Phase Transitions in Semidefinite Relaxations
(a fast & robust algorithm for community detection)

Federico Ricci-Tersenghi
(Sapienza University)

in collaboration with
Adel Javanmard and Andrea Montanari

arxiv:1511.08769 -> PNAS 2016
arXiv:1603.09045 -> J. Phys. Conf. Ser. 2016

Outline of the talk

- Hidden partition model (a.k.a. community detection pb.)
  - Spectral vs. optimization methods (ML & SDP)
  - Optimality vs. robustness
- Community detection algorithm based on SDP
  - It is simple, fast and robust
- Phase transitions in SDP
  - Statistical physics approach -> phase transitions
  - (Quasi-)optimality of SDP for (sparse) dense graphs

Communities detection problem

- Detecting communities/partitions/clusters in graphs is a widespread problem in many different disciplines
- Examples of applications: social networks mining, recommendation systems improvement, images segmentation and classification, and many more in biology…
- We need fast (linear and scalable) algorithms
    - robust (real datasets are very noisy and not random)
    - optimal (on random ensemble benchmarks)

Benchmark for community detection

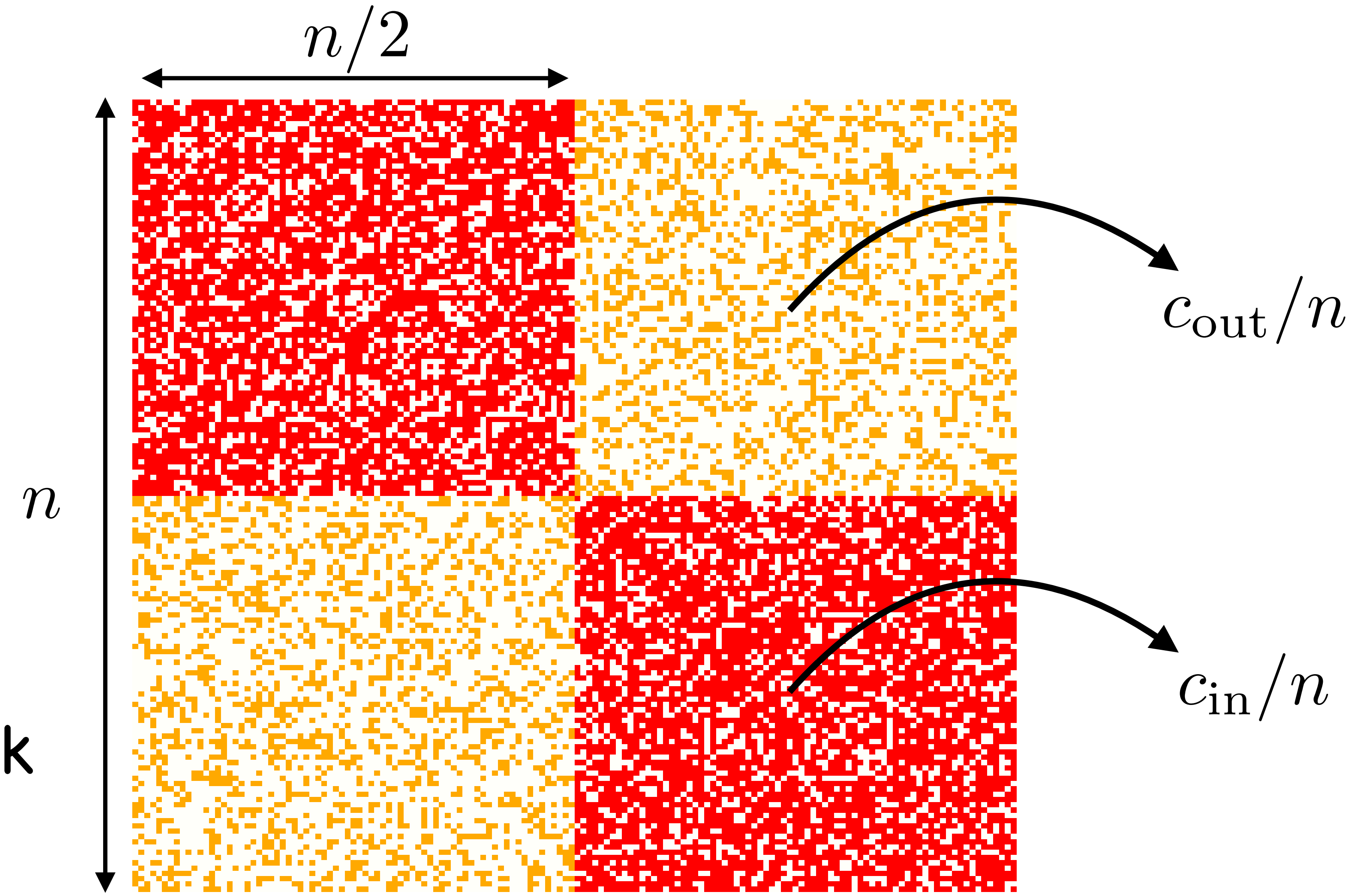
Hidden partition model or stochastic block model (SBM)

- 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

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

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

The hidden partition model



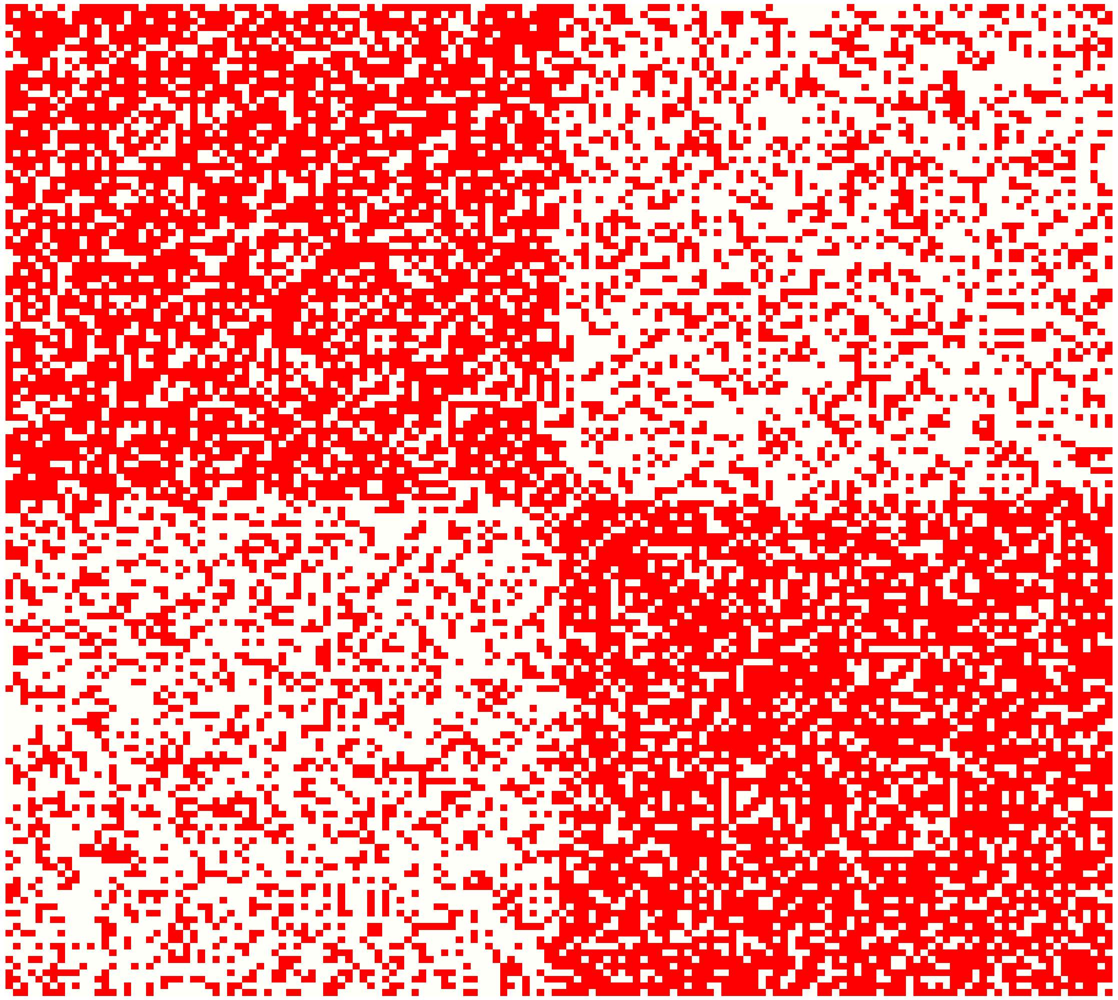
Stochastic block model (SBM) with $q = 2$

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

Assortative model:

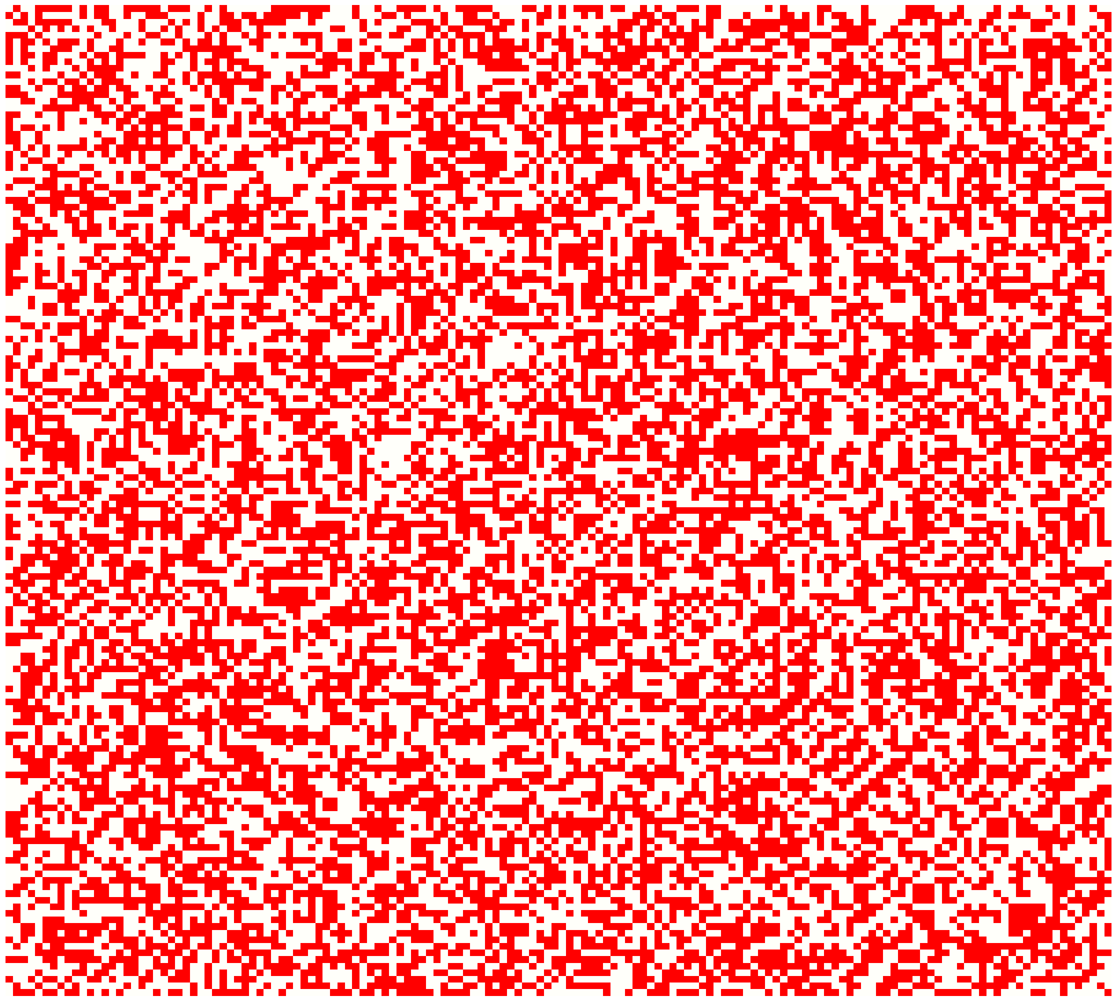
$$c_{\text{in}} > c_{\text{out}}$$

The hidden partition model



Colors are not provided !

