The Replica Symmetric Phase of Random Constraint Satisfaction Problems

Noela Müller

Goethe-University Frankfurt

Youth in High-dimensions
ICTP Triest, 2nd of July 2020

Joint work with Amin Coja–Oghlan and Tobias Kapetanopoulos

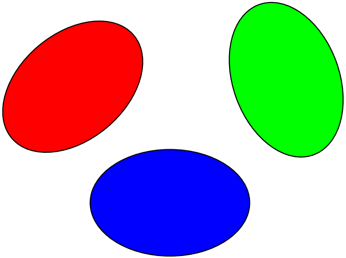
Overview

- Planted Coloring
- Solution space geometry
- Results

Planted Coloring: Step 1

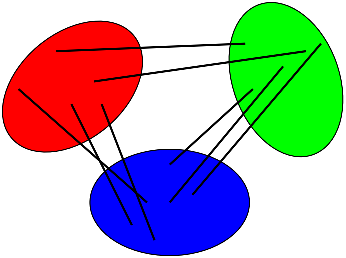
Given: $V = \{x_1, \ldots, x_n\}$ set of n vertices, and $q \in \mathbb{N}$ colours

For each $i \in [n]$ independently, draw a uniform colour $\boldsymbol{\sigma}^*(x_i)$ from $[q]$

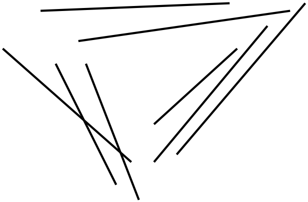


Planted Coloring: Step 2

For each $i, j \in [n]$ with $\boldsymbol{\sigma}(x_i) \neq \boldsymbol{\sigma}(x_j)$, let $\{x_i, x_j\} \in E$ w. p. d/n, $d > 0$.



The inference problem



Goal: Draw conclusions about $\boldsymbol{\sigma}^*$ from $G = (V, E)$.

The inference problem

I) Is $\boldsymbol{G}^*(\boldsymbol{\sigma}^*)$ distinguishable from the binomial random graph $\boldsymbol{G}(n,(q-1)p/(qn))$?

II) With

$$\omega(\tau,\boldsymbol{\sigma}^*) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{\tau_{x_i} = \boldsymbol{\sigma}^*_{x_i}\right\},$$

do there exist a polynomial-time algorithm $\mathcal{A}$ and $\varepsilon > 0$ such that $\mathcal{A}(\boldsymbol{G}^*)$ outputs a colouring τ with

$$\mathbb{P}\left(\omega(\tau,\boldsymbol{\sigma}^*) > (1+\varepsilon)/q\right) = 1 - o(1)?$$

Evolution of the solution space of planted colouring

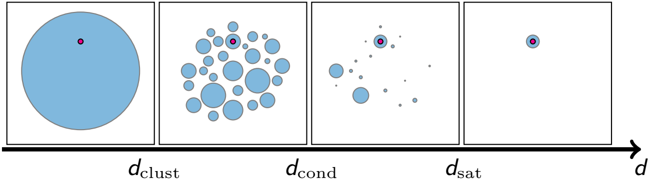


Figure: An adaption of Krząkała and Zdeborová '09.

Algorithmic and information-theoretic predictions

Figure: Predictions by Decelle, Krząkała, Moore and Zdeborová '13

Algorithmic result: Alon & Kahale 1997

Deterministic spectral algorithm with expected polynomial running time that finds a colouring correlated with the ground truth with probabilty $1 - o(1)$, if $d > c$ for some large absolute constant c.

The algorithm starts with a spectral approach and then performs some local improvement steps to yield a proper colouring.

Our results

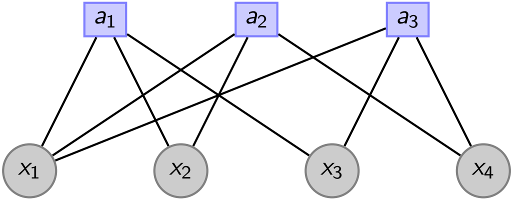
- verify prediction on the location of the condensation threshold
- identify d_{cond} as the threshold below which it is not information-theoretically possible to discern the planted and the null model

More general Random Constraint Satisfaction Problems

- variable set $V = \{x_1, \ldots, x_n\}$
- constraints $F = \{a_1, \ldots, a_m\}$, chosen randomly

More general Random Constraint Satisfaction Problems

- variable set $V = \{x_1, \ldots, x_n\}$
- constraints $F = \{a_1, \ldots, a_m\}$, chosen randomly



Evolution of *SOL*

If each constraint binds exactly k variables, consider

$$\boldsymbol{m} \sim Poi(dn/k)$$

many constraints.

How does the solution space evolve when the density parameter d increases?

Evolution of *SOL*

If each constraint binds exactly k variables, consider

$$\boldsymbol{m} \sim Poi(dn/k)$$

many constraints.

How does the solution space evolve when the density parameter d increases?
Important quantities:

- The number of solutions: *partition function* $Z(G)$
- The uniform measure on solutions: *Boltzmann distribution* μ_G:
  If $Z(G) > 0$, set

$$\mu_G(\sigma) = \frac{\mathbf{1}\{\sigma \text{ is a solution}\}}{Z(G)}$$

- Inspect $Z(\mathbb{G})$ and $\mu_{\mathbb{G}}$ for increasing *density* parameters $d > 0$.

