Information-theoretic thresholds

Amin Coja-Oghlan

Goethe University Frankfurt

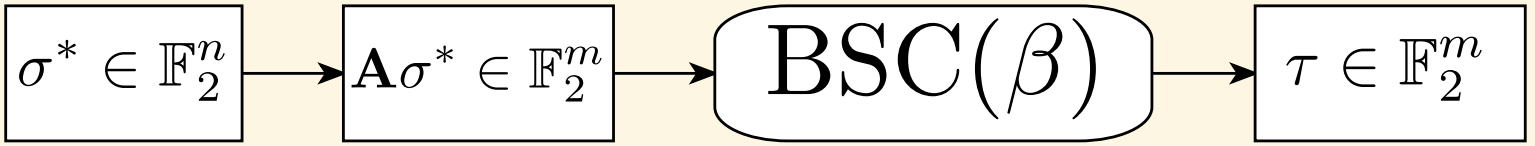
based on joint work with

Florent Krzakala (ENS Paris)
Will Perkins (Birmingham)
Lenka Zdeborová (CEA Saclay)

Inference from samples

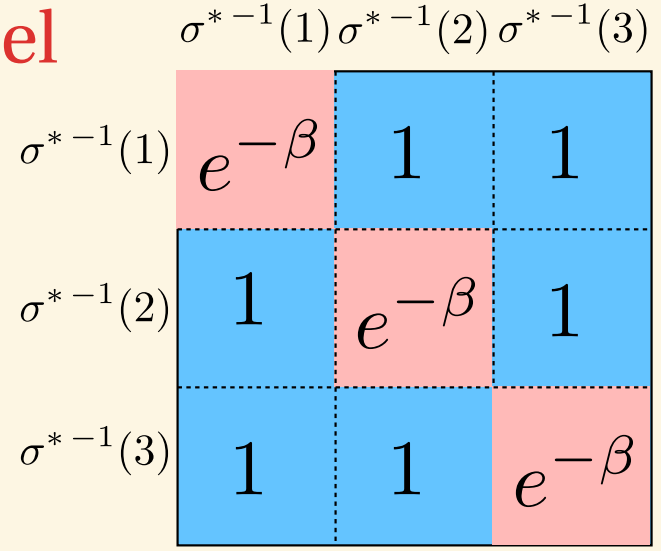
- to infer an unkown probability distribution from samples
- the distribution itself is random, determined by parameters $\boldsymbol{\sigma}^*$

Example: error-correcting codes



- $\boldsymbol{A} \in \mathbb{F}_2^{m \times n}$ is the generator matrix
- $\boldsymbol{A}\sigma^*$ is subjected to noise

Example: the stochastic block model

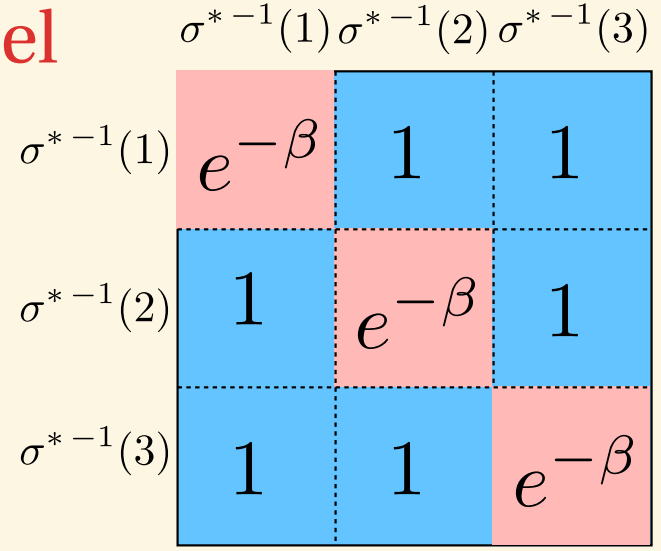


- random coloring $\boldsymbol{\sigma}^* : V \to \{1, \ldots, q\}$
- for each $e = \{v, w\}$ independently,

$$\mathrm{P}\left[e \in \mathbf{G}^* \middle| \boldsymbol{\sigma}^*\right] = \frac{d}{n} \cdot \frac{q}{q-1+e^{-\beta}} \cdot \begin{cases} e^{-\beta} & \text{if } \boldsymbol{\sigma}^*(v) = \boldsymbol{\sigma}^*(w), \\ 1 & \text{if } \boldsymbol{\sigma}^*(v) \neq \boldsymbol{\sigma}^*(w) \end{cases}$$

