

POLITECNICO
MILANO 1863

Graph Colouring and Belief Propagation

Seminary for the course of Physic of Complex System

Physics Engineering

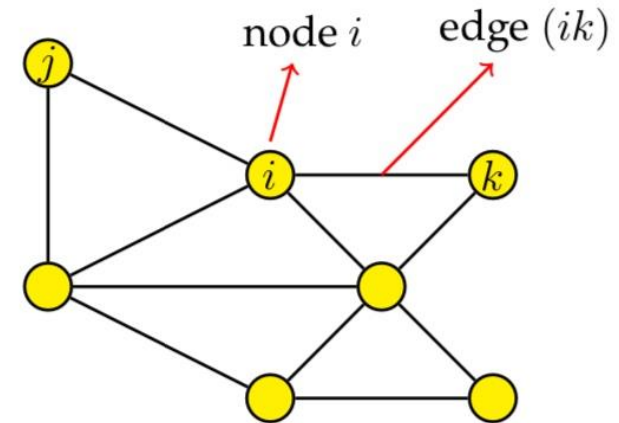
Author: D. Patricelli

Professor Paolo Biscari

Graph

A graph $G(V, E)$ is defined by:

- A set of nodes V , which we will index by $i \in V$.
- A set of edges E , which we will index by a pair of nodes $(ij) \in E$.



Neighbour and degree of a node

$$\partial i := \{j \in V \mid (ij) \in E\} \quad d_i = |\partial i|$$

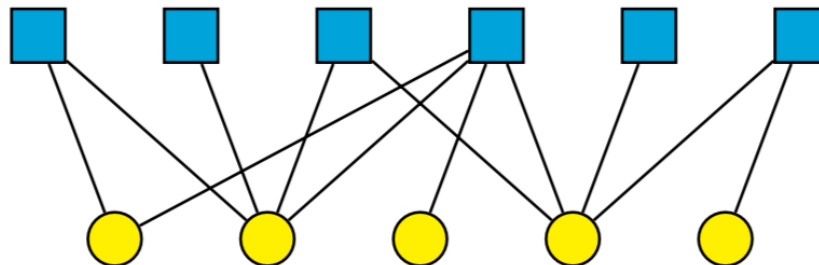
$$d_i = \sum_{j=1}^N A_{ij} = \sum_{j=1}^N \mathbb{1}_{(ij) \in E}.$$

Adjacency Matrix

$$A_{ij} = \begin{cases} 1 & \text{if } (ij) \in E \\ 0 & \text{if } (ij) \notin E \end{cases}$$

Factor Graph

A factor graph is bipartite: only edges of type 'circle'—'square' exist, no 'circle'—'circle' edges and no 'square'—'square' edges.



$$\begin{aligned}\partial i &:= \{a \mid (ia) \in E\}, & |\partial i| &= d_i, & \forall i \in \{1, \dots, N\} \\ \partial a &:= \{i \mid (ia) \in E\}, & |\partial a| &= d_a, & \forall a \in \{1, \dots, M\}\end{aligned}$$

- Every variable node represents a *variable* $s_i \in \Lambda$.
- Every factor node represents a *non-negative function* $f_a(\{s_i\}_{i \in \partial a})$.

Factor Graph

A graphical model represents a joint probability distribution over the variables $\{s_i\}_{i=1}^N$:

$$P\left(\{s_i\}_{i=1}^N\right) = \frac{1}{Z} \prod_{i=1}^N g_i(s_i) \prod_{a=1}^M f_a(\{s_i\}_{i \in \partial a})$$

