Counting the solutions of random regular NAE-SAT

Nike Sun

joint work with
Allan Sly and Yumeng Zhang

University of California, Berkeley

Analysis of Boolean Functions, 6 April 2016
Random CSPs

CSP = constraint satisfaction problem:
Random CSPs

CSP = constraint satisfaction problem:

Variables $x_1, \ldots, x_n$ in a discrete alphabet $\mathcal{X}$, subject to $m$ constraints. Seek $\mathbf{x} \in \mathcal{X}^n$ satisfying all constraints.
Random CSPs

CSP = constraint satisfaction problem:

Variables $x_1, \ldots, x_n$ in a discrete alphabet $\mathcal{X}$, subject to $m$ constraints. Seek $\mathbf{x} \in \mathcal{X}^n$ satisfying all constraints.

We are interested in probabilistic models of various CSPs — coloring, independent set, max-cut, $k$-SAT.
Random CSPs

CSP = constraint satisfaction problem:

Variables $x_1, \ldots, x_n$ in a discrete alphabet $\mathcal{X}$, subject to $m$ constraints. Seek $\underline{x} \in \mathcal{X}^n$ satisfying all constraints.

We are interested in probabilistic models of various CSPs — coloring, independent set, max-cut, $k$-SAT.

For example, let $G_n$ be $G_{\text{reg}}(n, d)$ or $G_{\text{ER}}(n, d/n)$. Asymptotics of its independence number, chromatic number, and so on?
Random CSPs

CSP = constraint satisfaction problem:

Variables $x_1, \ldots, x_n$ in a discrete alphabet $\mathcal{X}$, subject to $m$ constraints. Seek $\mathbf{x} \in \mathcal{X}^n$ satisfying all constraints.

We are interested in probabilistic models of various CSPs — coloring, independent set, max-cut, $k$-SAT.

For example, let $G_n$ be $G_{\text{reg}}(n, d)$ or $G_{\text{ER}}(n, d/n)$. Asymptotics of its independence number, chromatic number, and so on?

The random $k$-SAT model is of a similar nature, but defined by a random bipartite graph rather than a random graph.
Random $k$-SAT and $k$-NAE-SAT

CNF formula encoded by bipartite graph with $+/-$ edges:

$m = 3$ clauses, each degree $k = 4$

$n = 7$ binary variables $x_i \in \{+,-\}$
Random $k$-SAT and $k$-NAE-SAT

CNF formula encoded by bipartite graph with +/- edges:

$m = 3$ clauses, each degree $k = 4$

$n = 7$ binary variables $x_i \in \{+, -\}$

Leftmost clause: $(+x_1 \text{ OR } +x_3 \text{ OR } -x_5 \text{ OR } -x_7)$.
Random $k$-SAT and $k$-NAE-SAT

CNF formula encoded by bipartite graph with $+/-$ edges:

$m = 3$ clauses, each degree $k = 4$

$n = 7$ binary variables $x_i \in \{+, -\}$

Leftmost clause: $(+x_1 \lor +x_3 \lor -x_5 \lor -x_7)$.
SAT solution: $x \in \{+, -\}^n$ such that every clause evaluates to $+$. 
Random $k$-SAT and $k$-NAE-SAT

CNF formula encoded by bipartite graph with $+/-$ edges:

$m = 3$ clauses, each degree $k = 4$

$n = 7$ binary variables $x_i \in \{+, -\}$

Leftmost clause: $(+x_1 \lor +x_3 \lor -x_5 \lor -x_7)$.
SAT solution: $\overline{x} \in \{+, -\}^n$ such that every clause evaluates to $+$.
NAE-SAT solution: both $+\overline{x}$ and $-\overline{x}$ are SAT solutions.
Random $k$-SAT and $k$-NAE-SAT

CNF formula encoded by bipartite graph with $+/-$ edges:

$m = 3$ clauses, each degree $k = 4$

$n = 7$ binary variables $x_i \in \{+, -\}$

Leftmost clause: $(+x_1 \lor +x_3 \lor -x_5 \lor -x_7)$.

SAT solution: $\bar{x} \in \{+, -\}^n$ such that every clause evaluates to $+$. NAE-SAT solution: both $+\bar{x}$ and $-\bar{x}$ are SAT solutions.

