Geometric properties of satisfying assignments of random ϵ-1-in-k SAT

Gabriel Istrate [a]

[a] *eAustria Research Institute*
Bd. V. Pârvan 4, cam 045B,
Timişoara, RO-300223, Romania

Abstract

We study the geometric structure of the set of solutions of random ϵ-1-in-k SAT problem [ACIM01,RSZ07]. For $l \geq 1$, two satisfying assignments A and B are l-*connected* if there exists a sequence of satisfying assignments $A_0, A_1, \ldots A_l$, $A_0 = A$, $A_l = B$, A_i and A_{i+1} are at Hamming distance at most l.

We first prove that w.h.p. two satisfying assignments of a random ϵ-1-in-k SAT instance are $O(\log n)$-connected. Also, there exists $\epsilon_0 \in (0, \frac{1}{k-2})$ such that w.h.p. no two satisfying assignments at distance at least $\epsilon_0 \cdot n$ form a hole. We believe that this is true for all $\epsilon > 0$, and thus satisfying assignmennts of a random 1-in-k SAT instance form a single cluster.

Key words: ϵ-1-in-k SAT, overlaps, random graphs, phase transition.

1 Introduction

The geometric structure of solutions of random constraint satisfaction problems has lately become a topic of significant interest [ART06], [MMZ05], [KMRT+07], [Ist07]. The motivation is the study of *phase transitions in combinatorial optimization problems* [HW05,PIM06], particularly using methods from physics of Spin Glasses such as the so-called *replica method* and *cavity approach*. These methods, so far without a complete rigorous foundation, are largely responsible for our substantially increased understanding of structural properties of constraint satisfaction problems.

Email address: gabrielistrate@acm.org (Gabriel Istrate).

Of special interest are two special cases when the replica method [MPV87] applies, those characterized by so-called "replica symmetry" or "one-step replica symmetry breaking". These assumptions make predictions on (and have implications for) the typical geometry of the set of solutions of a random instance. Specifically, the two assumptions seem to constrain the set of solutions in the following way:

(1) For problems displaying replica-symmetry, the set of solutions forms a single cluster. The typical overlap is concentrated around a single value, and the distribution of overlaps has continuous support.
(2) In the presence of one-step replica-symmetry breaking, the solution space is no longer connected, but breaks into a number of clusters. These clusters correspond to the emergence of $\Omega(n)$-size *mini-backbones*, sets of variables taking the same value for all solutions in a cluster. The clusters do not possess further geometrical structure (hence the "one-step" RSB), and are separated by $\Omega(n)$ variable flips. The distribution of overlaps develops multiple peaks and has discontinuous support.

In this paper we study the geometric structure of the set of solutions of the *random 1-in-k SAT problem [ACIM01]*, and a generalization of this problem from [RSZ07], *random ϵ-1-in-k satisfiability.* This latter problem is parameterized by a real number $\epsilon \in [0, 1/2]$, and essentially coincides with 1-in-k SAT for $\epsilon = 1/2$. Results in the cited work suggest that for $\epsilon \in (\epsilon_c, 1/2]$, where $\epsilon_c \sim 0.2726$ is the solution of equation $2x^3 - 2x^2 + 3x - 1 = 0$, ϵ-1-in-k SAT behaves qualitatively "like 2-SAT". In particular, for both problems the threshold location can be predicted in both cases by a "percolation of contradictory cycles" argument, and the replica symmetry ansatz is correct.

For 2-SAT we have previously proved [Ist07] two results supporting replica symmetry: with high probability satisfying assignments of a random 2-CNF formula with clause/variable ratio $c < 1$ form a single cluster; also the overlap distribution has continuous support. From the heuristic similarity of the two problems, we expect similars results to also hold for ϵ-1-in-k SAT, $\epsilon \in (\epsilon_c, 1/2]$. Though the replica symmetric approach seems correct [RSZ07], we cannot rigorously prove such results. Instead:

- We first note (Theorem 1) that the replica symmetric picture holds in the subcritical regime of the formula hypergraph.
- We show (Theorem 2) that for any two given assignments A, B at sufficiently large Hamming distance, with probability $1-o(1)$ A, B are $O(\log n)$-connected (conditional on being satisfying assignments).
- We show (Theorem 4) that with probability $1 - o(1)$ (as $n \to \infty$) the set of satisfying assignments of a random instance of 1-in-k SAT with clause/variable ratio $\lambda < \frac{1}{\binom{k}{2}}$ does *not* have holes of size $> \epsilon_k n$, for some $\epsilon_k > 0$.

