On the phase transitions of inference problems

Guilhem Semerjian

LPT-ENS

28.02.2017 / Les Houches

with F. Krzakala, F. Ricci-Tersenghi, L. Zdeborova

Outline

1. Introduction : the Stochastic Block Model example
2. Qualitative message
3. Quantitative results

Outline

1. Introduction : the Stochastic Block Model example
2. Qualitative message
3. Quantitative results

The Stochastic Block Model

null model for community detection

generation of a graph G on N vertices by

- drawing labels $\underline{\tau} = (\tau_1, \dots, \tau_N) \in \{1, \dots, q\}^N$ i.i.d. with proba $\overline{\eta}_\tau$
- for each possible edge $\langle i, j \rangle$, include it in G with proba $\frac{1}{N} c_{\tau_i, \tau_j}$

Parameters :

- $q \geq 2$, the number of "communities"
- $\overline{\eta}$, a probability law on $\{1, \dots, q\}$ (prior on the communities)
- c, a $q \times q$ symmetric matrix (affinities)

Inference problem : infer $\underline{\tau}$ from G

- Assumption : $c = \sum_\sigma c_{\tau,\sigma}\,\overline{\eta}_\sigma$ independent of τ:
  same average degree c in all communities,
  i.e. no information on τ_i contained in the degree of i

- Symmetric case :

$$\overline{\eta}_\tau = \frac{1}{q}, \qquad c_{\tau,\sigma} = \begin{cases} c_{\rm in} & \text{if } \tau = \sigma \\ c_{\rm out} & \text{otherwise} \end{cases}$$

  Parameters : c and $\theta = \frac{c_{\rm in} - c_{\rm out}}{qc}$

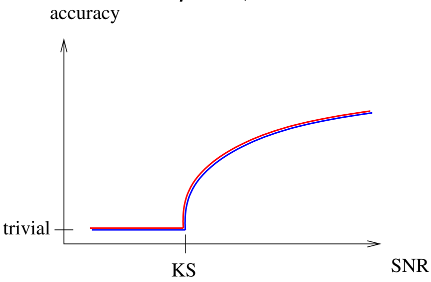
  $\theta < 0$: disassortative, $\theta > 0$: assortative
  $\theta = 0$: pure Erdos-Renyi

- Signal to Noise Ratio : c at fixed $|\theta|$, or $|\theta|$ at fixed c
- Accuracy : some distance (overlap) between $\underline{\tau}$ and estimator $\widehat{\underline{\tau}}(G)$

Conjectured phase transitions for the symmetric SBM

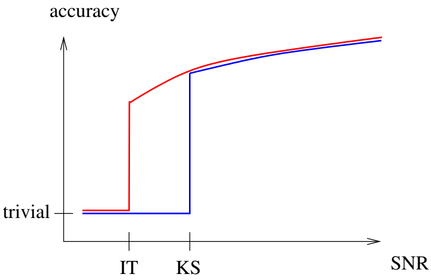
[Decelle, Krzakala, Moore, Zdeborova 11]

$q = 2, 3$



impossible / easy

$q \geq 5$



impossible / hard / easy

optimal (Bayes) easily achievable (BP)

Kesten-Stigum (KS) and Information Theoretic (IT) transitions partially proven rigorously

[Massoulié 14]
[Mossel, Neeman, Sly 14]
[Bordenave, Lelarge, Massoulié 15]
[Abbe, Sandon 16]
[Coja-Oghlan, Krzakala, Perkins, Zdeborova 16]

these two scenari found in many other inference problems, notably:

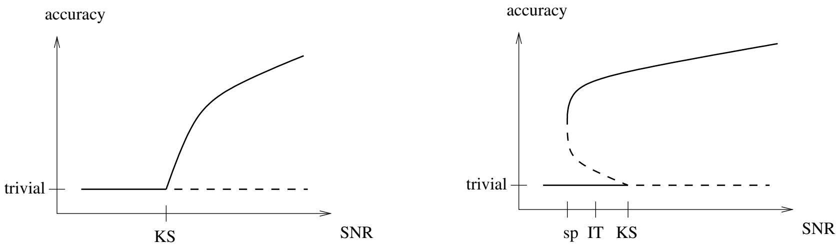
- low rank matrix factorization problems (dense version of SBM)
- planted CSPs

Outline

1. Introduction : the Stochastic Block Model example
2. Qualitative message
3. Quantitative results

Typology of phase transitions

Interpretation of the previous plots as bifurcation diagrams :