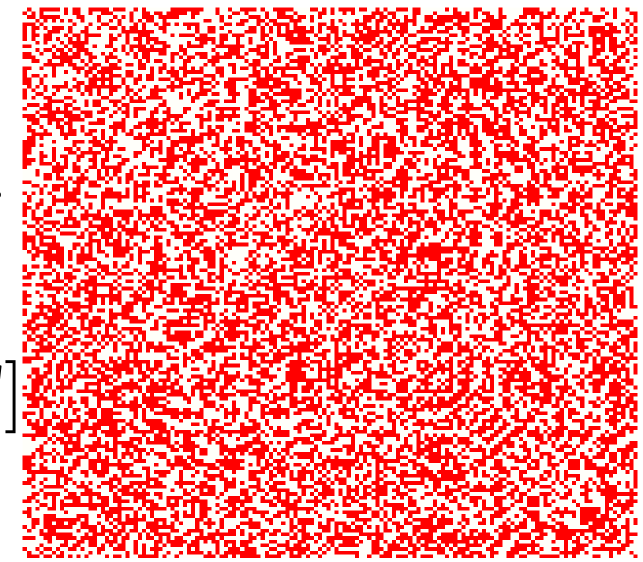
The hidden partition model



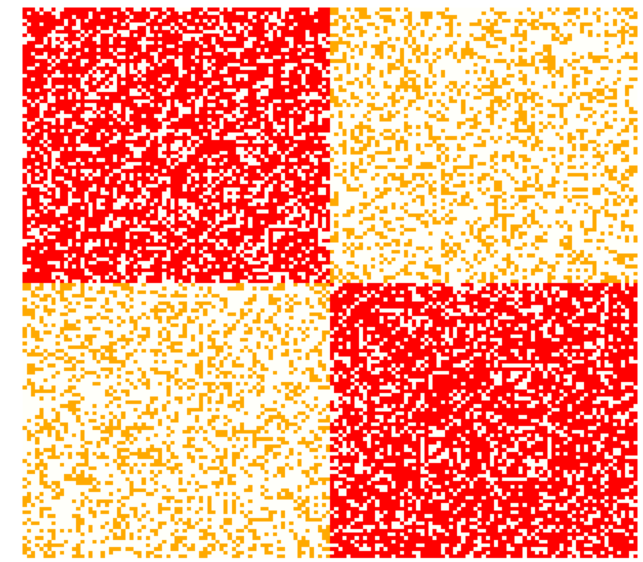
The right ordering neither !!

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 -> $\boldsymbol{x}_0 \in \{+1, -1\}^n$

Estimated partition -> $\hat{\boldsymbol{x}}(G) \in \{+1, -1\}^n$

Quality of inference via the overlap -> $Q = \frac{1}{n} |\langle \hat{\boldsymbol{x}}(G), \boldsymbol{x}_0 \rangle|$

Assortative SBM with 2 equal-size groups

Relevant parameters and threshold

- Mean degree $d = \frac{c_{\mathrm{in}} + c_{\mathrm{out}}}{2}$
- Signal-to-noise ratio $\lambda = \frac{c_{\mathrm{in}} - c_{\mathrm{out}}}{2\sqrt{d}}$
- Bayes optimal threshold $\lambda_c = 1$
  - Impossible detection for $\lambda < \lambda_c$
  - BP algorithm with $Q > 0$ for $\lambda > \lambda_c$

Very ingenious spectral methods

[Decelle, Krzakala, Moore, Zdeborova, 2011]
[Massoulie, 2013] [Mossel, Neeman, Sly, 2013]

Maximum Likelihood (ML)

- If no information on the generative model is given (apart being assortative and with 2 equal-size groups) a good choice is to maximize the likelihood

$$\text{maximize} \quad \sum_{i,j} A_{i,j} x_i x_j$$

$$\text{subject to} \quad x_i \in \{+1,-1\} \quad \text{and} \quad \sum_i x_i = 0$$

- NP-hard problem

Lagrangian formulation

- Maximize $\sum_{i,j} A_{ij} x_i x_j - \eta \Big(\sum_i x_i \Big)^2$ over $\boldsymbol{x} \in \{+1, -1\}^n$

- A good choice is $\eta \geq d/n$

- For $\eta = d/n$ the centered adjacency matrix appears

  $A^{\text{cen}}_{ij} = A_{ij} - d/n$

- Maximize $\sum_{i,j} A^{\text{cen}}_{ij} x_i x_j$ over $\boldsymbol{x} \in \{+1, -1\}^n$

Spectral relaxation (PCA)

- PCA on $\boldsymbol{A}^{\mathbf{cen}}$ relaxes the constraint $\boldsymbol{x} \in \{+1, -1\}^n$ and maximizes $\sum_{i,j} A_{ij}^{\text{cen}} x_i x_j$ over $\boldsymbol{x} \in \mathbb{R}^n$

- Computes the eigenvector of the largest eigenvalue

$$\boldsymbol{v}_1(\boldsymbol{A}^{\mathbf{cen}})$$

- Estimates the partition via a projection on $\boldsymbol{x} \in \{+1, -1\}^n$

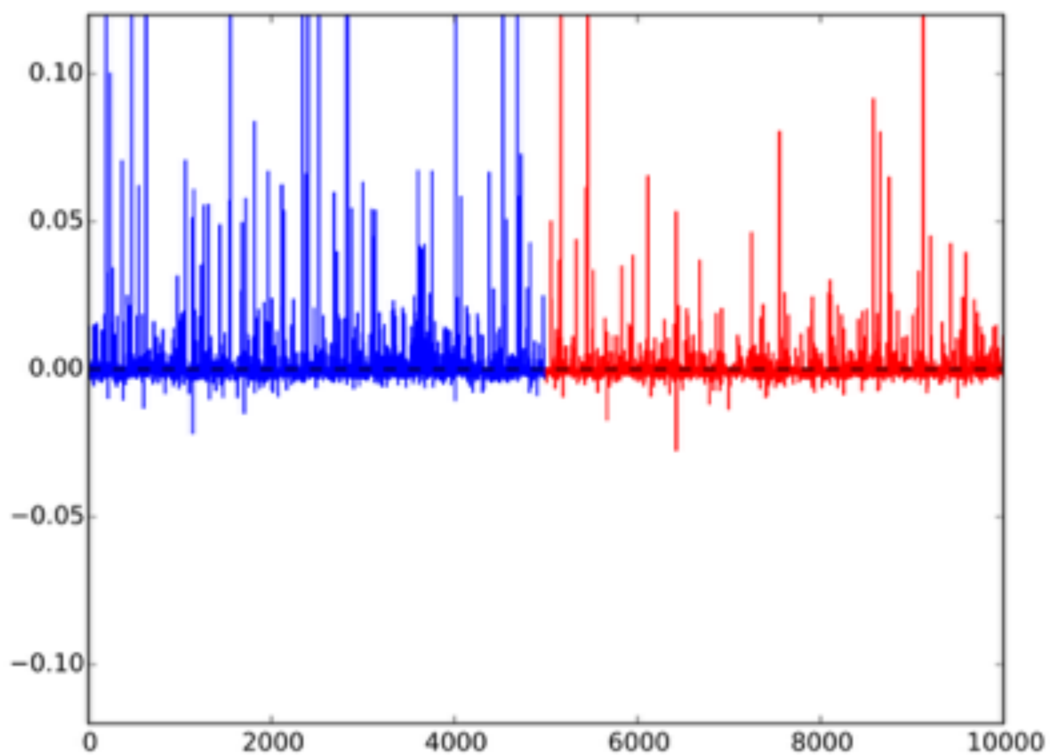
$$\hat{\boldsymbol{x}}^{\text{PCA}} = \text{sign}(\boldsymbol{v}_1(\boldsymbol{A}^{\mathbf{cen}}))$$

- It is good as long as components of $\boldsymbol{v}_1(\boldsymbol{A}^{\mathbf{cen}})$ have similar moduli/intensities

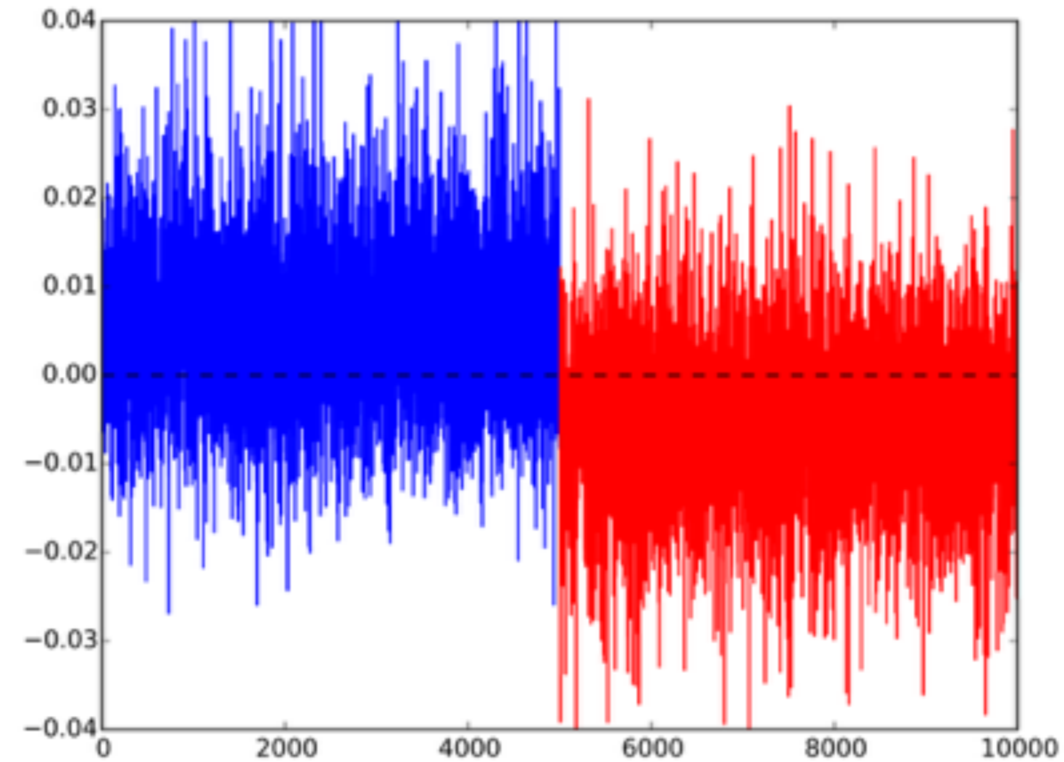
Spectral relaxation fails on sparse graphs

$$n = 10^4 \quad \lambda = 1.2$$

$d = 3$



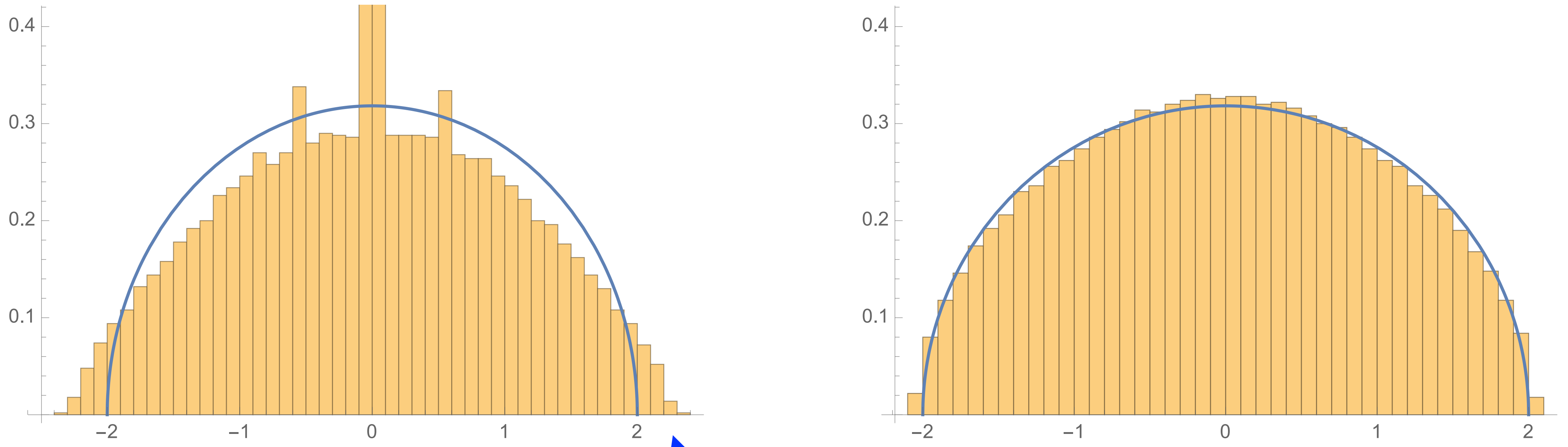
$d = 20$



$$\boldsymbol{v}_1(\boldsymbol{A}^{\mathbf{cen}})$$

Why PCA fails?