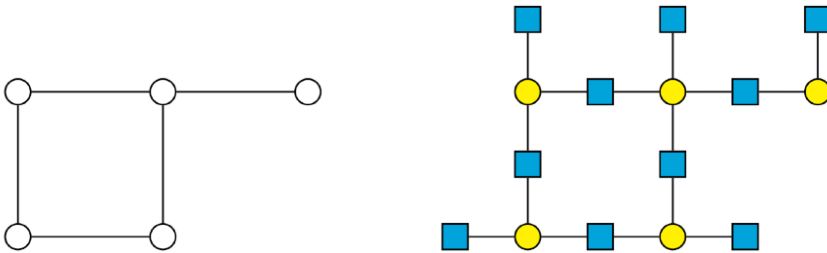
$$\mu_i(s_i) \equiv \sum_{\substack{\{s_j\}_{j=1}^N \\ j \neq i}} P\left(\{s_j\}_{j=1}^N\right)$$

$$Z \equiv \sum_{\{s_i\}_{i=1}^N} \prod_{i=1}^N g_i(s_i) \prod_{a=1}^M f_a(\{s_i\}_{i \in \partial a}) .$$

$$N\Phi = \log Z$$

Compute those quantities of interest is a really Hard task!!

Spin glass:



N Ising spins $s_i \in \Lambda = \{-1, +1\}$.

$$\mathcal{H}(\{s_i\}_{i=1}^N) = - \sum_{(ij) \in E} J_{ij} s_i s_j - \sum_i h_i s_i$$

$$P\left(\{s_i\}_{i=1}^N\right) = \frac{1}{Z_N} e^{-\beta \mathcal{H}(\{s_i\}_{i=1}^N)} = \frac{1}{Z_N} \prod_{i=1}^N e^{\beta h_i s_i} \prod_{(ij) \in E} e^{\beta J_{ij} s_i s_j},$$

Graph Coloring



$$P\left(\{s_i\}_{i=1}^N\right) = \frac{1}{Z_N} \prod_{(ij) \in E} (1 - \delta_{s_i, s_j})$$

q colors $s_i \in \{\text{red, blue, green, yellow, } \dots, \text{black}\} = \{1, 2, \dots, q\}$

$$Z_N = \sum_{\{s_i\}_{i=1}^N} \prod_{(ij) \in E} (1 - \delta_{s_i, s_j})$$

The graph is colorable if and only if $Z_N \geq 1$.

Softened Model

$$P\left(\{s_i\}_{i=1}^N, \beta\right) = \frac{1}{Z_N(\beta)} \prod_{(ij) \in E} e^{-\beta \delta_{s_i, s_j}}$$

- $\beta < 0$ attractive FM case
- $\beta > 0$ repulsive AF case
- $\beta \rightarrow \infty$ recover the original formulation

Compute the Free Entropy Density: Belief Propagation

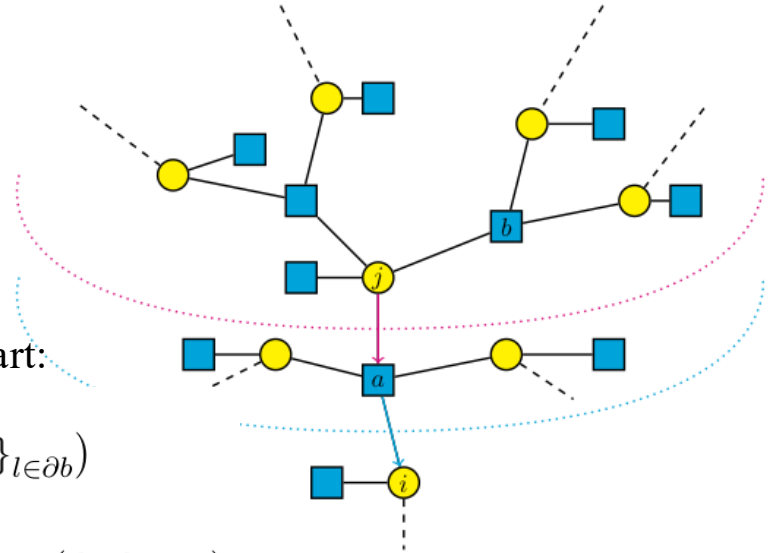
We want to compute the partition function:

$$Z \equiv \sum_{\{s_i\}_{i=1}^N} \prod_{i=1}^N g_i(s_i) \prod_{a=1}^M f_a(\{s_i\}_{i \in \partial a}) .$$

