Vive la satisfaisabilité !

(avec le soutien de J.-P. Allouche)

N. Creignou[1] H. Daudé[2] E. de Panafieu[3] V. Ravelomana[3]
R. Rossignol[4]

LIF, AMU
LATP, AMU
Paris 7
Grenoble

Paris, ANR Boole

Outline

Outline

1. Quick overview of previous results

2. Generalized Satisfiability
   - Random (Quantified) Formulas in (non-)normal Form
   - Sensitivity of generalized Boolean formulas

3. Variants and Robust Instances
   - Related work
   - 2-XOR-SAT a central problem
   - Conclusion and perspectives

A threshold phenomenon for 3-SAT

$Pr_{n,cn}$(3-SAT) : probability that a 3-CNF formula over n variables with cn clauses is satisfiable

- The transition from satisfiability to unsatisfiability is sharp (Friedgut, 1998)
- The critical ratio is *estimated* at around 4.25
- Only lower and upper bounds have been established

$$Pr_{n,c \cdot n}(\text{3-SAT}) \to 1 \text{ for } c \leq 3.52$$

(Kaporis, Kirousis, Lalas, 2003) use the analysis of algorithms

$$Pr_{n,c \cdot n}(\text{3-SAT}) \to 0 \text{ for } c \geq 4.4898$$

(Diaz, Kirousis, Mitsche, Pérez, 2009)
previously 4.506 (Dubois, Boufkhad, Mandler, 2000)

A threshold phenomenon for 2-SAT

- The transition for 2-SAT is sharp and the critical ratio is 1 (Chvatal & Reed, Goerdt, 1992)
  They use first and second moment methods
- The scaling window is known (Bollobàs *et al.*, 2001)
- The probability of satisfiablility of a random 2-CNF at the critical ratio $c = 1$ has been experimentally estimated to

$$Pr_{n,n}(\text{2-SAT}) \sim 0.9$$

(Deroulers, Monasson 2006)

What makes the difference (a posteriori) ?

- There is a simple combinatorial characterization of unsatisfiable 2-CNF formulas :
  - 2-SAT is in P
  - A linear time algorithm allows simulations at a very high scale
  - For the threshold, one can focus on the emergence of the most likely unsatisfiable formulas in random formulas
- Such a characterization is missing for 3-CNF formulas :
  - 3-SAT is NP-complete.
  - Simulations are therefore hard to run
  - For the threshold, no focus on typical unsatisfiable formulas is known

Nature of the transition for generalized satisfiability

Given a constraint satisfaction problem, depending on the size of the scaling window the transition SAT/UNSAT is either *sharp* or *coarse*.

Generalized satisfiability : Formulas can be seen are hypergraphs.

Theorem (Creignou, Daudé 2009)

If every tree-formula and every unicyclic formulas are satisfiable, then the satisfiability property has a sharp threshold .

Typical coarse transition : 1-SAT and 2-XOR-SAT.

What do we learn ?

How to make progress on the study of phase transition for monotone properties ?

1. Broaden the scope
2. Study variants on combinatorially robust instances

Outline

Random Formulas Generator

Random instances are useful for

- providing the basis for theoretical investigations
- testing SAT/QSAT/non-PCNF solvers
- evaluating the performance of these solvers (hard instances are at the threshold)

[q]bfGen : a general formula generator which creates formula instances by interpreting the random model specification

http ://fmv.jku.at/qbfgen/

(Creignou, Egly, Seidl 2012)

Constraints

$f\colon \{0,1\}^k \longrightarrow \{0,1\}$, $\tau \subseteq \{1,\ldots,k\}$

$f_\tau(x) = f(x^\tau)$ where x^τ is obtained by complementing those bits in x indexed by elements of τ.

$\{x_1,\ldots,x_n\}$ a set of variables

An n-$\{f\}$-constraint is given by :

- a one-to-one function $\varphi\colon \{1,\ldots,k\} \to \{1,\ldots,n\}$ (scope),
- a subset $\tau \subseteq \{1,\ldots,k\}$ (negated positions),

is denoted by

$$C = (f, \varphi, \tau)$$

and stands for

$$C = f_\tau(x_{\varphi(1)},\ldots x_{\varphi(k)})$$

Example

- $f\colon \{0,1\}^3 \longrightarrow \{0,1\}$ such that $f^{-1}(1) = \{0,1\}^3 \setminus \{000\}$, $f(x,y,z) = (x \vee y \vee z)$
- $\varphi(1) = 3$, $\varphi(2) = 5$, $\varphi(3) = 4$
- $\tau = \{1,2\}$

$f_\tau(x,y,z) = (\bar{x} \vee \bar{y} \vee z)$

The constraint $C = (f, \varphi, \tau)$ stands for $(\bar{x}_3 \vee \bar{x}_5 \vee x_4)$.

Constraint satisfied by an assignment

$C = (f, \varphi, \tau)$, $I \colon \{1, \ldots, n\} \longrightarrow \{0, 1\}$.

$m(I, \varphi)$ the motif of I pinpointed by φ :

$$m(I, \varphi) := (I(\varphi(1)), \ldots, I(\varphi(k)))$$

The status of the constraint C with respect to I.

$$C(I) := f_\tau(m(I, \varphi)),$$

The n-assignment I satisfies C if $C(I) = 1$.