$$\frac{1}{\sqrt{d}} \boldsymbol{A}^{\mathrm{cen}} = \frac{\lambda}{n} \boldsymbol{x_0} \boldsymbol{x_0}^{\mathsf{T}} + \boldsymbol{W}$$



$d = 3$ $d = 20$

localized eigenvectors

Other spectral methods

- Compute eigenvalues and eigenvectors of some matrix related to the adjacency matrix A
- Fail for the same reason -> eigenvectors localization
- Laplacian $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$
  with $\boldsymbol{D}$ being the diagonal matrix of degrees
- Normalized Laplacian $\boldsymbol{\mathcal{L}} = \boldsymbol{D}^{-1/2} \boldsymbol{L} \boldsymbol{D}^{-1/2}$
- Shown to be sub-optimal in the sparse regime
  [Kawamoto, Kabashima, 2015]

$$\lambda_c^{\mathrm{Lap}} = \sqrt{\frac{d}{d-1}} > 1$$

Improved spectral methods

- Non-backtracking matrix
  [Krzakala, Moore, Mossel, Neeman, Sly, Zdeborova, Zhang, 2013]
  - seems to avoid localization around large degree nodes
  - optimal for the SBM
  - complex spectrum, not easy to compute
- Bethe Hessian [Saade, Krzakala, Zdeborova, 2014]

$$H(r) = (r^2 - 1)\mathbb{1} - rA - D$$

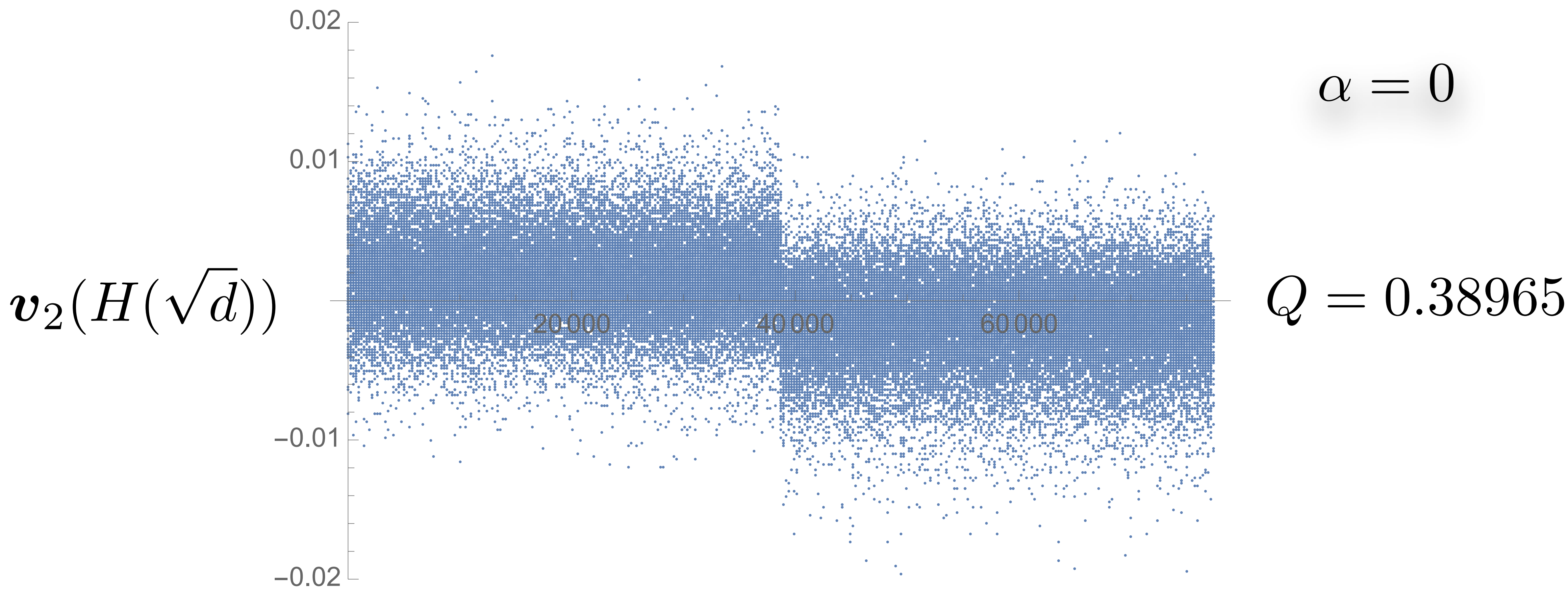
  - symmetric matrix, real spectrum
  - easier to use
  - optimal for the SBM with $r = \sqrt{d}$

Quasi-random graphs

- Generate a graph according to the SBM
- Choose a subset S of vertices of size $|S| = \alpha n$
- For each vertex in S connect all its neighbours
- The number of edges increases by $\sim \alpha d^2 n/2$
  i.e. by a fraction $\sim \alpha d$
- A robust inference method should work also for $\alpha > 0$
  at least in the regime $\alpha \ll 1/d$

Improved spectral methods fail on quasi-random graphs

$n = 10^5 \quad d = 3 \quad \lambda = 1.1$

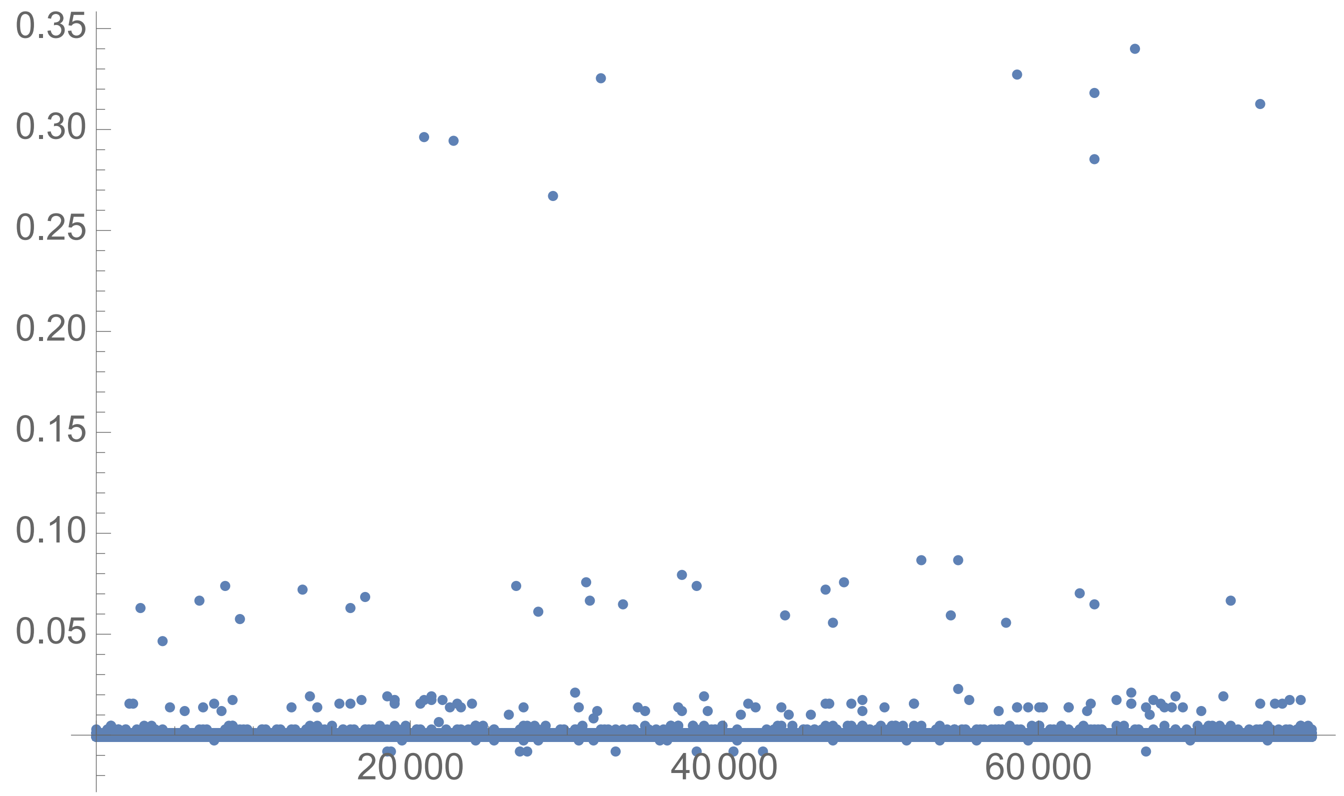


(on the 2-core 77336 nodes and 132763 edges)

Improved spectral methods fail on quasi-random graphs

$n = 10^5 \quad d = 3 \quad \lambda = 1.1$

$\boldsymbol{v}_2(H(\sqrt{d}))$



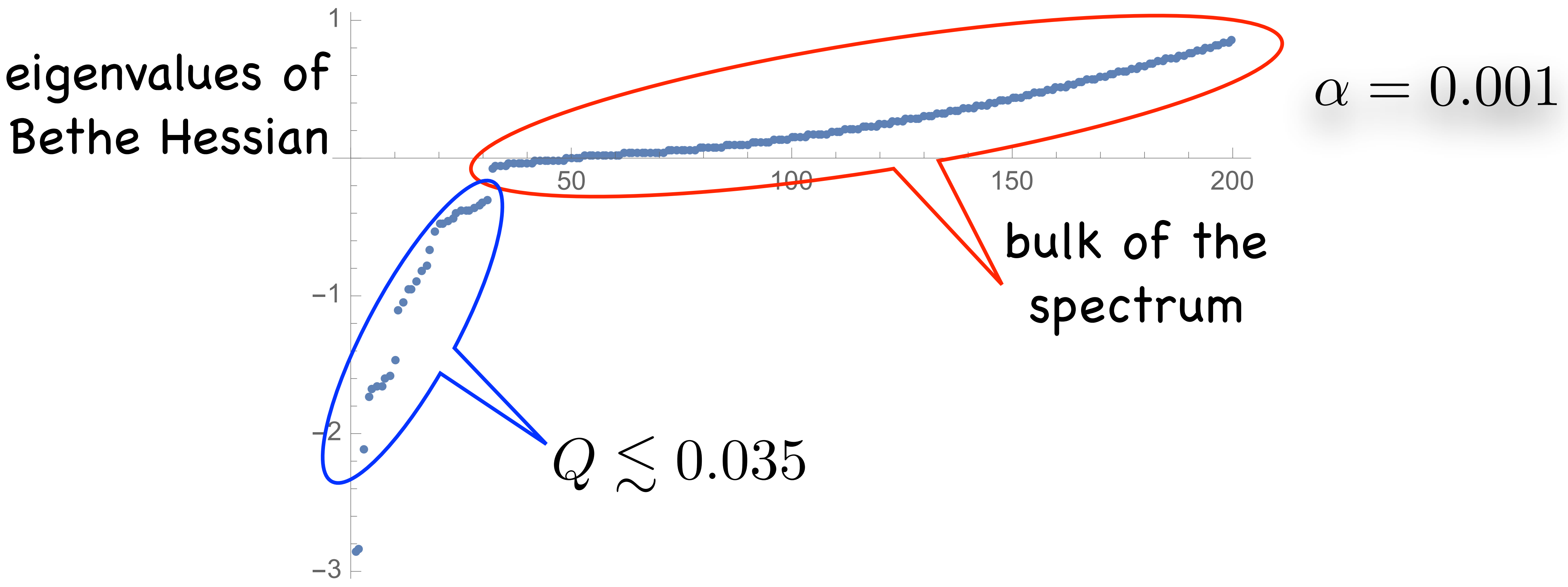
$\alpha = 0.001$

$Q \approx 0$

(on the 2-core 77336 nodes and 133185 edges)

Improved spectral methods fail on quasi-random graphs

$$n = 10^5 \quad d = 3 \quad \lambda = 1.1$$



SDP: a better relaxation?