Define two auxiliary partition functions, obtained fixing some node of the system and considering the remaining part:

$$R_{s_j}^{j \rightarrow a} = g_j(s_j) \sum_{\{s_k\}_{\text{all } k \text{ above } j}} \prod_{\text{all } k \text{ above } j} g_k(s_k) \prod_{\text{all } b \text{ above } j} f_b(\{s_l\}_{l \in \partial b})$$

$$V_{s_i}^{a \rightarrow i} = \sum_{\{s_j\}_{\text{all } j \text{ above } a}} f_a(\{s_k\}_{k \in \partial a}) \prod_{\text{all } j \text{ above } a} g_j(s_j) \prod_{\text{all } b \text{ above } a} f_b(\{s_k\}_{k \in \partial b})$$



For a Tree structure, once a node is fixed, the following branches are independent, thus one can rearrange the terms and express R and V as functions of each other, this hints a recursive procedure.

$$Z = \sum_{s_j} g_j(s_j) \prod_{b \in \partial j} V_{s_j}^{b \rightarrow j} .$$

Messages

$$\chi_{s_j}^{j \rightarrow a} \equiv \frac{R_{s_j}^{j \rightarrow a}}{\sum_s R_s^{j \rightarrow a}} \quad \text{so that} \quad \sum_s \chi_s^{j \rightarrow a} = 1, \quad \forall (ja) \in E$$

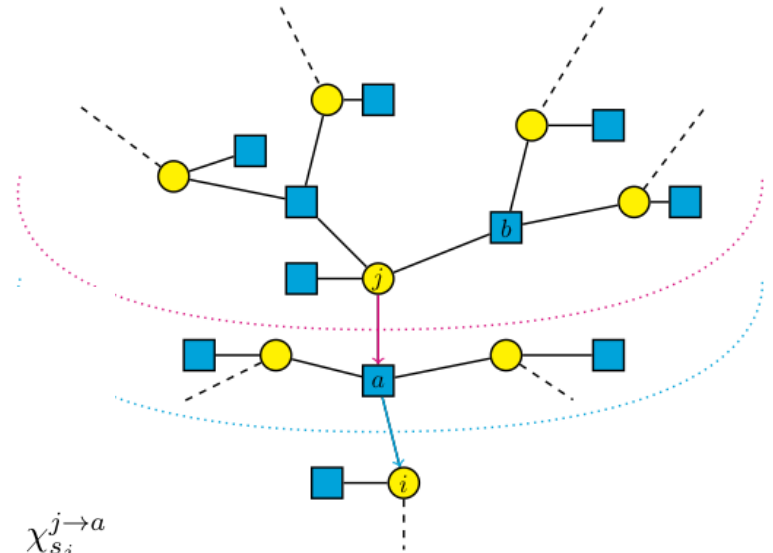
$$\psi_{s_i}^{a \rightarrow i} \equiv \frac{V_{s_i}^{a \rightarrow i}}{\sum_s V_s^{a \rightarrow i}} \quad \text{so that} \quad \sum_s \psi_s^{a \rightarrow i} = 1, \quad \forall (ia) \in E.$$

The Belief Propagation equations read

$$\chi_{s_j}^{j \rightarrow a} = \frac{1}{Z_{j \rightarrow a}} g_j(s_j) \prod_{b \in \partial j \setminus a} \psi_{s_j}^{b \rightarrow j}$$

$$\psi_{s_i}^{a \rightarrow i} = \frac{1}{Z_{a \rightarrow i}} f_a(\{s_j\}_{j \in \partial a}) \prod_{j \in \partial a \setminus i} \chi_{s_j}^{j \rightarrow a}$$

$$\mu_i(s_i) = \frac{1}{Z_i} g_i(s_i) \prod_{a \in \partial i} \psi_{s_i}^{a \rightarrow i}$$



In a tree one can start from the leafs, initialize the message sthat will depend only on the node, and then iteratively spread them to the root, and find all of them.

