Phase transitions in inference problems on sparse random graphs

Guilhem Semerjian

LPT-ENS

22.08.2018 / Cargese

with F. Ricci-Tersenghi and L. Zdeborova

arxiv:1806.11013

Outline

Outline

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, Bayesian posterior $\mathbb{P}(\underline{\tau}|G)$

The Stochastic Block Model

- Balanced 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

- Particular (symmetric) case :

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

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

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

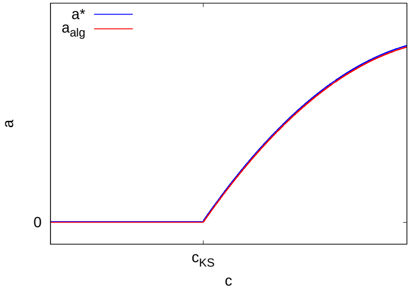
- Signal to Noise Ratio : c at fixed $|\theta|$, or $|\theta|$ at fixed c

- Accuracy : some distance (overlap) between $\underline{\tau}$ and estimator $\widehat{\underline{\tau}}(G)$

Conjectured phase transitions for the symmetric SBM

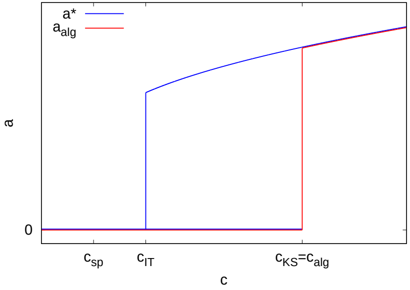
[Decelle, Krzakala, Moore, Zdeborova 11]

$q = 2, 3$



impossible / easy

$q \geq 5$



impossible / hard / easy

optimal (Bayes) easily achievable (BP)

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

Graph inference problems

partially proven rigorously

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

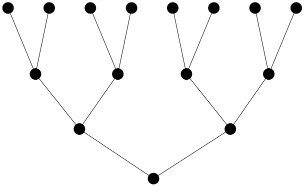
these two scenarios found in many other inference problems, notably:

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

Outline

The (robust) reconstruction problem on trees

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



- Draw the label τ of the root of a tree with proba $\overline{\eta}$
- Broadcast on each edge with Markov transition probability $M_{\tau\tau'}$

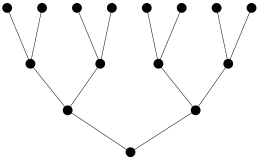
Assumption : M is reversible with respect to $\overline{\eta}$, irreducible, aperiodic

Observations: the labels at distance t from the root, $\underline{\tau}_{V_t}$

Can one guess τ as $t \to \infty$ better than with $\overline{\eta}$?

i.e. does $\underline{\tau}_{V_t}$ carry information on τ ?

The (robust) reconstruction problem on trees



posterior probability $\eta_\tau = \mathbb{P}(\tau | \underline{\tau}_{V_t})$

tree structure $\to$ recursive computation

dynamic programming, Belief Propagation

$P^{(t)}(\eta)$ its distribution, with respect to

- broadcast process
- Galton-Watson tree with offspring probability law q_ℓ

obeys the functional recurrence $P^{(t+1)} = V(P^{(t)}, \text{parameters})$

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

with initial condition : $\quad P^{(t=0)}(\eta) = \sum_\tau \overline{\eta}_\tau \delta(\eta - \delta_\tau)$

The (robust) reconstruction problem on trees

functional recurrence $P^{(t+1)} = V(P^{(t)}, \text{parameters})$

$P_{\text{triv}}(\eta) = \delta(\eta - \overline{\eta})$ trivial (uninformative) fixed point, $P_{\text{triv}} = V(P_{\text{triv}})$

reconstruction question : does $P^{(t)} \to P_{\text{triv}}$ as $t \to \infty$ or to a non-trivial fixed point ?

Robust variant : one observes a fraction ε of the vertices at distance t, with $\varepsilon \to 0$ after $t \to \infty$

idem with $$P^{(t=0)}(\eta) = \varepsilon \sum_{\tau} \overline{\eta}_\tau \delta(\eta - \delta_\tau) + (1-\varepsilon)\,\delta(\eta - \overline{\eta})$$

Outline

The connection between the SBM and the tree reconstruction

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

- a Galton Watson tree with offspring distribution Poisson(c)
- on which labels τ_i are broadcast with $M_{\tau\tau'} = c_{\tau\tau'}\eta_{\tau'}/c$

But no observation of the labels in the graph problem. . .