- Maximize $\sum_{i,j} A^{\text{cen}}_{ij} x_i x_j$ over $\boldsymbol{x} \in \{+1, -1\}^n$

  it is equivalent to maximize $\sum_{i,j} A^{\text{cen}}_{ij} X_{ij} \equiv \langle \boldsymbol{A}^{\text{cen}}, \boldsymbol{X} \rangle$

  subject to $\boldsymbol{X} \in \mathbb{R}^{n \times n}$, $\boldsymbol{X} \succeq 0$ (i.e. all eigenvalues >= 0)

  $X_{ii} = 1$ and $\boldsymbol{X}$ being of rank 1

- SDP relaxes the rank and maximizes $\langle \boldsymbol{A}^{\text{cen}}, \boldsymbol{X} \rangle$ over the convex space of positive semidefinite matrices

- The maximizer is a matrix of rank m with $m \in [1, n]$ to be projected back on a rank 1 matrix...

$$\boldsymbol{X}^{\text{opt}} \longrightarrow \hat{\boldsymbol{x}}^{\text{SDP}} (\hat{\boldsymbol{x}}^{\text{SDP}})^{\mathsf{T}}$$

SDP-based algorithm

- Maximize $\langle \boldsymbol{A}^{\rm cen}, \boldsymbol{X} \rangle$ over rank-m matrices = correlation matrices between m-components variables of unit norm

$$C_{ij} = \underline{x}_i \cdot \underline{x}_j \,, \quad \text{with } \underline{x}_i \in \mathbb{R}^m \,, \; \|\underline{x}_i\|^2 = \underline{x}_i \cdot \underline{x}_i = 1$$

- Maximize $\sum_{(ij)\in E} \underline{x}_i \cdot \underline{x}_j$ subject to $\sum_i \underline{x}_i = \underline{0}$

  by greedy T=0 dynamics (very fast! no gradient used)

- Given the maximizer $\underline{\boldsymbol{x}}^* = \{\underline{x}_1^*, \dots, \underline{x}_n^*\}$ compute the empirical covariance matrix (m x m)

$$\Sigma_{jk} = \frac{1}{n} \sum_{i=1}^{n} (\underline{x}_i^*)_j (\underline{x}_i^*)_k$$

- Project on its principal eigenvector $\hat{x}_i^{\rm SDP} = {\rm sign}(\underline{x}_i \cdot \underline{v}_1)$

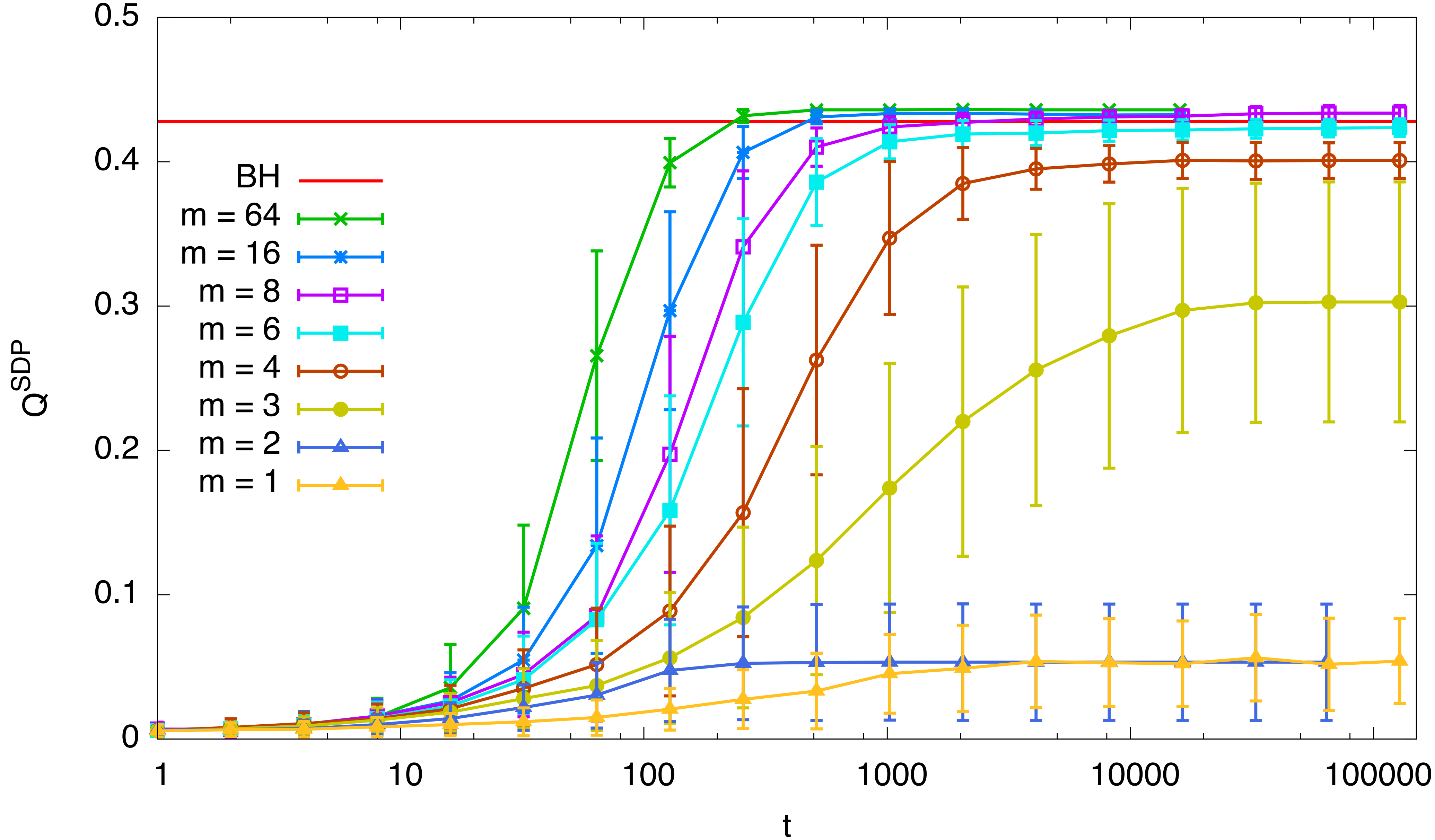
http://web.stanford.edu/~montanar/SDPgraph/

SDP-based algorithm

- Algorithm complexity $O(n\, m\, t_{\rm conv})$ and quality of inference do depend on m
    - m=1 -> ML, very rough objective function, NP-hard
    - m=n -> SDP, convex objective function
      no local maxima for $m > \sqrt{2n}$ [Burer, Monteiro, 2003]
    - m>1, but small -> smooth enough objective function ?
      local minima are "close enough" $O(m^{-1/2})$
      to global minimum [Montanari, 2016]
- Running times grows very mildly with m and n
  e.g. if stopping rule is max variation < 10^{-3} -> $t_{\rm conv} \propto n^{0.22}$

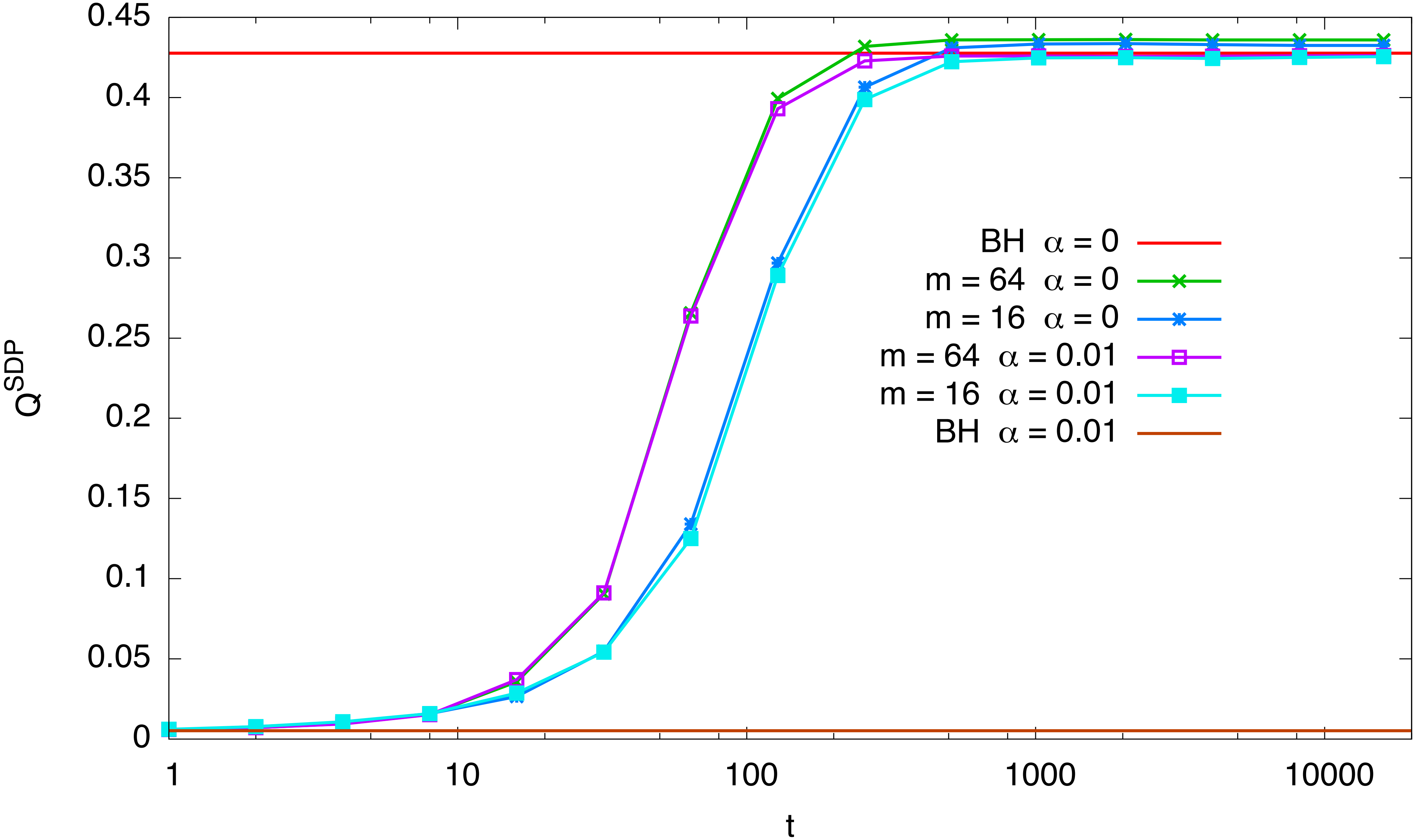
Small m values are fine!