2 Preliminaries

Definition 1 *Let $\epsilon \in [0, 1/2]$. An instance of the* ϵ-1-in-k SAT problem *is a propositional formula Φ in clausal form, with exactly k literals in each clause. A* satisfying assignment *for instance Φ is a mapping of variables in Φ to $\{0, 1\}$ such that in each clause of Φ* exactly *one literal is true.*

We will use two related models to *the constant probability model* to generate random instances of ϵ-1-in-k SAT.

(1) The *counting* model is parameterized by a real number $r > 0$. A random instance of ϵ-1-in-k SAT will have rn clauses, out of which $rn \cdot \epsilon^i (1 - \epsilon)^j$ have i negative and j positive variables (where $i + j = k$).
(2) The *constant probability* model is parameterized by a probability p. A random instance Φ is obtained by including independently with probability $p\epsilon^i (1 - \epsilon)^j$ each possible clause with i negative and j positive variables (where $i + j = k$).

Using standard methods ([Bol85], Chapter 2; see also a similar issue in [Mol02]) the two models we described above for ϵ-1-in-k SAT are equivalent:

Lemma 1 *Let $r > 0$ and let $p = p(n)$ be such that $p \cdot \binom{n}{k} = rn$. Let Φ_1 be a random instance of ϵ-1-in-k SAT with rn clauses generated according to the counting model, and let Φ_2 be a random instance of ϵ-1-in-k SAT, generated according to the constant probability model with probability p. Let B be an arbitrary monotone property and $\mu \in \{0, 1\}$. Then:*

$$\lim_{n \to \infty} Prob[\Phi_1 \models B] = \mu,$$

iff

$$\lim_{n \to \infty} Prob[\Phi_2 \models B] = \mu.$$

In the sequel we will liberally use one model or the other, depending on our goals.

Results in [ACIM01] and [RSZ07] imply the fact that for $\epsilon \in (\epsilon_c, 1/2]$ the threshold of satisfiability for the ϵ-1-in-k satisfiability (under the counting model) is located at critical value [1]

$$r_{k,\epsilon} = \frac{1}{4\epsilon(1 - \epsilon)} \cdot \frac{1}{\binom{k}{2}}$$

[1] in [RSZ07] the result is only stated and proved for $k = 3$, but the method outlined there works for any $k \geq 3$

The corresponding threshold for ϵ-1-in-k SAT under the constant probability model is

$$p_{k,\epsilon} = \frac{(k-2)!}{2\epsilon(1-\epsilon)} \cdot n^{1-k}$$

Definition 2 *The* overlap *of two assignments A and B for a formula Φ on n variables, denoted by overlap(A, B), is the fraction of variables on which the two assignments agree. Formally overlap$(A, B) = \frac{|\{i: A(x_i)=B(x_i)\}|}{n}$.*

The distribution of overlaps is, indeed, the original order parameter that was originally used to study the phase transition in random k-SAT [MZ97].

Definition 3 *Let $l \geq 1$ be an integer and let A, B be two satisfying assignments of an instance Φ of ϵ-1-in-k SAT. Pair (A, B) is called* l-connected *if there exists a sequence of satisfying assignments $A_0, A_1, \ldots A_r$, $A_0 = A$, $A_r = B$, with A_i and A_{i+1} at Hamming distance at most l.*

Definition 4 *Let A, B be arbitrary assignments for the variables of an instance Φ of 1-in-k SAT. Pair (A, B) is called a* hole *if:*

(1) A, B are satisfying assignments for Φ.
(2) There exists no satisfying assignment C with $d_H(A, C) + d_H(C, B) = d_H(A, B)$ (where d_H is the Hamming distance).

The number $\lambda = d_H(A, B)$ is called the size of hole (A, B).

3 Results

First, we prove that for low enough clause/variable ratios the set of satisfying assignments of a random instance of 1-in-k SAT behaves in the way predicted by the replica symmetry ansatz:

Theorem 1 *Let $k \geq 3$ and $c < 1/k(k-1)$. Then there exists $\gamma > 0$ such that, with probability $1 - o(1)$ (as $n \to \infty$), a random instance of 1-in-k SAT with n variables and cn clauses has all its satisfying assignments $\gamma \log(n)$-connected.*

We believe (and would like to prove) that the result in Theorem 1 is valid for values of c up to $2/k(k-1)$ (the satisfiability threshold of 1-in-k SAT [ACIM01]). We cannot prove this statement. Instead, we prove a result that implies a weaker claim for 1-in-k SAT but is valid, more generally, for ϵ-1-in-k SAT:

Theorem 2 *Let $0 \leq \epsilon \leq \frac{1}{2}$, let $c < 1$, let Φ be a random instance of ϵ-1-in-k SAT with clause/variable ratio $\frac{1}{\max[4\epsilon(1-\epsilon), \epsilon^2+(1-\epsilon)^2]} \cdot \frac{c}{\binom{k}{2}}$, and let $(A_n, B_n) \in$*

$\{0,1\}^n \times \{0,1\}^n$ *such that*

$$2 \cdot [overlap(A_n, B_n)(1-\epsilon)]^{k-2} \leq 1. \quad (1)$$

Then there exists $\lambda_{c,\epsilon} > 0$ *such that*

$$\Pr[(A_n, B_n) \text{ are not } \lambda_{c,\epsilon} \cdot \log(n)\text{-connected} \mid A_n, B_n \models \Phi\} < 1/n.$$

for large enough n.

In other words, every *single pair* of assignments is likely to be $O(\log n)$-connected, conditional on being a pair of satisfying assignments, *and being far enough*. The remarkable thing about condition (1) is that it depends on ϵ and k but **not** c. For certain values of ϵ (we specifically believe this is the case in the region $[0, \epsilon_c)$) it might simply signal the fact that there are no satisfying assignments of a certain overlap. This is not a problem for $\epsilon = 1/2$ (i.e. for the 1-in-k SAT), since the condition (1) is trivially satisfied for every overlap value. For this problem, the results in Theorems 1 and 2 are highly reminiscent of the results for 2-SAT in [Ist07]. On the other hand for any $c < 1$ and all $q \in (0,1)$ a random 2-CNF formula has w.h.p. two satisfying assignments of overlap approximately q. Despite 2-SAT and 1-in-k SAT being similar in other ways (see e.g. [IPB05]), the corresponding statement is *not* true for 1-in-k SAT:

Theorem 3 *For any* $c > 0$ *there exists* $q_c \in (0,1)$ *such that w.h.p. a random instance of 1-in-k SAT with clause/variable ratio c has, with probability* $1-o(1)$ *no satisfying assignments of overlap less than* q_c.

We next consider an alternative approach to characterizing the geometry of satisfying assignments of 1-in-k SAT by considering *holes* in the set of such assignments. For other problems, e.g. k-SAT, $k \geq 9$, that display clustering the set of satisfying assignments has large holes. Indeed [MMZ05], for certain values of $q_1 < q_2 < q_3$ and $c > 0$, a random instance of k-SAT of constraint density c will have, with high probability, satisfying assignments of overlap q_3, but no satisfying assignments of overlap λ, $q_1 \leq \lambda \leq q_2$. Consider A, B two satisfying assignments of overlap q_3. Then the set of assignments C between A, B contains a hole of size at least $(q_2 - q_1)n$.

We would like to stat that for any $\lambda > 0$ a random instance Φ of 1-in-k SAT as in Theorem 2 has *no* hole of size at least $\lambda \cdot n$. We cannot, however prove this result (we leave it as an intriguing open problem). Instead we prove a weaker result:

Theorem 4 *For any* $k \geq 3$ *there exists* $\epsilon_k \in (0, 1/k-1)$ *such that with probability* $1-o(1)$ *(as* $n \to \infty$*) a random instance of 1-in-k SAT of clause/variable ratio* $c < 1/\binom{k}{2}$ *has no holes of size* $\geq \epsilon_k \cdot n$.

4 Proofs

4.1 Proof of Theorem 1

First, note that location $c = 1/k(k-1)$ in Theorem 1 is the phase transition location for the random k-uniform hypergraph [SPS85]. For smaller values of c, by results in [SPS85] there exists $\gamma > 0$ such that w.h.p. the largest connected component of H has size no larger than $\gamma \log(n)$.

This argument immediately implies the desired result. Indeed, let P, Q be two arbitrary satisfying assignments, and let (P_1, Q_1), (P_2, Q_2), $\ldots, (P_v, Q_v)$ represent the restrictions of P and Q on the connected components of Φ on which $P \neq Q$. One can obtain a path from P to Q by starting at P and then obtain the next satisfying assignments by replacing P_i by Q_i for $i = 1, \ldots, v$. In this way we are constructing satisfying assignments for Φ, since we change assignments consistently on connected components of the formula hypergraph. We are changing at most $\gamma \log(n)$ values at a time, since this is the upper bound on the component size of H.

