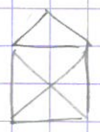
MONT AUDUBON

03/07/2017 C. MOORE INTRODUCTION TO COMPUTATIONAL COMPLEXITY.

let's start by looking at two problems:

1 COMPUTATIONAL COMPLEXITY CLASSES

- Can you draw the house in one try, without going twice trough one edge?

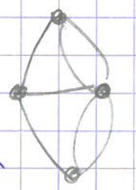


→ you can try exhaustive search ... but long. (actually possible).

now with

→ ?

- Can you cross all the bridges exactly once?



→ another way of answering to the question than exhaustive search is to notice that each node needs even number of edges for the problem to be solvable! (have an Eulerian cycle).

Rk two nodes might have odd number (start and end of path).

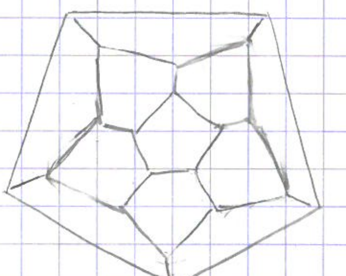
non deterministic

NP

Hamiltonian path → if you are given a solution, it can be checked in polynomial time. (also could not check in poly time that it is the shortest ↳ decision pb rather than optimisation!)

Eulerian path P → the total computation time needed to answer is growing at most as $O(n^c)$
↓
size of the pb

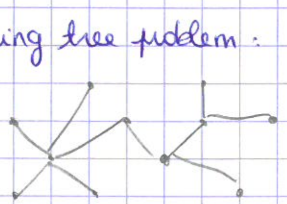
- Can you visit exactly once each node?
  ~ traveling sales man problem.



It looks similar to the previous pbs, but actually believed to be much harder!

↳ . believed that the exhaustive search is necessary
  . reasons to believe that there is no shortcut, if there was, many other problem will become much easier than what we think.

- Minimum spanning tree problem: Electricity network for Bohemia. connect in a network every body with minimal total edge length
  → no need for loops, will be a tree



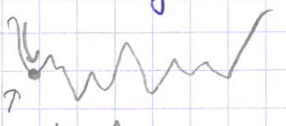
- greedy algorithm: connect shortest pairs successively
  ≡ quench in phys. don't connect if forming a loop.

↳ does it find the shortest? proof by contradiction.
  - Suppose you want to add an edge that was no in the solution
  - will create a loop, from which the edge to delete will necessarily be the one previously proposed.

landscape:

The greedy algorithm is working in this case because there are no local minima!

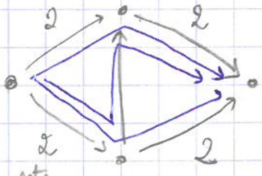
BUT it would not be working for the traveling salesman problem.
→ basically it is just the same minimum spanning tree problem to which we forbid any branching. And this little change makes the landscape go crazy:



not optimal.

Rk: notion of landscape not always well defined.

- The max-flow:



but the greedy algorithm might only get 3 students to the final destination.

while obvious way of getting 4 through.

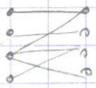
Amusingly, if we get students going backward along the arrows as anti-students, then greedy works.
↳ The point that will be made later is that everything depends on which states are neighboring (think single spin flip).

Problems can be translated into one another sometimes; which has some consequences in terms of analysis of the complexity.

if $A \xrightarrow[\text{translation}]{} B$ meaning a way to easily transform A to B while preserving the answer to the question "yes" or "no" of interest. the reverse translation might not be possible.

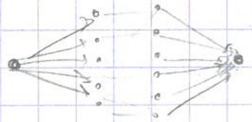
$\Rightarrow$ { if B is easy, the A is easy
if A is hard, then B is hard.

example: BIPARTITE MATCHING



from graph of compatibility, can you deduce whether a perfect matching exists?

⇒ well can make it a max flow!



so if one can answer max-flow → can answer the perfect matching.

