The Large Deviations of the Whitening Process in Random Constraint Satisfaction Problems
and of the bootstrap percolation

Guilhem Semerjian
LPT-ENS
05.05.2016 / Simons Institute

based on Braunstein, Dall’Asta, S, Zdeborová
Outline

1. Hypergraph bicoloring and its phase transitions
2. Rigidity and freezing
3. Main results
4. Minimal contagious sets of random regular graphs
5. Conclusions and perspectives
An example of CSP

- Hypergraph bicoloring (positive NAE-\(k\)-SAT):
  - \(N\) variables: \(\sigma = (\sigma_1, \ldots, \sigma_N) \in \{-1, 1\}^N\)
  - \(M\) constraints on the hyperedges of a \(k\)-uniform hypergraph

\[
\psi_a(\{\sigma_i\}_{i \in \partial a}) = \begin{cases} 
1 & \text{at least one +1 and one -1} \\
0 & \text{all +1 or all -1}
\end{cases}
\]

solutions: \(S = \{\sigma : \psi_a(\sigma_{\partial a}) = 1 \ \forall a\}\)
random hypergraph with $M$ edges (regular or Erdös-Rényi)
density of constraints $\alpha = M/N$, thermodynamic limit $N, M \to \infty$

- Satisfiability threshold at $\alpha_{\text{sat}}(k) \sim 2^{k-1} \ln 2$
- Shattering of solutions in clusters at $\alpha_{\text{d}}(k) \sim \alpha_{\text{sat}}(k) \frac{\ln k}{k \ln 2}$
- Reconstruction threshold on the tree
- Condensation, sub-exponential nb. of clusters at $\alpha_{\text{c}}(k) \sim \alpha_{\text{sat}}(k)$
Phase transitions for random CSPs (also $k$-SAT, $q$-COL, ...)

Recent rigorous results on hypergraph bicoloring/random NAESAT:

- Satisfiability threshold  
  [Ding, Sly, Sun 13]

- Condensation at positive temperature  
  [Bapst, Coja-Oghlan, Rassmann 14]

- Typical number of solutions  
  [Sly, Sun, Zhang 16]

- Fluctuations of the number of solutions  
  [Rassman 16]

- Failure of Survey Propagation for $\alpha > \alpha_d$  
  [Hetterich 16]

- ...
One more phase transition: rigidity

Coarse-grained description of a cluster: $\sigma^* \in \{-1, 1, 0\}^N$

with $\sigma_i^* = \begin{cases} 
1 & \text{if } \sigma_i = 1 \text{ in all solutions of the cluster} \\
-1 & \text{if } \sigma_i = -1 \text{ in all solutions of the cluster} \\
0 & \text{otherwise}
\end{cases}$

Frozen variables of a cluster: the ones with $\sigma_i^* = \pm 1$
One more phase transition: rigidity

- Alternative definition of frozen variables:
  - start with a solution $\sigma$
  - a constraint $a$ blocks a variable $\sigma_i = \pm 1$ iff $\sigma_j = -\sigma_i$ for all $j \in \partial a \setminus i$
  - if $i$ is not blocked by any constraint, “whiten” it, $\sigma_i \to 0$
  - repeat until fixed point $\sigma^*$ is reached

Procedure known as whitening, peeling, coarsening...

Largest subcube containing $\sigma$ with no solutions at Hamming distance 1

$\theta$: fraction of frozen variables ($\sigma_i^* = \pm 1$) in a fixed point

Either $\theta = 0$ or $\theta \geq \theta_{\text{min}} > 0$  

[Maneva, Mossel, Wainwright 07]

unfrozen / frozen solutions
One more phase transition: rigidity

Typical fraction of frozen variables (solution chosen u.a.r.):

\[ \theta \]

\[ 1 \]

\[ \theta_r \]

\[ \alpha_r \]

\[ \alpha_{sat} \]

\[ \alpha \]

\[ \alpha_d(k) \leq \alpha_r(k) \]: stronger form of correlation (naive reconstruction)

At large \( k \), \( \alpha_r(k) \sim \alpha_d(k) \)
Frozen variables and algorithmic difficulty

- Frozen solutions should be hard to find: need to set collectively order $N$ variables

- Indeed heuristic algorithms output unfrozen solutions

- Algorithmic barrier: no known algorithm finds solutions in polynomial time for
  \[ \alpha > \alpha_d(k) \sim \alpha_r(k) \quad \text{(at large } k) \]

- Up to which densities do (atypical) unfrozen solutions exist?
  Called freezing transition, $\alpha_f(k)$
Main results (I)

\[ \alpha_{\text{sat}} \]

\[ \alpha_{r} \]

\[ \theta_{r} \]

\[ \theta_{\text{min}} \]

\[ \alpha_{l,=} \]

\[ \alpha_{l,u} \]

\[ \alpha_{f} \]

\[ \alpha_{\text{sat}} \]
Unfrozen solutions exist up to

\[ \alpha_f(k) \sim \frac{1}{2} \alpha_{\text{sat}}(k) \]

previously, \( \alpha_f(k) \leq \frac{4}{5} \alpha_{\text{sat}}(k) \) [Achlioptas, Ricci-Tersenghi 06]

Recall \( \alpha_r(k) \sim \alpha_d(k) \sim \frac{\ln k}{k \ln 2} \alpha_{\text{sat}}(k) \)
 Locked solutions ($\theta = 1$, all variables frozen, sol. = whitening f.p.)