Evolution of the solution space of rCSPs

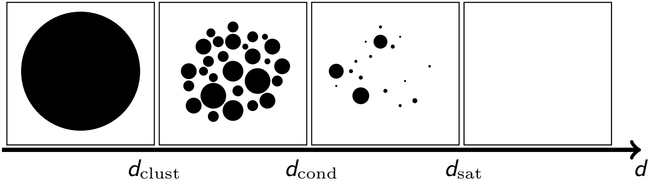


Figure: An adaption of a conjecture by Krząkała, Montanari, Ricci-Tersenghi, Semerjian and Zdeborová '07.

Theorem (Coja-Oghlan, Kapetanopoulos, Müller '20)

Let $d > 0$. With $\boldsymbol{\gamma}$ a $\mathrm{Po}(d)$-random variable, $\boldsymbol{\rho}_1^{(\pi)}, \boldsymbol{\rho}_2^{(\pi)}, \ldots$ chosen from $\pi \in P(\Omega)$ and $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots \in \Psi$ chosen from P, all mutually independent, let

$$\mathcal{B}(d, P, \pi) = \mathrm{E}\left[q^{-1}\xi^{-\boldsymbol{\gamma}}\Lambda\left(\sum_{\sigma\in\Omega}\prod_{i=1}^{\boldsymbol{\gamma}}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=\sigma\}\boldsymbol{\psi}_i(\tau)\prod_{j=1}^{k-1}\boldsymbol{\rho}_{ki+j}^{(\pi)}(\tau_j)\right) - \frac{d(k-1)}{k\xi}\Lambda\left(\sum_{\tau\in\Omega^k}\boldsymbol{\psi}_1(\tau)\prod_{j=1}^{k}\boldsymbol{\rho}_j^{(\pi)}(\tau_j)\right)\right],$$

$$d_{\mathrm{cond}} = \inf\left\{d > 0 : \sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}(d, P, \pi) > \ln q + \frac{d}{k}\ln\xi\right\}.$$

Then $1/(k-1) \le d_{cond}(q) < \infty$ and for all $d < d_{cond}(q)$ we have

$$\sqrt[n]{Z(\mathbb{G})} \quad \stackrel{n\to\infty}{\longrightarrow} \quad q\xi^{d/k} \qquad \text{in probability.}$$

By constrast, for any $d > d_{cond}(q)$ there exists $\varepsilon > 0$ such that

$$\limsup_{n\to\infty} \mathrm{P}\left[\sqrt[n]{Z(\mathbb{G})} > q\xi^{d/k} - \varepsilon\right]^{\frac{1}{n}} < 1 - \varepsilon.$$

Limiting Distribution

Let $(K_\ell)_{\ell\geq 1}$ be Poisson variables with means $\mathrm{E}[K_\ell] = \frac{1}{2\ell}(d(k-1))^\ell$ and let $(\boldsymbol{\psi}_{\ell,i,j})_{\ell,i,j\geq 1}$ be a sequence of samples from P, all mutually independent.

Theorem (Coja-Oghlan, Kapetanopoulos, Müller '20)

Suppose that $0 < d < d_{\mathrm{cond}}$ and $\mathbb{G}$ is a random factor graph model that satisfies certain assumptions. Let

$$\mathcal{K} = \exp\left(\frac{d(k-1)(1-\mathrm{tr}(\Phi))}{2} + \mathbf{1}\{k=2\}\frac{d^2(1-\mathrm{tr}(\Phi^2))}{4}\right)$$

$$\times \prod_{\ell=2+\mathbf{1}\{k=2\}}^{\infty} \exp\left(\frac{(d(k-1))^\ell}{2\ell}\left(1-\mathrm{tr}(\Phi^\ell)\right)\right)\prod_{i=1}^{K_\ell}\mathrm{tr}\prod_{j=1}^{\ell}\Phi_{\boldsymbol{\psi}_{\ell,i,j}}$$

Then $\mathcal{K} > 0$ almost surely and

$$\frac{Z(\mathbb{G})}{q^{n+\frac{1}{2}\xi\boldsymbol{m}}}\prod_{\lambda\in\mathrm{Eig}(\Phi)\setminus\{1\}}\sqrt{1-d(k-1)\lambda} \quad \xrightarrow{n\to\infty} \quad \mathcal{K}$$

in distribution.

Previous results

- Coja-Oghlan, Krząkała, Perkins and Zdeborová '17
- Coja-Oghlan, Efthymiou, Jaafari, Kang and Kapetanopoulos '18

Open Questions for Planted Coloring

- Does a variant of the Alon-Kahale algorithm work up to condensation?
- For three communities: Prove $d_{\mathrm{rec}} = d_{\mathrm{cond}} = d_{\mathrm{KS}}$.

INFERENCE PROBLEMS:
ALGORITHMS AND LOWER BOUNDS

31st of August to 4th of September

Registration:
`fellingh@math.uni-frankfurt.de`