4.2 Proof of Theorem 3

We prove the theorem by a simple first moment bound. We will work with the constant probability model.

Definition 5 *Let Φ be a formula. A* cover of Φ *is a set of variables W such that every clause of Φ contains at least one variable in W.*

The theorem now follows from the following two lemmas:

Lemma 2 *Let A,B be satisfying assignments of an instance Φ of 1-in-k SAT. Then the set $\{x : A(x) = B(x)\}$ is a cover of Φ.*

Proof.

Suppose this was not the case, and there exists a clause C of Φ consisting entirely of variables in the set $\{x : A(x) \neq B(x)\}$. Then clause C has two satisfying assignments at distance k. But this is not possible, since all satisfying assignments of a given 1-in-k clause have Hamming distance two. □

Lemma 3 *For any $c > 0$ there exists a $q_c > 0$ such that a random instance of 1-in-k SAT of clause/variable ratio c has, w.h.p. no cover of size at most $q_c n$.*

Proof. Let $\lambda < 1/2$. The probability that Φ has a cover of size $i \leq \lambda n$ is at most

$$\sum_{i=1}^{\lambda n} \binom{n}{i} (1-p)^{\binom{n-i}{k}} \leq \lambda n \cdot (1-p)^{\binom{n(1-\lambda)}{k}} \cdot \left[\sum_{i=1}^{\lambda n} \binom{n}{i} \right] \leq$$
$$\leq (\lambda n)^2 \cdot e^{-p\binom{n(1-\lambda)}{k}} \cdot \binom{n}{\lambda n} \leq (1+w) \cdot (\lambda n)^2 \cdot e^{-p\binom{n(1-\lambda)}{k}} \cdot \left(\frac{1}{\lambda^{\lambda}(1-\lambda)^{1-\lambda}} \right)^n \cdot$$
$$\cdot \frac{1}{\sqrt{2\pi\lambda(1-\lambda)n}}.$$

for some $w > 0$ (we have applied the fact that $\lambda < 1/2$ and Stirling's formula) So the probability is at most

$$\frac{(1+w)(\lambda n)^2}{\sqrt{2\pi\lambda(1-\lambda)n}} \cdot e^{-p(n(1-\lambda))^k/k! + n[\lambda \ln(1/\lambda) + (1-\lambda)\ln(1/(1-\lambda))]} =$$
$$= \frac{(1+w)(\lambda n)^2}{\sqrt{2\pi\lambda(1-\lambda)n}} \cdot e^{-n[c(1-\lambda))^k - \lambda \ln(1/\lambda) - (1-\lambda)\ln(1/(1-\lambda))]}$$

Since $c > 0$ and $\lim_{\lambda \to 0} \lambda \ln(1/\lambda) - (1-\lambda)\ln(1/(1-\lambda)) = 0$, there exists $q_c > 0$ such that for $\lambda < q_c$, $c(1-\lambda))^k - \lambda \ln(1/\lambda) + (1-\lambda)\ln(1/(1-\lambda)) > 0$. Thus, for $q < q_c$ the probability that a random instance of 1-in-k SAT has a cover of size at most qn is exponentially small. □

4.3 Proof of Theorem 2

For a pair of assignments (A, B) define

$$V_0 = \{x : A(x) = B(x) = 0\}, V_1 = \{x : A(x) = 0, B(x) = 1\},$$
$$V_2 = \{x : A(x) = 1, B(x) = 0\}, V_3 = \{x : A(x) = B(x) = 1\}.$$

Pair (A, B) has *type* (a, b, c, d) if $|V_0| = a, |V_1| = b, |V_2| = c, |V_3| = d$. Also denote $\alpha = a/n, \beta = b/n, \gamma = c/n, \delta = d/n$.

Conditioning on A, B being satisfying assignments, define a graph H on the set of variables in $A \neq B$ as follows: x and y are connected if there exists a clause C of Φ consisting of $k-2$ literals whose variables are from $V_0 \cup V_3$ and x, y. Since both A and B must be satisfying assignments, only four combinations are possible for the literal combination present in C:

type	V_0 (a)	V_3 (d)	V_1 (b)	V_2 (c)	number	probability
C_1	$k-i-2$ [0]	i [i]	1 [0]	1 [0]	$\binom{a}{k-i-2}\binom{d}{i}$	$p\epsilon^i(1-\epsilon)^{k-i}$
C_2	$k-i-2$ [0]	i [i]	1 [1]	1 [1]	$\binom{a}{k-i-2}\binom{d}{i}$	$p\epsilon^{i+2}(1-\epsilon)^{k-i-2}$
C_3	$k-i-2$ [0]	i [i]	2 [1]	0 [0]	$2\binom{a}{k-i-2}\binom{d}{i}$	$p\epsilon^{i+1}(1-\epsilon)^{k-i-1}$
C_4	$k-i-2$ [0]	i [i]	0 [0]	2 [1]	$2\binom{a}{k-i-2}\binom{d}{i}$	$p\epsilon^{i+1}(1-\epsilon)^{k-i-1}$