This is actually the notion of Reduction $A \leq B$, B is at least as hard as A.

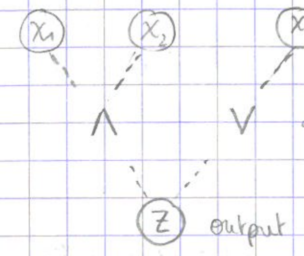
That leads to the definition of NP-completeness: if B is NP-complete, $A \leq B$ for any A problem in NP.

meaning that B as some kind of great expressivity.
↳ trivial NP-complete problem is writing a checking programm on a computer

The class of NP-complete problems is obviously included in the class of NP problems.

if the decision problem is hard the optimisation is at least as hard.

A concrete example: Boolean circuits → K-SAT



any program, e.g. hamiltonian circuit checker, can be translated to the Boolean circuits down to the basic computer language.

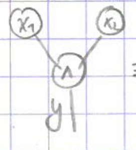
e.g. 3-SAT: take a series of variables $x_1, \dots, x_n$, and build a formula:

e.g. $(x_1 \vee \bar{x}_2 \vee \bar{x}_3) \wedge (\dots$ — is there a satisfying assignement?

true or false ...

Well we can embed the claim that a boolean circuit is functionning properly into a 3-SAT problem.

↳ Add a series of y variables on edges



$\equiv (\bar{x}_1 \vee \bar{x}_2 \vee y) \wedge (x_1 \vee \bar{y}) \wedge (x_2 \vee \bar{y})$

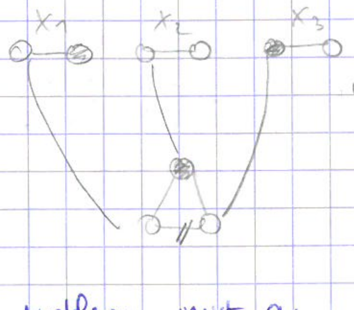
EXERCIZE: Prove that 2-SAT is in P.

NP ≡ if the answer is yes, there is an easily checkable proof of that fact. Also finding the proof (the satisfying instance) might not be easy.

- Max-Cut a.k.a. Ising anti-ferromagnet.
  From a graph G, can it be divided into two sets of vertices with at least k edges crossing between two sets.
  ↳ NP-complete.

  This problem can be translated into NAE-3-SAT.
  In this slightly modifying CSP, clauses like (x_1, x_2, x_3) asks for everybody not to be equal.
  ↓
  EXERCIZE: • 3-SAT ≤ NAE-3-SAT (difficult direction) → 4 variables
  • NAE-3-SAT ≥ 3-SAT $\quad (x_1, x_2, x_3) = (x_1 \vee x_2 \vee x_3) \wedge (\bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_3)$

· gadgets



what should be the target k, n variables, m clauses.

$k = n + 5m$

- another funny problem: input $a_1, \dots a_n$ n integers.

  is $\int_{-\pi}^{\pi} d\theta \, (\cos \theta a_1)(\cos \theta a_2) \dots (\cos \theta a_n) \neq 0$?

04/07/2017

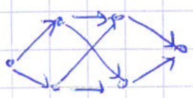
Recall: We have shown that some problem can be hard.
Is there a "thing" (which can be checked easily)? ≤ circuit satisfiability ≤ 3-SAT.

And furthermore: 3-SAT ≤ NAE-3-SAT ≤ MAX-CUT
≡ minimizing the energy of a ferromagnet.
$\equiv \sum_{ij} J_{ij} s_i s_j$ → NP-complete

Another problem that NAE-3-SAT can be reduced to is 3-coloring → Potts anti-ferromagnet

INDEPENDENT-SET → set of vertices that are not neighbor of one another.
↳ target size of independent-set does it exist?
→ relating to the packing of hard-spheres

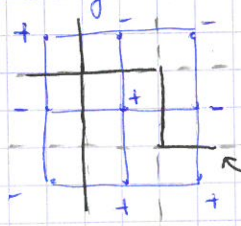
Recall max-flow → in P but it is not clear at first sight.
A related problem is MIN-CUT: what is the smallest number of edges to be cut to prevent any flow?