$$n = 4 \cdot 10^4 \quad d = 3 \quad \lambda = 1.1 \quad \alpha = 0.0$$



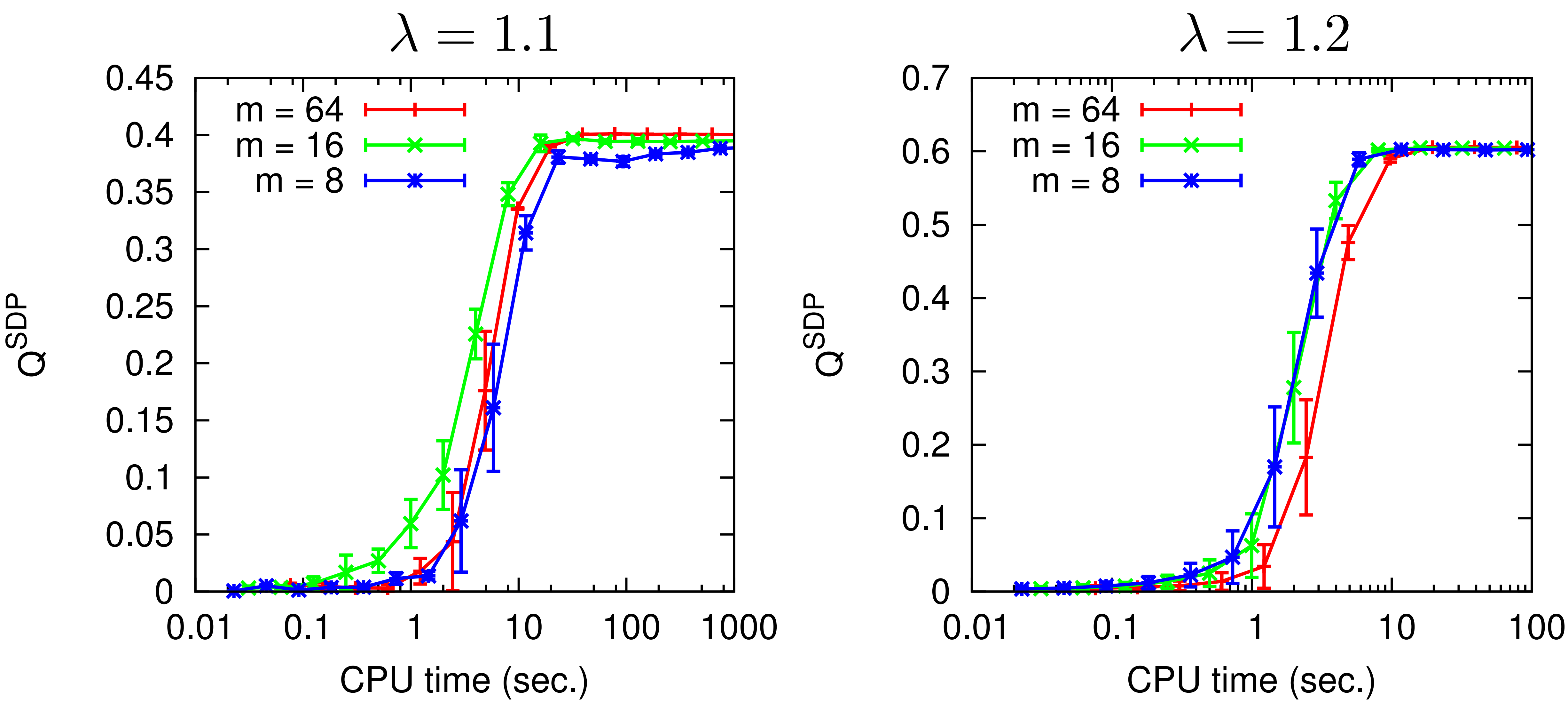
The algorithm is very robust!

$$n = 4 \cdot 10^4 \quad d = 3 \quad \lambda = 1.1$$



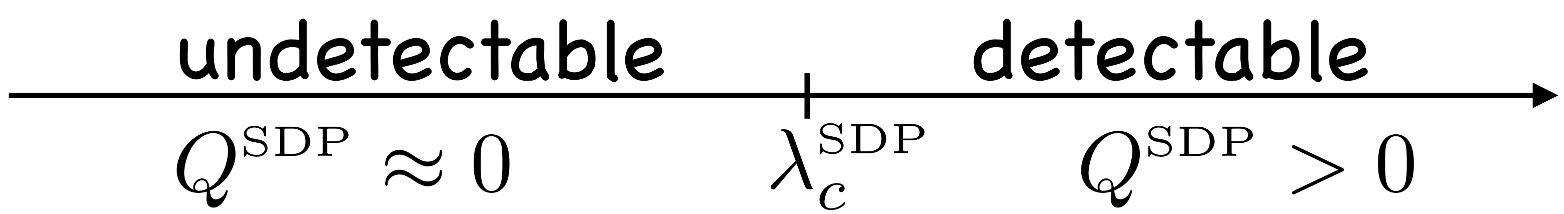
The algorithm is very fast!

$n = 10^5 \quad d = 3$



SDP optimality

- Analytical predictions on SDP-based hidden partition detection (signal recovery) in the limit $m \to \infty$



- A simpler synchronization problem:

$$\boldsymbol{Y} = \frac{\lambda}{n} \boldsymbol{x_0} \boldsymbol{x_0}^* + \boldsymbol{W}$$

- Recovery x_0 given the noisy relative positions in $\boldsymbol{Y}$
- Different models: $\boldsymbol{x}_0 \in \mathbb{R}^n \quad W_{ij} \sim \mathsf{N}(0, 1/n)$

$$\boldsymbol{x}_0 \in \mathbb{C}^n \quad W_{ij} \sim \mathsf{CN}(0, 1/n)$$

Different estimators

- Bayes optimal $\hat{\boldsymbol{x}}^{\text{Bayes}}(\boldsymbol{Y}) = \mathbb{E}\{\, \boldsymbol{x} \,|\, (\lambda/n)\boldsymbol{x}\boldsymbol{x}^* + \boldsymbol{W} = \boldsymbol{Y} \,\}$

- Maximum likelihood $\hat{\boldsymbol{x}}^{\text{ML}}(\boldsymbol{Y}) = c(\lambda)\, \text{argmax}_{\boldsymbol{x}\in\{+1,-1\}^n} \, \langle \boldsymbol{x}, \boldsymbol{Y}\boldsymbol{x}\rangle$

- SDP

$$\begin{aligned} &\text{maximize} \quad \langle \boldsymbol{X}, \boldsymbol{Y}\rangle\,, \\ &\text{subject to} \quad \boldsymbol{X} \succeq 0\,, \quad X_{ii} = 1 \;\; \forall i \in [n] \end{aligned}$$

$$\hat{\boldsymbol{x}}^{\text{SDP}}(\boldsymbol{Y}) = \sqrt{n}\, c^{\text{SDP}}(\lambda)\, \boldsymbol{v}_1(\boldsymbol{X}_{\text{opt}}(\boldsymbol{Y}))$$

$$\text{argmax}_{\underline{x}} \sum_{i,j} \text{Re}(Y_{ij}\underline{x}_i \cdot \underline{x}_j) \qquad \underline{x}_i \in \mathbb{F}^m \;, ||\underline{x}_i|| = 1$$

Statistical physics approach

- Unified framework: statistical physics models with m-component variables: $\underline{x}_i \in \mathbb{F}^m \,, ||\underline{x}_i|| = 1$

$$P(\underline{\boldsymbol{x}}) = \frac{1}{Z} \exp\left[2m\beta \sum_{i<j} \mathrm{Re}(Y_{ij}\underline{x}_i \cdot \underline{x}_j)\right]$$

- Bayes: $m = 1, \quad \beta = \begin{cases} \lambda/2 & \text{if } \mathbb{F} = \mathbb{R} \\ \lambda & \text{if } \mathbb{F} = \mathbb{C} \end{cases}$

- ML: $m = 1, \quad \beta \to \infty$

- SDP: $m \to \infty, \quad \beta \to \infty$

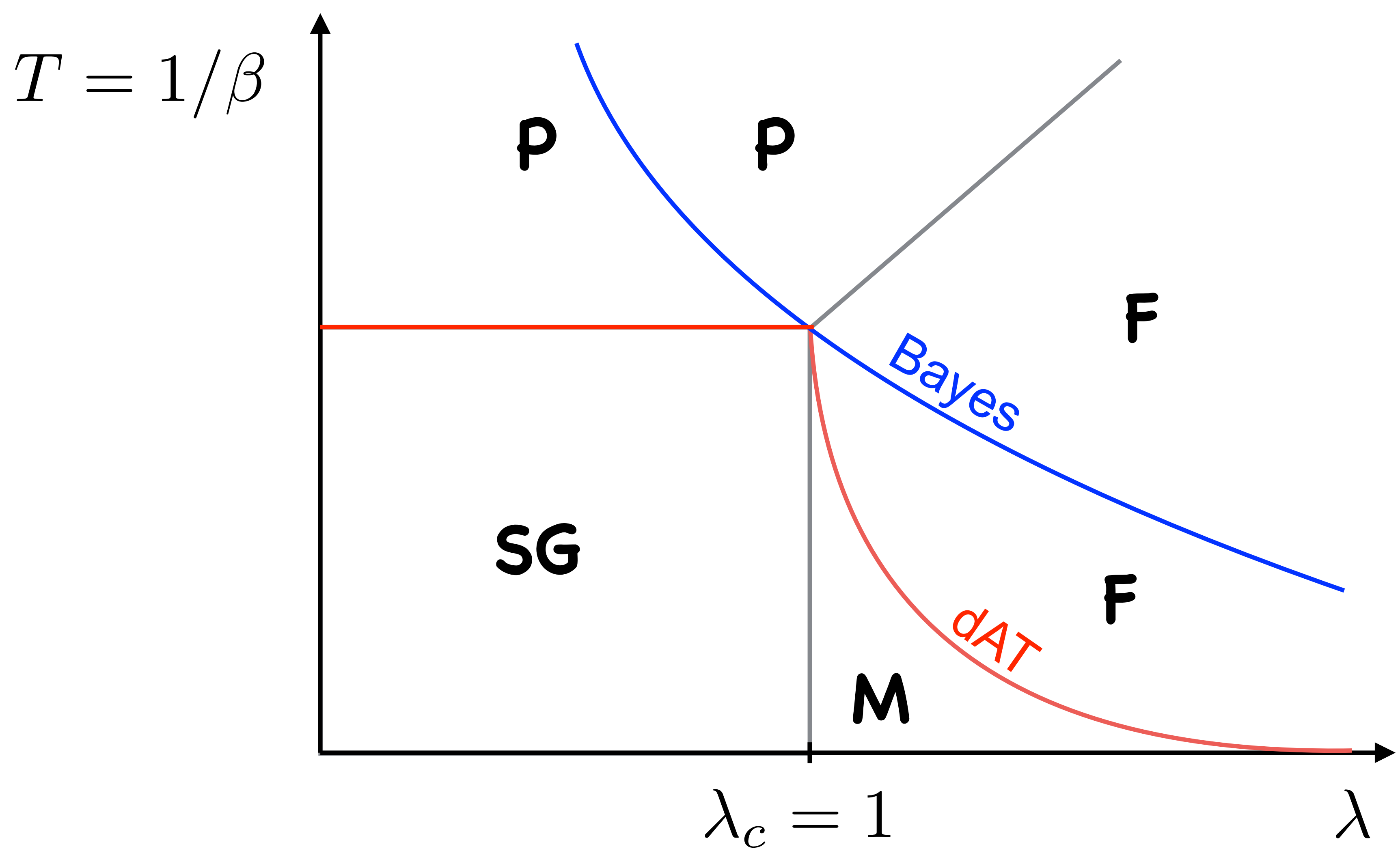
Statistical physics approach

Ising variables (m=1), dense graph

Sherrington-Kirkpatrick model

$$H = -\sum_{i<j} J_{ij} s_i s_j \qquad s_i \in \{-1, 1\}$$

$$J_{ij} \sim \mathrm{N}(\lambda/n, 1/n)$$



P: Q=0 easy

F: Q>0 easy

SG: Q=0 hard

M: Q>0 hard

Statistical physics approach

- Ansatz for the marginals in m-component dense models

$$P_i(\underline{x}_i) = \frac{1}{Z_i} \exp\left[2m\beta(\underline{\xi}_i^{\mathsf{T}} \underline{x}_i + \underline{x}_i^{\mathsf{T}} \boldsymbol{C}_i \underline{x}_i)\right]$$