Bethe Free Entropy

The free entropy density Φ , which is exact on trees and is called the Bethe free entropy on more general graphs, reads

$$\begin{aligned} N\Phi = \log Z &= \sum_{i=1}^N \log Z^i + \sum_{a=1}^M \log Z^a - \sum_{(ia)} \log Z^{ia} \\ Z^i &\equiv \sum_s g_i(s) \prod_{a \in \partial i} \psi_s^{a \rightarrow i} \\ Z^a &\equiv \sum_{\{s_i\}_{i \in \partial a}} f_a(\{s_i\}_{i \in \partial a}) \prod_{i \in \partial a} \chi_{s_i}^{i \rightarrow a} \\ Z^{ia} &\equiv \sum_s \chi_s^{i \rightarrow a} \psi_s^{a \rightarrow i} \end{aligned}$$

The Bethe Free Entropy depends by definition on the messages, a nice property is that that at the BP fixed point, i.e. when the messages obey BP equations, Bethe Free entropy is stationary so for instance:

$$\frac{d\Phi_{\text{Bethe}}(\beta)}{d\beta} = \underbrace{\frac{\partial \Phi_{\text{Bethe}}}{\partial \chi}}_{=0 \text{ at BP fixed point}} \cdot \frac{\partial \chi}{\partial \beta} + \frac{\partial \Phi_{\text{Bethe}}(\beta)}{\partial \beta}$$

Sparse Random Graph

BP is a powerful tool for calculations and algorithms. However if we are not in a tree the Bethe Free Entropy is not equal to True Free Entropy. The temptation is to assume that the former is a good approximation of the latter, albeit this in general is not true, it can provide asymptotically exact result in Sparse Random Graphs:

Sparse random graphs are locally tree-like

Informally, a graph is *locally tree-like* if, for almost all nodes, the neighborhood up to distance d is a tree, and $d \rightarrow \infty$ as $N \rightarrow \infty$.

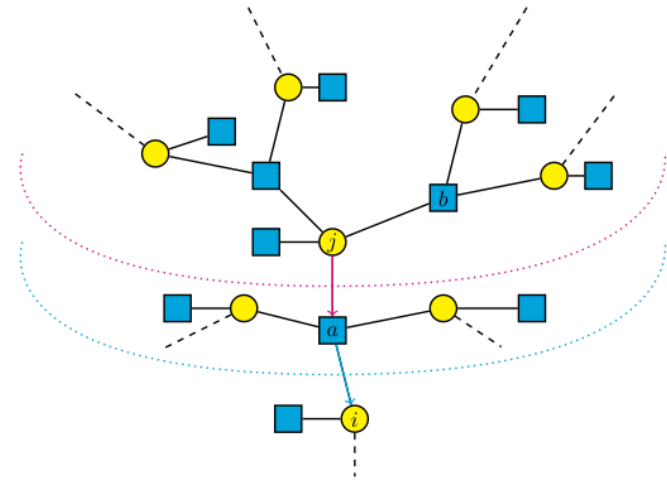
Importantly, sparse random factor graphs are locally tree-like. Sparse here refers to the fact that the average degree of both the variable node and the factor nodes are constants while the size of the graph $N, M \rightarrow \infty$.

The result presented in the following, obtained through BP, holds in particular for a sparse random graph. So rather to imagine a country we should rather think about a big archipelago connected by few naval lines

BP for colouring

$$\chi_{s_j}^{j \rightarrow i} = \frac{1}{Z^{j \rightarrow i}} \prod_{k \in \partial j \setminus i} \left[1 - (1 - e^{-\beta}) \chi_{s_k}^{k \rightarrow j} \right]$$

$$\psi_{s_i}^{(ij) \rightarrow i} = \frac{1}{Z^{(ij) \rightarrow i}} \left[1 - (1 - e^{-\beta}) \chi_{s_i}^{j \rightarrow (ij)} \right]$$



→ $1 - (1 - e^{-\beta}) \chi_{s_j}^{k \rightarrow i} = \sum_{s_k \neq s_j} \chi_{s_k}^{k \rightarrow i} + e^{-\beta} \chi_{s_j}^{k \rightarrow i}$, is the probability that neighbor k allows node j to take color s_j .

