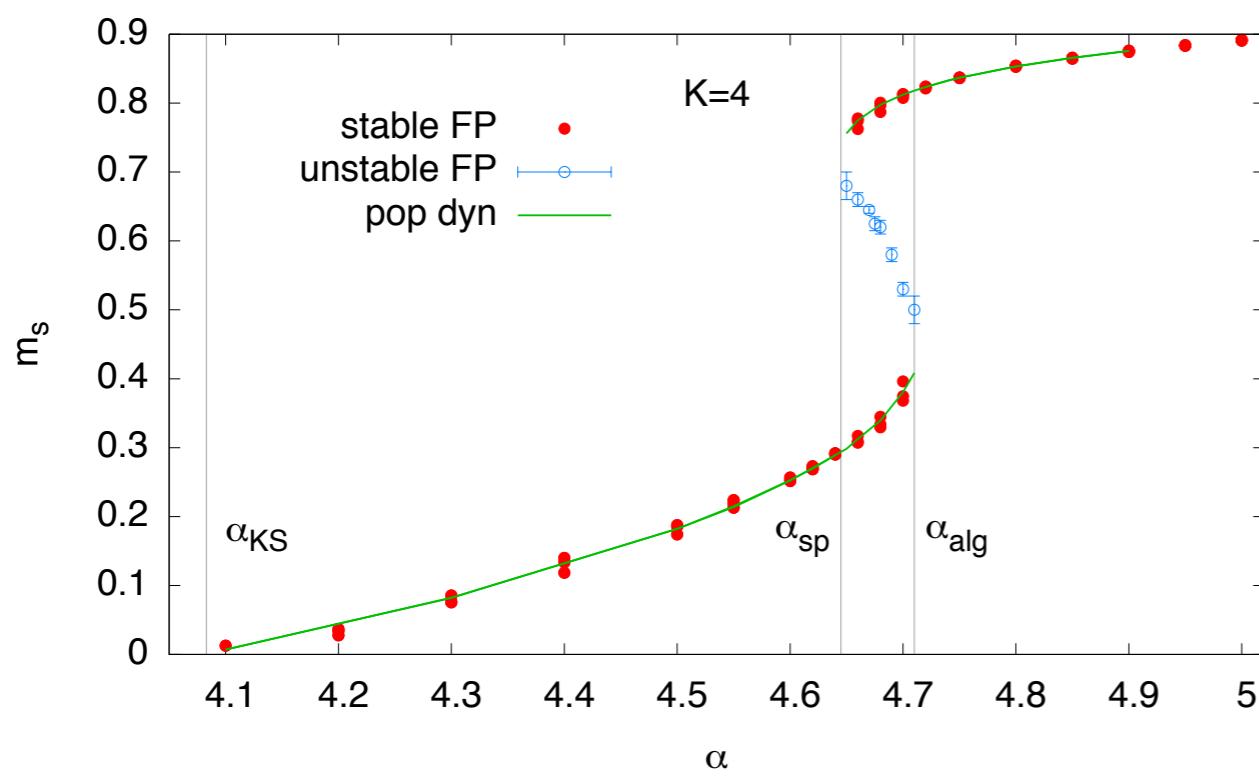
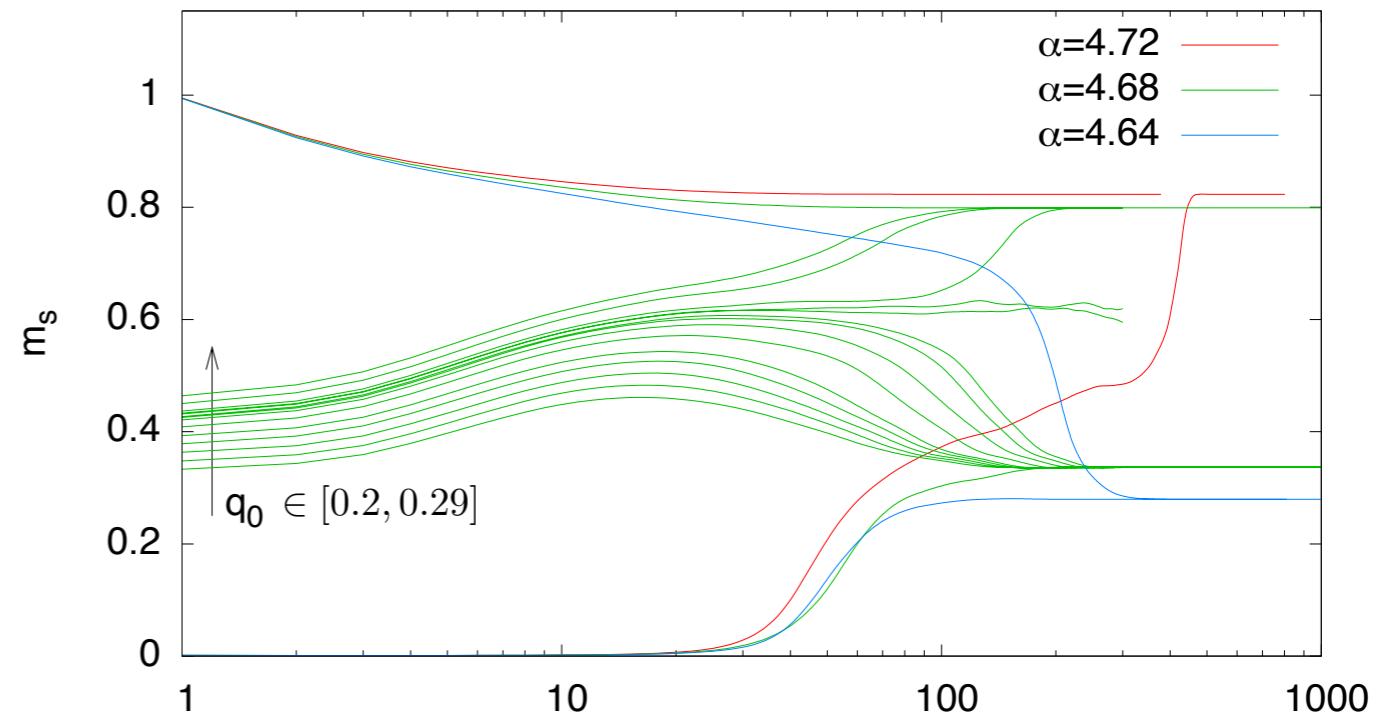
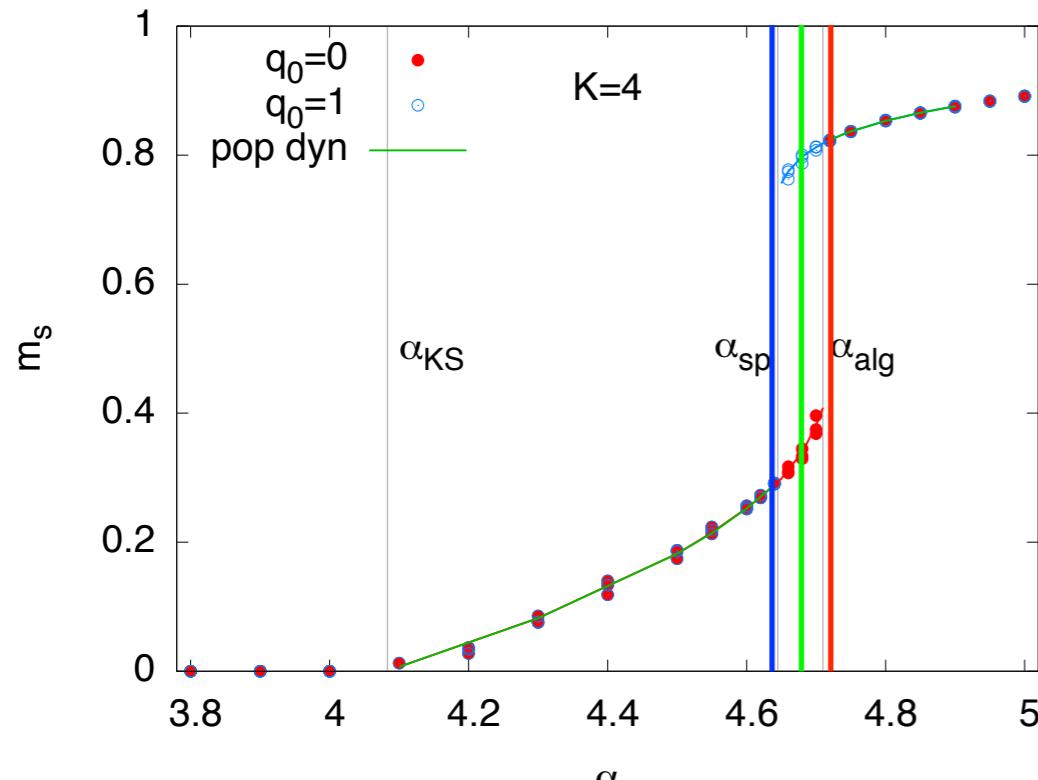
Finding the unstable fixed point via BP

Hard (impossible?) to do via PopDyn



Finding the unstable fixed point via BP

Hard (impossible?) to do via PopDyn

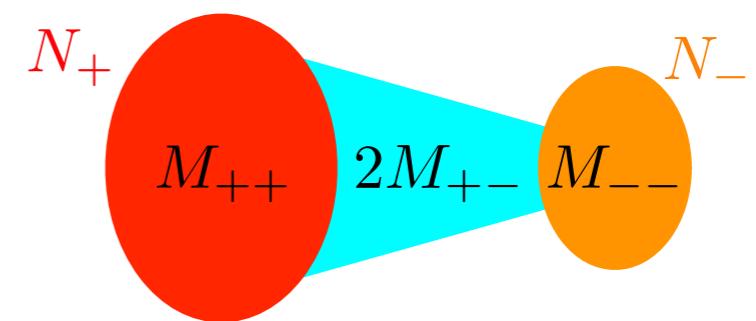


Asymmetric SBM

Application 1 in Guilhem's talk

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

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



- 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 \rightarrow Ising model on a sparse graph