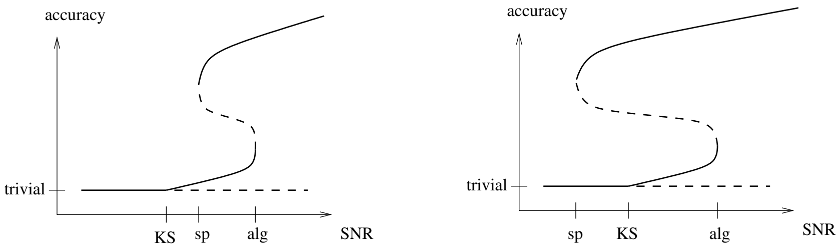


Solid / dashed : stable / unstable fixed point
(of BP/AMP, or cavity method/state evolution)

- KS : instability of trivial fixed point
- sp : spinodal for disappearance of non-trivial fixed point
- IT : crossing of free-energies associated to the two fixed points

Typology of phase transitions

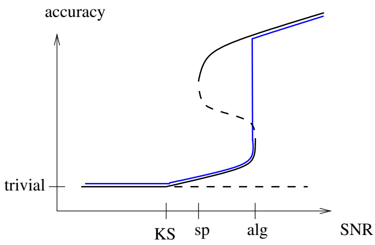
Slightly more complicated bifurcation scenario :

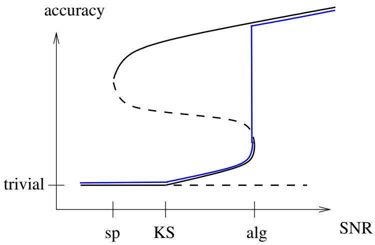


alg for algorithmic spinodal

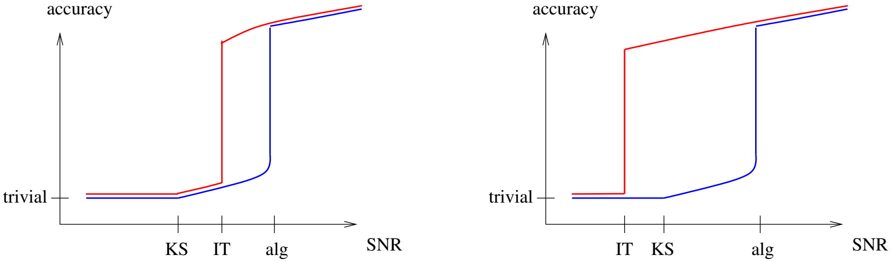
Where is the IT transition ? Which curves are blue and red ?

Easily achievable : smallest stable fixed point





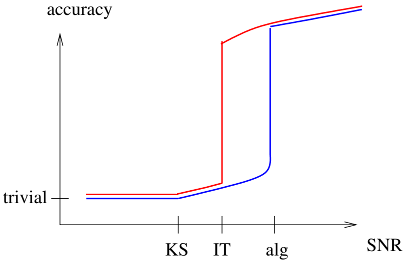
Typology of phase transitions



KS < IT	IT < KS
impossible / easy / "easy" / easy	impossible / hard / "easy" / easy

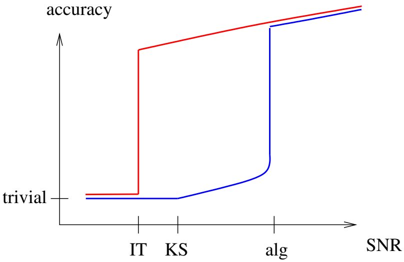
- impossible : to beat trivial accuracy
- easy : to achieve optimal accuracy
- "easy" : to beat trivial accuracy, but hard to achieve optimal one
- hard : to beat trivial accuracy

Typology of phase transitions



KS < IT

impossible / easy / “easy” / easy



IT < KS

impossible / hard / “easy” / easy

- Found in [Lesieur, Krzakala, Zdeborova 17]
  for matrix factorization with $\{0, +1, -1\}$ prior
- mentioned in [Lelarge, Miolane 16]

Outline

1. Introduction : the Stochastic Block Model example
2. Qualitative message
3. Quantitative results

Scalar bifurcations

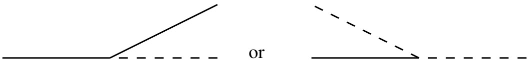
Analytical expansions can determine (perturbatively) the bifurcation diagrams

- Bifurcation analysis of a scalar recursion
  $q^{t+1} = V(q^t, \theta)$ (with $q \geq 0$)
- $V(0, \theta) = 0$ for all θ : trivial, uninformative fixed point
- $V(q, \theta) = a(\theta)q + b(\theta)q^2 + c(\theta)q^3 + \dots$
- linearized analysis, location of Kesten-Stigum found with $a(\theta_{\mathrm{KS}}) = 1$:

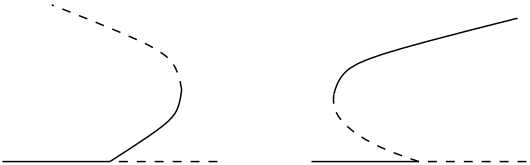
Scalar bifurcations

- $a(\theta_{\mathrm{KS}}) = 1$:

- next order, depending on the sign of $b(\theta_{\mathrm{KS}})$:



- next next order, depending on the signs of $b(\theta_{\mathrm{KS}})$ and $c(\theta_{\mathrm{KS}})$, one can also have :



hence θ_{alg} and θ_{sp} (when they are close to θ_{KS})

Bifurcation of the uninformative fixed point in inference problems

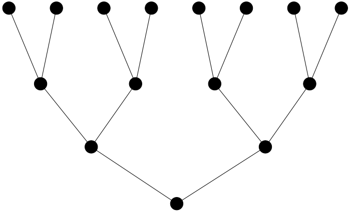
- for fully-connected models analysed with AMP/state evolution, scalar or low-dimensional recursions on $q^{t+1} = V(q^t)$
- for sparse problems like the SBM, functional recursion on a probability distribution

In the SBM, $(\underline{\tau}, G)$ converges locally to

- a Galton Watson tree with Poisson offspring distribution
- on which labels τ_i are Markovian (broadcast)

Expansions around Kesten-Stigum on a tree

Inference problem of the SBM is thus tightly liked to the tree (robust) reconstruction problem



[Janson, Mossel 04]
[Mézard, Montanari 06]
[Sly 08]

functional recursion $P^{t+1} = V(P^t, \theta)$

$P(\eta) = \delta(\eta - \overline{\eta})$ its trivial fixed point

$$P^{(t+1)}(\eta) = \sum_{\ell=0}^{\infty} q_\ell \int \mathrm{d}P^{(t)}(\eta^1) \dots \mathrm{d}P^{(t)}(\eta^\ell)\, \delta(\eta - f(\eta^1, \dots, \eta^\ell))\, z(\eta^1, \dots, \eta^\ell)$$

fortunately can be reduced to a finite dimensional bifurcation analysis on the low order moments

Expansions around Kesten-Stigum on a tree

$$\delta_\sigma = \eta_\sigma - \overline{\eta}_\sigma \ , \qquad \widehat{\delta}_\sigma = \sum_{\sigma'} M_{\sigma\sigma'} \frac{1}{\overline{\eta}_{\sigma'}} \delta_{\sigma'}$$

$$a_{\sigma\tau} = \mathbb{E}[\delta_\sigma \delta_\tau] \ , \qquad \widehat{a}_{\sigma\tau} = \mathbb{E}[\widehat{\delta}_\sigma \widehat{\delta}_\tau]$$

$$\begin{aligned} a_{\sigma\tau} &= \mathbb{E}[\ell] \, \overline{\eta}_\sigma \overline{\eta}_\tau \, \widehat{a}_{\sigma\tau} \\ &+ \frac{1}{2} \mathbb{E}[\ell(\ell-1)] \, \overline{\eta}_\sigma \overline{\eta}_\tau \left[\widehat{a}^2_{\sigma\tau} - \sum_\gamma \overline{\eta}_\gamma (\widehat{a}_{\sigma\gamma} + \widehat{a}_{\gamma\tau})^2 + \sum_{\gamma\beta} \overline{\eta}_\gamma \overline{\eta}_\beta \widehat{a}^2_{\gamma\beta} \right] \end{aligned}$$

Expansions around Kesten-Stigum on a tree

$$\delta_\sigma = \eta_\sigma - \overline{\eta}_\sigma \,, \qquad a_{\sigma\tau} = \mathbb{E}[\delta_\sigma \delta_\tau] \,, \qquad b_{\sigma\tau\gamma} = \mathbb{E}[\delta_\sigma \delta_\tau \delta_\gamma] \,, \qquad c_{\sigma\tau\gamma\beta} = \mathbb{E}[\delta_\sigma \delta_\tau \delta_\gamma \delta_\beta]$$