Fig. 1. The four types of clauses leading to an edge (x, y) in graph H

(1) $(x, y \in C$ or $\overline{x}, \overline{y} \in C)$ and $A(x) \neq A(y)$, or
(2) $(x, \overline{y} \in C$ or $\overline{x}, y \in C)$ and $A(x) = A(y)$.

We can rewrite conditions (1) and (2) as

(1) $(x, y \in C$ or $\overline{x}, \overline{y} \in C)$ and $(x \in V_1 \wedge y \in V_2) \vee (x \in V_2 \wedge y \in V_1)$, or
(2) $(x, \overline{y} \in C$ or $\overline{x}, y \in C)$ and $(x, y \in V_1) \vee (x, y \in V_2)$.

To summarize this discussion, there are four types of clauses that imply the existence of an edge (x, y) in graph H. They are described in the table from Figure 4.3. The semantics of columns in the table is the following: first column (type) lists the four types of clauses, labeled C_1 to C_4. Columns labeled V_0 to V_3 contain two numbers. The first one is the number of literals of the given clause type that are in the set V_j. The second number (in square brackets) lists the number of negated variables in the set V_j. Column labeled "number" computes the total number of clauses of type C_i. The column labeled "Probability" lists the probability that a fixed clause of type C_i be in Φ.

The probability that an edge is present in graph H is the same for all pairs (x, y) such that $A(x) = A(y)$. Similarly the probability that an edge is present in graph H is the same for all pairs (x, y) such that $A(x) \neq A(y)$. We denote by $\mu_= = \mu_=(n, a, b, c, d)$ and $\mu_{\neq} = \mu_{\neq}(n, a, b, c, d)$ these two probabilities.

$$\mu_= \leq p \cdot \sum_{i=0}^{k-2} \binom{a}{k-i-2}\binom{d}{i} \cdot \left[\epsilon^i(1-\epsilon)^{k-i} + \epsilon^{i+2}(1-\epsilon)^{k-i-2}\right]$$

$$\mu_{\neq} \leq p \cdot \left\{ \sum_{i=0}^{k-2} \binom{a}{k-i-2}\binom{d}{i} \cdot \left[2\epsilon^{i+1}(1-\epsilon)^{k-i-1} + 2\epsilon^{i+1}(1-\epsilon)^{k-i-1}\right]\right\}$$

Applying inequality $\binom{a}{i} \leq \frac{a^i}{i!}$ and rewriting the second term of the previous inequalities we get

$$\mu_{=} \leq \frac{p[\epsilon^2+(1-\epsilon)^2]}{(k-2)!} \cdot \left\{ \sum_{i=0}^{k-2} \binom{k-2}{i} \cdot a^{k-i-2} d^i \cdot \epsilon^i (1-\epsilon)^{k-i-2} \right\} =$$
$$= \frac{p[\epsilon^2+(1-\epsilon)^2]}{(k-2)!} \cdot [a(1-\epsilon)+d\epsilon]^{k-2}$$

$$\mu_{\neq} \leq \frac{4p\epsilon(1-\epsilon)}{(k-2)!} \cdot \left\{ \sum_{i=0}^{k-2} \binom{k-2}{i} \cdot a^{k-i-2} d^i \cdot \epsilon^i (1-\epsilon)^{k-i-2} \right\} =$$
$$= \frac{4p\epsilon(1-\epsilon)}{(k-2)!} \cdot [a(1-\epsilon)+d\epsilon]^{k-2}$$

The equation $\epsilon^2+(1-\epsilon)^2 = 4\epsilon(1-\epsilon)$ has a solution $\epsilon_0 = \frac{3-\sqrt{3}}{6} \sim 0.2113....$ For $\epsilon \in (\epsilon_0, 1/2]$ we have $\epsilon^2+(1-\epsilon)^2 < 4\epsilon(1-\epsilon)$.