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

prior
likelihood edges
likelihood non-edges

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

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

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

- 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

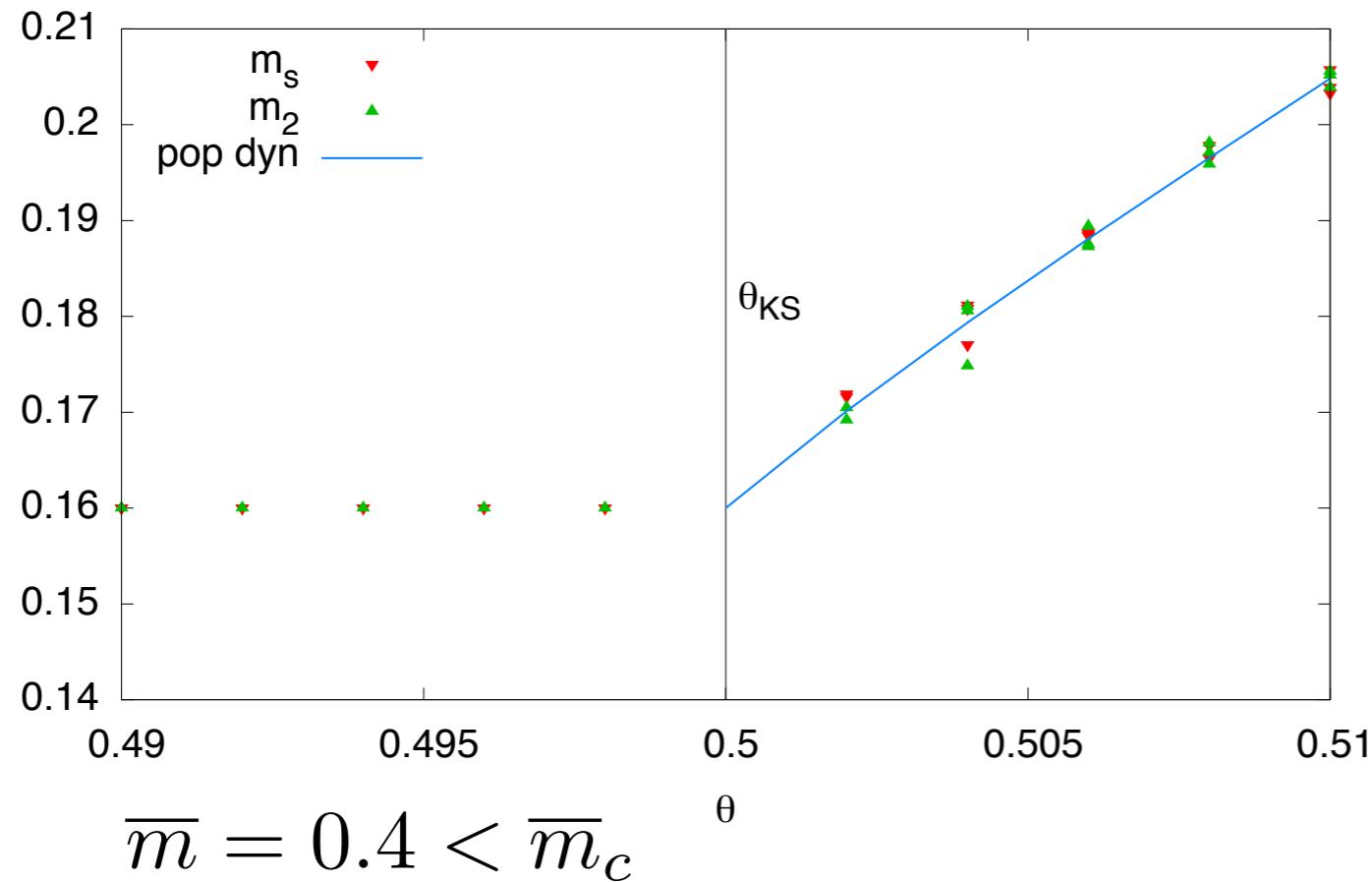
$$u_{i \rightarrow j}^{(t+1)} = \gamma u_{i \rightarrow j}^{(t)} + (1 - \gamma) \operatorname{atanh} \left[\tanh(J) \tanh \left(\tilde{H}_i^{(t)} + \sum_{k \in \partial i \setminus j} u_{k \rightarrow i}^{(t)} \right) \right]$$

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

- Local magnetizations (marginals) $m_i^{(t)} = \tanh \left(\tilde{H}_i^{(t)} + \sum_{k \in \partial i} u_{k \rightarrow i}^{(t)} \right)$
- Convergence criterion $|m_i^{(t+1)} - m_i^{(t)}| < 10^{-8} \quad \forall i$
- BP initialized close to a configuration $\underline{\tau}^0$ with

$$\mathbb{P}[\tau_i^0 = s] = q_0 \delta_{s, \sigma_i^*} + (1 - q_0) \frac{1 + \bar{m} s}{2}$$
- Uninformative fixed point has all messages $u_{i \rightarrow j}^* = K$
 It becomes locally unstable at $\theta_{\text{KS}} = 1/\sqrt{d}$

Asymmetric SBM (d=4)

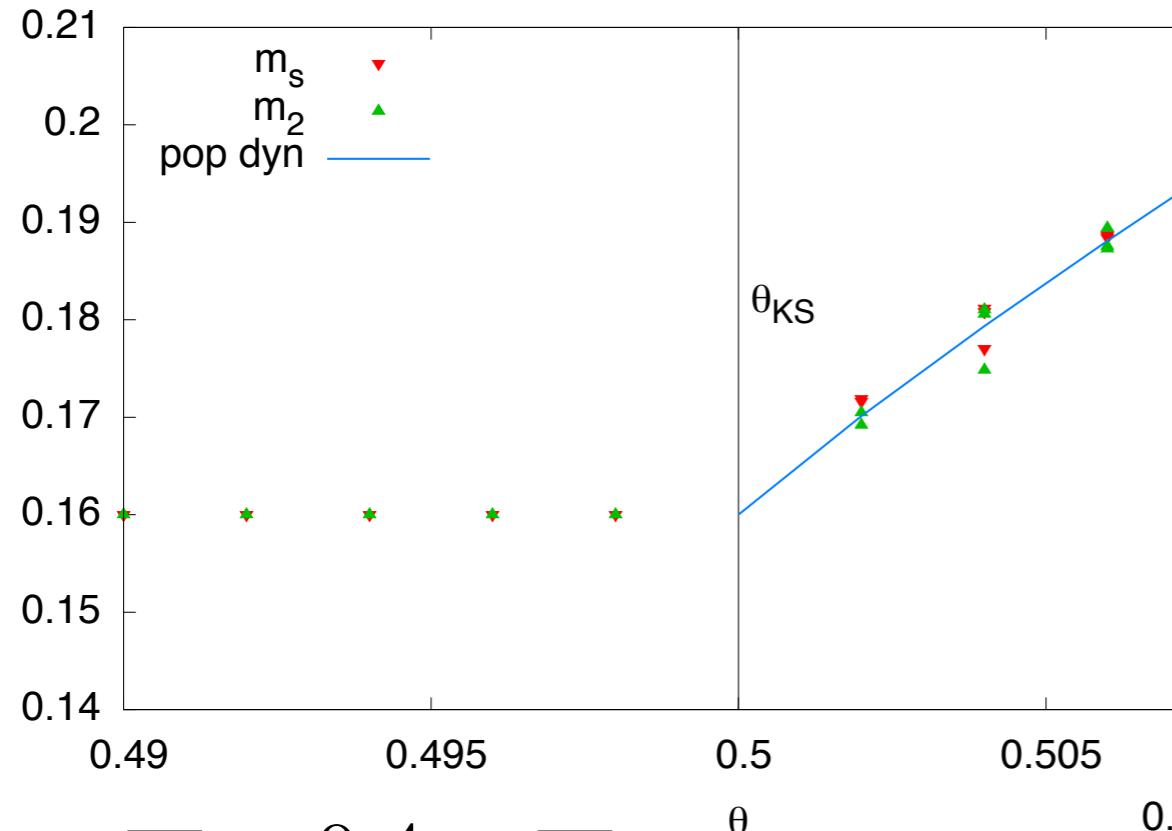


3 samples of
size $N=10^7$

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

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

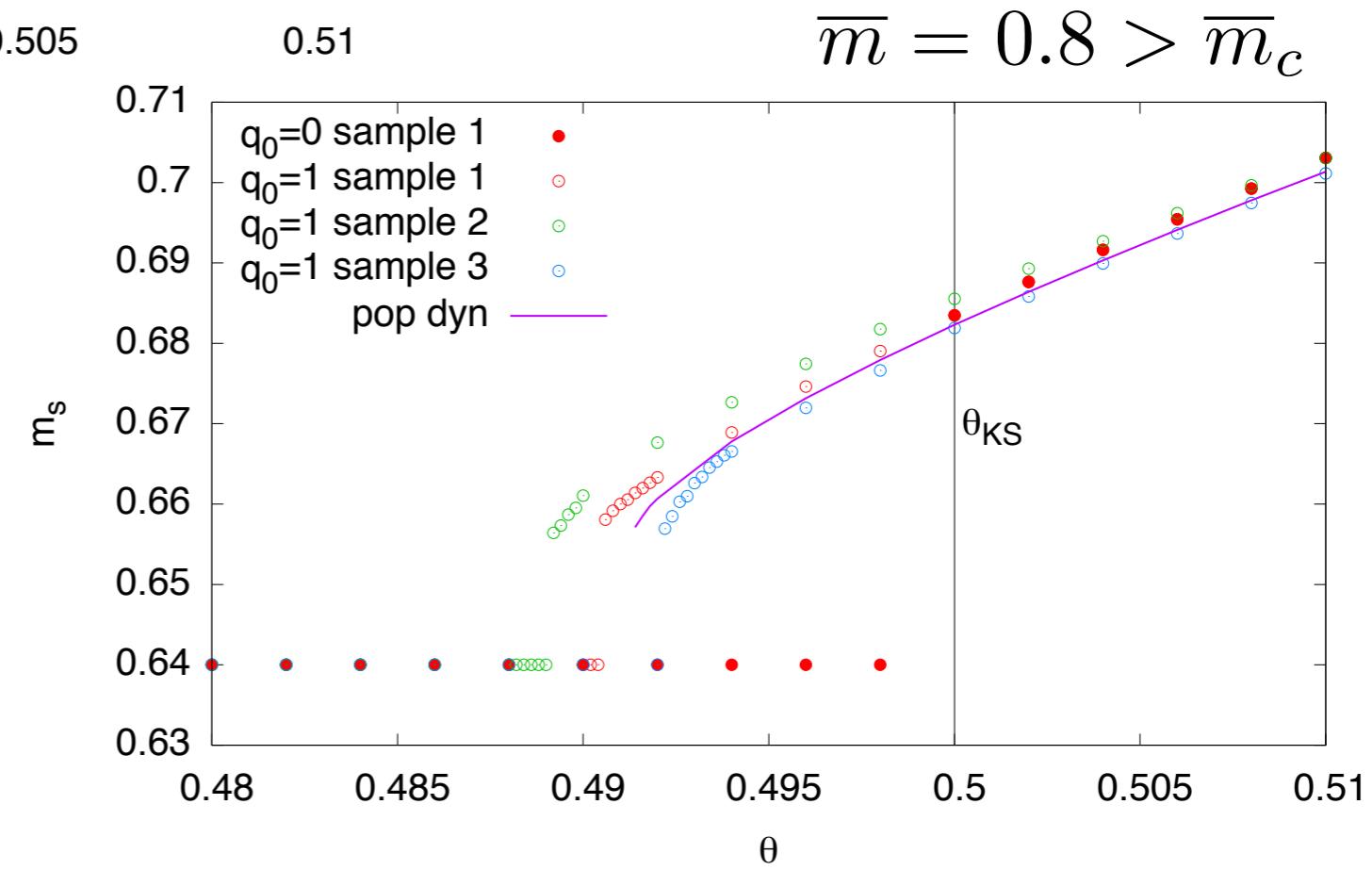
Asymmetric SBM (d=4)