→ $\prod_{k \in \partial j \setminus i} [\dots]$ is the the probability that all the neighbors let node j take color s_j (i was excluded, as the edge (ij) is removed). The product is used because of the implicit assumption of BP, about the independence of the neighbors when conditioning on the value s_j .

Statistical Physics formulation of the problem

$$\text{energy } e = \sum_{(ij) \in E} \delta_{s_i, s_j} / N \quad \text{number of colorings of a given energy}$$
$$\mathcal{N}(e) = \mathbb{e}^{Ns(e)}$$

Letting Φ be the free entropy density, we can then rewrite the partition function as

$$\mathbb{e}^{N\Phi(\beta)} = Z_G = \sum_{\{s_i\}_{i=1}^N} \mathbb{e}^{-\beta \sum_{(ij) \in E} \delta_{s_i, s_j}} = \sum_e \sum_{\text{all coloring of energy } e} \mathbb{e}^{-N\beta e} = \int de \mathbb{e}^{Ns(e) - N\beta e}$$

The saddle-point method then gives us

$$\left. \frac{\partial s(e)}{\partial e} \right|_{e=e^*} = \beta, \quad \Phi(\beta) = s(e^*) - \beta e^*$$

A random configuration (sampled from the Boltzmann distribution at temperature β) will have energy concentrating at e^* w.h.p. $\Rightarrow \langle e \rangle_{\text{Boltz}} = e^*$. We also remind that the Boltzmann distribution is convenient because

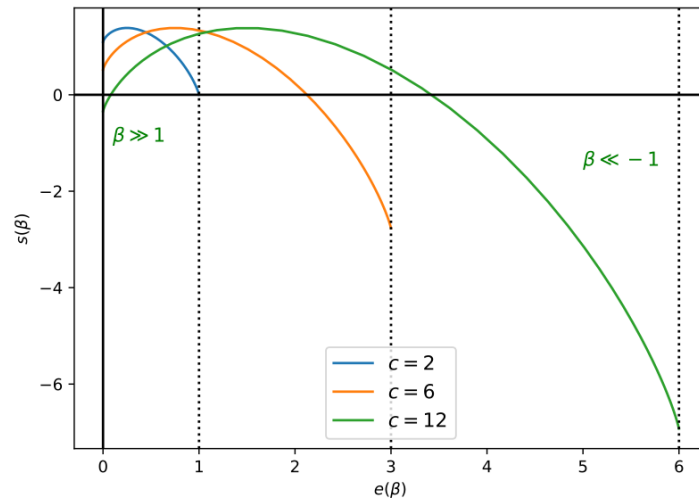
$$\frac{d\Phi(\beta)}{d\beta} = -\langle e \rangle_{\text{Boltz}} \quad \Rightarrow \quad s(e^*) = \Phi(\beta) + \beta e^*.$$

Paramagnetic Fixed Point

The Paramagnetic fixed point is given by:

$$\chi_{s_j}^{j \rightarrow i} = 1/q$$

- It is a fixed point in the sense that plugging this expression in the BP equations, one finds back this value.
- It is called paramagnetic because it relates to uniform distributed marginal among all the colours.