For $p = \lambda \cdot k! \cdot n^{1-k}$, with $\lambda = \frac{c}{4\epsilon(1-\epsilon)} \cdot \frac{2}{k(k-1)}$, with $c<1$ we have $\max(\mu_{=}, \mu_{\neq}) = \frac{2c}{n}[\alpha(1-\epsilon)+\delta\epsilon]^{k-2} \leq \frac{c}{n} \cdot 2 \cdot [overlap(A,B)(1-\epsilon)]^{k-2} \leq \frac{c}{n}$. It follows that the graph H has all its connected components of size at most $\lambda_c \log n$, where [JLR00]

$$\lambda_c = \frac{3}{(1-c)^2} \qquad (2)$$

By the discussion of clause types in Formula 4.3, edges of type (1) correspond to a constraint $x \neq y$ between a variable in V_1 and one in V_2, while clauses of type (2) correspond to a constraint $x = y$ between two variables, both in V_1 or both in V_2. H does not contain contradictory cycles, since we have conditioned on A, B being satisfying assignments.

It is easy to see that setting one value of a given variable in H uniquely determines the values on its whole connected component. Similarly, different values of x lead to opposite assignments on the connected component of x. Given the small size of the connected components, the statement of the theorem immediately follows. □

4.4 Proof of Theorem 4

We first prove a simple result about the connectivity of a random graph that we will use in the sequel.

Lemma 4 *Let $0 < c < 1$ and let G be a random graph from $G(n, c/n)$. Then*

$$Pr[G \textit{ is connected }] \leq \frac{c^{n-1}}{n}.$$

Proof. There are $u = n^{n-2}$ labeled trees on the set of vertices of G. Denote by $T_1, \ldots T_u$ the edge sets of these trees, and by W_i the event "$T_i \subseteq E[G]$". It is easy to see that G is *not* connected if and only if $\bigwedge_{i=1}^{u} \overline{W_i}$.

By Janson's inequality [AES92]

$$\Pr[\bigwedge_{i=1}^{u} \overline{W_i}] \geq \prod_{i=1}^{u} \Pr[\overline{W_i}] = (1 - (\frac{c}{n})^{n-1})^u.$$

So, by applying inequality $1 - (1-x)^a \leq ax$ we get:

$$\Pr[G \text{ connected }] \leq 1 - (1 - (\frac{c}{n})^{n-1})^{n^{n-2}} \leq \frac{c^{n-1}}{n},$$

□

We will work with the constant probability model. Each clause will be included in formula Φ with probability p, where $p \cdot 2^k \cdot \binom{n}{k} = \lambda \cdot \frac{1}{\binom{k}{2}} n$, with $\lambda < 1$.

Lemma 5 *The probability that there exist two satisfying assignments A and B of overlap i that form a hole is at most*

$$\frac{\binom{n}{i} \cdot 2^i \cdot [2^{2-k} \cdot (\frac{i}{n})^{k-2} \cdot (1 - \frac{i}{n})]^{n-i-1}}{(n-i)} \cdot e^{-\frac{\lambda n}{\binom{k}{2}}[1 - \frac{k\binom{i}{k} + 2\binom{i}{k-2}\binom{n-i}{2}}{2^k \cdot \binom{n}{k}}]}$$

Proof.

For two assignments A, B of overlap i to be satisfying assignments of a formula Φ, all clauses of Φ must fall into one of the following two categories:

(1) Clause C contains $k-1$ literals from $A = B = 0$ and one literal from $A = B = 1$.
(2) Clause C contains $k-2$ literals from $A = B = 0$ and two literals from $A \neq B$, of opposite signs in A.

There are $k \cdot \binom{i}{k}$ clauses of the first type and $2\binom{i}{k-2} \cdot \binom{n-i}{2}$ clauses of the second type. So the probability that all clauses of Φ fall into these two categories is

$$(1-p)^{2^k \cdot \binom{n}{k} - k \cdot \binom{i}{k} - 2\binom{i}{k-2}\binom{n-i}{2}} \leq e^{-p \cdot [2^k \binom{n}{k} - k \cdot \binom{i}{k} - 2\binom{i}{k-2}\binom{n-i}{2}]} =$$
$$= e^{-\frac{\lambda n}{\binom{k}{2}}[1 - \frac{k\binom{i}{k} + 2\binom{i}{k-2}\binom{n-i}{2}}{2^k \cdot \binom{n}{k}}]}.$$

The probability is at most $\binom{n}{i} \cdot 2^i$ times the probability that giving specific values to i variables we simplify the formula Φ to one for which the following graph H is connected: two variables $x, y \in \{\lambda : A(\lambda) \neq B(\lambda)\}$ are connected if there exists a clause C of Φ that contains both of them (and no other variables in that set).