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

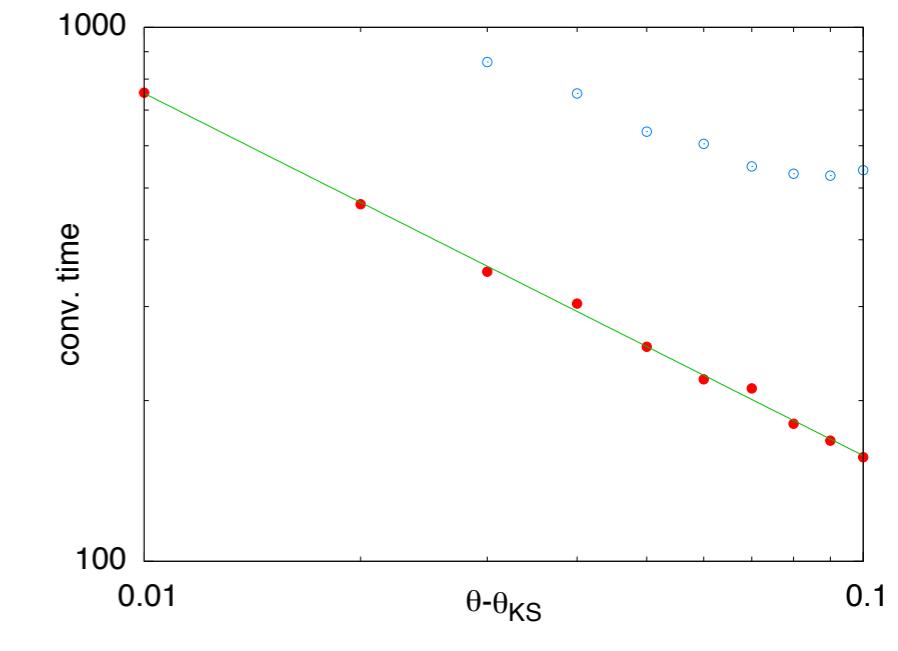
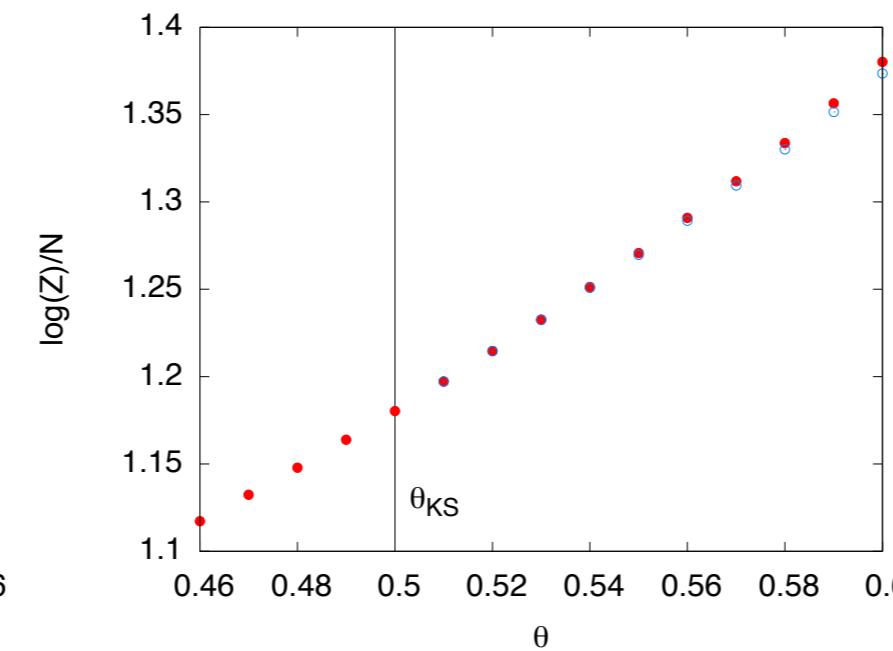
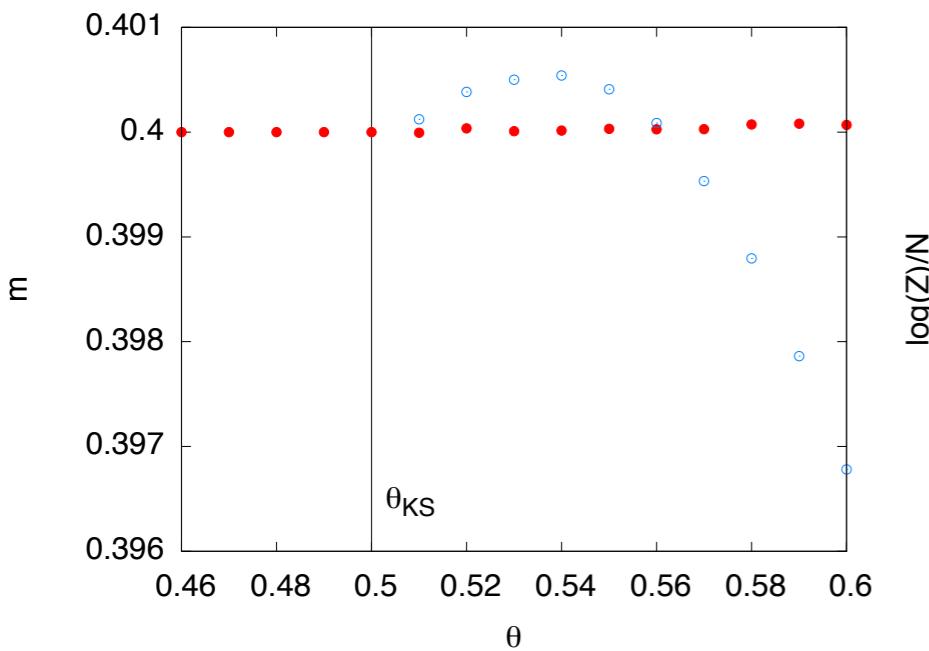
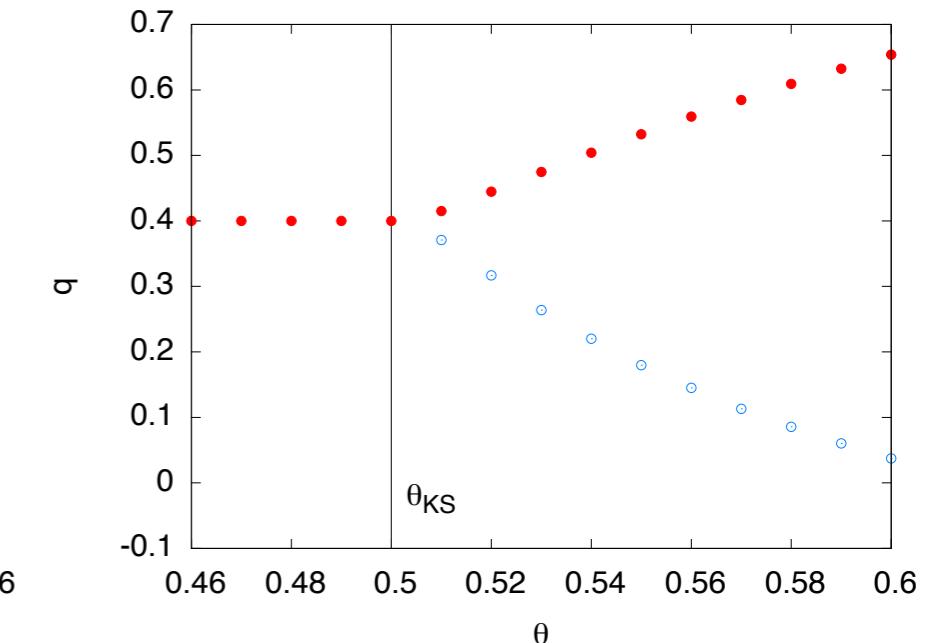
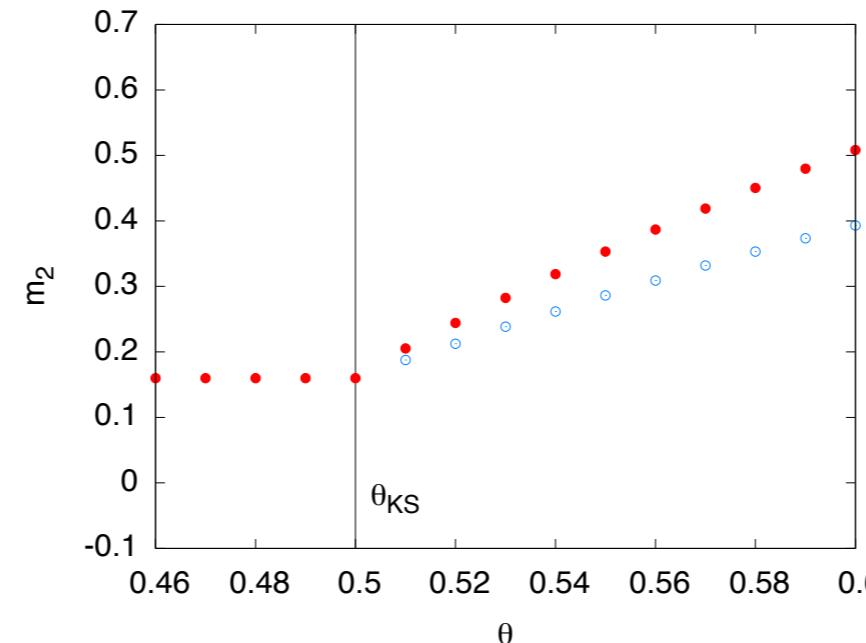
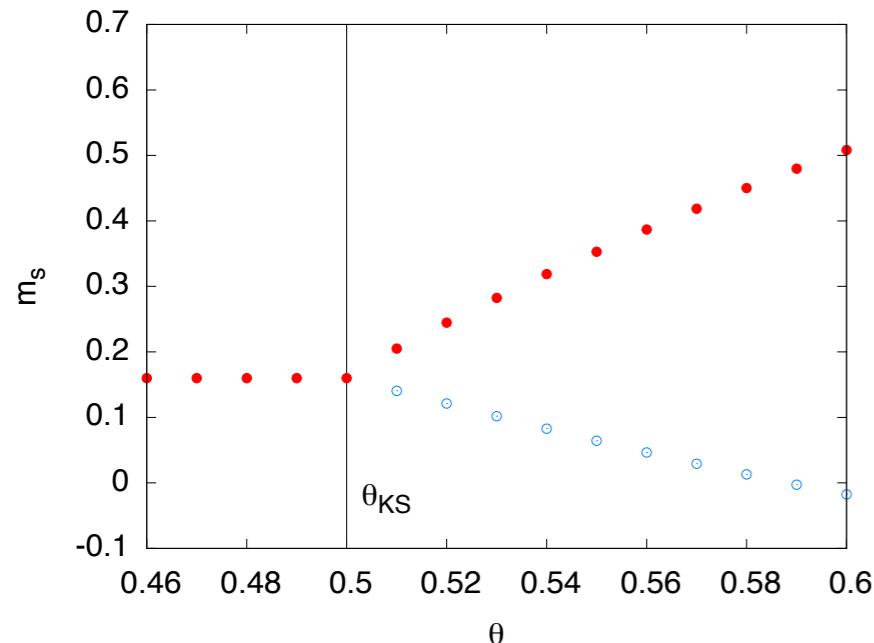
critical asymmetry
 $\bar{m}_c = 1/\sqrt{3} \simeq 0.577$