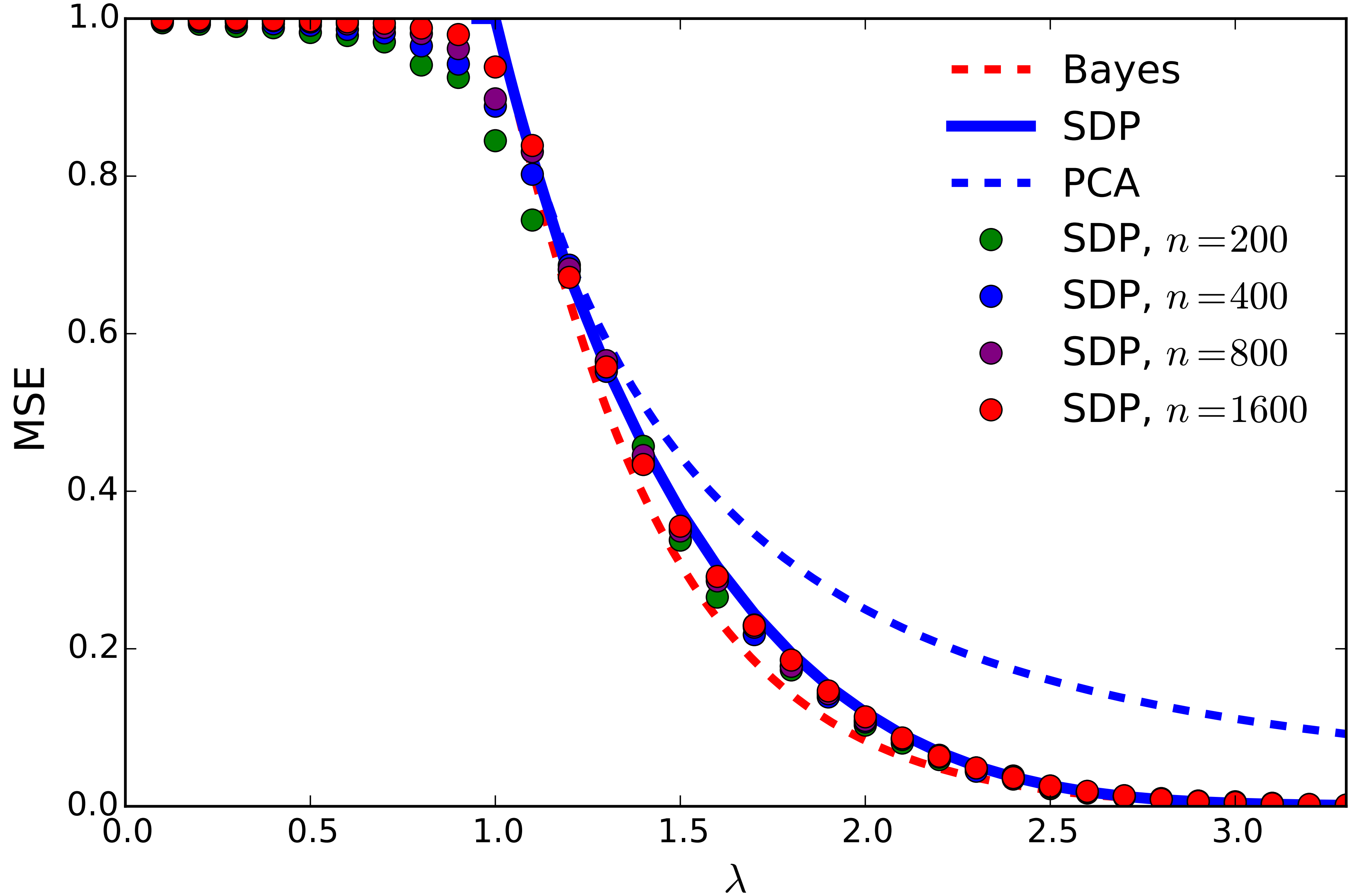
$$\underline{x}_i \in \mathbb{F}^m \,, ||\underline{x}_i|| = 1 \qquad \underline{\xi}_i \sim \mathcal{N}(\underline{\mu}, \boldsymbol{Q}) \qquad \boldsymbol{C}_i = \boldsymbol{C}$$

- Self consistency equations in the dense case

$$\begin{aligned} \underline{\mu} &= \lambda\, \mathbb{E}[\langle \underline{x} \rangle] \\ \boldsymbol{Q} &= \mathbb{E}[\langle \underline{x} \rangle \langle \underline{x}^{\mathsf{T}} \rangle] \\ \boldsymbol{C} &= \beta\, m\, \mathbb{E}[\langle \underline{x}\, \underline{x}^{\mathsf{T}} \rangle - \langle \underline{x} \rangle \langle \underline{x}^{\mathsf{T}} \rangle] \end{aligned}$$

Analytical solution: dense real case

$$\mathrm{MSE}_n(\hat{\boldsymbol{x}}) \equiv \frac{1}{n}\mathbb{E}\{\min_{s\in\{+1,-1\}} \|\hat{\boldsymbol{x}}(\boldsymbol{Y}) - s\,\boldsymbol{x_0}\|_2^2\}$$



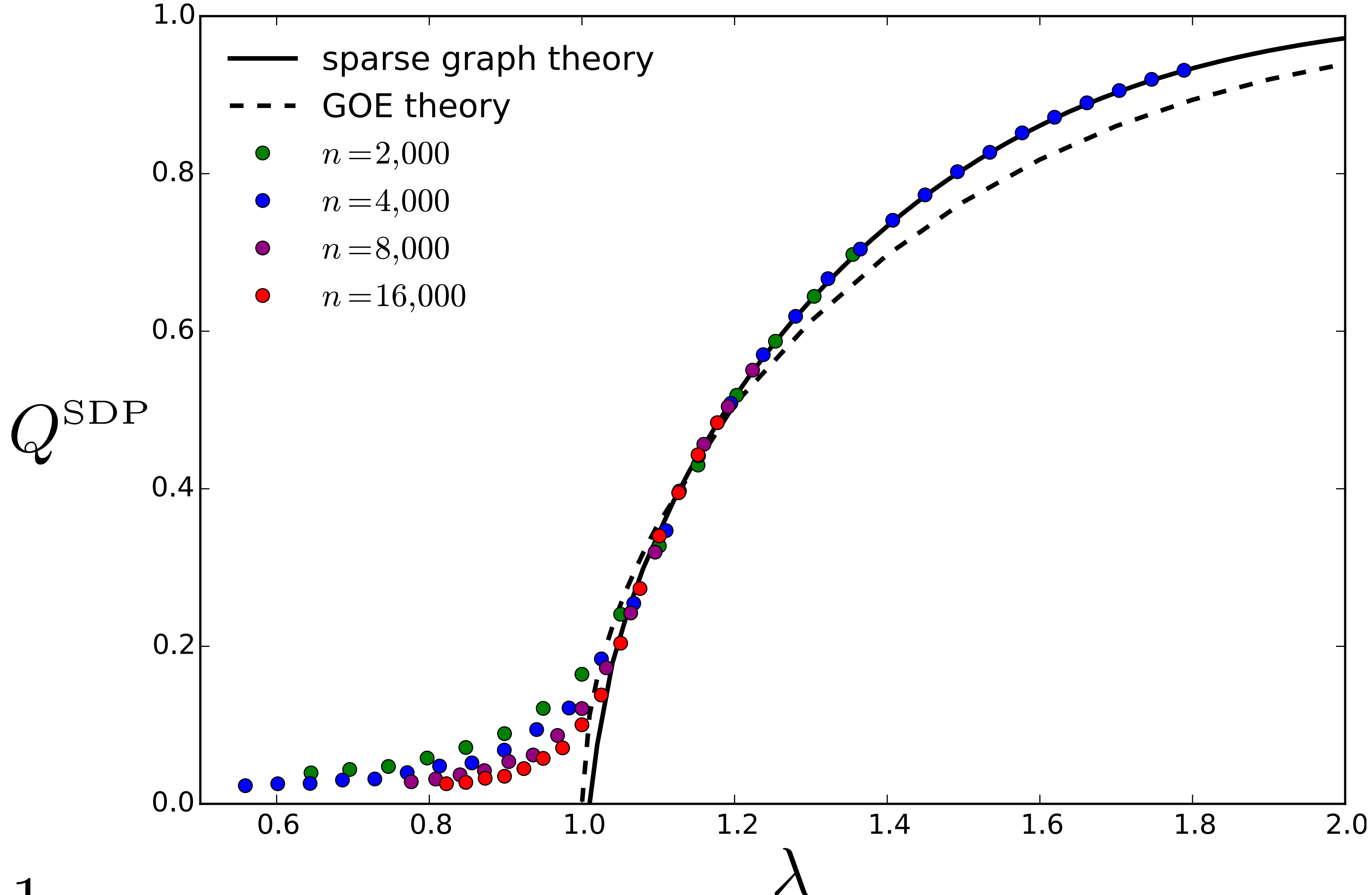
Analytical solution: sparse case (SBM)

- Run the SDP-based algorithm for very large m values

- Approximate ansatz (exact in the large d limit)

Analytical solution: sparse case (SBM)

$d = 5 \quad N_{\text{samples}} = 500$

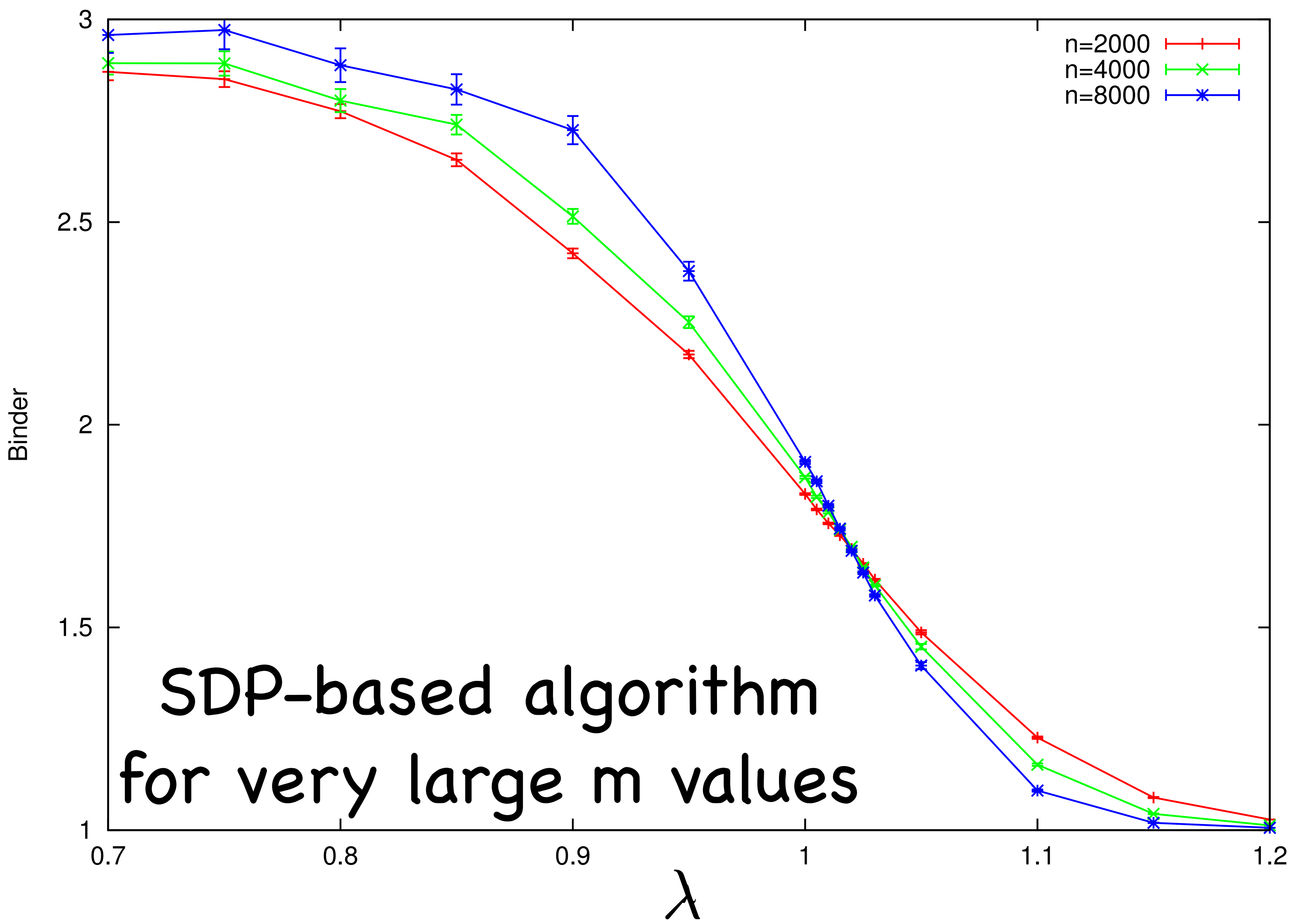


$$Q^{\text{SDP}} = \frac{1}{n}\mathbb{E}\{|\langle \boldsymbol{x}^{\text{SDP}}, \boldsymbol{x}_0 \rangle|\}$$