↓
They are actually <u>dual</u>: one needs to cut the edges which capacities are saturated in the answer to max-flow.

Yet another: Ferromagnetic random fields Ising model RFIM: $H = -\sum_{(ij)} s_i s_j - \sum_i h_i s_i \quad s_i = \pm 1$.
→ Finding the ground state is actually easy.
— It can be reduced to min-cut. EXERCIZE: min-cut weighted graph → another graph for which Ising model neighbors.
total weight of min cut = ground state energy
in spirit: choose $s=+1$ / $s=-1$ vs choose right part graph / left part.

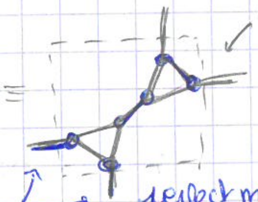
2D-Ising model on square lattice:



— dual graph → each node in this dual has 0, 2 or 4 edges

← through unaligned edges.

=



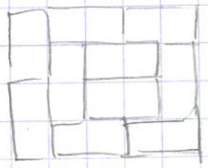
+ weights to be added to perfect matching to get partition function.
= # perfect matching

2 edges only → perfect matching

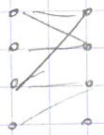
Class of problems #P "count P" or "sharp P" → counting solutions

(simplification to unweighted matchings on planar graphs. → sharp P complete problem.

Also related to the filling of a checker board with dominos



Perfect matching on the dating problem



— not hard to say whether there a matching
— hard to count how many there are
#P-hard.

Adjacency matrix $A_{ij} = \begin{cases} 1 \\ 0 \end{cases}$

$\mathrm{Per}(A) = \sum_{\pi} \prod_i A_{i,\pi(i)}$ = # perfect matching

$\mathrm{Det}(A) = \sum_{\pi} (-1)^{\pi} \prod_i A_{i,\pi(i)}$

parity of the permutation

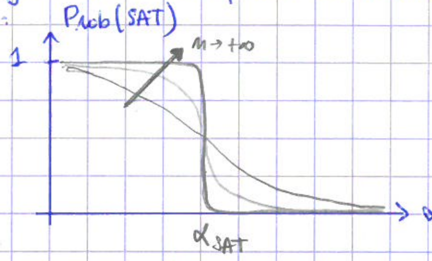
In the case of <u>planar</u> graph, explained weird way to map both. And Det is computable in polynomial time!
↳ not the case for all kinds of graph.

<u>2 PHASE TRANSITIONS</u>

* RANDOM 3-SAT: n variables $\quad (x_i \vee \bar{x_j} \vee x_k) \rightarrow$ 3 spins spin glass.
m clauses
flipped a coin to decide whether negated or not.

We define $\alpha = \frac{m}{n}$, the density of clauses, that we suppose constante for $n \to +\infty$.

(So that we consider a problem closely related to sparse random graph.)
In the limit $n \to +\infty$ one observes:



tuning of the diff. of the problem.

$= \alpha_{critical} \approx 4,267$

Let's look at how one could attempt to locate this transition:

Let's define $Z = \#$ satisfying assignement. $\equiv$ the zero temperature limit of the partition function for $H = \sum_{clause} \delta(\text{clause satisfied})$.

which is a very complicated r.v with very complicated distribution.

What we can start with, is the annealed computation:

$$\mathbb{E}(Z) = \underset{\substack{\uparrow \\ \text{total \# assignments}}}{2^n} \Big(\underbrace{1 - 2^{-k}}_{\text{prob one clause happy}} \Big)^{m} \quad \begin{array}{l} \leftarrow k\text{-SAT} \\ \uparrow \text{ total \# of clauses} \end{array}$$

So that $\mathbb{E}(Z) = \left(2(1-2^{-k})^{\alpha} \right)^n \ll 1 \iff \alpha > \frac{\ln 2}{\ln(1-2^{-k})} \approx 2^k \ln 2$. for large k.