3 samples of
size $N=10^7$



Asymmetric SBM: more BP fixed points

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

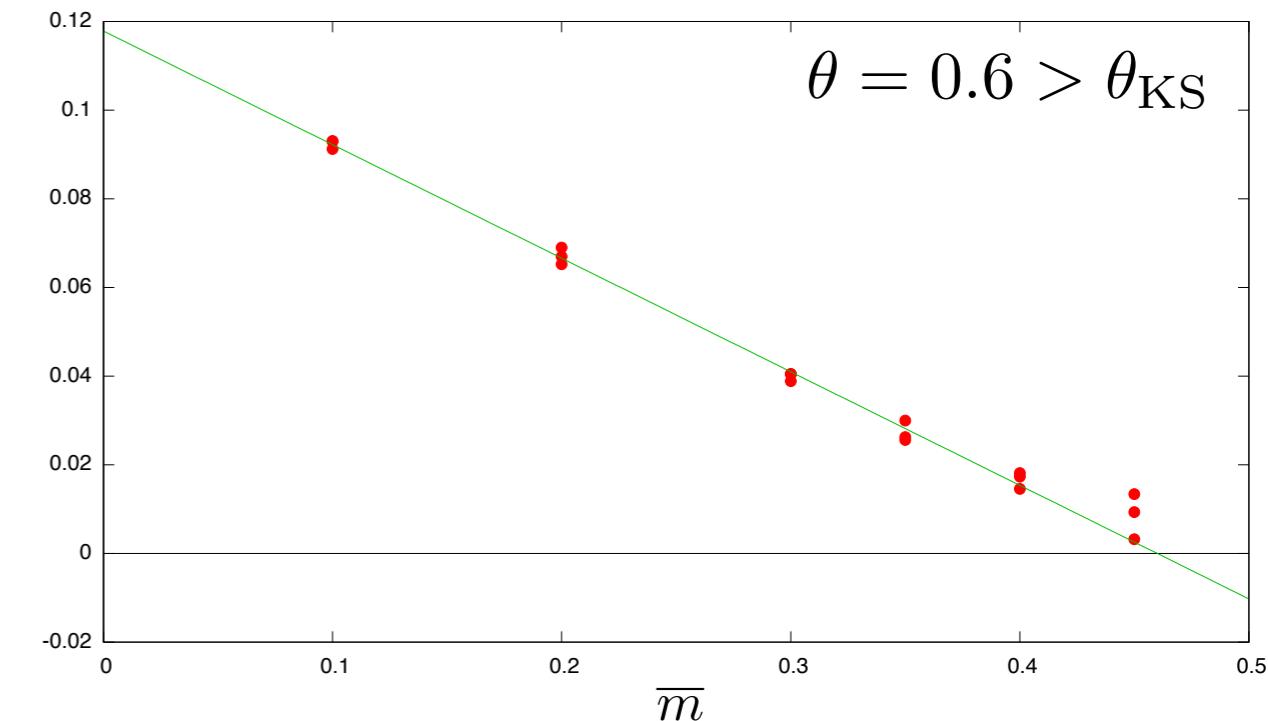
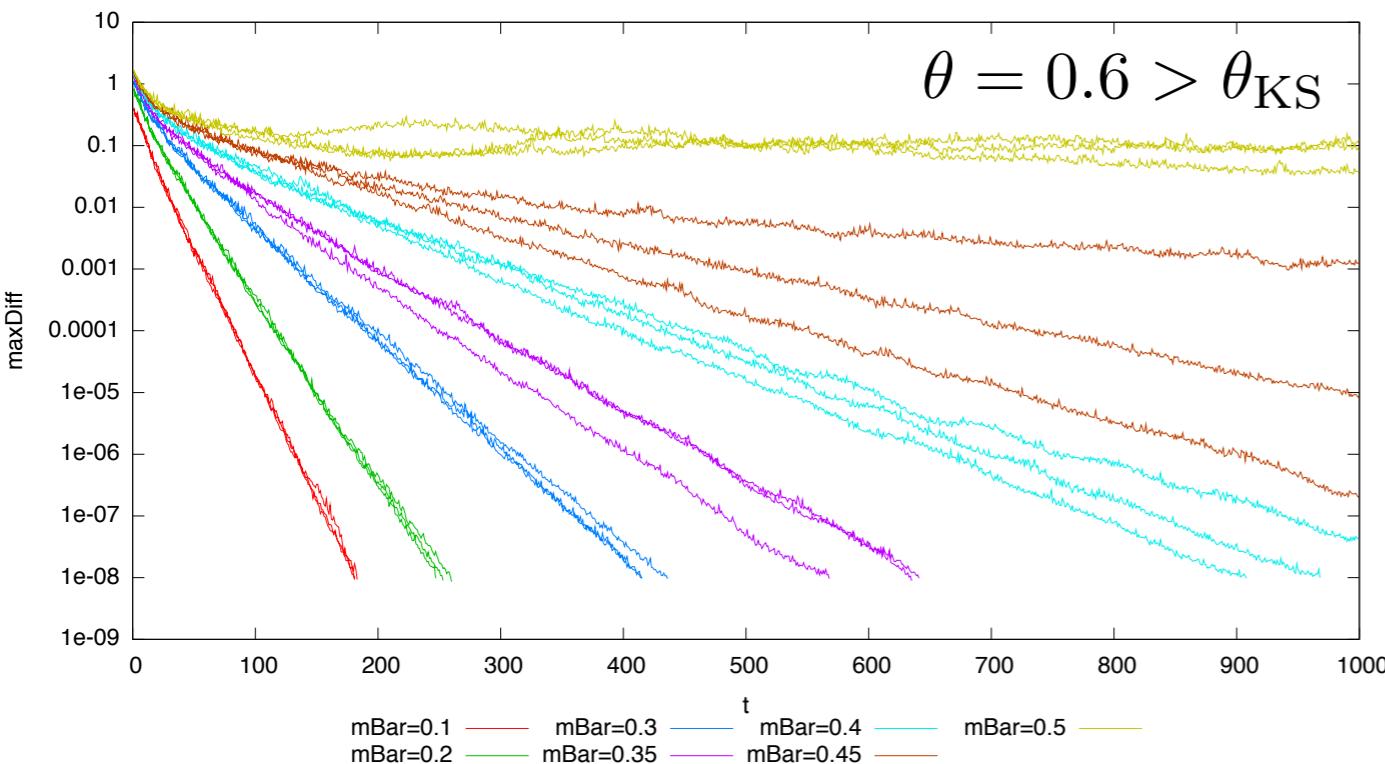


dominating fixed point

sub-dominating fixed point

Asymmetric SBM: RSB states?

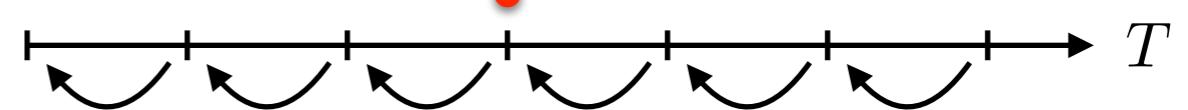
- The sub-dominating fixed point does not satisfy the Bayes/Nishimori condition
 - > may undergo a transition to a RSB phase
 - > BP does not converge anymore...



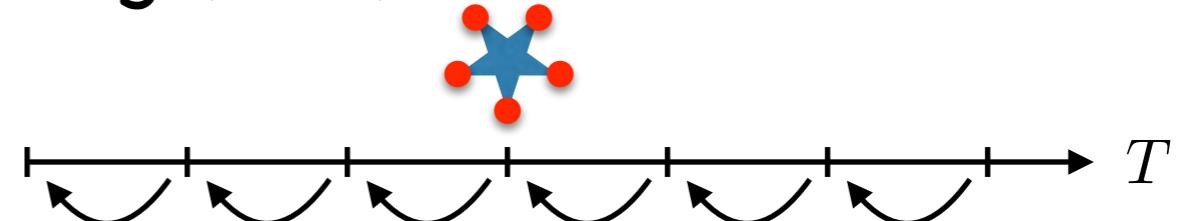
- From the algorithmic/dynamic point of view the asymmetric SBM seems more complex than expected

Monte Carlo based algorithms

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



- Replicated Simulated Annealing (RSA)

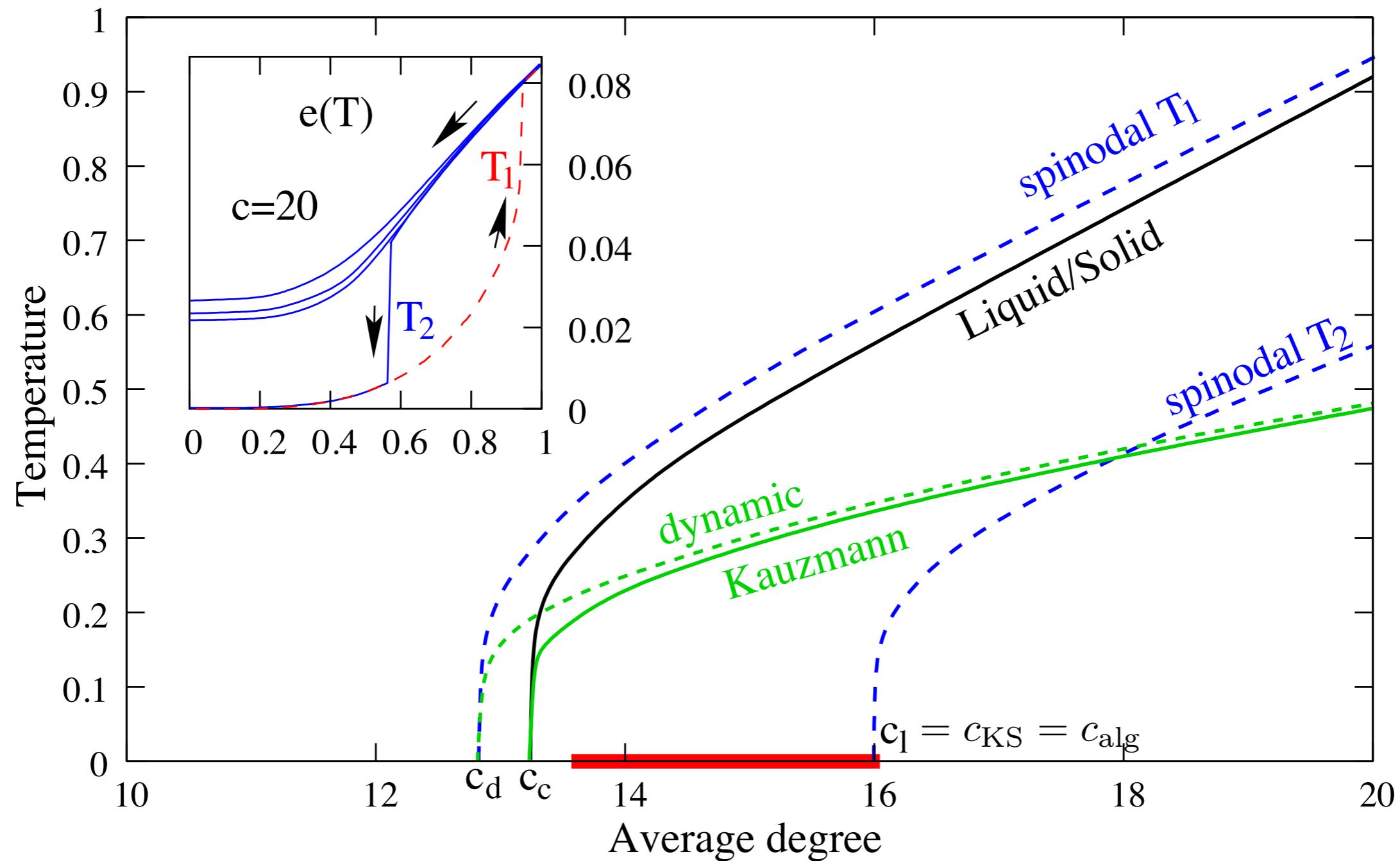


- Parallel Tempering (PT)