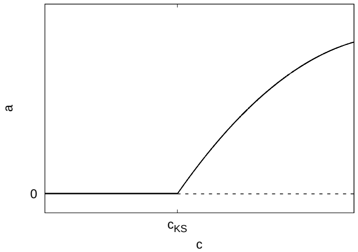
Global connection is subtle, conjectures (cavity method) :

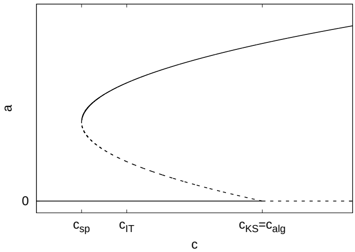
- Efficiently achievable accuracy corresponds to robust reconstruction fixed point
- Mutual information between $\underline{\tau}$ and G (hence information theoretically optimal accuracy) related to $\sup \phi(P)$ over all fixed points, $\phi(P)$ functional free-entropy

Outline

Typology of phase transitions

Interpretation of the previous plots as bifurcation diagrams :



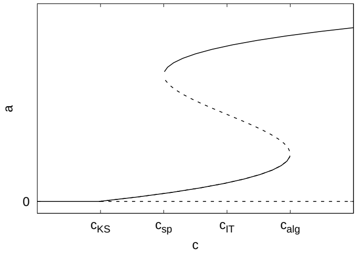


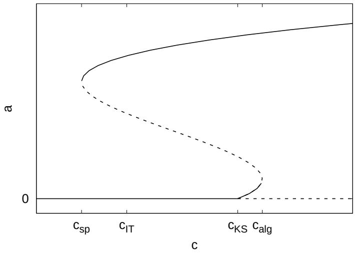
Solid / dashed : stable / unstable fixed point (of $P = V(P)$)

- KS : instability of trivial fixed point / robust reconstruction transition
- sp : spinodal for existence of non-trivial fixed point / reconstruction transition
- IT : crossing of the free-entropies $\phi(P)$ of the two fixed points

Typology of phase transitions

Slightly more complicated bifurcation scenario :

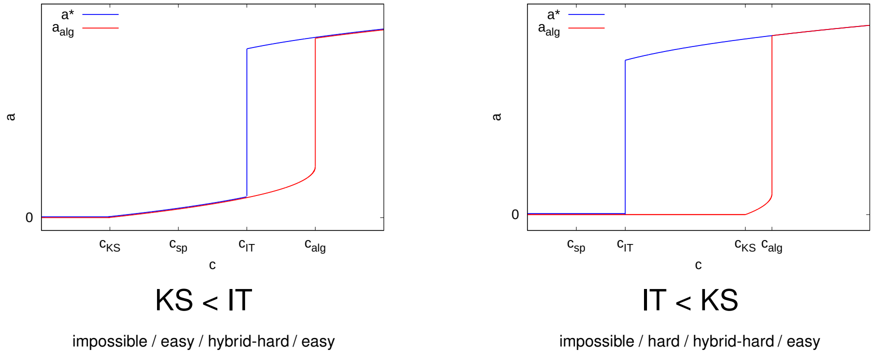




alg for algorithmic spinodal

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

Typology of phase transitions



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

matrix factorization with $\{0, \pm 1\}$ prior[Lesieur, Krzakala, Zdeborova 17]

Outline

Scalar bifurcations

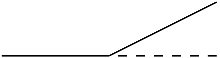
Analytical expansions can determine (perturbatively) the bifurcation diagrams

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

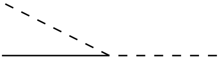
Scalar bifurcations

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

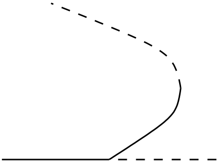
- next order, depending on the sign of $b(\theta_{\rm KS})$:

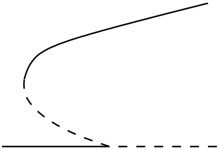


or



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





hence $\theta_{\rm alg}$ and $\theta_{\rm sp}$ (when they are close to $\theta_{\rm KS}$)

Bifurcation of the uninformative fixed point in inference problems

- for "fully-connected models" (dense or large degree limit), scalar or low-dimensional recursions on $q^{t+1} = V(q^t)$
- for sparse problems like the SBM/tree reconstruction, functional recursion on a probability distribution $P^{(t+1)} = V(P^{(t)})$

infinite-dimensional bifurcation, but $P(\eta) \approx P_{\rm triv}(\eta) = \delta(\eta - \overline{\eta})$, hence one can close it on a finite number of moments

[Borgs, Chayes, Mossel, Roch 06]
[Sly 08]

Also in [Coja-Oghlan, Efthymiou, Jaafari, Kang, Kapetanopoulos 18]
expansion of $\phi(P)$ for $P \approx P_{\rm triv}$