And $\langle \log Z \rangle \leq \log \langle Z \rangle$ and $Pr(Z>0) \leq \langle Z \rangle$ (first moment method Markov inequality)

Rk: For 3-SAT, $\alpha_{RS} = 5.19$, above which the number of solution is exponentially small. Notice that between α_c and α_{RS}, there are typically no solution, but there exist some instance with exponentially small probability with an exponential number of it

Furthermore, Cauchy-Schwarz tells us: $Pr(Z>0) \geqslant \frac{(\mathbb{E} Z)^2}{\mathbb{E} Z^2}$

indicator r.v $\mathbb{1}(\sigma \text{ is sat})$
Rk: $\mathbb{E}(\mathbb{1}(\sigma \text{ is SAT})) = Pr(\sigma \text{ is SAT})$

Let's compute $\mathbb{E} Z^2$. We rewrite $Z^2 = \left(\sum_{\sigma} \mathbb{1}(\sigma \text{ is sat}) \right)^2$

$$= \sum_{\sigma, \tau} \mathbb{1}(\sigma \text{ and } \tau \text{ both satisfying})$$

$$\Rightarrow \mathbb{E} Z^2 = \sum_{\sigma, \tau} \text{Prob}(\sigma \text{ and } \tau \text{ both satisfying})$$

This depends on how many variables σ and τ have in common.

$$\mathbb{E} Z^2 = 2^n \sum_{\underset{\substack{\uparrow \\ \text{number of variables they agree}}}{y=0}}^{n} \binom{n}{y} f\left(\frac{y}{n}\right)^m \rightarrow$$

probability that a random clause is satisfied as a function of the fraction on which they agree.

EXERCIZE $f(\zeta) = 1 - \underset{\substack{\uparrow \\ \text{only one}}}{2} \cdot 2^{-k} + \zeta^k 2^{-k}$ $\quad \leftarrow$ both satisfy

+ $\binom{n}{y} \sim e^{n H(y/n)}$ $\quad \leftarrow$ Shanon entropy

And finally $\mathbb{E}Z^2 \simeq 2^n \int_0^1 d\zeta\, e^{n[h(\zeta) \;\; + \alpha \log(f(\zeta))]}$

Rk: $f(1/2) = (1-2^{-k})^2$ → act as independent assignments
↘ are they dominating the above integral? in which case $\mathbb{E}Z^2 \sim (\mathbb{E}Z)^2$.

↳ saddle point / Laplace method
$\begin{cases} h(\zeta) \text{ maximum for } \zeta = 1/2 \\ f(\zeta) \text{ maximum for } \zeta = 1 \end{cases}$ ⟹ if $\zeta^* = 1/2$ → then satisfiability would come out of both bounds.
↳ DOES NOT WORK FOR K-SAT!
$\zeta^* > 1/2$ for any $\alpha > 0$
Rk: NAE-SAT with symmetry, ok up to some k

A SEARCH ALGORITHM TO SOLVE K-SAT:

$(x_3 \vee \bar{x}_5 \vee x_7) \wedge (\ldots)$ → stupidest: explore branch tree of possible assignment and back track when contradiction.

↓

our algorithm: will never stop: → if there are clauses with only one variable
↳ run until everybody set. left to assign (unit clause) for it not to be broken
↳ choose one randomly and satisfy it
→ else:
↳ choose a not yet assigned variable
↳ flip a coin and set it.