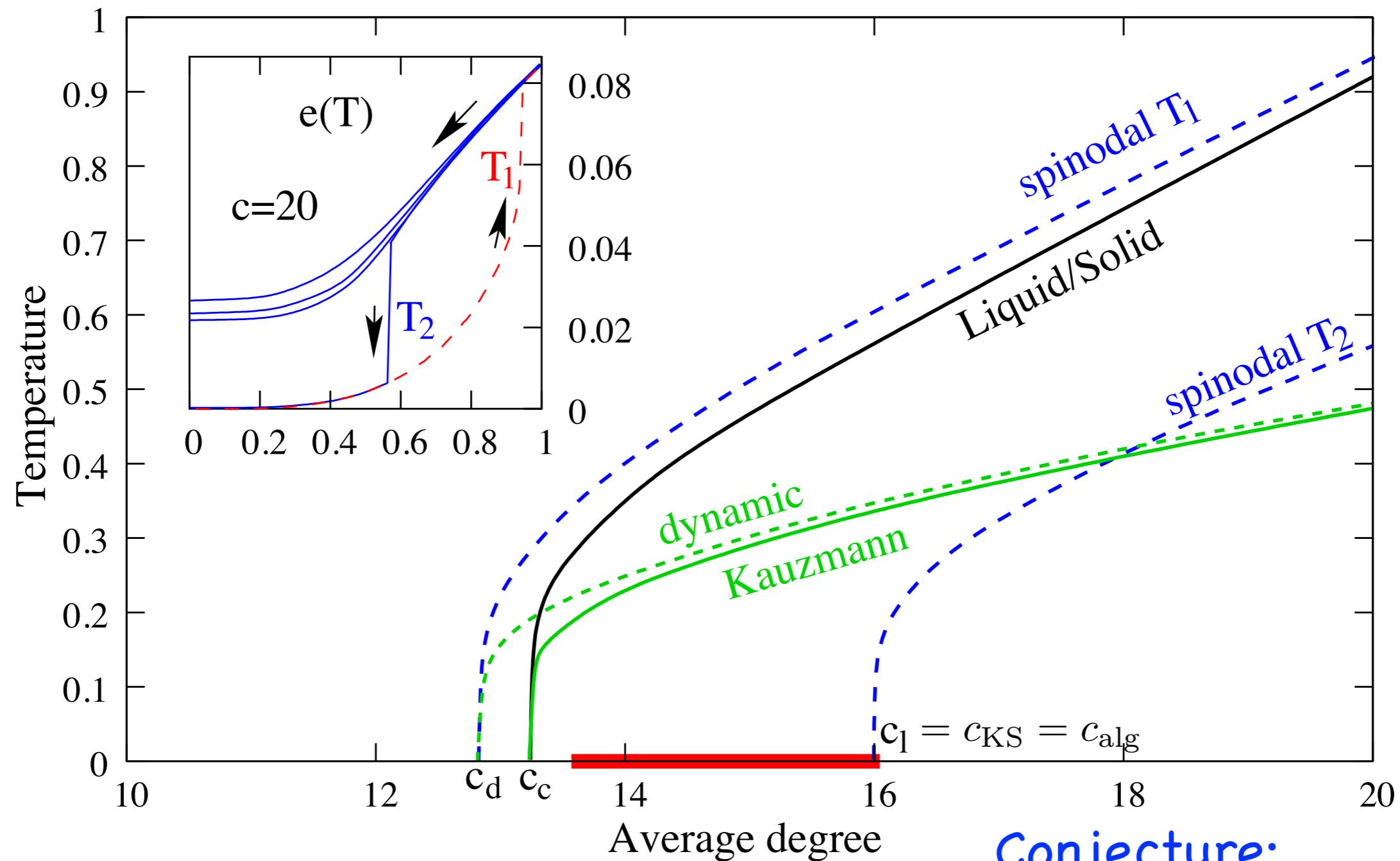
SA in planted models ($q=5$ coloring)

Krzakala & Zdeborova, PRL 2009



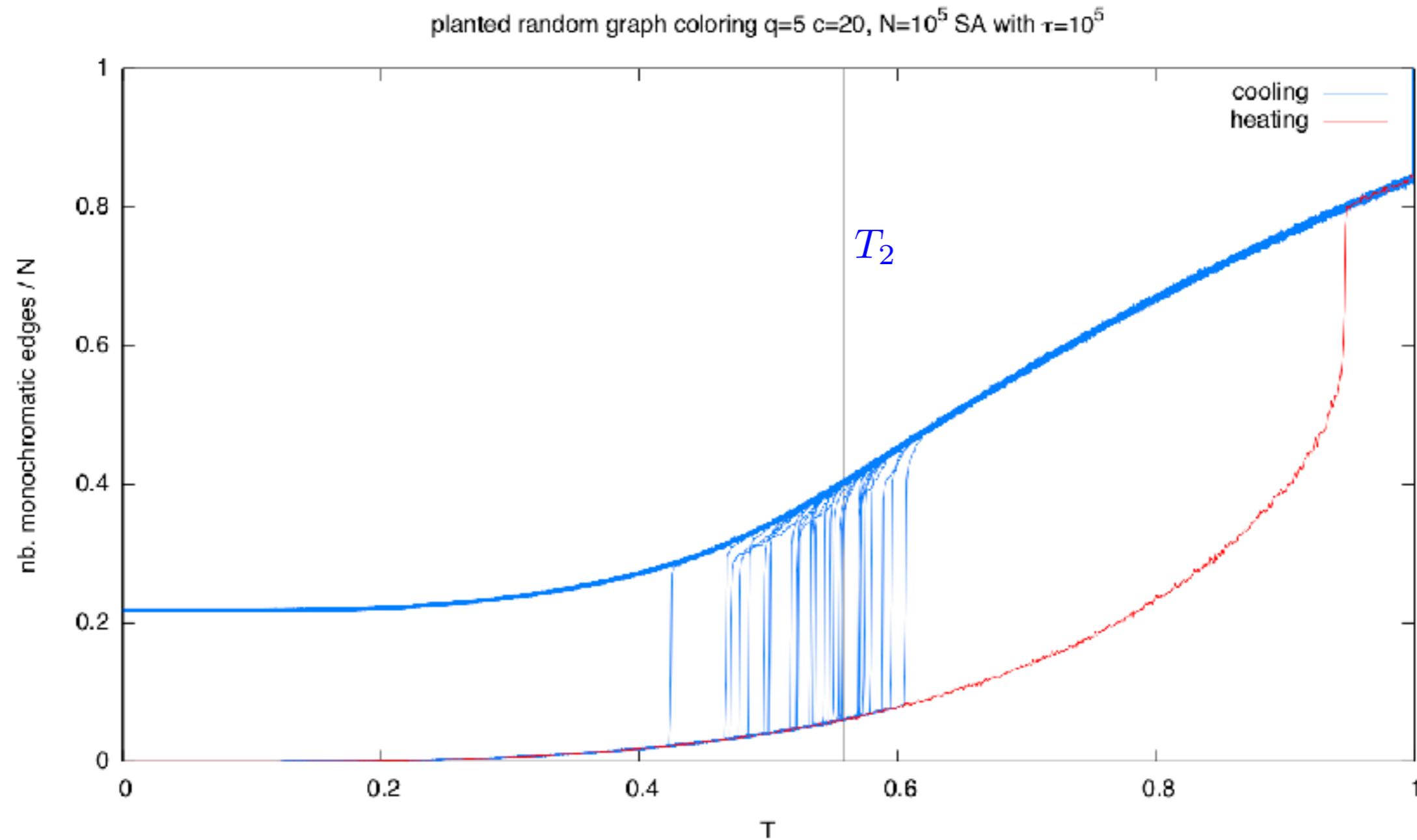
SA in planted models ($q=5$ coloring)