- d = signal strength; $e^{-\beta}$ = noise

Example: the stochastic block model



- the agreement of $\sigma, \tau : V \to \{1, \dots, q\}$ is

$$\alpha(\sigma, \tau) = \frac{1}{q-1} \max_{\kappa \in S_q} \left\{ \frac{q}{n} \sum_{v \in V} \mathbf{1}\{\sigma(v) = \kappa \circ \tau(v)\} - 1 \right\}.$$

- for what d, β is it possible to recover $\tau_{\mathbf{G}^*}$ such that

$$\mathrm{E}[\alpha(\boldsymbol{\sigma}^*, \tau_{\mathbf{G}^*})] \geq \Omega(1) \ ?$$

Example: the stochastic block model

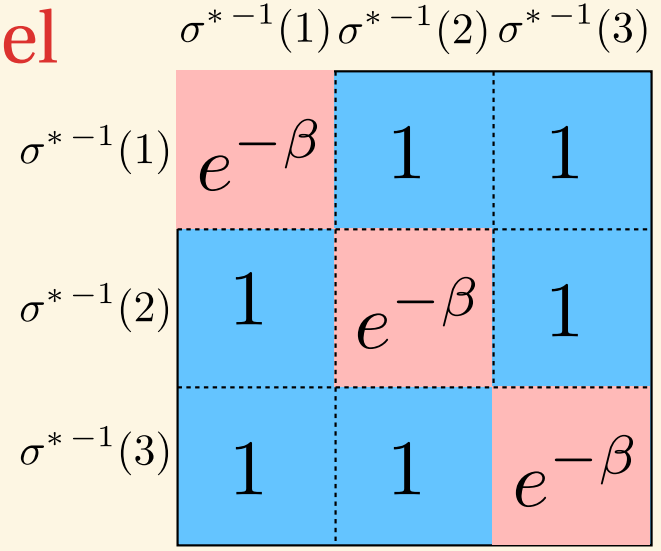
	$\sigma^{*-1}(1)$	$\sigma^{*-1}(2)$	$\sigma^{*-1}(3)$
$\sigma^{*-1}(1)$	$e^{-\beta}$	1	1
$\sigma^{*-1}(2)$	1	$e^{-\beta}$	1
$\sigma^{*-1}(3)$	1	1	$e^{-\beta}$

Easy–hard–impossible

- for large d efficient algorithms should detect $\boldsymbol{\sigma}^*$
- for very small d there is nothing to detect
- in-between the problem may be well-posed but hard

$$0 < d_{\text{inf}}(\beta) < d_{\text{alg}}(\beta)$$

Example: the stochastic block model

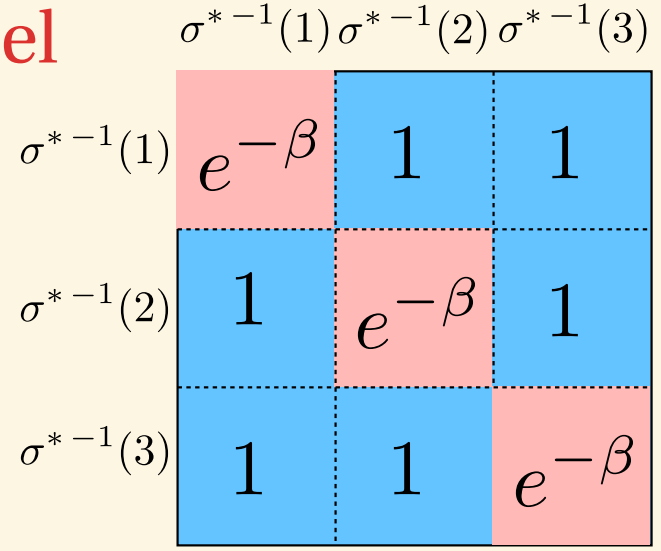


The algorithmic threshold

- combinatorial algorithms for large d [1980s]
- spectral algorithms for moderate d [1990s, 2000s]
- the Kesten-Stigum threshold [AS15]

$$d_{\mathrm{alg}}(\beta) \stackrel{?}{=} \left(\frac{q-1+e^{-\beta}}{1-e^{-\beta}} \right)^2$$