Random $k$-SAT or $k$-NAE-SAT: sample a random bipartite graph with $m/n \approx \alpha$
Random $k$-SAT and $k$-NAE-SAT

CNF formula encoded by bipartite graph with +/- edges:

$m = 3$ clauses, each degree $k = 4$

$n = 7$ binary variables $x_i \in \{+,-\}$

Leftmost clause: $(+x_1 \text{ OR } +x_3 \text{ OR } -x_5 \text{ OR } -x_7)$.

SAT solution: $\underline{x} \in \{+,-\}^n$ such that every clause evaluates to $+$.  
NAE-SAT solution: both $+\underline{x}$ and $-\underline{x}$ are SAT solutions.

Random $k$-SAT or $k$-NAE-SAT: sample a random bipartite graph with $m/n = \alpha$ (two flavors — regular and Erdős–Rényi).
Satisfiability thresholds

Each model is parametrized by some “difficulty parameter” $\alpha$: in $k$-SAT $\alpha$ is the clause-variable ratio $m/n$; in independent set $\alpha$ is the fraction of occupied vertices.
Satisfiability thresholds

Each model is parametrized by some “difficulty parameter” $\alpha$: in $k$-SAT $\alpha$ is the clause-variable ratio $m/n$; in independent set $\alpha$ is the fraction of occupied vertices.

Would like to understand, how does the model behave as $\alpha$ varies?
Satisfiability thresholds

Each model is parametrized by some “difficulty parameter” $\alpha$: in $k$-SAT $\alpha$ is the clause-variable ratio $m/n$; in independent set $\alpha$ is the fraction of occupied vertices.

Would like to understand, how does the model behave as $\alpha$ varies?

Ding, Sly, S.: in some models, the probability for an instance to be satisfiable has a sharp transition at some critical $\alpha_{\text{sat}}$:

- (’13a) random regular $k$-NAE-SAT,
- (’13b) independence ratio of random $k$-regular graphs,
- (’14) random $k$-SAT — in all cases for $k \geq k_0$. 
Satisfiability thresholds

Each model is parametrized by some “difficulty parameter” $\alpha$:
in $k$-SAT $\alpha$ is the clause-variable ratio $m/n$;
in independent set $\alpha$ is the fraction of occupied vertices.

Would like to understand, how does the model behave as $\alpha$ varies?

Ding, Sly, S.: in some models, the probability for an instance to be
satisfiable has a sharp transition at some critical $\alpha_{\text{sat}}$:

- (13a) random regular $k$-NAE-SAT,
- (13b) independence ratio of random $k$-regular graphs,
- (14) random $k$-SAT — in all cases for $k \geq k_0$.

The threshold $\alpha_{\text{sat}}$ is an explicit function of $k$, predicted by
methods of statistical physics (Mertens–Mézard–Zecchina ’03).
Conjectural phase diagram

The satisfiability threshold is only one aspect of a much richer picture predicted by physicists for a broad class of CSPs.

Krząkała, Montanari, Ricci-Tersenghi, Semerjian, Zdeborová ’07-’08
Conjectural phase diagram

The satisfiability threshold is only one aspect of a much richer picture predicted by physicists for a broad class of CSPs.

Each box depicts geometry of solution space (subgraph of $\mathcal{X}^n$) for a typical CSP instance, in a certain regime of $\alpha$. 

Krząkała, Montanari, Ricci-Tersenghi, Semerjian, Zdeborová '07-'08
Conjectural phase diagram

The satisfiability threshold is only one aspect of a much richer picture predicted by physicists for a broad class of CSPs.

(\(\alpha\) increases to the right)

Krząkała, Montanari, Ricci-Tersenghi, Semerjian, Zdeborová ’07-’08

Each box depicts geometry of solution space (subgraph of \(\mathcal{X}^n\)) for a typical CSP instance, in a certain regime of \(\alpha\).

Between boxes 4&5 is satisfiability transition \(\alpha_{\text{sat}}\).

Conjectural phase diagram (4/10)
Conjectural phase diagram

The satisfiability threshold is only one aspect of a much richer picture predicted by physicists for a broad class of CSPs.