$$q = 4$$
$$e^* = c/2q$$
$$s(c/2) \leq 0$$

Ferromagnetic Fixed Point

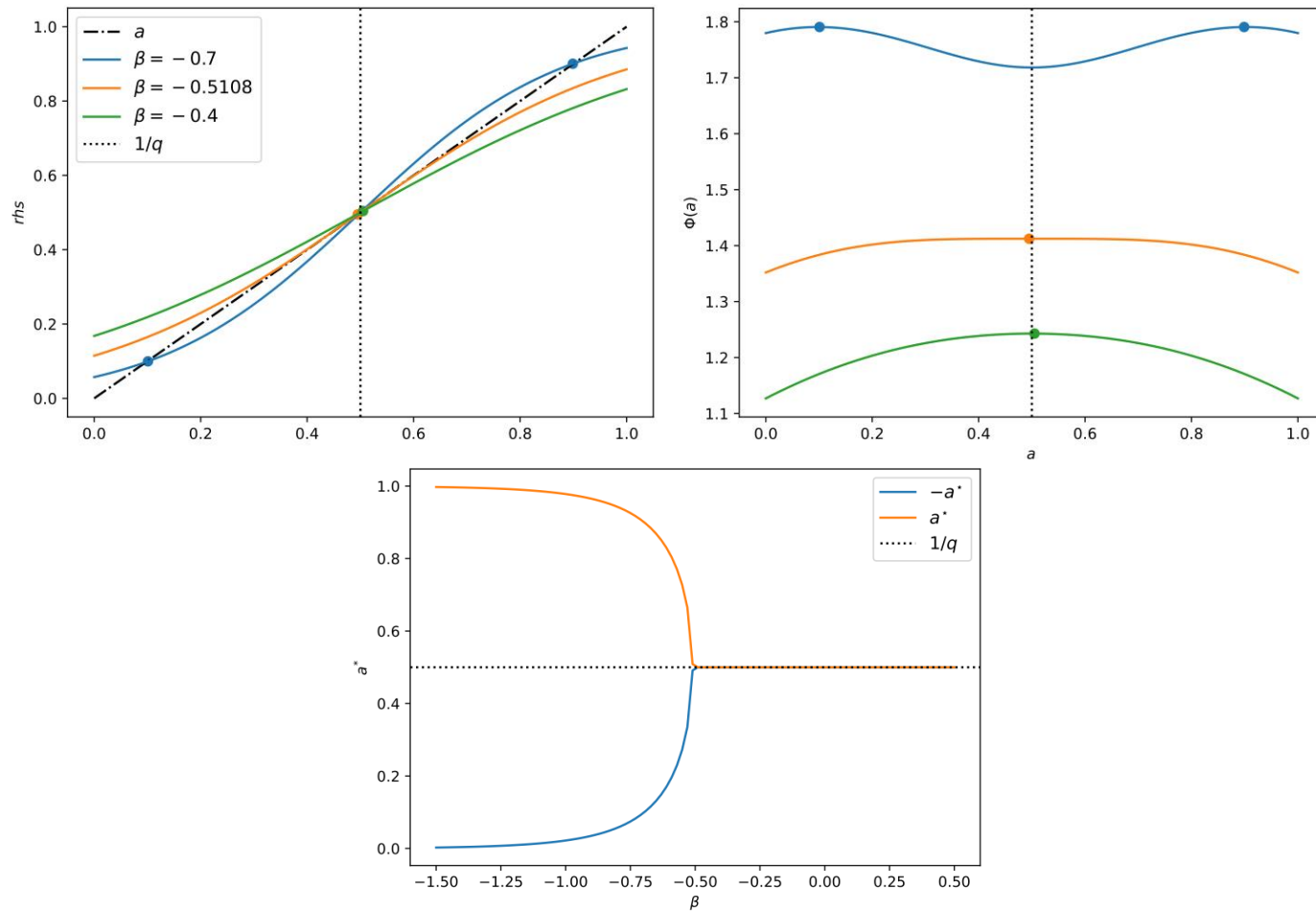
Ansatz: the message corresponding to a certain color take a different value, i.e. it will have a different marginal probability

$$\chi_1^{i \rightarrow j} = a, \quad \chi_s^{i \rightarrow j} = b = \frac{1-a}{q-1}, \quad \forall s \neq 1$$