- appear at $\alpha_{1,-}(k) \sim \frac{1}{k} \alpha_{\text{sat}}(k)$
- disappear at $\alpha_{1,+}(k) \sim \alpha_{\text{sat}}(k)$
- are the only frozen solutions up to $\alpha_{1,u}(k) \sim \alpha_{d}(k)$

Recall $\alpha_{r}(k) \sim \alpha_{d}(k) \sim \alpha_{\text{sat}}(k) \frac{\ln k}{k \ln 2}$
The idea of the computation

- Parallel version of the whitening process:
  - initial condition $\sigma^0 = \sigma$ a solution
  - discrete time parallel evolution:
    $$\sigma_i^{t+1} = \begin{cases} 
    \sigma_i & \text{iff } \exists a \in \partial i, \ \forall j \in \partial a \setminus i, \ \sigma_j^t = -\sigma_i \\
    0 & \text{otherwise}
    \end{cases}$$

- Monotonic evolution, fixed-points obtained as $\sigma^* = \lim_{t \to \infty} \sigma^t$

- For a finite time horizon $T$, biased measure over solutions:
  $$\mu(\sigma, T, \epsilon) = \frac{1}{Z(T,\epsilon)} \mathbb{I}(\sigma \in S) e^\epsilon \sum_i |\sigma_i^T|$$

- $Z(T, \epsilon)$: generating function of the number of solutions classified by the number of white variables after $T$ steps
The idea of the computation

- $\sigma^T_i$ depends on $\sigma$ through variables at distance $\leq T$ from $i$
- $\mu(\sigma, T, \epsilon)$ has interactions at distance $T$
- they can be made local with additional variables (whitening times)
- then graphical model on a sparse random factor graph
  $\Rightarrow$ “routine” cavity method computation
- Large $T$ limit can be taken analytically to get the fixed points

Very similar to previous works on minimal contagious sets for bootstrap percolation

[Altarelli, Braunstein, Dall’Asta, Zecchina 13]
[Guggiola, S. 15]
Main results (II)

For each $T$, threshold $\alpha_T(k)$ such that for $\alpha < \alpha_T(k)$,
typical configurations of $\mu(\underline{\sigma}, T, \epsilon)$ are unfrozen (for a well-chosen $\epsilon$)

- $\alpha_T(k)$ grows with $T$, $\alpha_f(k)$ obtained as $\lim_{T \to \infty} \alpha_T(k)$

- For fixed $T$, at large $k$:
  - $\alpha_1(k) \sim \frac{\alpha_{sat}(k)}{\ln k}$
  - $\alpha_2(k) \sim \frac{\alpha_{sat}(k)}{\ln \ln k}$
  - in general $\alpha_T(k) \sim \frac{\alpha_{sat}(k)}{\ln^\circ T k}$ \text{ $T$-times iterated logarithm}

  recall $\alpha_d(k) \sim \alpha_{sat}(k) \frac{\ln k}{k \ln 2}$
Minimal contagious sets

- bootstrap percolation dynamics: inactive vertices become active if they have $\geq l$ active neighbors

- $\theta_{\min}(k, l)$: minimal fraction of active vertices in order to activate completely a $k + 1$ regular random graph

- for $l = k$, corresponds to the decycling number (Feedback Vertex Set)

- for $l = k - 1$, corresponds to the de-3-coring number

Analytic results for (lowerbounds on) $\theta_{\min}(k, l)$ (RS and 1RSB)

[Guggiola, S. 15]
Minimal contagious sets

Special cases:

- **decycling of 3- and 4-regular graphs**:
  \[ \theta_{\text{min}}(2, 2) = \frac{1}{4}, \quad \theta_{\text{min}}(3, 3) = \frac{1}{3} \]
  First (second) one proven (conjectured) \[\text{[Bau, Wormald, Zhou 02]}\]

- **de-3-coring of 5- and 6-regular graphs**:
  \[ \theta_{\text{min}}(4, 3) = \frac{1}{6}, \quad \theta_{\text{min}}(5, 4) = \frac{1}{4} \]

Conjecture: these 4 cases are the only ones that saturate the lowerbound:

for all \( k, l \), \( \theta_{\text{min}}(k, l) \geq \frac{2l - k - 1}{2l} \) \[\text{[Dreyer, Roberts 09]}\]

Conjecture for the decycling number at large degree:

\[ \theta_{\text{min}}(k, k) = 1 - \frac{2 \ln k}{k} - \frac{2}{k} + O\left(\frac{1}{k \ln k}\right) \]
ok with rigorous bound \[\text{[Haxell, Pikhurko, Thomason 08]}\]
Definition as a problem about processes on infinite trees:

- \( C_\theta = \) probability measures \( \mu \) on \( \{0, 1\}^{\mathbb{T}_{k+1}} \) that are translationally invariant (ergodic), with \( \mu[\sigma_0 = 1] = \theta \)

- \( \max\{\theta : \exists \mu \in C_\theta \text{ with } \mu[0 \leftrightarrow \infty] = 0\} ? \)
Conclusions and perspectives

- Freezing transition rather close to the satisfiability

- Done on the regular hypergraph bicoloring, should generalize to other CSPs

- RS computation, RSB effects should not spoil large $k$ asymptotics

- Biasing the measure, with interactions between variables at finite distance, can turn atypical properties into typical ones, in a large density range

Could it help to break the algorithmic barrier?