(Krząkała, Montanari, Ricci-Tersenghi, Semerjian, Zdeborová ’07-’08)

Each box depicts geometry of solution space (subgraph of \( \mathcal{X}^n \)) for a typical CSP instance, in a certain regime of \( \alpha \).

Between boxes 4&5 is \textit{satisfiability transition} \( \alpha_{\text{sat}} \).
Between boxes 3&4 is \textit{condensation} or \textit{Kauzmann transition} \( \alpha_{\text{cond}} \).
Let $\text{SOL} \subseteq \mathcal{X}^n$ be the solution space of the (random) instance. Let $Z \equiv \#\text{SOL}$. 
Condensation

Let $\text{SOL} \subseteq \mathcal{X}^n$ be the solution space of the (random) instance. Let $Z \equiv \#\text{SOL}$. The physics conjecture is that

$$Z \asymp \mathbb{E} Z = \exp(\Theta(n)) \quad \text{for } \alpha \in (0, \alpha_{\text{cond}}),$$
$$Z \leq \mathbb{E} Z \exp(-\Omega(n)) \quad \text{for } \alpha \in (\alpha_{\text{cond}}, \alpha_{\text{sat}}).$$
Condensation

Let \( \text{SOL} \subseteq \mathcal{X}^n \) be the solution space of the (random) instance. Let \( Z \equiv \#\text{SOL} \). The physics conjecture is that

\[
Z \asymp \mathbb{E}Z = \exp(\Theta(n)) \quad \text{for} \quad \alpha \in (0, \alpha_{\text{cond}}),
\]
\[
Z \leq \mathbb{E}Z \exp(-\Omega(n)) \quad \text{for} \quad \alpha \in (\alpha_{\text{cond}}, \alpha_{\text{sat}}).
\]

For \( \alpha < \alpha_{\text{cond}} \), either one large cluster of solutions (boxes 1&2), or \( \exp(n) \) many clusters (box 3). For \( \alpha \in (\alpha_{\text{cond}}, \alpha_{\text{sat}}) \), most mass in bounded number of clusters.
Condensation

Let \( \text{SOL} \subseteq \mathcal{X}^n \) be the solution space of the (random) instance. Let \( Z \equiv \#\text{SOL} \). The physics conjecture is that

\[
\begin{align*}
Z & \asymp \mathbb{E}Z = \exp(\Theta(n)) \quad \text{for } \alpha \in (0, \alpha_{\text{cond}}), \\
Z & \leq \mathbb{E}Z \exp(-\Omega(n)) \quad \text{for } \alpha \in (\alpha_{\text{cond}}, \alpha_{\text{sat}}).
\end{align*}
\]

For \( \alpha < \alpha_{\text{cond}} \), either one large cluster of solutions (boxes 1&2), or \( \exp(n) \) many clusters (box 3). For \( \alpha \in (\alpha_{\text{cond}}, \alpha_{\text{sat}}) \), most mass in \textit{bounded} number of clusters.

Let \( \nu \) be uniform measure on \text{SOL}, and sample \( X \) from \( \nu \).
Condensation

Let $\text{SOL} \subseteq \mathcal{X}^n$ be the solution space of the (random) instance. Let $Z \equiv \#\text{SOL}$. The physics conjecture is that

\[
\begin{align*}
Z &\geq \mathbb{E}Z = \exp(\Theta(n)) \quad \text{for } \alpha \in (0, \alpha_{\text{cond}}), \\
Z &\leq \mathbb{E}Z \exp(-\Omega(n)) \quad \text{for } \alpha \in (\alpha_{\text{cond}}, \alpha_{\text{sat}}).
\end{align*}
\]

For $\alpha < \alpha_{\text{cond}}$, either one large cluster of solutions (boxes 1&2), or $\exp(n)$ many clusters (box 3). For $\alpha \in (\alpha_{\text{cond}}, \alpha_{\text{sat}})$, most mass in bounded number of clusters.