↳ idea: unit clause are the dangerous thing to keep track of.
↳ 2 of the variables were already assigned in opposite way.
so needs the last one to be set in right value

Let's analyse the algorithm: $\begin{cases} s_3 = \text{density of 3-clauses} & s_3(0) = \alpha \\ s_2 = \text{density of 2-clauses} & s_2(0) = 0 \end{cases}$ → extensive whereas s_1 should not be extensive
↳ unsat almost surely.

we also rescale the time: $t = \frac{T}{n}$ with T = total # of steps.

At any time, remains to be assigned: $(1-t)n$ variables

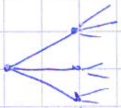
DE: $\frac{ds_3}{dt} = -3 \cdot \frac{s_3}{1-t}$ $\qquad \frac{ds_2}{dt} = \frac{3/2\, s_3 - 2 s_2}{1-t}$ $\quad \Rightarrow \begin{cases} s_3(t) = \alpha(1-t)^3 \\ s_2(t) = \frac{3}{2}\alpha t (1-t)^2 \end{cases}$

3 variables belonging to it → 3 way to pick it

unset variables

half a chance to set variable the wrong way in a three clause

unit clauses: branching process:



expected λ children

$\lambda < 1$ = "subcritical" → geometric series $1 + \lambda + \ldots + = \frac{1}{1-\lambda}$
$\lambda > 1$ = supercritical → $+\infty$
still some proba to die but $\neq 1$.

finite expected total number of descendent
↳ will die out.

↳ so that for our problem:
expected number of unit clause formed when setting value of a variable from a unit clause

$\lambda = 1/2 \; 2\, s_2/1-t$
$= \frac{3}{2}\alpha t(1-t) \leq \frac{3}{8}\alpha$

↳ keep $\alpha < 8/3$ → to work.
$= \alpha^{2}/k$!

So this algorithm was actually found to succeed up to $\alpha = O\left(\frac{2^k \log k}{k}\right) \equiv \alpha_d$.

What is going on for larger α, smaller than α_{sat}? → exponential time algorithm.
α_d = dynamical transition, clustering, RSB.

How come the clustering is hindering so much the finding of a solution?
→ What is definetly hard is to sample solutions, as the MCMC must cross high energy barrier

Well actually, what makes it difficult to find solutions is actually what's called the rigidity transition, happening right after. → Not found algorithm in poly time above
proofs that classes of algorithm must fail.

The true thermodynamics transition = is the condensation = Kauzmann transition. RSB.

3 NETWORKS GRAPHS COMMUNITIES

Statistical significance → Are the communities really there?
→ One wants to understand the Gibbs distribution between many possible fits of the data → Bayesian view. In this view, when no communities actually exist, it is obvious to see.

Reproach to Bayesians: we need for such analysis to have a prior opinion on how the data were generated.

THE STOCHASTIC BLOCK MODEL: (k state potts model ≈)

$\begin{cases} k \text{ groupes} \\ n \text{ nodes with labels } \sigma_i \in \{1, \dots k\} \end{cases}$ → $(i,j) \in E$ with probability $p_{\sigma_i \sigma_j}$ ~ generalization ER $\quad P = {}_k\left[\quad\right]^k$

planted version: Nature generated $\{\sigma_i\}$ then $G = (V, E)$ and you only observe $G = (V, E)$.

we → work in the special case: – groups in equal size ≈ uniformly picked initial labels
– $p_{rs} = \frac{1}{n}\begin{pmatrix} c_{in} & & c_{out} \\ & \diagdown & \\ c_{out} & & c_{in} \end{pmatrix}$ ⇒ avg degree $= \frac{c_{in} + (k-1)c_{out}}{k}$
sparse (the $\frac{1}{n}$) ; poisson distributed

BAYESIAN INFERENCE:
probability to get a graph given labels: $P(G|\sigma) = \prod_{(ij) \in E} \frac{c_{\sigma_i \sigma_j}}{n} \prod_{(ij) \notin E} \left(1 - \frac{c_{\sigma_i \sigma_j}}{n}\right)$
prob edge exist