Example: the stochastic block model



The information-theoretic threshold

- statistical physics prediction [DKMZ11]
- the case $q = 2$ [MNS13, MNS14, M14]
- bounds on $d_{\text{inf}}(q, \beta)$ [BMNN16]

The information-theoretic threshold

Theorem [COKPZ16]

For $\beta > 0$, $d > 0$ let

$$\mathscr{B}^*_{q,\beta}(d) = \sup\left\{\mathscr{B}_{q,\beta,d}(\pi) : \pi = \mathscr{T}_{q,\beta,d}(\pi), \int \mu(i)\,d\pi(\mu) = 1/q\right\} \qquad \text{where}$$

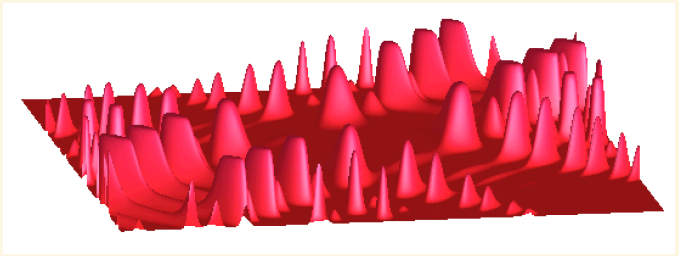
$$\mathscr{T}_{q,\beta,d} : \pi \mapsto \sum_{\gamma=0}^{\infty} \frac{d^\gamma \exp(-d)}{q(1-(1-e^{-\beta})/q)^\gamma \gamma!} \int \left[\sum_{h=1}^{q} \prod_{j=1}^{\gamma} 1-(1-e^{-\beta})\mu_j(h)\right] \delta_{\mathrm{BP}_{\mu_1,\dots,\mu_\gamma}}\,\mathrm{d}\pi^{\otimes\gamma}(\mu_1,\dots,\mu_\gamma),$$

$$\mathrm{BP}_{\mu_1,\dots,\mu_\gamma}(i) = \frac{\prod_{j=1}^{\gamma} 1-(1-e^{-\beta})\mu_j(i)}{\sum_{h=1}^{q}\prod_{j=1}^{\gamma} 1-(1-e^{-\beta})\mu_j(h)},$$

$$\mathscr{B}_{q,\beta,d}(\pi) = \mathrm{E}\left[\frac{\Lambda(\sum_{\sigma=1}^{q}\prod_{i=1}^{\boldsymbol{\gamma}} 1-(1-e^{-\beta})\boldsymbol{\mu}_i^{(\pi)}(\sigma))}{q(1-(1-e^{-\beta})/q)^{\boldsymbol{\gamma}}} - \frac{d}{2}\frac{\Lambda(1-(1-e^{-\beta})\sum_{\sigma=1}^{q}\boldsymbol{\mu}_1^{(\pi)}(\sigma)\boldsymbol{\mu}_2^{(\pi)}(\sigma))}{1-(1-e^{-\beta})/q}\right].$$

Then $d_{\mathrm{inf}}(q,\beta) = \inf\left\{d > 0 : \mathscr{B}^*_{q,\beta}(d) > \ln q + \frac{d}{2}\ln(1-(1-e^{-\beta})/q)\right\}$.

The posterior distribution



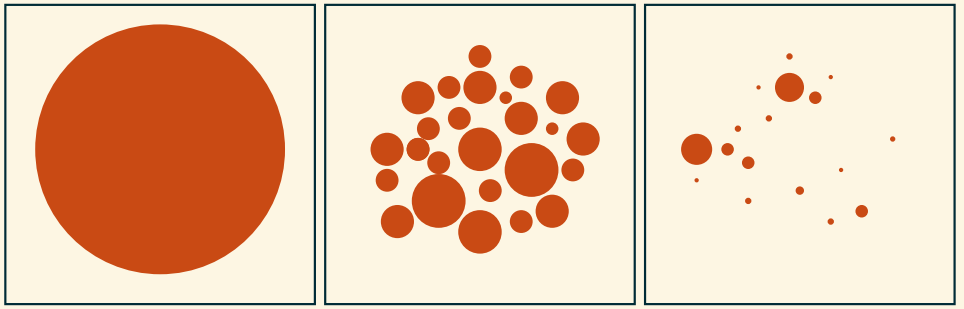
- define

$$\psi_{\mathbf{G}^*}(\sigma) = \prod_{\{v,w\}\in E(G)} \exp(-\beta \mathbf{1}\{\sigma(v) = \sigma(w)\}),$$

$$Z(\mathbf{G}^*) = \sum_{\sigma\in\Omega^V} \psi_{\mathbf{G}^*}(\sigma).$$

- then $\quad \mathrm{P}\left[\boldsymbol{\sigma}^* = \sigma | \mathbf{G}^*\right] \asymp \mu_{\mathbf{G}^*}(\sigma) = \psi_{\mathbf{G}^*}(\sigma)/Z(\mathbf{G}^*)$