Let $\nu$ be uniform measure on $\text{SOL}$, and sample $X$ from $\nu$. It is predicted that $X$ has correlation decay for $\alpha \leq \alpha_{\text{cond}}$, long-range correlations for $\alpha > \alpha_{\text{cond}}$. 
Complexity function $\Sigma(s)$

Conjectural function $\Sigma(s) \equiv \Sigma(s; \alpha)$ such that number of clusters of size $\exp\{ns\}$ concentrates around a mean value $\exp\{n\Sigma(s)\}$:
Complexity function $\Sigma(s)$

Conjectural function $\Sigma(s) \equiv \Sigma(s; \alpha)$ such that number of clusters of size $\exp\{ns\}$ concentrates around a mean value $\exp\{n\Sigma(s)\}$:

$$\mathbb{E}Z \equiv \sum_s \exp\{ns\} \exp\{n\Sigma(s)\}$$
**Complexity function $\Sigma(s)$**

Conjectural function $\Sigma(s) \equiv \Sigma(s; \alpha)$ such that number of clusters of size $\exp\{ns\}$ concentrates around a mean value $\exp\{n\Sigma(s)\}$:

$$\mathbb{E}Z \doteq \sum_s \exp\{ns\} \exp\{n\Sigma(s)\}$$

$\mathbb{E}Z$ is dominated by $s_1$ where $\Sigma'(s_1) \equiv -1$ (depending on $\alpha$).
Complexity function $\Sigma(s)$

Conjectural function $\Sigma(s) \equiv \Sigma(s; \alpha)$ such that number of clusters of size $\exp\{ns\}$ concentrates around a mean value $\exp\{n\Sigma(s)\}$:

$$\mathbb{E}Z \equiv \sum_s \exp\{ns\} \exp\{n\Sigma(s)\}$$

$\mathbb{E}Z$ is dominated by $s_1$ where $\Sigma'(s_1) \equiv -1$ (depending on $\alpha$).
Complexity function $\Sigma(s)$

Conjectural function $\Sigma(s) \equiv \Sigma(s; \alpha)$ such that number of clusters of size $\exp\{ns\}$ concentrates around a mean value $\exp\{n\Sigma(s)\}$:

$$\mathbb{E}Z \doteq \sum_s \exp\{ns\} \exp\{n\Sigma(s)\}$$

$\mathbb{E}Z$ is dominated by $s_1$ where $\Sigma'(s_1) \equiv -1$ (depending on $\alpha$).
Conjectural function $\Sigma(s) \equiv \Sigma(s; \alpha)$ such that number of clusters of size $\exp\{ns\}$ concentrates around a mean value $\exp\{n\Sigma(s)\}$:

$$\mathbb{E}Z = \sum_s \exp\{ns\} \exp\{n\Sigma(s)\}$$

$\mathbb{E}Z$ is dominated by $s_1$ where $\Sigma'(s_1) \equiv -1$ (depending on $\alpha$).
Conjectural function $\Sigma(s) \equiv \Sigma(s; \alpha)$ such that number of clusters of size $\exp\{ns\}$ concentrates around a mean value $\exp\{n\Sigma(s)\}$:

$$\mathbb{E}Z \doteq \sum_s \exp\{ns\} \exp\{n\Sigma(s)\}$$

$\mathbb{E}Z$ is dominated by $s_1$ where $\Sigma'(s_1) \equiv -1$ (depending on $\alpha$).
Complexity funtion $\Sigma(s)$

Conjectural function $\Sigma(s) \equiv \Sigma(s; \alpha)$ such that number of clusters of size $\exp\{ns\}$ concentrates around a mean value $\exp\{n\Sigma(s)\}$:

$$\mathbb{E}Z \equiv \sum_{s} \exp\{ns\} \exp\{n\Sigma(s)\}$$

$\mathbb{E}Z$ is dominated by $s_1$ where $\Sigma'(s_1) \equiv -1$ (depending on $\alpha$).
Complexity function $\Sigma(s)$

Conjectural function $\Sigma(s) \equiv \Sigma(s; \alpha)$ such that number of clusters of size $\exp\{ns\}$ concentrates around a mean value $\exp\{n\Sigma(s)\}$:

$$\mathbb{E} Z \triangleq \sum_s \exp\{ns\} \exp\{n\Sigma(s)\}$$

$\mathbb{E} Z$ is dominated by $s_1$ where $\Sigma'(s_1) \equiv -1$ (depending on $\alpha$).
**Complexity function \( \Sigma(s) \)**