⇒ $P(\sigma|G) = \frac{1}{Z} P(G|\sigma) = \frac{P(G|\sigma) P(\sigma)}{P(G)}$
↳ which corresponds: $H(\sigma) = -\sum_{(ij)} \log c_{\sigma_i \sigma_j}$ + terms explaining non edges.
in stat phys

PHENOMENOLOGY:

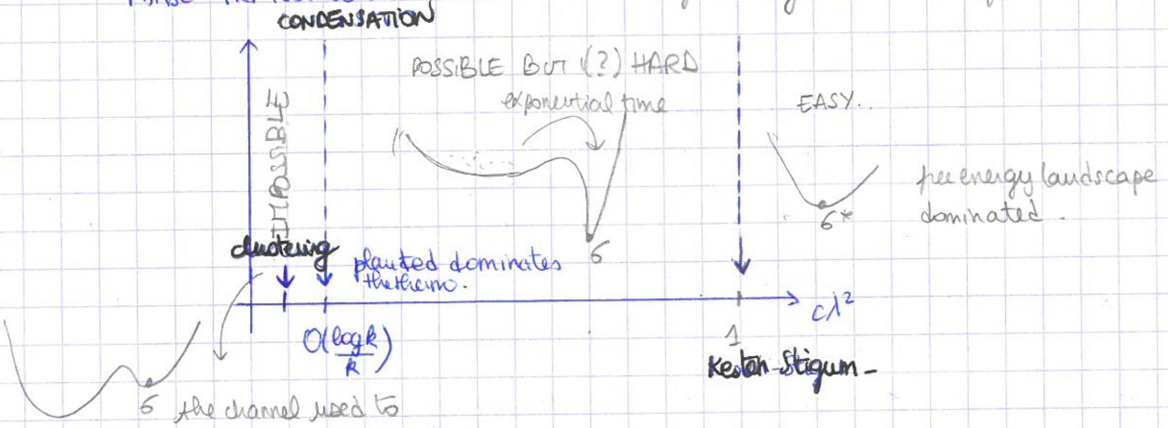
Think of the effect of one edge: $\sigma_i = r$ $\quad i \bullet\!\!-\!\!-\!\!\bullet j$ ← knowing group + knowing edge → info on j?

Markov chain related to $[C] = \begin{pmatrix} c_{in} & c_{out} \\ c_{out} & c_{in} \end{pmatrix}$

↳ transfer: $\frac{1}{kc} \begin{pmatrix} c_{in} & c_{out} \\ c_{out} & c_{in} \end{pmatrix} = \mathbb{1} \lambda + \frac{(1-\lambda)}{k} J$ $\quad [\,^{111}_{111}]$ $\quad \lambda = \frac{c_{in} - c_{out}}{kc}$ → second e.v.

proba to copy label of i to j. → measure of the strength of the community.

PHASE TRANSITIONS



6 the channel used to convey the message is actually not carrying information (noise dominates)

06/07/2017

A small parenthesis and remark:
↳ A slightly different problem: XORSAT $(x_1 \vee \bar{x}_2 \vee x_3)$ "exclusive OR"

⇒ still has a bumpy landscape but can be solved in polynomial time as it can be mapped to set of coupled linear equations:

$$(x_1 \vee \bar{x}_2 \vee x_3) \equiv x_1 \oplus x_2 \oplus x_3 = 1 \mod \quad . \Rightarrow$$

GOING BACK TO THE STOCHASTIC BLOCH MODEL:

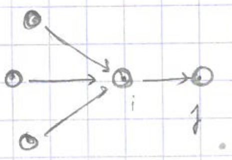
We define our null model as the Erdös-Reyni graphs. $G(n, p = \frac{c}{n}) \to pr\ ij \in E = p.$

Rk: The problem of determining whether or not communities exist and the problem of finding communities become doable at the same threshold.