Krzakala & Zdeborova, PRL 2009

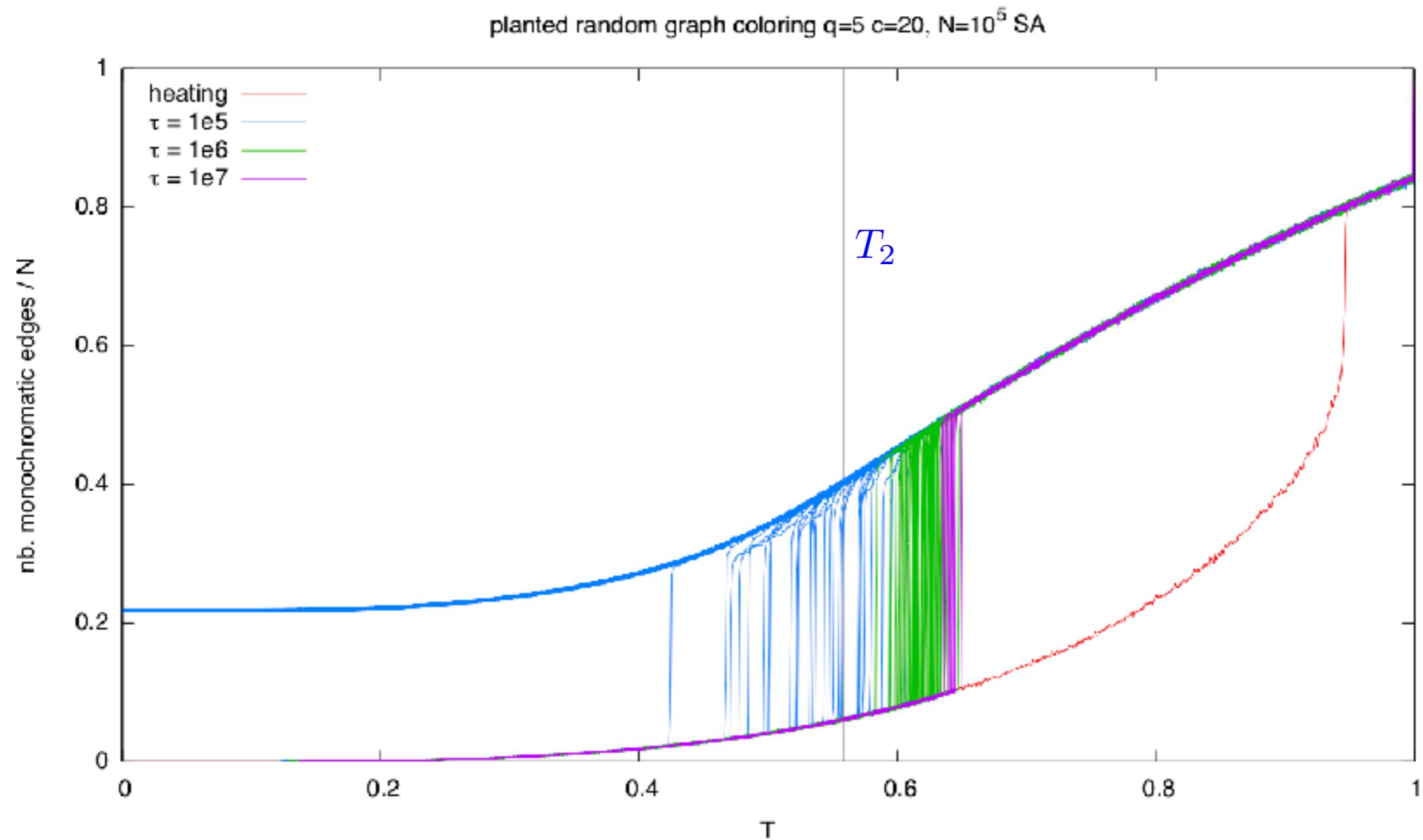


Conjecture:
SA works for $c > c_{KS}$

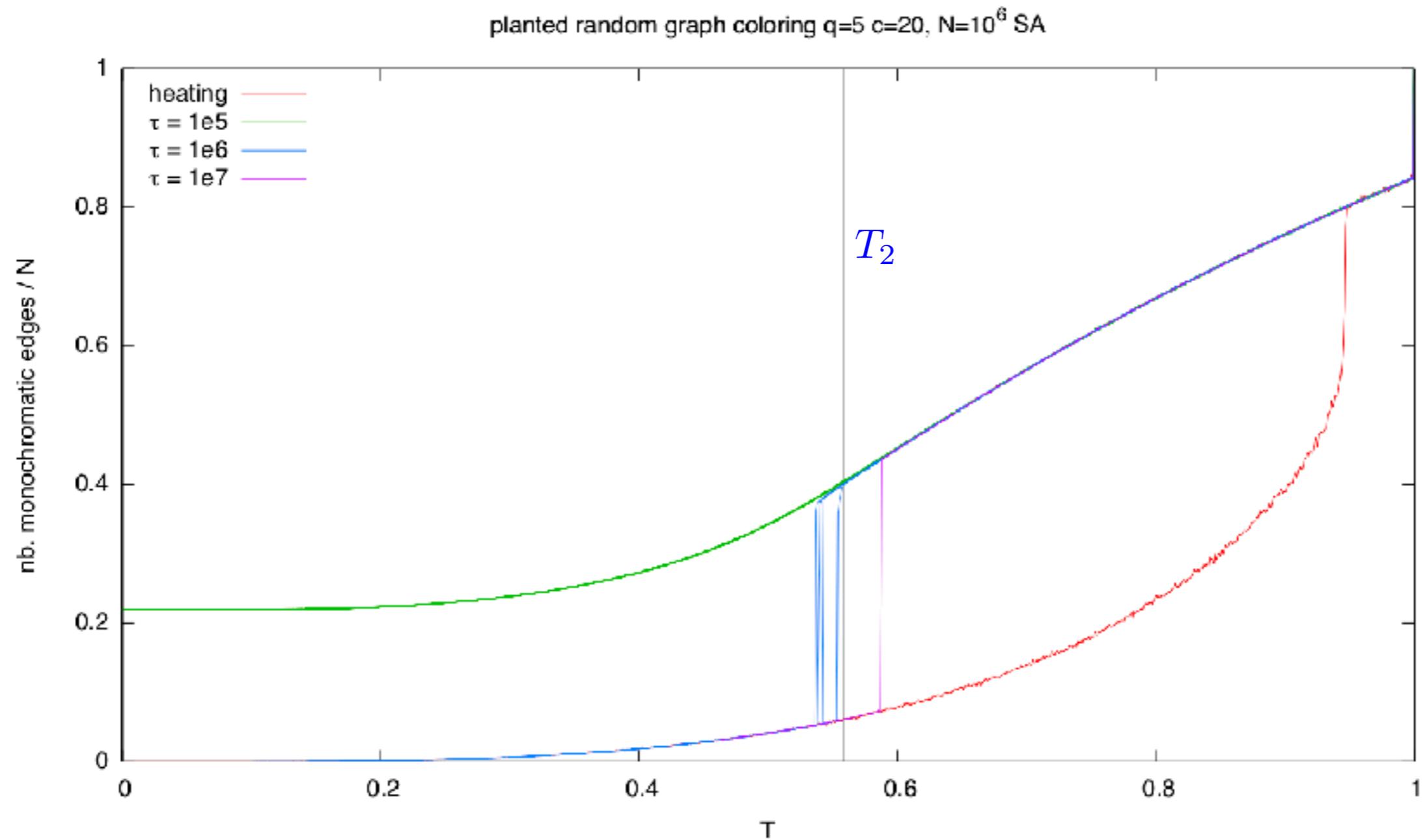
SA in planted models ($q=5$ coloring)



SA in planted models ($q=5$ coloring)

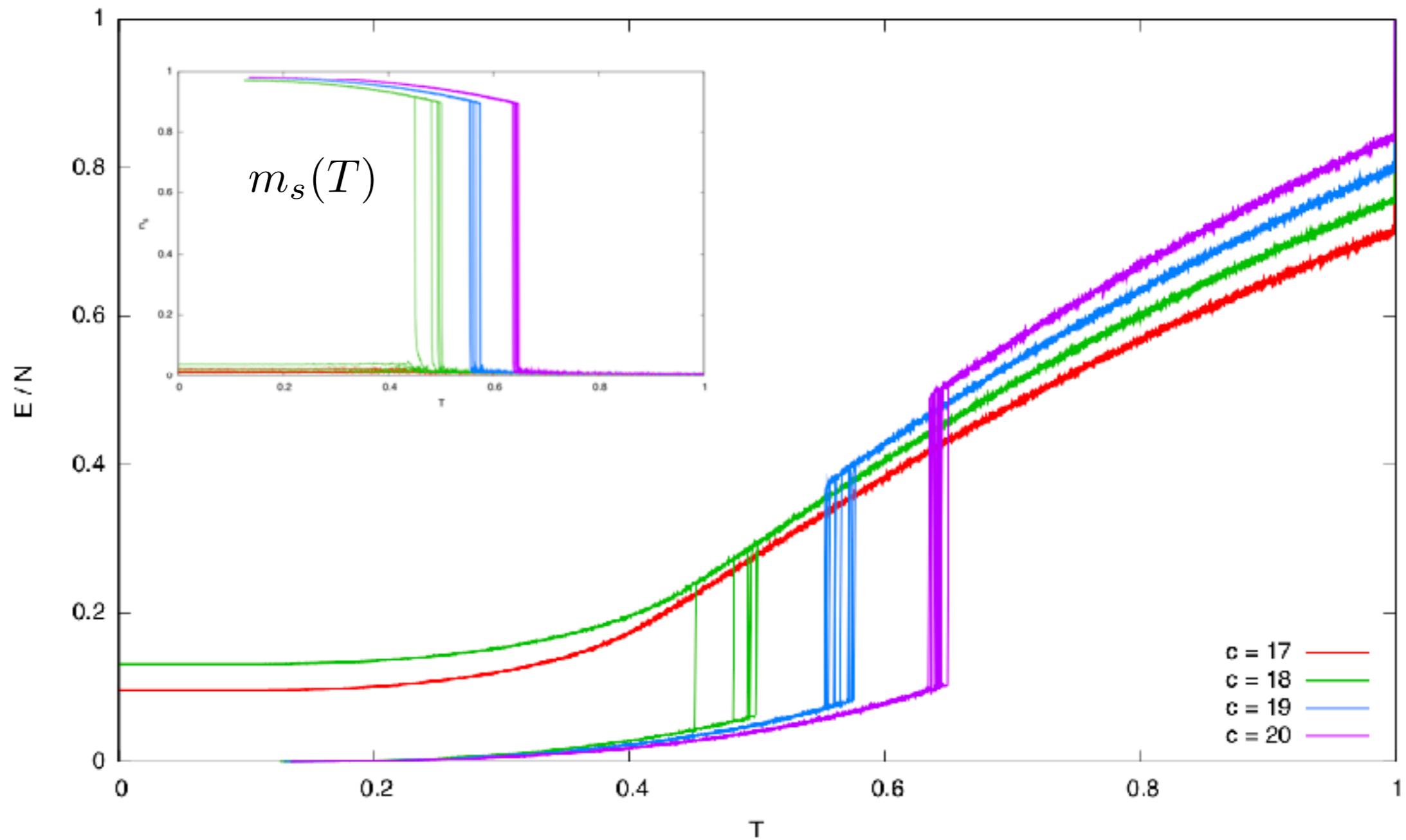


SA in planted models ($q=5$ coloring)

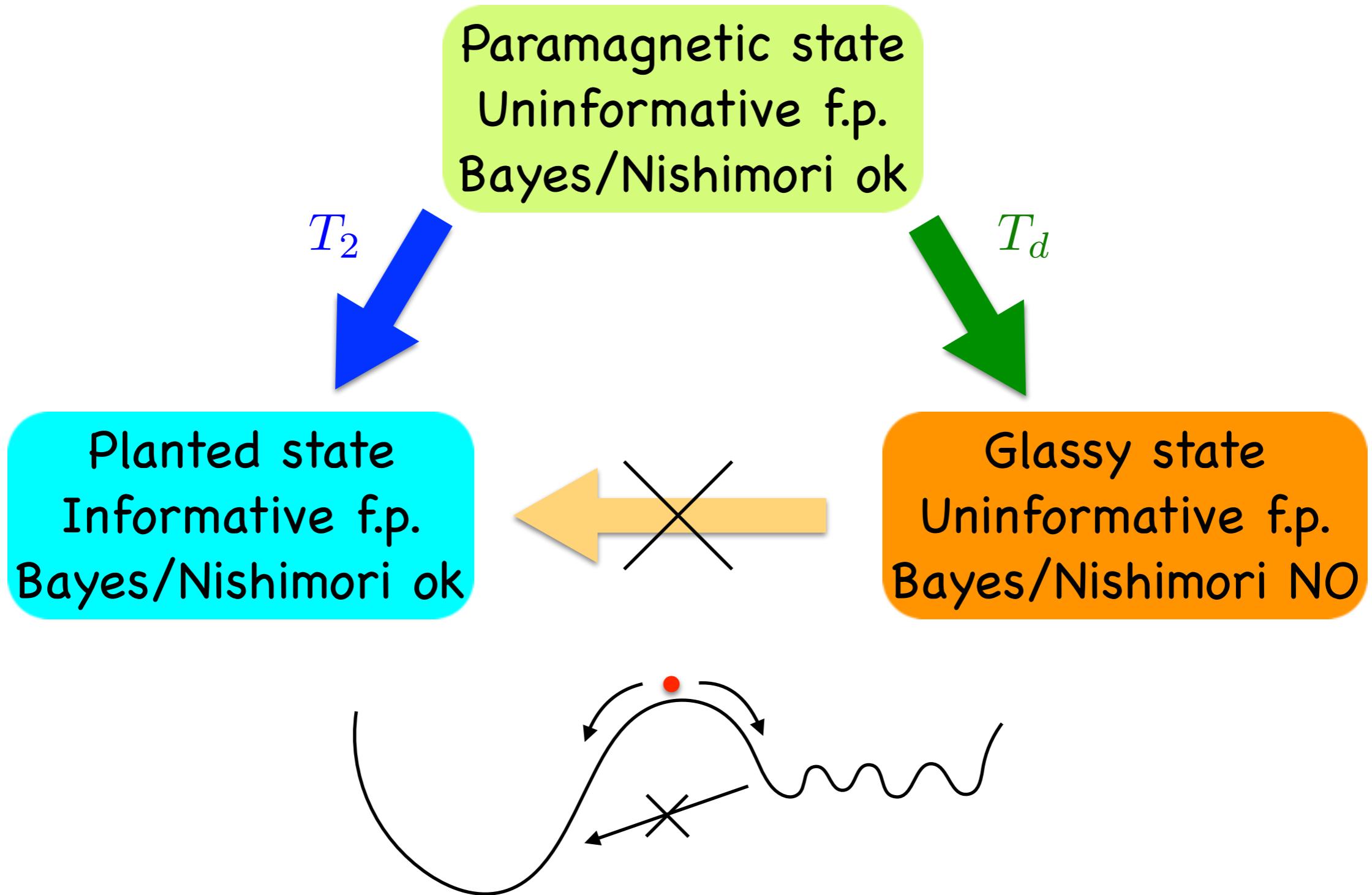


SA in planted models ($q=5$ coloring)