Conjectural function \( \Sigma(s) \equiv \Sigma(s; \alpha) \) such that number of clusters of size \( \exp\{ns\} \) concentrates around a mean value \( \exp\{n\Sigma(s)\} \):

\[
\mathbb{E} Z = \sum_{s} \exp\{ns\} \exp\{n\Sigma(s)\}
\]

\( \mathbb{E} Z \) is dominated by \( s_1 \) where \( \Sigma'(s_1) \equiv -1 \) (depending on \( \alpha \)).
Complexity function $\Sigma(s)$

Conjectural function $\Sigma(s) \equiv \Sigma(s; \alpha)$ such that number of clusters of size $\exp\{ns\}$ concentrates around a mean value $\exp\{n\Sigma(s)\}$:

$$\mathbb{E} Z \doteq \sum_s \exp\{ns\} \exp\{n\Sigma(s)\}$$

$\mathbb{E} Z$ is dominated by $s_1$ where $\Sigma'(s_1) \equiv -1$ (depending on $\alpha$).

Conjectured that $n^{-1} \log Z \rightarrow \arg\max_s \{s + \Sigma(s) : \Sigma(s) \geq 0\}$.
Main result

We consider random regular $k$-NAE-SAT: $n$ variables of degree $d$, $m$ clauses of degree $k$, $\alpha = m/n = d/k$. 
Main result

We consider random regular $k$-NAE-SAT: $n$ variables of degree $d$, $m$ clauses of degree $k$, $\alpha = m/n = d/k$.

Appeal of this model is that it is easiest in this class of problems (binary and highly symmetric), yet is believed to exhibit the same qualitatively interesting behavior.
Main result

We consider random regular $k$-NAE-SAT: $n$ variables of degree $d$, $m$ clauses of degree $k$, $\alpha = m/n = d/k$.

Appeal of this model is that it is easiest in this class of problems (binary and highly symmetric), yet is believed to exhibit the same qualitatively interesting behavior.

**Theorem** (Sly, S., Zhang '16). In random regular $k$-NAE-SAT for $k \geq k_0$, the number of solutions $Z$ has asymptotics

$$n^{-1} \log Z \longrightarrow f(\alpha) \text{ in probability as } n \to \infty,$$

for explicit $f(\alpha) = f(\alpha; k)$ that matches the physics prediction.
Main result

We consider random regular $k$-NAE-SAT: $n$ variables of degree $d$, $m$ clauses of degree $k$, $\alpha = m/n = d/k$.

Appeal of this model is that it is easiest in this class of problems (binary and highly symmetric), yet is believed to exhibit the same qualitatively interesting behavior.

**Theorem** (Sly, S., Zhang ’16). In random regular $k$-NAE-SAT for $k \geq k_0$, the number of solutions $Z$ has asymptotics

$$n^{-1} \log Z \longrightarrow f(\alpha) \text{ in probability as } n \rightarrow \infty,$$

for explicit $f(\alpha) \equiv f(\alpha; k)$ that matches the physics prediction. This locates a threshold $\alpha_{\text{cond}} < \alpha_{\text{sat}}$ such that $f(\alpha)$ agrees with $2(1 - 2/2^k)^\alpha$ for $\alpha \leq \alpha_{\text{cond}}$, and diverges thereafter.
Explicit formula for NAE-SAT free energy

\( \forall \lambda \in [0, 1], \) we construct probability measures \( \mu^\lambda, \hat{\mu}^\lambda \) on \([0, 1]\) such that

\[
\mu^\lambda(B) = \mathcal{Z}^\lambda^{-1} \int \left( 2 - \prod_{i=1}^{k-1} x_i - \prod_{i=1}^{k-1} (1 - x_i) \right)^\lambda \prod_{i=1}^{k-1} \frac{1}{2 - \prod_{i=1}^{k-1} x_i - \prod_{i=1}^{k-1} (1 - x_i)} \in B \prod_{i=1}^{k-1} \hat{\mu}^\lambda(dx_i)
\]

\[
\hat{\mu}^\lambda(B) = \mathcal{Z}^\lambda^{-1} \int \left( \prod_{i=1}^{d-1} y_i + \prod_{i=1}^{d-1} (1 - y_i) \right)^\lambda \prod_{i=1}^{d-1} y_i + \prod_{i=1}^{d-1} (1 - y_i) \in B \prod_{i=1}^{d-1} \mu^\lambda(dy_i)
\]