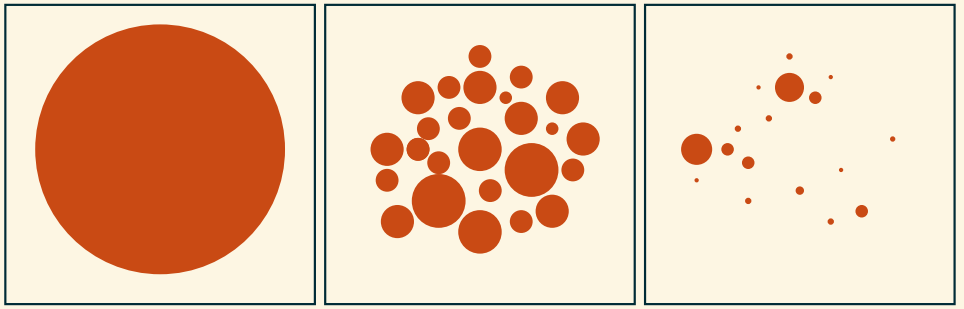
The posterior distribution



- reconstruction is impossible iff

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{v,w} \mathrm{E} \left\| \mu_{\mathbf{G}^*,v,w} - \mu_{\mathbf{G}^*,v} \otimes \mu_{\mathbf{G}^*,w} \right\|_{\mathrm{TV}} = 0$$

The posterior distribution



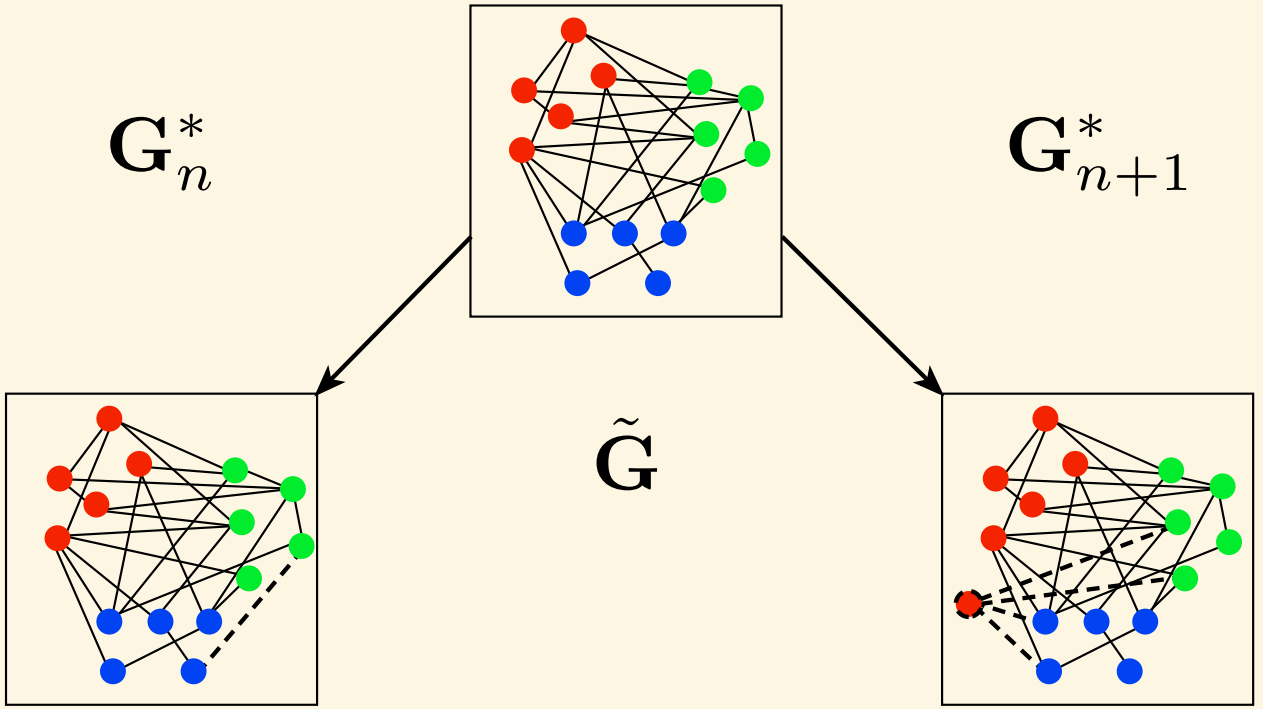
$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{v,w} \mathrm{E} \left\| \mu_{\mathbf{G}^*,v,w} - \mu_{\mathbf{G}^*,v} \otimes \mu_{\mathbf{G}^*,w} \right\|_{\mathrm{TV}} = 0$$