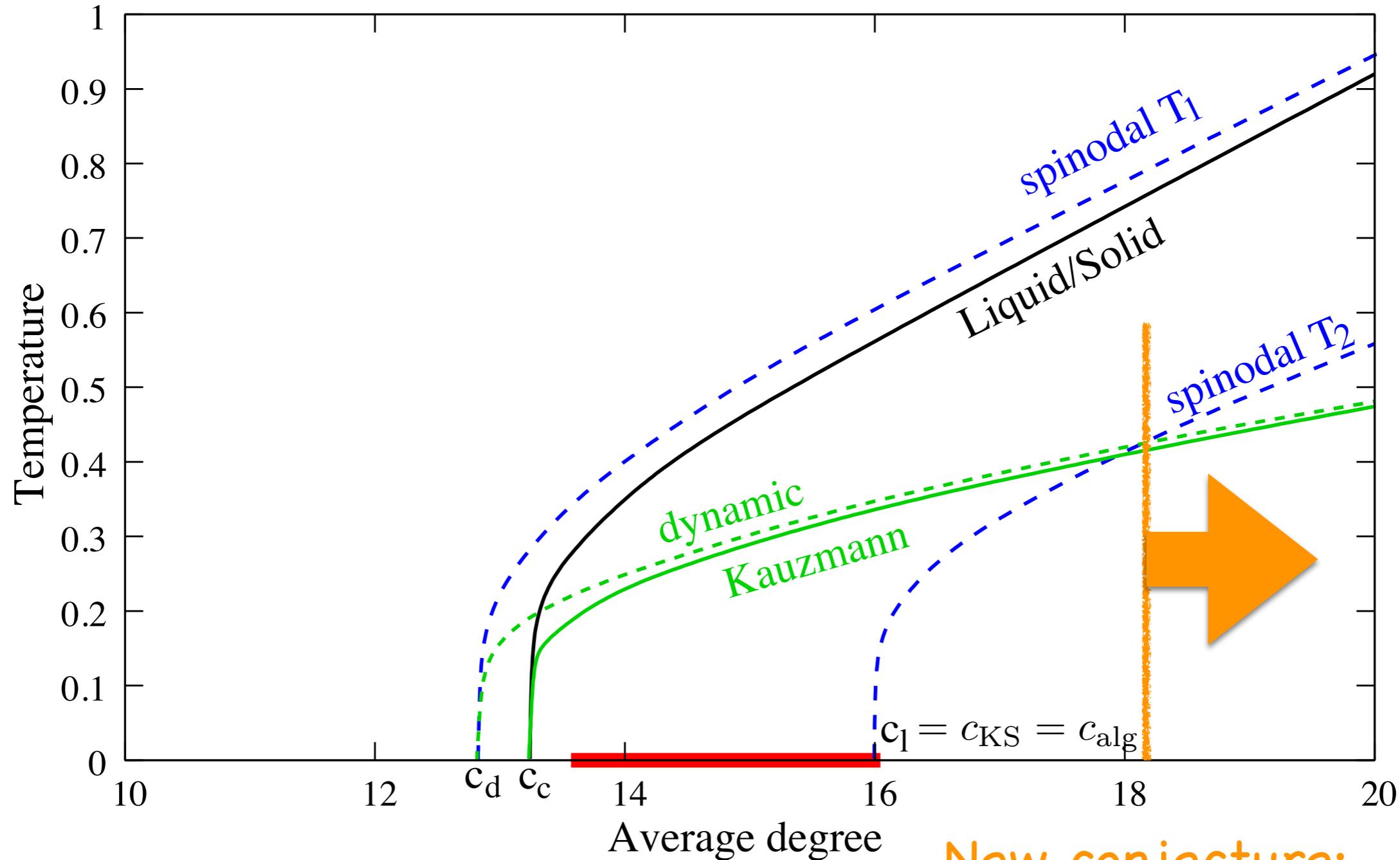
$$N = 10^5 \quad \tau = 10^7$$



Picture for SA in planted models



SA in planted models ($q=5$ coloring)



New conjecture:
SA works if $T_2 > T_d$

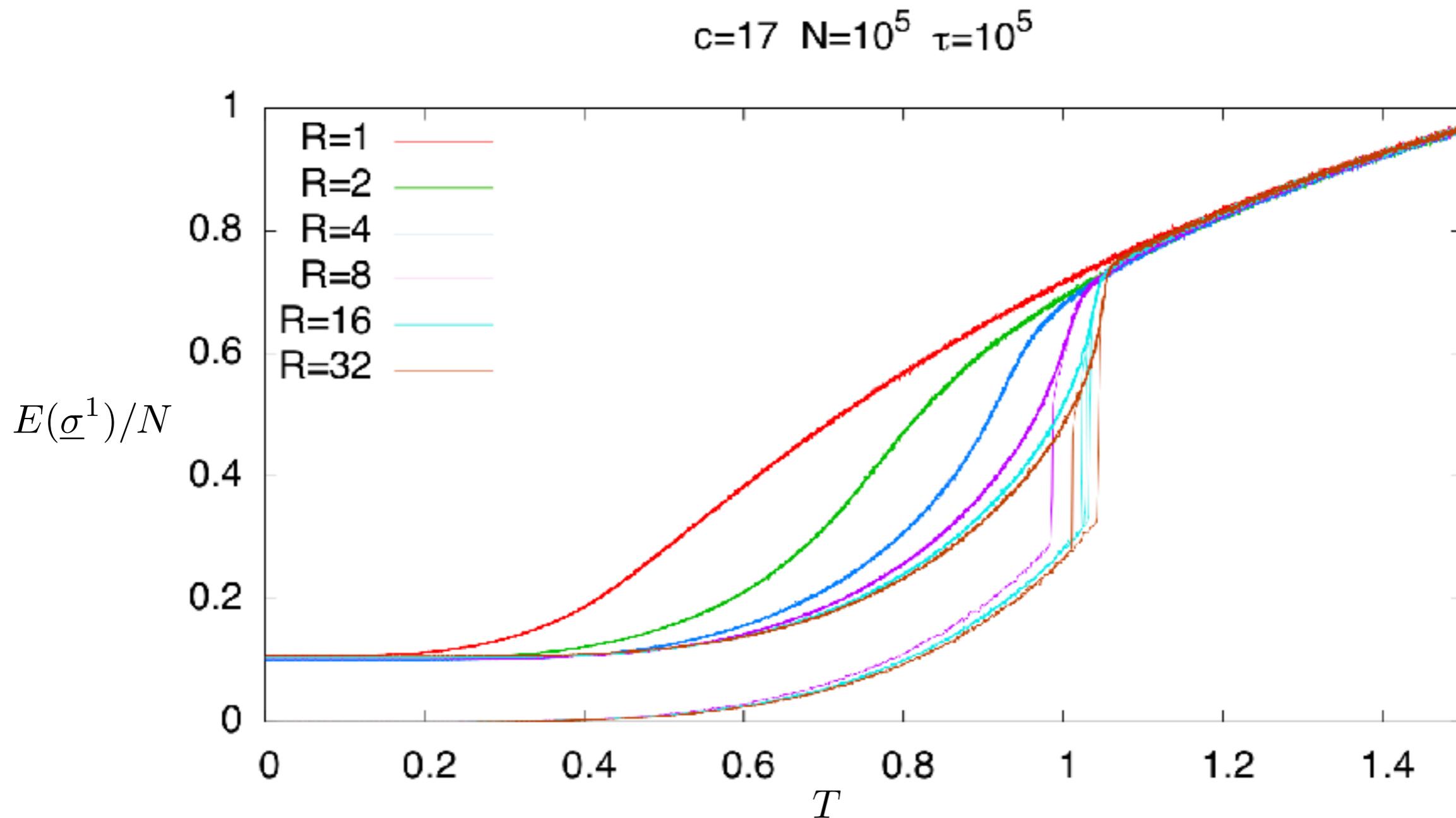
Replicated SA

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

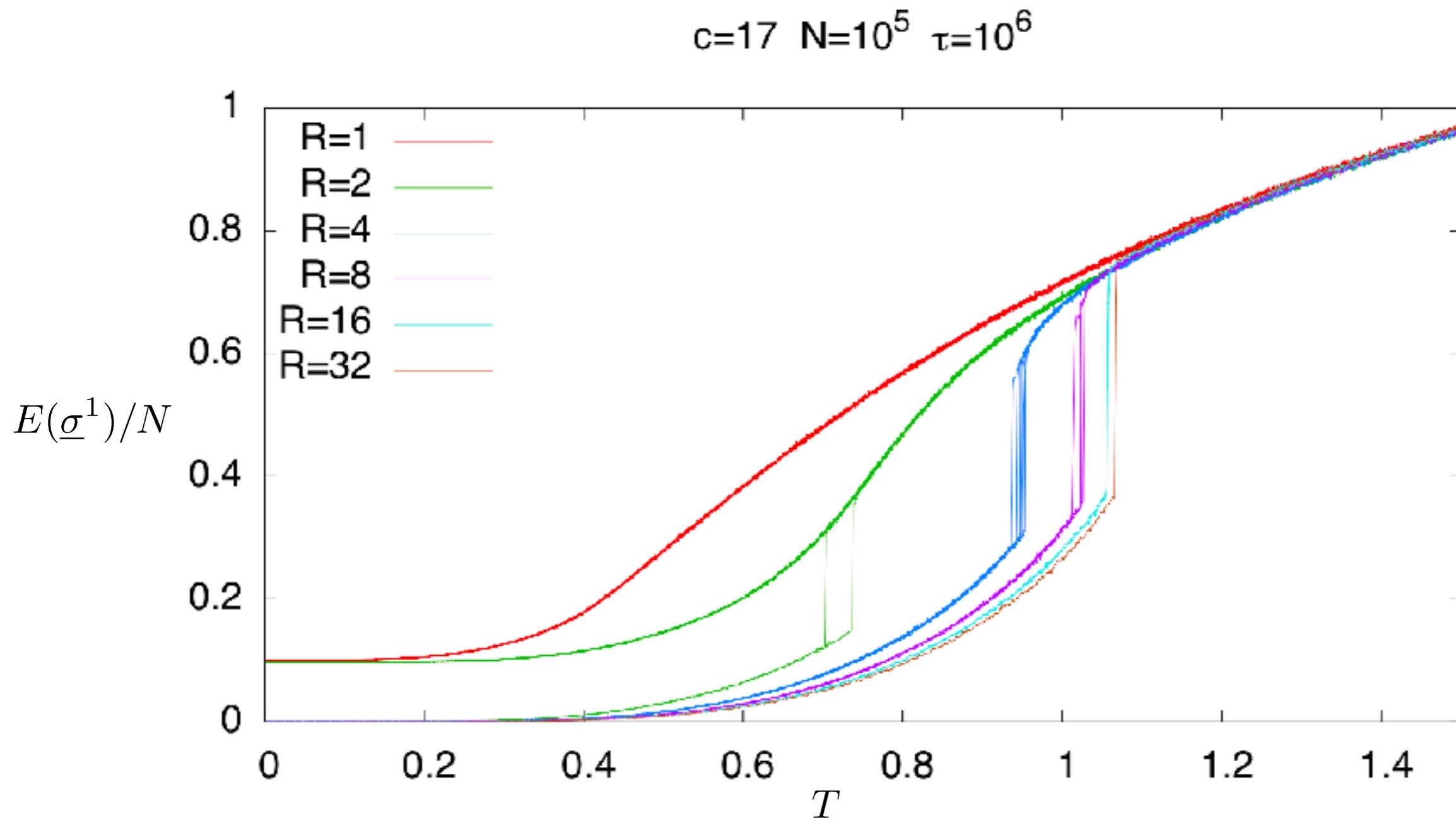
$$\sum_a E(\underline{\sigma}^a) - \frac{\gamma}{R} \sum_{a < b} \sum_i \delta_{\sigma_i^a, \sigma_i^b}$$