Define \( \mathcal{F}(\lambda) \equiv \log \mathcal{Z}^\lambda + k \log \hat{\mathcal{Z}}^\lambda - d \log \mathcal{Z}^\lambda \) where

\[
\dot{\omega}^\lambda(B) = \mathcal{Z}^\lambda^{-1} \int \left( \prod_{i=1}^{d} y_i + \prod_{i=1}^{d} (1 - y_i) \right)^\lambda \prod_{i=1}^{d} y_i + \prod_{i=1}^{d} (1 - y_i) \in B \prod_{i=1}^{d} \hat{\mu}^\lambda(dy_i)
\]

\[
\hat{\omega}^\lambda(B) = \mathcal{Z}^\lambda^{-1} \int \left( 1 - \prod_{i=1}^{k} x_i - \prod_{i=1}^{k} (1 - x_i) \right)^\lambda \prod_{i=1}^{k} x_i - \prod_{i=1}^{k} (1 - x_i) \in B \prod_{i=1}^{k} \mu^\lambda(dx_i)
\]

\[
\tilde{\omega}^\lambda(B) = \mathcal{Z}^\lambda^{-1} \int \left( xy + (1 - x)(1 - y) \right)^\lambda \prod_{i=1}^{d} xy + (1 - x)(1 - y) \in B \prod_{i=1}^{d} \mu^\lambda(dx) \hat{\mu}^\lambda(dy).
\]

Let \( s^\lambda \equiv \langle \dot{\omega}^\lambda, \log x \rangle + k \langle \hat{\omega}^\lambda, \log x \rangle - dk \langle \hat{\omega}^\lambda, \log x \rangle, \) and for \( s = s^\lambda \) let \( \Sigma(s) = \mathcal{F}(\lambda) - \lambda s^\lambda = -\mathcal{F}^*(s) \) (Legendre dual). We prove that

\[
\lim_{n \to \infty} n^{-1} \log Z = f(\alpha) = \arg\max_s \{ s + \Sigma(s) : \Sigma(s) \geq 0 \}.
\]
Let $G$ be a $(d, k)$-regular NAE-SAT graph on $n$ variables.
Long-range correlations

Let $G$ be a $(d, k)$-regular NAE-SAT graph on $n$ variables.

Variables $v_1, \ldots, v_k$ (random): remove each $v_i$ with its clauses to form $G^\circ$, which has $dk(k - 1)$ variables $u_j$ of degree $d - 1$. 
Let $\mathcal{G}$ be a $(d, k)$-regular NAE-SAT graph on $n$ variables.

Variables $v_1, \ldots, v_k$ (random): remove each $v_i$ with its clauses to form $\mathcal{G}^\circ$, which has $dk(k - 1)$ variables $u_j$ of degree $d - 1$. Add $d(k - 1)$ random clauses to $\mathcal{G}^\circ$ to form a new $(d, k)$-regular instance $\mathcal{G}'$ with $n - k$ variables.
Long-range correlations

Let $\mathcal{G}$ be a $(d, k)$-regular NAE-SAT graph on $n$ variables.

Variables $v_1, \ldots, v_k$ (random): remove each $v_i$ with its clauses to form $\mathcal{G}^\circ$, which has $dk(k-1)$ variables $u_j$ of degree $d-1$. Add $d(k-1)$ random clauses to $\mathcal{G}^\circ$ to form a new $(d, k)$-regular instance $\mathcal{G}'$ with $n-k$ variables.

(if free energy exists) $Z = Z(\mathcal{G}) \equiv \left( \frac{Z(\mathcal{G})}{Z(\mathcal{G}^\circ)} \frac{Z(\mathcal{G}^\circ)}{Z(\mathcal{G}')}) \right)^{n/k}$. 