$$\Leftrightarrow \lim_{n\to\infty} \frac{1}{n} \mathrm{E}[\log Z(\mathbf{G}^*)] = \log q + \frac{d}{2} \log(1 - (1 - e^{-\beta})/q).$$

The Aizenman-Sims-Starr scheme

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{E}[\log Z(\mathbf{G}^*)] = \lim_{n\to\infty} \mathrm{E}\left[\log \frac{Z(\mathbf{G}^*_{n+1})}{Z(\mathbf{G}^*_n)}\right]$$

The Aizenman-Sims-Starr scheme



$$\frac{Z(\tilde{\mathbf{G}} + \overline{vw})}{Z(\tilde{\mathbf{G}})} = \sum_{\sigma,\tau \in [q]} e^{-\beta \mathbf{1}\{\sigma=\tau\}} \mu_{\mathbf{G}^*,v,w}(\sigma,\tau)$$

Correlations

- $\mathscr{X}$ = fixed finite set
- $\mu \in \mathscr{P}(\mathscr{X}^n)$ for some large integer n

Correlations

Definition

A probability measure $\mu \in \mathscr{P}(\mathscr{X}^n)$ is ε-symmetric if

$$\frac{1}{n^2} \sum_{i,j=1}^{n} \left\| \mu_{i,j} - \mu_i \otimes \mu_j \right\|_{\mathrm{TV}} < \varepsilon$$

Correlations

The magic lemma [COKPZ16]

For any $\varepsilon > 0$ there is a bounded random variable T such that for all $\mu \in \mathscr{P}(\mathscr{X}^n)$ the following is true:

- choose $\boldsymbol{U} \subset \{1, \dots, n\}$ of size T randomly
- sample $\hat{\boldsymbol{\sigma}}$ from μ
- let

$$\hat{\boldsymbol{\mu}}(\tau) = \mu[\tau | \forall i \in \boldsymbol{U} : \tau(i) = \hat{\boldsymbol{\sigma}}(i)];$$

then

$$\mathrm{P}\left[\hat{\boldsymbol{\mu}} \text{ is } \varepsilon\text{-symmetric}\right] > 1 - \varepsilon$$

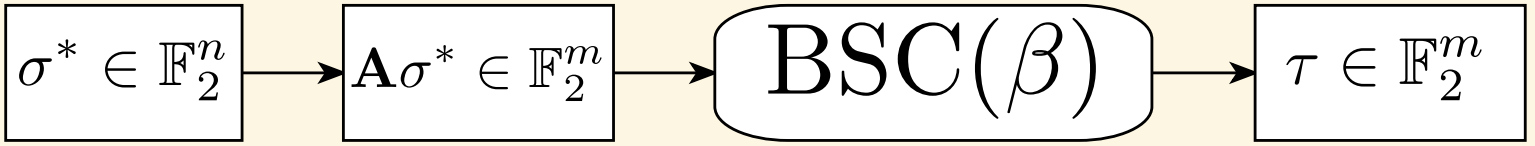
Correlations

Lemma [BCO15]

For any $\varepsilon > 0, k \geq 3$ there is $\delta > 0$ s.t. for $n > 1/\delta$ for δ-symmetric μ,