SDP-based algorithm
for very large m values

Analytical solution: sparse case (SBM)

- Crossing of the Binder cumulants to locate exactly λ_c^{SDP}



$$d = 5$$

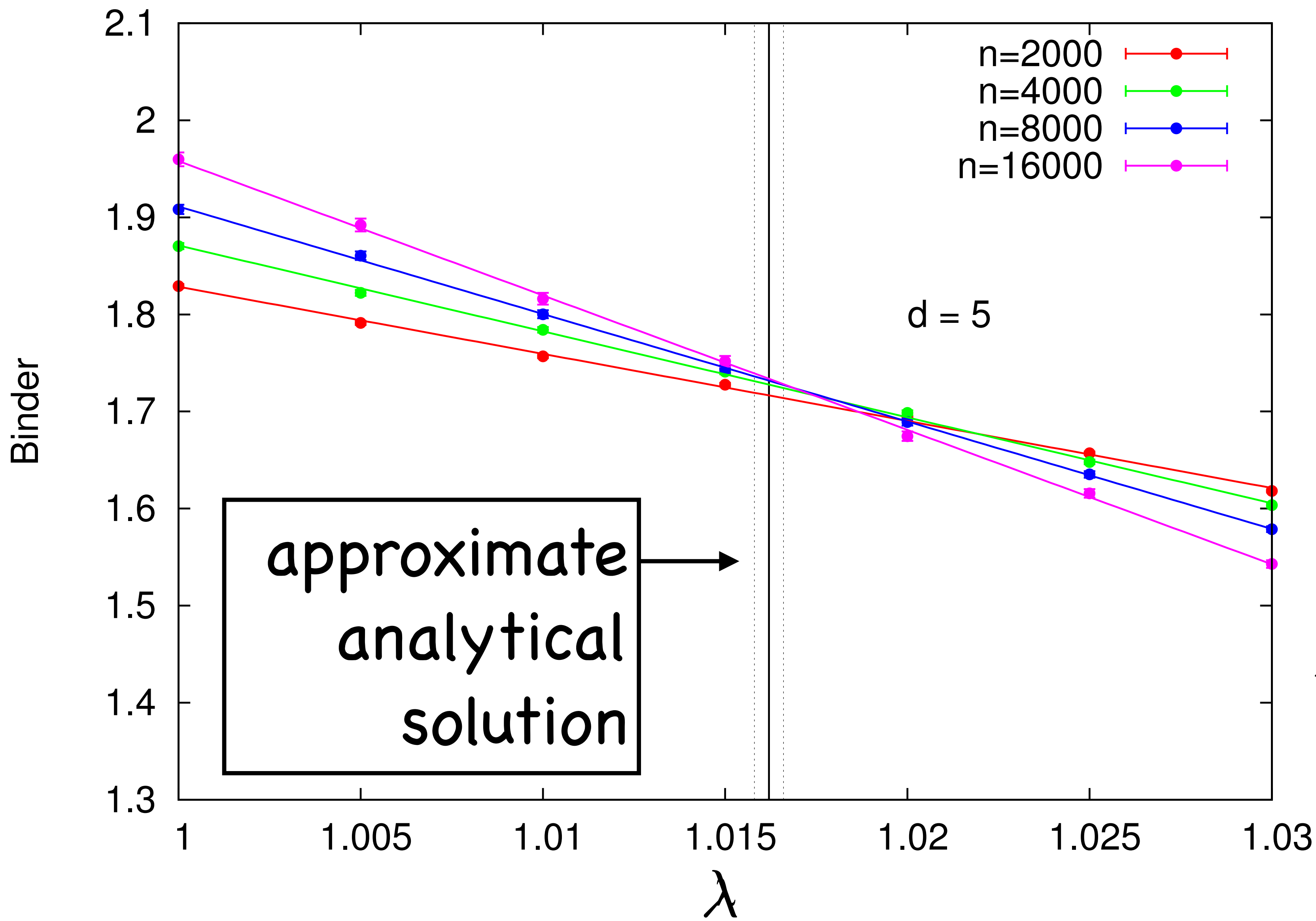
$$N_{\text{samples}} \geq 10^5$$

roughly 2 years of CPU time

$$B = \frac{n\,\mathbb{E}\{\langle \boldsymbol{x}^{\mathrm{SDP}}, \boldsymbol{x}_0 \rangle^4\}}{(\mathbb{E}\{\langle \boldsymbol{x}^{\mathrm{SDP}}, \boldsymbol{x}_0 \rangle^2\})^2}$$

Analytical solution: sparse case (SBM)

- Crossing of the Binder cumulants to locate exactly λ_c^{SDP}



$$d = 5$$

$$N_{\mathrm{samples}} \geq 10^5$$

roughly 2 years of CPU time

$$B = \frac{n\,\mathbb{E}\{\langle \boldsymbol{x}^{\mathrm{SDP}}, \boldsymbol{x}_0 \rangle^4\}}{(\mathbb{E}\{\langle \boldsymbol{x}^{\mathrm{SDP}}, \boldsymbol{x}_0 \rangle^2\})^2}$$

SDP-based algorithm for very large m values

Approximate analytical solution (SBM)

- In the recovery phase we assume the O(m) symmetry to break along the first component, while preserving O(m-1)

$$\underline{x}_i = (s_i, \boldsymbol{\tau}_i),\ s_i \in \mathbb{R},\ \boldsymbol{\tau}_i \in \mathbb{R}^{m-1}$$

- We write the marginal for $\underline{x}_i$ as

$$\exp\left\{2\beta\sqrt{m\mathsf{c}_i}\,\langle \boldsymbol{z}_i, \boldsymbol{\tau}_i\rangle + 2\beta m h_i\, s_i - \beta m r_i s_i^2 + O_m(1)\right\}\delta\Big(s_i^2 + \|\boldsymbol{\tau}_i\|_2^2 - 1\Big)$$

with $\boldsymbol{z}_i \sim \mathsf{N}(0, \mathrm{I}_{m-1})$

- Approximate because the z_i are correlated
- It should be valid in the limits $d \to 1$ and $d \to \infty$

Approximate analytical solution (SBM)

$$\exp\left\{2\beta\sqrt{m\mathsf{c}_i}\,\langle \boldsymbol{z}_i, \boldsymbol{\tau}_i\rangle + 2\beta m h_i\, s_i - \beta m r_i s_i^2 + O_m(1)\right\}\delta\Big(s_i^2 + \|\boldsymbol{\tau}_i\|_2^2 - 1\Big)$$

- Cavity method -> self consistency equation for marginals

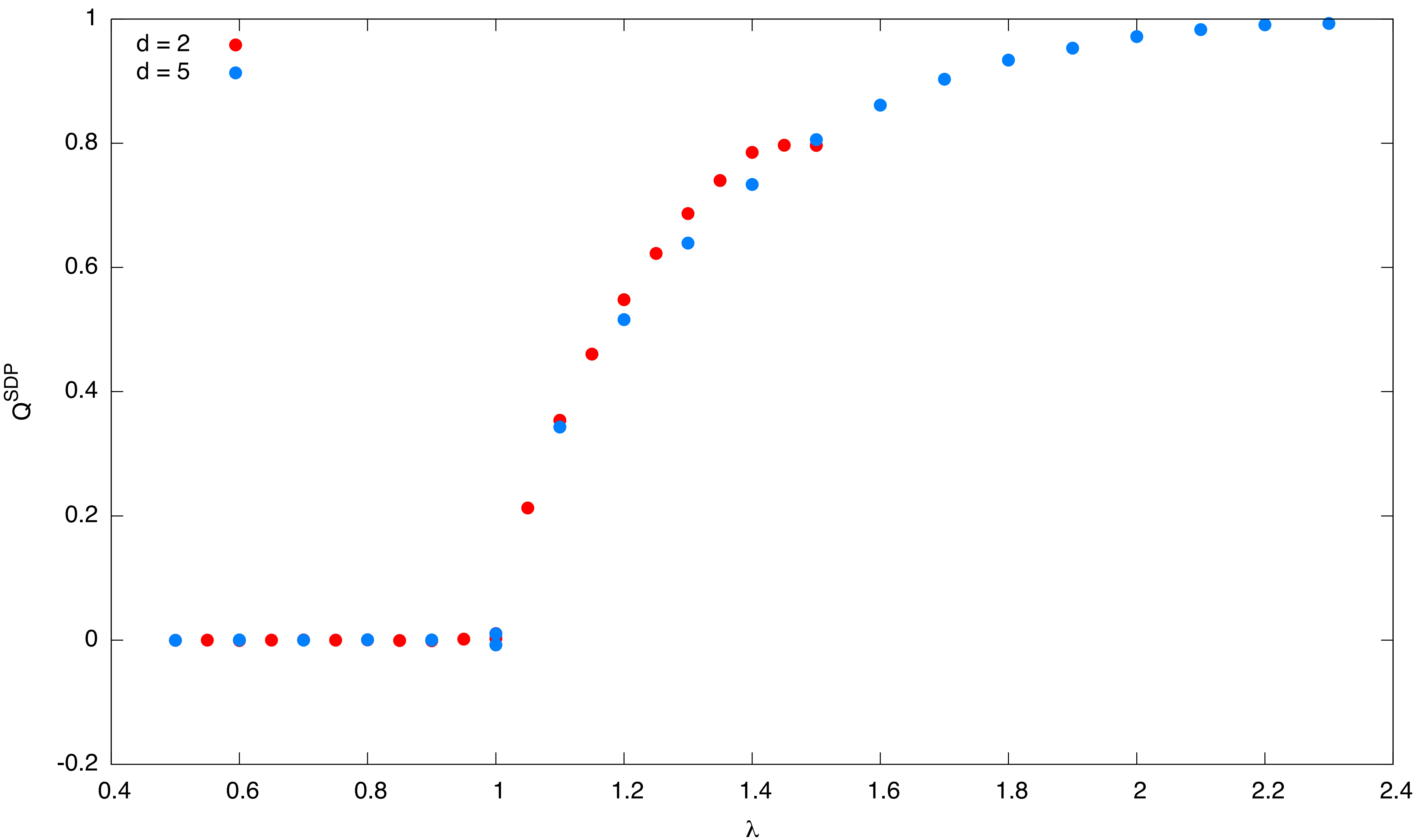
$$\mathsf{c}_0 = \sum_{i=1}^{k} \frac{\mathsf{c}_i}{\rho_i^2}\,, \qquad 1 = \frac{h_i^2}{(\rho_i + r_i)^2} + \frac{1+\mathsf{c}_i}{\rho_i^2}$$

$$h_0 = \sum_{i=1}^{k} \frac{h_i}{\rho_i + r_i}\,,$$

$$r_0 = \sum_{i=1}^{k} \left\{ \frac{1}{\rho_i} - \frac{1}{\rho_i + r_i} + \left(1 + \frac{(1+\mathsf{c}_i) r_i}{\rho_i^3}\right)^{-1} \frac{h_i^2}{(\rho_i + r_i)^3} \right\}$$

- Solve by population dynamics
- At the fixed point $Q^{\mathrm{SDP}} = \mathbb{E}[\mathrm{sign}(h^*)]$

Approximate analytical solution (SBM)



Approximate analytical solution (SBM)