This probability is at most $2 \cdot \binom{i}{k-2} \cdot p$. So an uper bound for the probability is

$$2 \cdot \frac{i^{k-2}}{(k-2)!} \cdot 2^{-k} \cdot \frac{k!}{n^k} \cdot \lambda \cdot \frac{1}{\binom{k}{2}} n = \frac{(\frac{i}{2n})^{k-2} \cdot \lambda}{2n} =$$
$$= \frac{2 \cdot (\frac{i}{2n})^{k-2} \cdot (1-\frac{i}{n}) \cdot \lambda}{n-i} \leq \frac{2^{2-k} \cdot (\frac{i}{n})^{k-2} \cdot (1-\frac{i}{n})}{n-i}.$$

Since connectivity is an increasing property, applying Lemma 4, the probability that H is connected is at most $\frac{[2^{2-k} \cdot \lambda \cdot (\frac{i}{n})^{k-2} \cdot (1-\frac{i}{n})]^{n-i-1}}{(n-i)}$. □

Let $\alpha = i/n$. Then the upper bound in the result above reads:

$$\frac{\binom{n}{\alpha \cdot n} \cdot 2^{\alpha \cdot n} \cdot [2^{2-k} \cdot \lambda \cdot (\frac{\alpha \cdot n}{n})^{k-2} \cdot (1-\frac{\alpha \cdot n}{n})]^{n-\alpha \cdot n-1}}{n(1-\alpha)} \cdot e^{-\frac{\lambda n}{\binom{k}{2}} \cdot [1-\frac{k\alpha^k + k(k-1)\alpha^{k-2}(1-\alpha)^2}{2^k}]}$$

Applying Stirling's formula for the factorial, the above expression simplifies to

$$\frac{(\frac{n}{e})^n \sqrt{2\pi n} \cdot 2^{\alpha \cdot n} \cdot [2^{2-k} \lambda (\frac{\alpha n}{n})^{k-2} (1-\frac{\alpha n}{n})]^{n-\alpha n-1} \cdot e^{-\frac{\lambda n}{\binom{k}{2}} \cdot [1-\frac{k\alpha^k + k(k-1)\alpha^{k-2}(1-\alpha)^2}{2^k}]}}{(\frac{\alpha \cdot n}{e})^{\alpha \cdot n} \sqrt{2\pi \alpha n} \cdot (\frac{(1-\alpha) \cdot n}{e})^{(1-\alpha) \cdot n} \sqrt{2\pi(1-\alpha)n} \cdot n(1-\alpha)} =$$
$$= \theta(1) \cdot \frac{2^{\alpha \cdot n} \cdot [2^{2-k} \lambda \alpha^{k-2} (1-\alpha)]^{n(1-\alpha)} \cdot e^{-\frac{\lambda n}{\binom{k}{2}} \cdot [1-\frac{k\alpha^k + k(k-1)\alpha^{k-2}(1-\alpha)^2}{2^k}]}}{\alpha^{\alpha n} \sqrt{\alpha} \cdot (1-\alpha)^{(1-\alpha)n} \sqrt{2\pi(1-\alpha)n} \cdot n(1-\alpha) \cdot (\alpha/2)^{k-2} \lambda (1-\alpha)} =$$
$$= \theta(1) \cdot \frac{n^{-3/2}}{\alpha^{k-3/2} \lambda (1-\alpha)^{5/2}} \cdot$$
$$\cdot \left\{ \frac{2^\alpha [(\alpha/2)^{k-2} \lambda (1-\alpha)]^{(1-\alpha)} e^{-\lambda/\binom{k}{2}} (1-\frac{k\alpha^k + k(k-1)\alpha^{k-2}(1-\alpha)^2}{2^k})}{\alpha^\alpha \cdot (1-\alpha)^{1-\alpha}} \right\}^n =$$
$$= \theta(1) \cdot \frac{n^{-3/2}}{\alpha^{k-3/2} \lambda (1-\alpha)^{5/2}} \cdot f_k(\alpha)^n,$$

where

$$f_k(x) = \lambda^{1-x} \cdot (x/2)^{(k-2)(1-x)-x} \cdot e^{-\lambda(1-\frac{kx^k + k(k-1)x^{k-2}(1-x)^2}{2^k})/\binom{k}{2}}$$

and the $\theta(1)$ factor does *not* depend on α or λ.