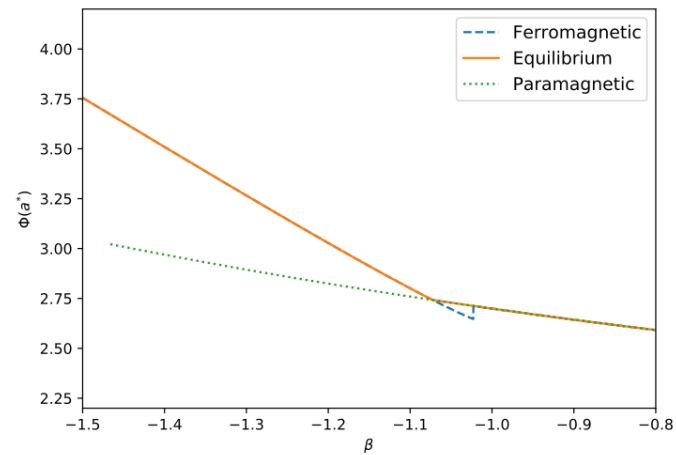
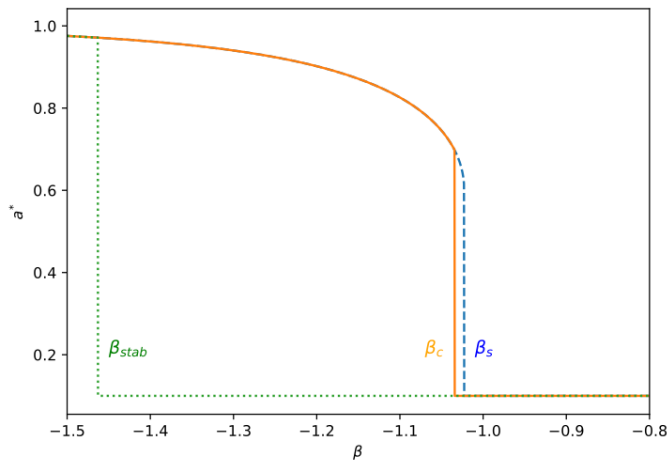
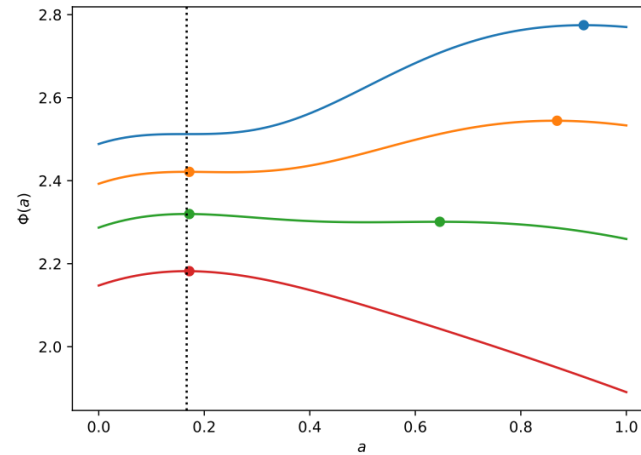
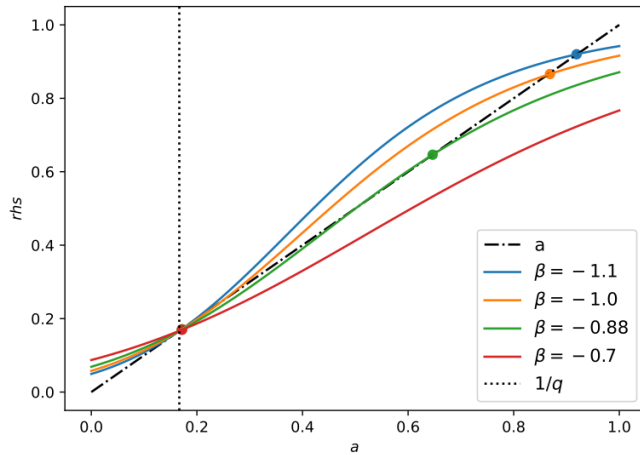
For a random *d-regular graphs*, namely a graph in which each node has same degree, one finds the following self consistent equation:

$$a = \frac{[1 - (1 - e^{-\beta}) a]^{d-1}}{[1 - (1 - e^{-\beta}) a]^{d-1} + (q-1) \left[1 - (1 - e^{-\beta}) \frac{1-a}{q-1}\right]^{d-1}} =: \text{RHS}(a; \beta, d, q)$$