Recall $P[6/G] \propto P[G/6] = e^{-\beta H(6)}$ where $H(6) = -\sum_{(ij) \in E} \log c_{6_i 6_j}$

We wish to compute the marginals $\mu_r^i = \Pr(6_i = r)$ → have a sense of how certain we are about the assignments, not only the best assignment.

- first method: Monte Carlo sampling / Heat bath eg
- second method. Belief propagation



message $\mu_r^{i\to j}$ = what i tells j of $\Pr(6_i = r)$

↳ recursive relation $\mu_r^{i\to j} = \prod_{\ell \in \partial i \setminus j} \sum_s \mu_s^{\ell \to j} c_{rs}$ $\quad P([i,j] \mid {}^{6_i = r}_{6_j = s})$

can be thought of progressive learning of information incorporated following Bayes rule.

Rk: Belief propagation assumes conditional independence: my neighbors are only correlated through me.
- ⇒ diluted models, with very few / no short loops. Locally tree like.
- ⇒ but also correlation decaying with length.

↳ In practice, even on real networks (full of triangles), it seems to be working.

<u>Algorithm</u>:
- initialize μs randomly
- update them according to some schedule
- until reaching a fixed point.

The two point probability: $\mu^{ij}_{rs} = P_r(\sigma_i = r, \sigma_j = s) \propto \mu_r^{i\to j} \mu_s^{j\to i} c_{rs}$

<u>Rk</u>: Why do we not just use marginals but use messages? → assuming one's marginal derives from the product of the marginal of neighbors
- → called naive Bayes, works well to some extent
- → corresponds to assuming a factorized distribution $P(\sigma) = \prod_i \mu^i_{\sigma_i}$

↳ where $\mu^i_{\sigma_i}$ are the marginals minimizing the KL divergence to the Gibbs (variational inference, Naive Mean-field).

Compared to vanilla mean-field
↳ BP corresponds to minimizing a KL-divergence with a more complicating ansatz taking into account correlations between neighbors; namely:

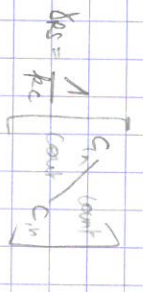
$$P(\sigma) = \prod_i \mu^i_{\sigma_i} \prod_{(ij)\in E} \frac{\mu^{ij}_{\sigma_i\sigma_j}}{\mu^i_{\sigma_i}\mu^j_{\sigma_j}}$$

⚠ this is properly normalized only if the graph is a tree...

⇒ BETHE-FREE ENERGY

<u>Rk</u>: This can be seen as a first step to a cluster approximation → KAKEYA APPROXIMATION

Belief propagation typically have many fixed points.



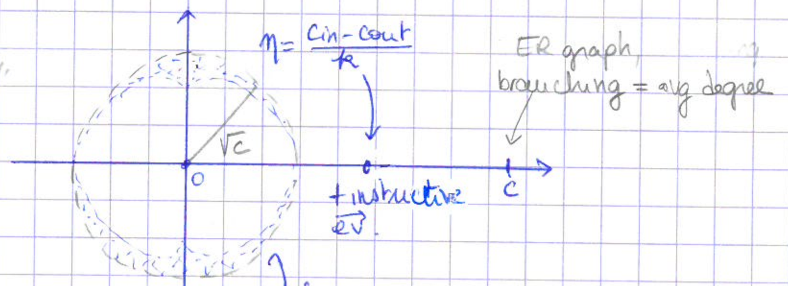
→ one clearly is $\mu_r^{i\to j} = 1/k$ (trivial fixed point, that is not really interesting...)

well is it stable? Study the perturbed message evolution $\mu_r^{i\to j} = 1/k + \varepsilon_r^{i\to j}$

$$\frac{\partial \mu_r^{i\to j}}{\partial \mu_s^{\ell\to m}} = \underbrace{\delta_{im}(1-\delta_{\ell j})}_{\text{non-backtraking maxtrix } B} \gamma_{RS} = B \otimes \gamma_{RS} \text{ tensor}$$