Let

$$
\begin{aligned}
g_k(x) &= \ln(f_k(x)) = \\
&= (1-x)\ln\lambda + [(k-2)(1-x)-x]\cdot\ln(x/2) - \\
&- \frac{\lambda}{\binom{k}{2}}(1 - \frac{kx^k + k(k-1)x^{k-2}(1-x)^2}{2^k}).
\end{aligned}
$$

For $x \in (0, \frac{k-1}{k-2}]$, since $\lambda < 1$, $\ln\lambda < 0$ and $1-x > 0$. Also $\ln(x/2) < 0$ while $(k-2)(1-x)-x > 0$. Finally,

The conclusion of this argument is that $x \in (0, \frac{k-1}{k-2}] \implies g_k(x) < 0$.

On the other hand $g_k(1) = \ln 2 - \frac{\lambda}{\binom{k}{2}}\left(1 - \frac{k}{2^k}\right) > \ln 2 - \frac{1}{\binom{k}{2}} > 0$, since $k \geq 3$. Thus the equation $g_k(x) = 0$ has a (smallest) root $x_k \in (\frac{k-1}{k-2}, 1)$.

For $\alpha < x_k$, $f(\alpha) < 1$ and the upper bound is asymptotically equal to zero. □

5 Conclusions

Theorem 1 connects the percolation of the giant component in the random formula hypergraph to the existence of a single cluster of satisfying assignments. Of course, since the phase transition in 1-in-k SAT is determined [ACIM01] by a "giant component" phenomenon in a directed version of the formula hypergraph, the main open question raised by this work is to prove that the statement of Theorem 1 holds up to critical threshold $c = \frac{2}{k(k-1)}$. Theorem 2 provides further evidence that this might be true.

We believe that it might be possible to prove this statement using a more robust generalization of the notion of "hole" in the set of satisfying assignments.

Acknowledgements

I thank Romeo Negrea for useful discussions. This work has been supported by a Marie Curie International Reintegration Grant within the 6th European Community Framework Programme.

References

[ACIM01] D. Achlioptas, A. Chtcherba, G. Istrate, and C. Moore. The phase transition in 1-in-k SAT and NAE3SAT. In *Proceedings of the 12th ACM-SIAM Symposium on discrete algorithms*, 2001.

[AES92] N. Alon, P. Erdős, and J. Spencer. *The probabilistic method*. John Wiley and Sons, second edition, 1992.

[ART06] D. Achlioptas and F. Ricci-Tersenghi. On the solution space geometry of random constraint satisfaction problems. In *Proceedings of the 36th ACM Symposium on Theory of Computing*, pages 130–139, 2006.

[Bol85] B. Bollobás. *Random Graphs*. Academic Press, 1985.

[HW05] A. Hartmann and M. Weigt. *Phase transitions in combinatorial optimization problems*. Wiley-VCH, 2005.

[IPB05] G. Istrate, A. Percus, and S. Boettcher. Spines of random constraint satisfaction problems: Definition and connection with computational complexity. *Annals of Mathematics and Artificial Intelligence*, 44(4):353–372, 2005.

[Ist07] G. Istrate. Satisfiability of boolean random constraint satisfaction: Clusters and overlaps. *Journal of Universal Computer Science*, 13(11):1655–1670, 2007.

[JLR00] S. Janson, T. Luczak, and A. Ruczinski. *Random Graphs*. John Wiley & Sons, 2000.

[KMRT+07] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborova. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104:10318–10323, 2007.

[MMZ05] M. Mézard, T. Mora, and R. Zecchina. Clustering of solutions in the random satisfiability problem. *Physical Review Letters*, 94(197205), 2005.

[Mol02] M. Molloy. Models for random constraint satisfaction problems. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, 2002.

[MPV87] M. Mézard, G. Parisi, and M. Virasoro. *Spin glass theory and beyond*. World Scientific, 1987.

[MZ97] R. Monasson and R. Zecchina. Statistical mechanics of the random k-SAT model. *Physical Review E*, 56:1357, 1997.

[PIM06] A. Percus, G. Istrate, and C. Moore, editors. *Computational Complexity and Statistical Physics*. Oxford University Press, 2006.

[RSZ07] J. Raymond, A. Sportiello, and L. Zdeborová. The phase diagram of random 1-in-3 satisfiability. *Phys. Rev. E*, 76(011101), 2007.

[SPS85] J. Schmidt-Pruznan and D. Shamir. Component structure in the evolution of random hypergraphs. *Combinatorica*, 5:81–94, 1985.