- To linear order in $h \implies r_i = 0$

$$\mathsf{c}_0 = \sum_{i=1}^{k} \frac{\mathsf{c}_i}{\rho_i^2}\,,$$

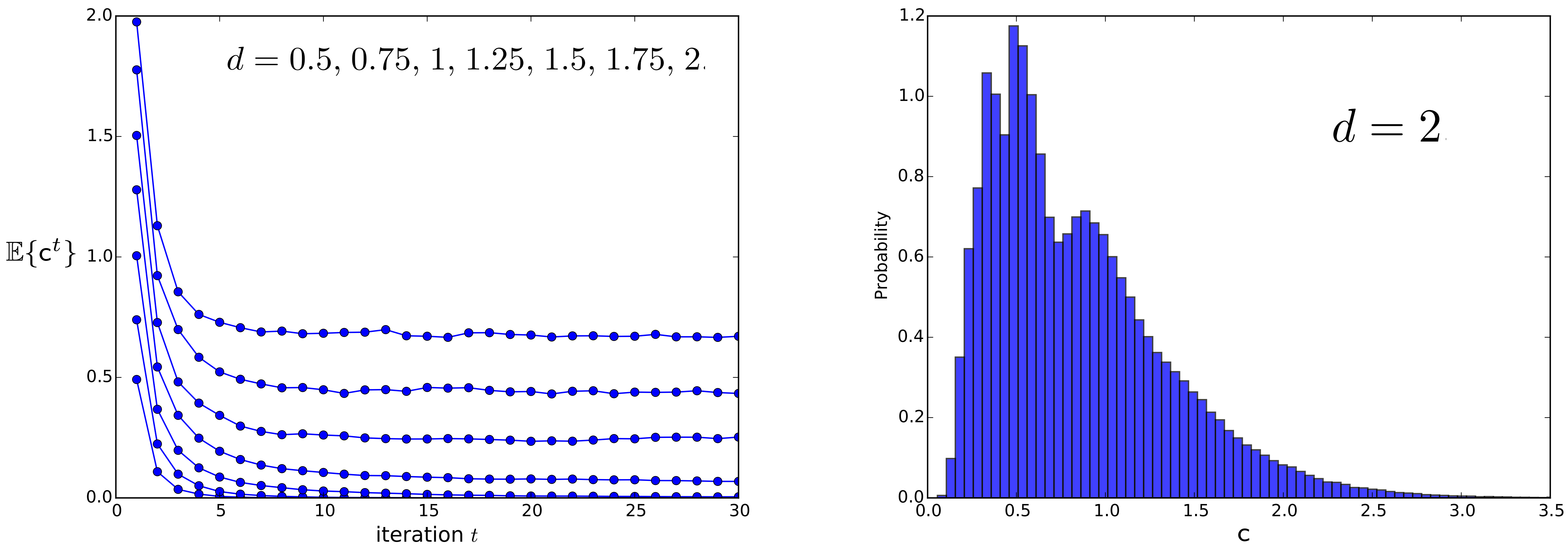
$$h_0 = \sum_{i=1}^{k} \frac{h_i}{\rho_i + r_i}\,,$$

$$r_0 = \sum_{i=1}^{k} \left\{ \frac{1}{\rho_i} - \frac{1}{\rho_i + r_i} + \left(1 + \frac{(1+\mathsf{c}_i) r_i}{\rho_i^3}\right)^{-1} \frac{h_i^2}{(\rho_i + r_i)^3} \right\}$$

$$1 = \frac{h_i^2}{(\rho_i + r_i)^2} + \frac{1 + \mathsf{c}_i}{\rho_i^2}$$

$$\mathsf{c}_0 = \sum_{i=1}^{k} \frac{\mathsf{c}_i}{1 + \mathsf{c}_i}\,, \qquad h_0 = \sum_{i=1}^{k} \frac{h_i}{\sqrt{1 + \mathsf{c}_i}}$$

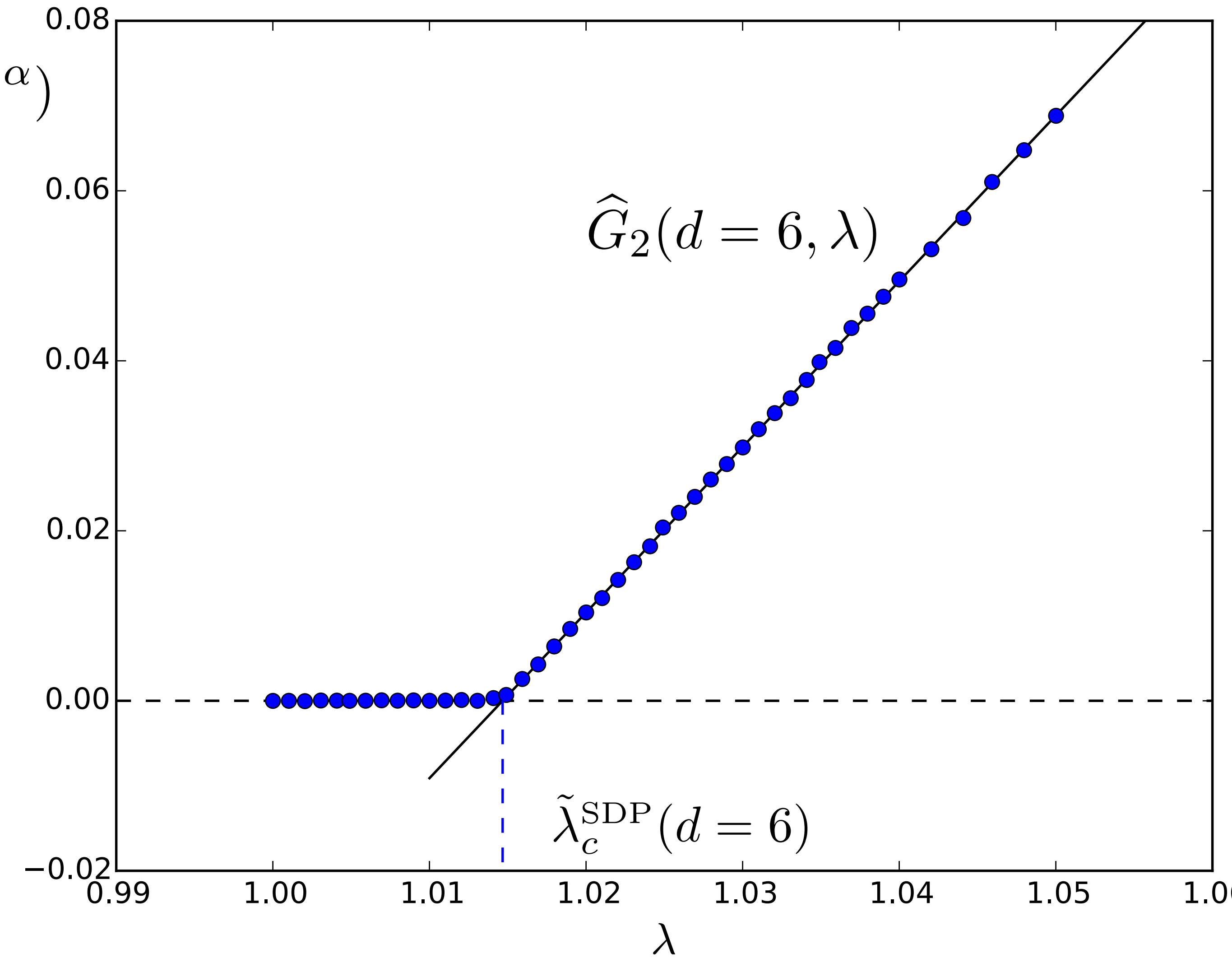
Approximate analytical solution (SBM)



$$\left(\mathsf{c}^{t+1}; h^{t+1}\right) \stackrel{\mathrm{d}}{=} \left(\sum_{i=1}^{L_+ + L_-} \frac{\mathsf{c}_i^t}{1+\mathsf{c}_i^t}; \sum_{i=1}^{L_+ + L_-} \frac{s_i h_i^t}{\sqrt{1+\mathsf{c}_i^t}} \right)$$

Approximate analytical solution (SBM)

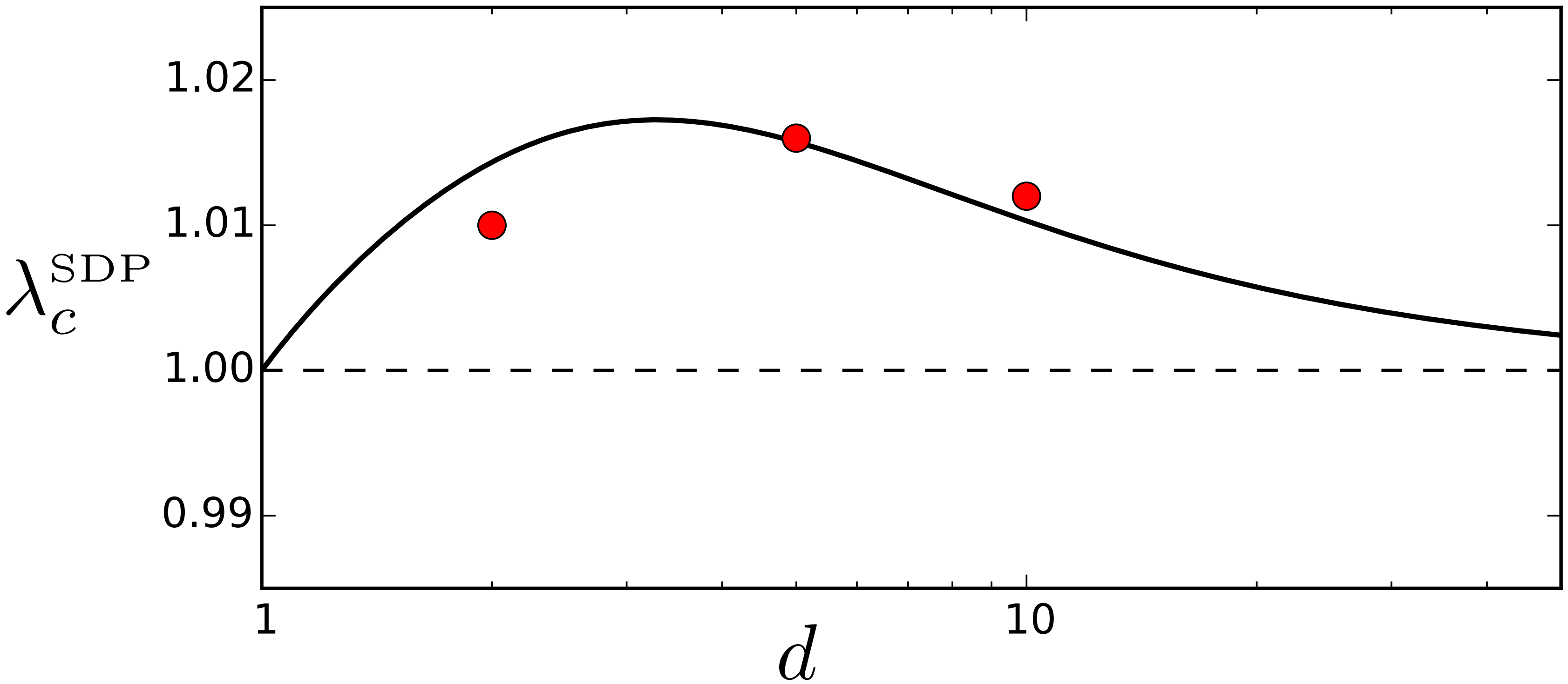
$$G_\alpha(d,\lambda) \equiv \liminf_{t\to\infty} \frac{1}{t\alpha} \log \mathbb{E}(|h^t|^\alpha)$$



$$\left(\mathsf{c}^{t+1}; h^{t+1}\right) \stackrel{\mathrm{d}}{=} \left(\sum_{i=1}^{L_+ + L_-} \frac{\mathsf{c}_i^t}{1+\mathsf{c}_i^t}; \sum_{i=1}^{L_+ + L_-} \frac{s_i h_i^t}{\sqrt{1+\mathsf{c}_i^t}}\right)$$

Analytical solution: sparse case (SBM)

- SDP at most 2% sub-optimal!



- Red points: numerical solution of the replica/cavity equations (crossing of Binder cumulants)
- Black line: approximated analytical solution

Take-home messages

- SDP relaxations are very effective:
  - robust and quasi-optimal
  - may outperform spectral relaxations
- Better than SDP are SDP-inspired algorithms (small m) `http://web.stanford.edu/~montanar/SDPgraph/`
- It is worth studying the statistical physics of models with m-component variables:
  - unifying framework to study and solve several estimators in statistical inference
  - different physics, better algorithms