Let $G$ be a $(d, k)$-regular NAE-SAT graph on $n$ variables.

Variables $v_1, \ldots, v_k$ (random): remove each $v_i$ with its clauses to form $G^\circ$, which has $dk(k - 1)$ variables $u_j$ of degree $d - 1$. Add $d(k - 1)$ random clauses to $G^\circ$ to form a new $(d, k)$-regular instance $G'$ with $n - k$ variables.

\[ \text{(if free energy exists)} \quad Z = Z(G) = \left( \frac{Z(G)}{Z(G^\circ)} \frac{Z(G^\circ)}{Z(G')} \right)^{n/k}. \]

To compute RHS, enough to understand joint law on $u_j$'s in $G^\circ$. 
Long-range correlations

Let $G$ be a $(d, k)$-regular NAE-SAT graph on $n$ variables.

Variables $v_1, \ldots, v_k$ (random): remove each $v_i$ with its clauses to form $G^\circ$, which has $dk(k - 1)$ variables $u_j$ of degree $d - 1$. Add $d(k - 1)$ random clauses to $G^\circ$ to form a new $(d, k)$-regular instance $G'$ with $n - k$ variables.

\[
\begin{align*}
\text{(if free energy exists)} \quad Z &= Z(G) = \left( \frac{Z(G)}{Z(G^\circ)} \frac{Z(G^\circ)}{Z(G')} \right)^{n/k}.
\end{align*}
\]

To compute RHS, enough to understand joint law on $u_j$'s in $G^\circ$.

Taking $u_j$ iid Bernoulli(1/2) gives the prediction $Z \doteq \mathbb{E} Z$. 
Long-range correlations

Let $\mathcal{G}$ be a $(d, k)$-regular NAE-SAT graph on $n$ variables.

Variables $v_1, \ldots, v_k$ (random): remove each $v_i$ with its clauses to form $\mathcal{G}^\circ$, which has $dk(k - 1)$ variables $u_j$ of degree $d - 1$. Add $d(k - 1)$ random clauses to $\mathcal{G}^\circ$ to form a new $(d, k)$-regular instance $\mathcal{G}'$ with $n - k$ variables.

$$Z = Z(\mathcal{G}) \doteq \left( \frac{Z(\mathcal{G})}{Z(\mathcal{G}^\circ)} \frac{Z(\mathcal{G}^\circ)}{Z(\mathcal{G}')} \right)^{n/k}.$$

To compute RHS, enough to understand joint law on $u_j$’s in $\mathcal{G}^\circ$.

Taking $u_j$ iid Bernoulli(1/2) gives the prediction $Z \doteq \mathbb{E}Z$.

Above $\alpha_{\text{cond}}$ this is false — determining the true behavior of $Z$ requires to understand the non-trivial dependencies among the $u_j$’s.
Open question

It remains an open question to establish the full statistical physics prediction for the typical structure of the random measure $\nu$ over $\{+,-\}^n$ (uniform on the CSP solution space SOL).
Open question

It remains an open question to establish the full statistical physics prediction for the typical structure of the random measure $\nu$ over $\{+,-\}^n$ (uniform on the CSP solution space $\text{SOL}$).

For example: show that for $\alpha \in (\alpha_{\text{cond}}, \alpha_{\text{sat}})$, the distribution of weights under $\nu$ forms a Poisson–Dirichlet process with parameter $\zeta = \zeta(\alpha) = -\Sigma'(s_{\text{max}})$.

analogies with Sherrington–Kirkpatrick model
(Parisi, Talagrand, Panchenko)
Open question

It remains an open question to establish the full statistical physics prediction for the typical structure of the random measure $\nu$ over $\{+, -\}^n$ (uniform on the CSP solution space $\text{SOL}$).

For example: show that for $\alpha \in (\alpha_{\text{cond}}, \alpha_{\text{sat}})$, the distribution of weights under $\nu$ forms a Poisson–Dirichlet process with parameter $\zeta = \zeta(\alpha) = -\Sigma'(s_{\text{max}})$.

analogies with Sherrington–Kirkpatrick model
(Parisi, Talagrand, Panchenko)

Thanks!