$$\begin{aligned}
a_{\sigma\tau} &= \mathbb{E}[\ell]\overline{\eta}_\sigma\overline{\eta}_\tau\hat{a}_{\sigma\tau} + \frac{1}{2}\mathbb{E}[\ell(\ell-1)]\overline{\eta}_\sigma\overline{\eta}_\tau\left[\hat{a}^2_{\sigma\tau} - \sum_\gamma \overline{\eta}_\gamma(\hat{a}_{\sigma\gamma}+\hat{a}_{\gamma\tau})^2 + \sum_{\gamma\beta}\overline{\eta}_\gamma\overline{\eta}_\beta\hat{a}^2_{\gamma\beta}\right] \\
&+ \mathbb{E}[\ell(\ell-1)]\overline{\eta}_\sigma\overline{\eta}_\tau\left[-\sum_\gamma\overline{\eta}_\gamma\hat{b}_{\sigma\tau\gamma}(\hat{a}_{\sigma\gamma}+\hat{a}_{\tau\gamma}) + \sum_{\gamma\beta}\overline{\eta}_\gamma\overline{\eta}_\beta(\hat{b}_{\sigma\gamma\beta}+\hat{b}_{\tau\gamma\beta})\hat{a}_{\gamma\beta} + \sum_{\gamma\beta}\overline{\eta}_\gamma\overline{\eta}_\beta\hat{c}_{\sigma\tau\gamma\beta}\hat{a}_{\gamma\beta}\right] \\
&+ \mathbb{E}[\ell(\ell-1)(\ell-2)]\overline{\eta}_\sigma\overline{\eta}_\tau\left[\frac{1}{6}\hat{a}^3_{\sigma\tau} - \frac{1}{6}\sum_\gamma\overline{\eta}_\gamma(\hat{a}_{\sigma\gamma}+\hat{a}_{\tau\gamma})^3 - \frac{1}{2}\hat{a}_{\sigma\tau}\sum_\gamma\overline{\eta}_\gamma(\hat{a}_{\sigma\gamma}+\hat{a}_{\tau\gamma})^2 + \frac{1}{6}\sum_{\gamma\beta}\overline{\eta}_\gamma\overline{\eta}_\beta\hat{a}^3_{\gamma\beta}\right. \\
&\qquad + \frac{1}{2}\sum_{\gamma\beta}\overline{\eta}_\gamma\overline{\eta}_\beta\hat{a}^2_{\gamma\beta}(\hat{a}_{\sigma\tau}+\hat{a}_{\sigma\gamma}+\hat{a}_{\tau\gamma}+\hat{a}_{\sigma\beta}+\hat{a}_{\tau\beta}) + \sum_{\gamma\beta}\overline{\eta}_\gamma\overline{\eta}_\beta\hat{a}_{\gamma\beta}(\hat{a}_{\sigma\gamma}+\hat{a}_{\tau\gamma})(\hat{a}_{\sigma\beta}+\hat{a}_{\tau\beta}) \\
&\qquad \left. - \sum_{\gamma\beta\alpha}\overline{\eta}_\gamma\overline{\eta}_\beta\overline{\eta}_\alpha\hat{a}_{\gamma\beta}\hat{a}_{\beta\alpha}\hat{a}_{\alpha\gamma}\right] \\
b_{\sigma\tau\gamma} &= \mathbb{E}[\ell]\overline{\eta}_\sigma\overline{\eta}_\tau\overline{\eta}_\gamma\hat{b}_{\sigma\tau\gamma} \\
&+ \mathbb{E}[\ell(\ell-1)]\overline{\eta}_\sigma\overline{\eta}_\tau\overline{\eta}_\gamma\left[\hat{a}_{\sigma\tau}\hat{a}_{\tau\gamma}+\hat{a}_{\sigma\gamma}\hat{a}_{\gamma\tau}+\hat{a}_{\tau\sigma}\hat{a}_{\sigma\gamma} - \sum_\beta\overline{\eta}_\beta(\hat{a}_{\sigma\beta}\hat{a}_{\beta\gamma}+\hat{a}_{\sigma\beta}\hat{a}_{\beta\tau}+\hat{a}_{\tau\beta}\hat{a}_{\beta\gamma})\right] \\
c_{\sigma\tau\gamma\beta} &= \mathbb{E}[\ell]\overline{\eta}_\sigma\overline{\eta}_\tau\overline{\eta}_\gamma\overline{\eta}_\beta\hat{c}_{\sigma\tau\gamma\beta} + \mathbb{E}[\ell(\ell-1)]\overline{\eta}_\sigma\overline{\eta}_\tau\overline{\eta}_\gamma\overline{\eta}_\beta(\hat{a}_{\sigma\tau}\hat{a}_{\gamma\beta}+\hat{a}_{\sigma\gamma}\hat{a}_{\tau\beta}+\hat{a}_{\sigma\beta}\hat{a}_{\tau\gamma})
\end{aligned}$$