Ising Ferromagnet ($q=2$)



Potts Ferromagnet ($q > 2$)



Back to the original Problem

Coming back to the original antiferromagnetic case:

We have $\beta > 0$ and we recover the original problem for $\beta \rightarrow +\infty$

We can ask, considering a certain number of colors q :

- Which is the maximum value of the average degree c for which the graph is colorable with high probability?
- When the coloring is possible, how many of them are there? (i.e. determine $s(e=0)$)

Back to the original Problem

The free entropy $\Phi_G = \frac{1}{N} \log(Z_G)$ then depends explicitly on the graph G . We expect the free entropy to be self-averaging, i.e. concentrating around its mean as

$$\forall \varepsilon > 0, \quad \Pr (|\Phi_G - \mathbb{E}_G[\Phi_G]| > \varepsilon) \xrightarrow{N \rightarrow \infty} 0$$

$$\text{quenched free entropy} \quad \Phi_{\text{quench}} \equiv \mathbb{E}_G \left[\frac{1}{N} \log(Z_G) \right]$$

$$\text{annealed free entropy} \quad \Phi_{\text{anneal}} \equiv \frac{1}{N} \log (\mathbb{E}_G[Z_G])$$

$$\Phi_{\text{anneal}} \geq \Phi_{\text{quench}}$$

$$\Phi_{\text{anneal}} = \Phi_{\text{Bethe}} \Big|_{\chi \equiv \frac{1}{q}}$$

Back to the original Problem

The annealed entropy is positive and vanishes at average degree

$$c_{\text{ann}}(\beta) = -\frac{2 \log q}{\log \left[1 - (1 - e^{-\beta}) \frac{1}{q} \right]}$$

However some more detailed analysis suggest that this is not actually the case

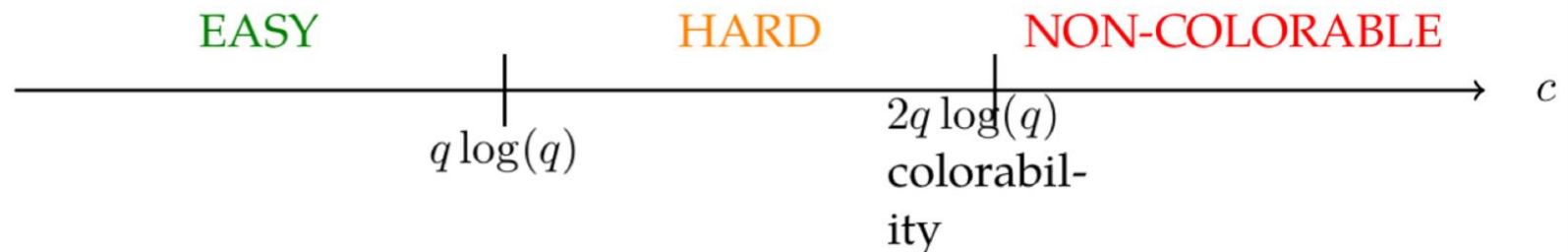
$q = 3$ and $\beta \rightarrow \infty$ we have that $c_{\text{KS}}(q = 3) = 4 < c_{\text{ann}}(q = 3) = 5.4$.

Kesten, H. and Stigum, B. P. (1967). Limit theorems for decomposable multi-dimensional galton-watson processes. *Journal of Mathematical Analysis and Applications*, 17(2):309–338.

Coja-Oghlan, A. (2013). Upper-bounding the k-colorability threshold by counting covers. *arXiv preprint arXiv:1305.0177*.

Back to the original Problem

For $q \geq 4$ and $\beta \rightarrow \infty$ we have that $c_{\text{KS}} > c_{\text{ann}}$



Achlioptas, D. and Coja-Oghlan, A. (2008). Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 793–802. IEEE.