- $\gamma = 1$ in all next plots

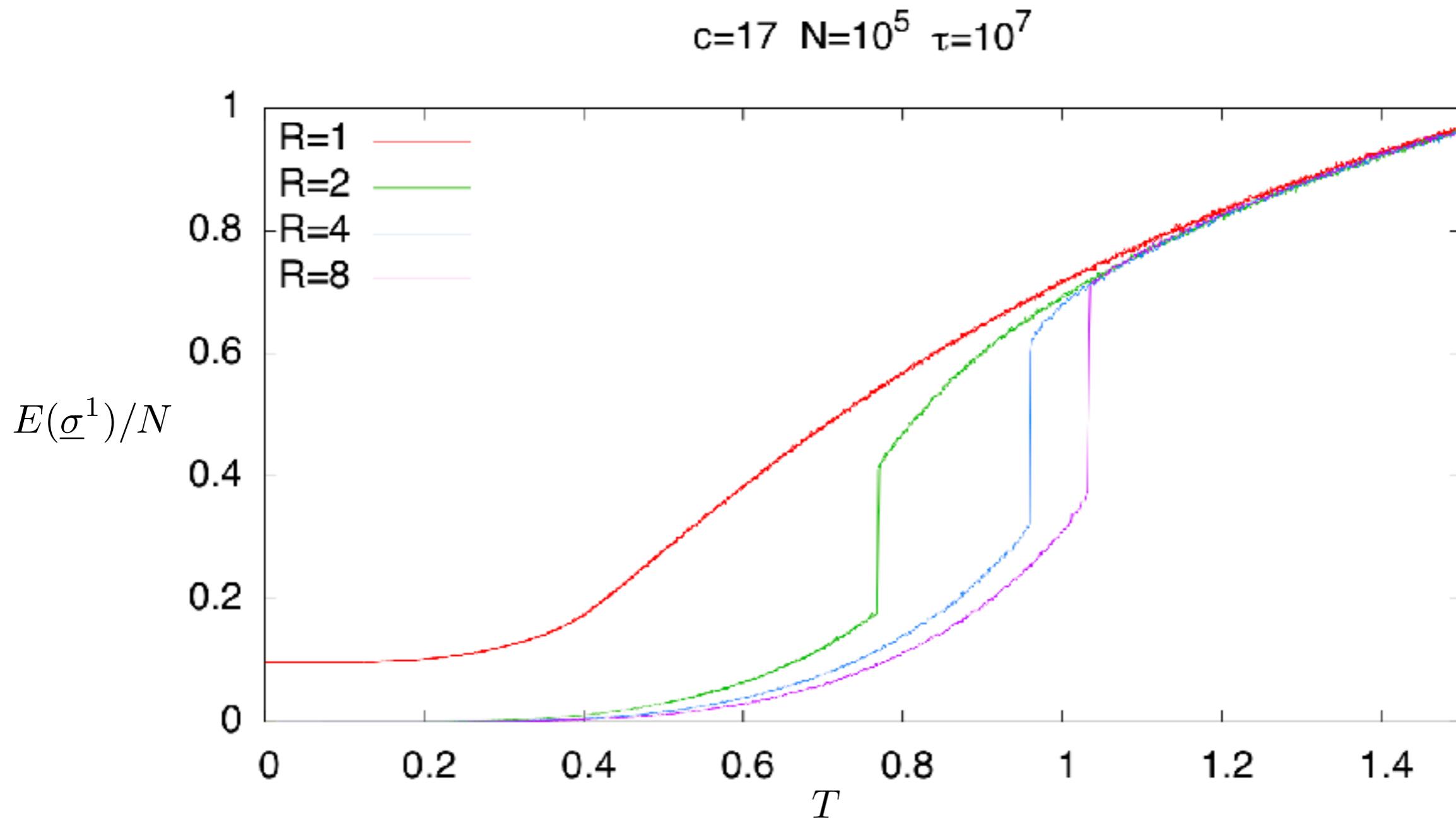
RSA in planted random coloring ($q=5$)



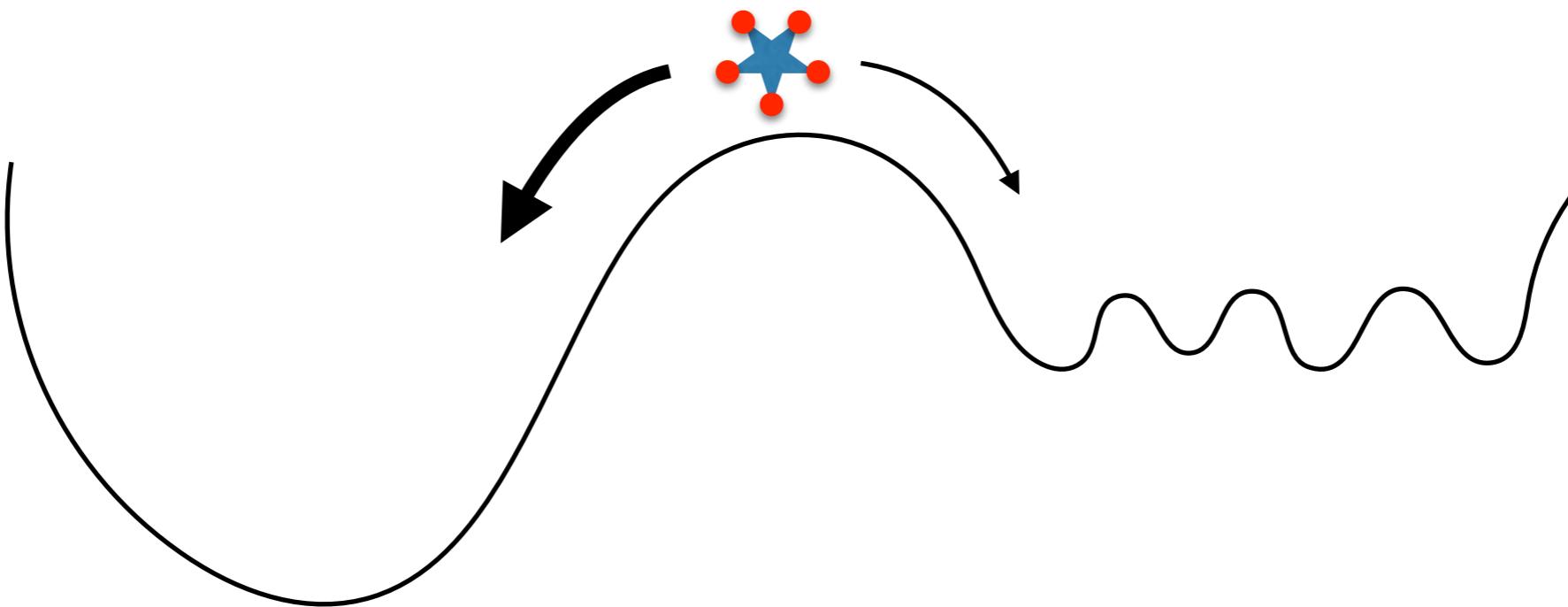
RSA in planted random coloring ($q=5$)



RSA in planted random coloring ($q=5$)



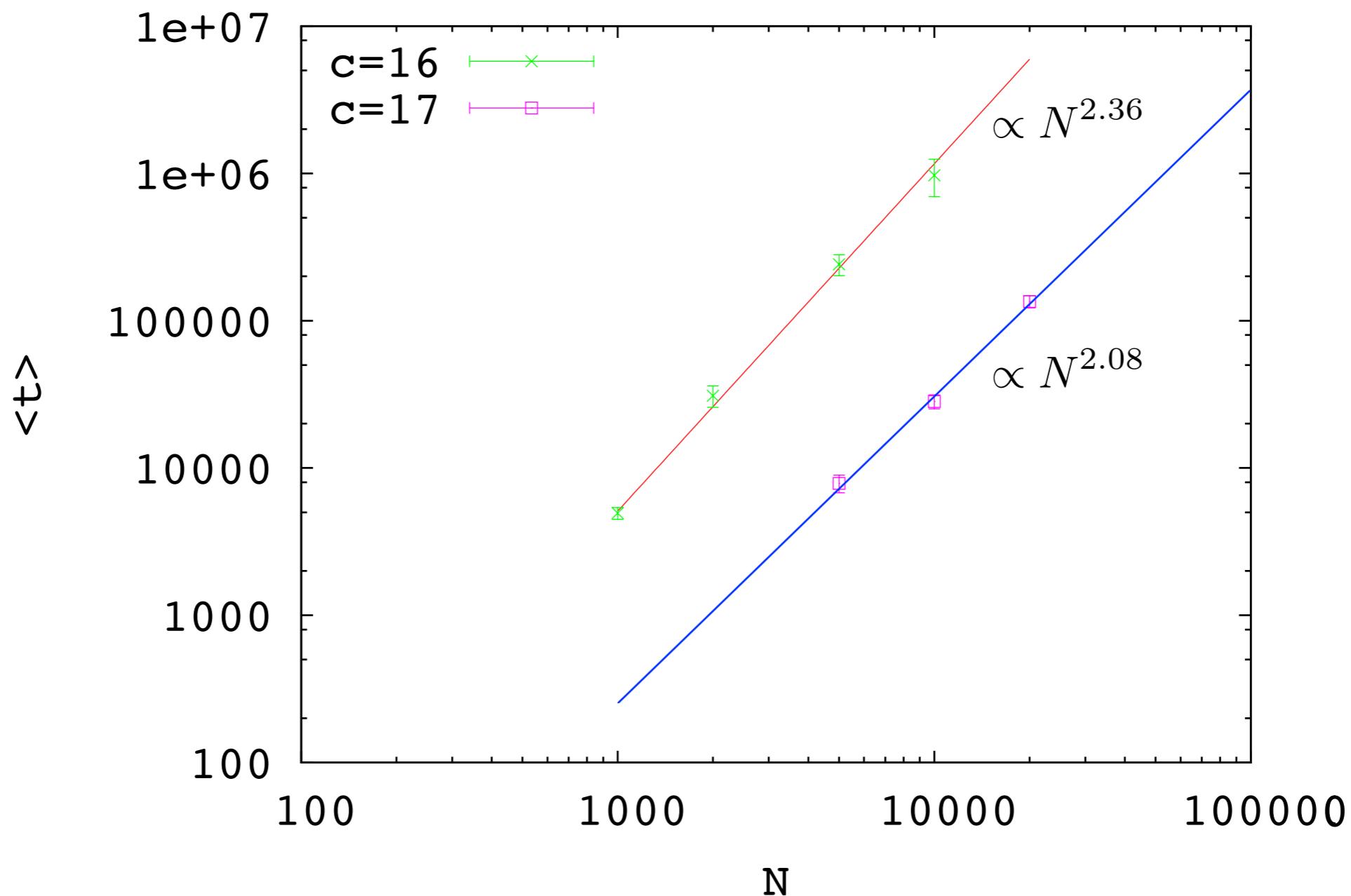
Picture for RSA in planted models



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 = $30 * \langle t \rangle$ (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_{alg} when $T_2 > T_d$
 - Replicated Simulated Annealing -> starting from c_{alg}
 - Parallel Tempering -> even below c_{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!