Application 1 : SBM with 2 asymmetric communities

$q = 2, \qquad \overline{\eta}_1 = \frac{1+\overline{m}}{2}, \qquad \overline{\eta}_2 = \frac{1-\overline{m}}{2}$

a.k.a. reconstruction of the asymmetric Ising model, $\overline{m}$: magnetization

Previous results :

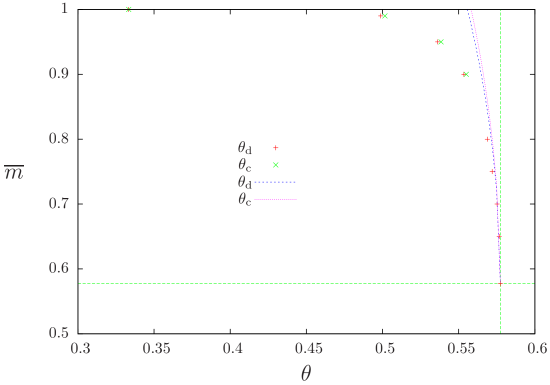
- KS is tight for small $|\overline{m}|$ [Borgs, Chayes, Mossel, Roch 06]
- KS is not tight for $|\overline{m}|$ close to 1 [Mossel 01]

new conjecture on the critical asymmetry : $\overline{m}_c = \frac{1}{\sqrt{3}}$

independence of the critical asymmetry on the degree distribution
hence $\overline{m}_c$ coincides with

- a fully connected equivalent problem
  [Barbier, Dia, Macris, Krzakala, Lesieur, Zdeborova 16]
- the large degree limit result [Caltagirone, Lelarge, Miolane 16]

Application 1 : SBM with 2 asymmetric communities



+ analytical expansions of the spinodal and IT line for $\overline{m} \to \overline{m}_{\rm c}^{+}$

Application 2 : symmetric SBM with $q = 4$

recall :

- $q = 2, 3$ has a continuous transition
- $q \geq 5$ has a discontinuous transition [Sly 08]
- $q = 4$ is marginal : first non-linear term at KS proportional to $q - 4$

the cubic term in the expansion yields

$$-\frac{7}{3}\frac{\mathbb{E}[\ell(\ell-1)(\ell-2)]}{\mathbb{E}[\ell]^3} + 3\left(\frac{\mathbb{E}[\ell(\ell-1)]}{\mathbb{E}[\ell]^2}\right)^2 \left(\frac{9}{7}\frac{1}{\mathbb{E}[\ell]-1} - 4\frac{1}{\mathrm{sign}(\theta)\sqrt{\mathbb{E}[\ell]}-1}\right)$$

analysis of the sign of this expression :

- in the assortative case ($\theta > 0$), continuous transition
- in the disassortative case ($\theta < 0$)
  - for small degrees, discontinuous transition
  - for large degrees, continuous transition

Poisson degree distribution : changes at $c \approx 21.1767$

For d-regular trees (d offspring), changes between $d \leq 23$ and $d \geq 24$

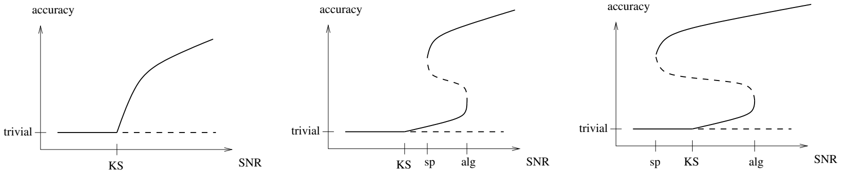
Application 3 : planted occupation models CSPs

i.e. $w(\sigma_1, \ldots, \sigma_k)$ depends only on $\sum \sigma_i$, for $\sigma_i = \pm 1$

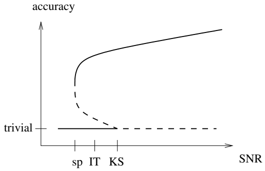
- if trivial message is a fixed point ($\Leftarrow$ balanced, $w(-\underline{\sigma}) = w(\underline{\sigma})$)
- and KS happens at finite degree (no 2-wise independence)

then the Kesten-Stigum transition is always continuous,

i.e. possible scenari :

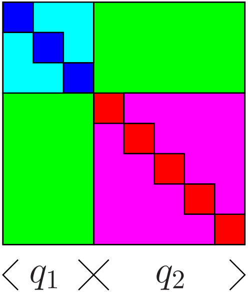


but no :



Application 4 : SBM with $q = q_1 + q_2$ communities

$c =$



symmetric inside each of the two “mega-communities”

should exhibit all possible scenari mentioned above

(work in progress)