↙ "adjacency of directed edges"

Spectrum of B in complex plan:



bulk that is a complexe version with 0,1 entries sparse matrices of semi-circle law → random!
↳ random eigenvectors

<u>Rk</u>: This relates to spectral clustering: usually done with the adjacency matrix. Same kind of philosophy, where the second eigenvector has coordinates which sign provide the communities labels.
YET, usually does not work for sparse graph → fooled by localized eigenvector

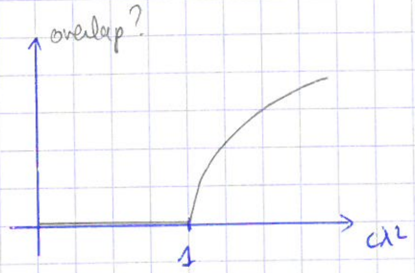
So finally in the stability analysis $\eta = \frac{c_{in} - c_{out}}{k} = c\lambda$

we have stability if $c\lambda^2 > 1$ → point also where η comes out of the bulk

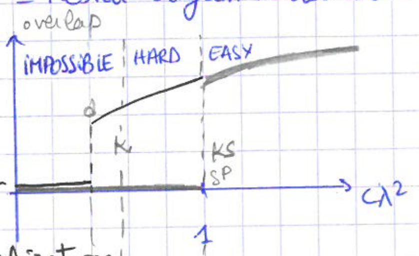
BP converges to fixed point correlated to planted assignement.

≡ Kesten - Stigum threshold.

$k = 2$



$k \geq 5$



initialization correlated to the ground truth

no communities can be detected.

exponentially small basin of attraction of the good fixed point

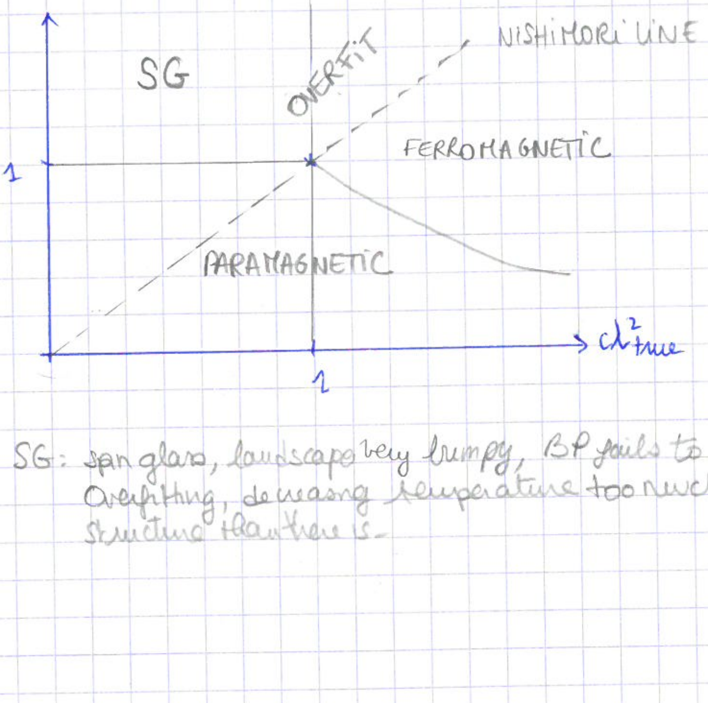
actually the free energy of accurate fixed point is higher than the one of the bulk before condensation

CONCLUSION ON THE PROBLEM OF OVERFITTING:

let's stop assuming you knew the value of λ. → so you run BP with a guess $c\lambda^2_{alg}$

$\beta = c\lambda^2_{alg}$.

$k = 2$



SG: spin glass, landscape very bumpy, BP fails to converge!
Overfitting, decreasing temperature too much, assuming more structure than there is.