Expansions around KS for tree reconstruction

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

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

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

Expansions around KS for tree reconstruction

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

$$\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

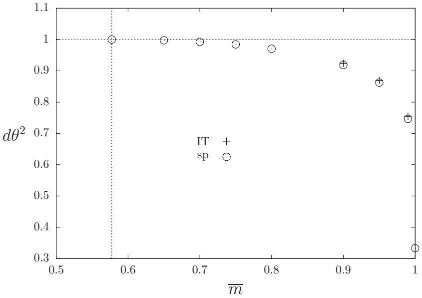
Previous results :

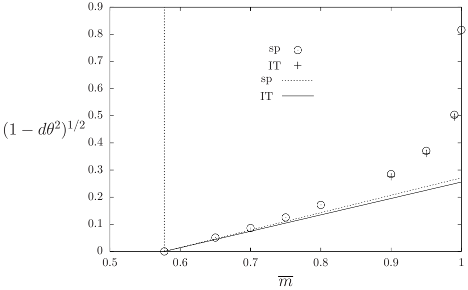
- KS is tight for small $|\overline{m}|$ [Borgs, Chayes, Mossel, Roch 06]
- KS is not tight for $|\overline{m}|$ close to 1 [Mossel 01]

new conjecture on the critical asymmetry : $\overline{m}_{\rm c} = \frac{1}{\sqrt{3}}$ [Liu, Ning 18]

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

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





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

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 $q - 4$

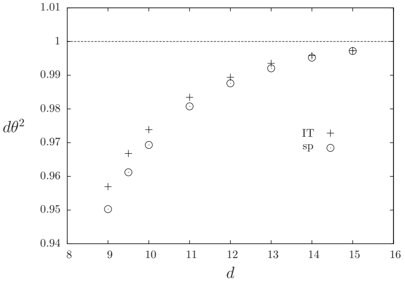
the cubic term in the expansion yields

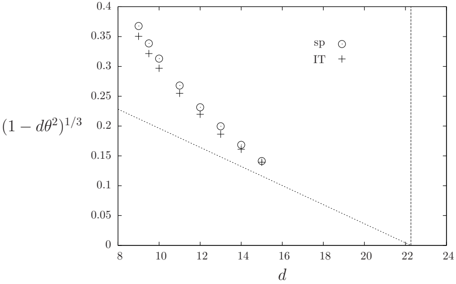
$$-\frac{7}{3}\frac{\mathbb{E}[\ell(\ell-1)(\ell-2)]}{\mathbb{E}[\ell]^3} + \left(\frac{\mathbb{E}[\ell(\ell-1)]}{\mathbb{E}[\ell]^2}\right)^2 \left(\frac{5}{\mathbb{E}[\ell]-1} - \frac{12}{\text{sign}(\theta)\sqrt{\mathbb{E}[\ell]}-1}\right)$$

analysis of the sign of this expression :

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

Poisson degree distribution : changes at $d \approx 22.2694$





For d-regular trees (d offspring), changes between $d \leq 24$ and $d \geq 25$

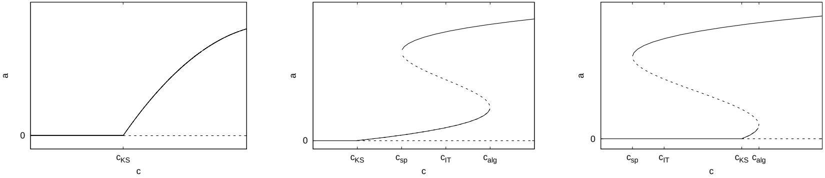
Application 3 : planted symmetric CSPs

$\tau_i = \pm 1$ with proba $1/2$
hyperedges between k variables with proba dependent on $\sum \tau_i$

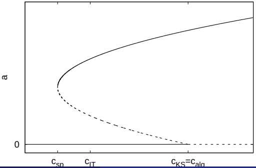
- if symmetry $+/-$
- and KS happens at finite degree (no 2-wise independence)

then the Kesten-Stigum transition is always continuous,

i.e. possible scenarios :



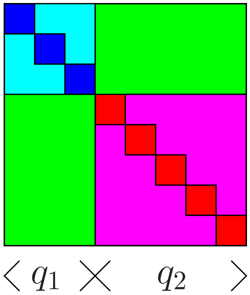
but no :



contrary to what was seen previously, but c_{alg} very close to c_{KS}

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

Simplest pattern to break the permutation symmetry between q states :

$c =$ 

symmetric inside each of the two groups

exhibits the hybrid-hard phase for some parameters