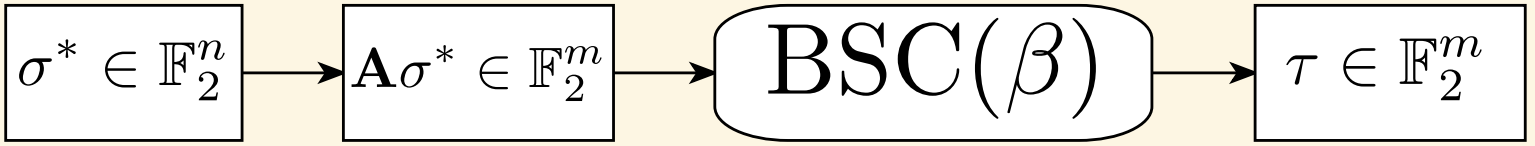
$$\frac{1}{n^k} \sum_{i_1,\ldots,i_k=1}^{n} \left\| \mu_{i_1,\ldots,i_k} - \mu_{i_1} \otimes \cdots \otimes \mu_{i_k} \right\|_{\mathrm{TV}} < \varepsilon$$

Low density generator matrix codes



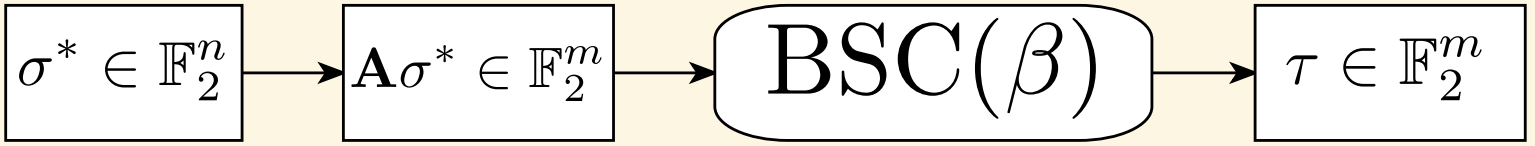
- $\boldsymbol{A} \in \mathbb{F}_2^{m \times n}$ with $k \geq 3$ ones per row
- signal $d = km/n$, noise β

Low density generator matrix codes



$$I(\boldsymbol{\sigma}^*, \boldsymbol{\tau} | \boldsymbol{A}) = \sum_{s,t} \mathrm{P}\left[\boldsymbol{\sigma}^* = s, \boldsymbol{\tau} = t\right] \log \frac{\mathrm{P}\left[\boldsymbol{\sigma}^* = s, \boldsymbol{\tau} = t\right]}{\mathrm{P}\left[\boldsymbol{\sigma}^* = s\right] \mathrm{P}\left[\boldsymbol{\tau} = t\right]}$$

Low density generator matrix codes



- non-rigorous statistical physics analysis [KS99]
- upper bound on the mutual information, even k [M05]
- existence of $\lim_{n\to\infty} \frac{1}{n} I(\boldsymbol{\sigma}^*, \boldsymbol{\tau}|\boldsymbol{A})$, even k [AM15]

Low density generator matrix codes

Theorem [CKPZ16]

For $k \geq 2$, $\beta > 0$, $d > 0$ and $\pi \in \mathscr{P}_0([-1,1])$ let

$$\mathscr{B}_{d,\beta}(\pi) = \mathrm{E}\left[\frac{1}{2}\Lambda\left(\sum_{\sigma=\pm 1}\prod_{i=1}^{r} 1 + (1-2\beta)\sigma \boldsymbol{J}_i \prod_{j=1}^{k-1} \boldsymbol{\theta}^{\pi}_{i,j}\right) - \frac{d(k-1)}{k}\Lambda\left(1 + (1-2\beta)\boldsymbol{J}\prod_{j=1}^{k}\boldsymbol{\theta}^{\pi}_{j}\right)\right]$$

Then

$$\lim_{n\to\infty}\frac{1}{n}I(\boldsymbol{\sigma}^*, \boldsymbol{\tau}|\boldsymbol{A}) = (1+d/k)\log 2 + \beta\log\beta + (1-\beta)\log(1-\beta) - \sup_{\pi\in\mathscr{P}_0([-1,1])}\mathscr{B}_{d,\beta}(\pi)$$

The information-theoretic threshold is equal to

$$d_{\mathrm{inf}}(\beta) = \inf\left\{d > 0 : \sup_{\pi\in\mathscr{P}_0([-1,1])}\mathscr{B}_{d,\beta}(\pi) > \log 2\right\}$$

Conclusions

- generalisation: the "teacher-student scheme"
- justification of the 'replica symmetric cavity method'
- other applications:
  - random graph colouring
  - Goldreich's one-way function
  - the diluted p-spin model