Example

- $f\colon \{0,1\}^3 \longrightarrow \{0,1\}$ such that $f^{-1}(1) = \{0,1\}^3 \setminus \{000\}$,
  $f(x,y,z) = (x \vee y \vee z)$
- $\varphi(1) = 3$, $\varphi(2) = 5$, $\varphi(3) = 4$
- $\tau = \{1,2\}$

$f_\tau(x,y,z) = (\bar{x} \vee \bar{y} \vee z)$

The constraint $C = (f, \varphi, \tau)$ stands for $(\bar{x}_3 \vee \bar{x}_5 \vee x_4)$

I the all-one assignment, $C(I) = f_\tau(1,1,1) = 1$

Schaefer's framework : $\mathcal{F}$-formulas

Given $\mathcal{F}$ a finite set of Boolean functions, an $\mathcal{F}$-constraint is a constraint using a function f from $\mathcal{F}$.

A sequence of $\mathcal{F}$-constraints,
over n variables is an (n, L)-$\mathcal{F}$-formula

- If $\mathcal{F} = \{f\}$ where $f(x, y, z) = (x \vee y \vee z)$
  - an $\mathcal{F}$-constraint is a 3-clause
  - an $\mathcal{F}$-formula is 3-CNF
- If $\mathcal{F} = \{g\}$ where $g(x, y) = (x \oplus y)$
  - an $\mathcal{F}$-constraint is a 2-XOR-clause
  - an $\mathcal{F}$-formula is 2-XOR-CNF

Sensitivity of Boolean functions

The sensitivity set of f at a particular input x, $\mathrm{S}(f,x)$:

$$\mathrm{S}(f,x) = \{t : 1 \leq t \leq k, f(x) \neq f(x^t)\}.$$

The sensitivity of f at x :

$$s(f,x) = |\mathrm{S}(f,x)|$$

Sensitivity of constraints

Let $C = (f, \varphi, \tau)$ be an n-constraint and I be a truth assignment. The sensitivity set of C at I :

$$\mathrm{S}(C, I) := \{t \,:\, 1 \leq t \leq n,\ C(I) \neq C(I^t)\}.$$

The sensitivity of C at the truth assignment I is :

$$s(C, I) := |\mathrm{S}(C, I)|.$$

Sensitivity of formulas

$\Phi = (C_1, \cdots, C_L)$ a formula, I a truth assignment

$$\Phi(I) = (C_1(I), \ldots, C_L(I))$$

The sensitivity set of Φ at I :

$$\mathrm{S}(\Phi, I) := \{t : 1 \leq t \leq n,\ \Phi(I) \neq \Phi(I^t)\}.$$

Enumeration

Let I be a truth assignment.

$$|\{\Phi : \Phi \text{ is an}(n, L)\text{-}\mathcal{F}\text{- formula,}\Phi(I) = (1, \ldots, 1) \text{ and } \mathrm{S}(\Phi, I) = S\}|$$
$$= \Gamma^1_{k,n,L,\mathcal{F}}(|S|)$$

The number of formulas satisfied by I having a given sensitivity set S at I is <u>independent from I</u> , depends on the cardinality of S only and can be expressed according to the <u>sensitivity of the functions in $\mathcal{F}$</u>.

(Creignou,Daudé 2013)

Applications

This result allows the enumeration of $\mathcal{F}$-formulas

- having a given truth assignment as a sufficiently isolated solution
- having a given truth assignment as a locally maximal solution (useful for threshold upperbounds, method of O. Dubois)

Outline

1. Quick overview of previous results
2. Generalized Satisfiability
   - Random (Quantified) Formulas in (non-)normal Form
   - Sensitivity of generalized Boolean formulas
3. Variants and Robust Instances
   - Related work
   - 2-XOR-SAT a central problem
   - Conclusion and perspectives

Previous studies

Problem	Scale	Nature	critical ratio	complexity
2-SAT	$L = \theta(n)$	Sharp	$c_2 = 1$	P
3-XOR-SAT	$L = \theta(n)$	Sharp	$c_{XOR,3} \sim 0.918$	P
2-XOR-SAT	$L = \theta(n)$	Coarse		P

((Chvatal & Reed, Goerdt, 1992)), (Dubois, Mandler 2002)

(1,2)-QSAT

A (1,2)-QCNF-*formula* is a closed quantified formula of the following type

$$\forall X \exists Y \, \varphi(X, Y)$$

- X has $m = \lfloor \alpha \log n \rfloor$ variables
- Y has n variables
- each clause in φ has 1 literal from X and 2 from Y

Sharp threshold : For any value of α, we give the exact location of the associated critical ratio, $a(\alpha)$.
(Creignou, Daudé, Egly, Rossignol 2009)

The 2-XOR-SAT problem and its graphical representation

- An instance : $(\mathbf{v_1} \vee \mathbf{v_2}) \wedge (\neg \mathbf{v_1} \vee \mathbf{v_3}) \wedge (\neg \mathbf{v_1} \vee \neg \mathbf{v_2})$
- A solution : SAT with $(\mathbf{v_1} = \mathbf{1}, \mathbf{v_2} = \mathbf{0}, \mathbf{v_3} = \mathbf{1})$.
- Localization of the threshold : n variables, $m = c \times n$ clauses randomly picked from the set of $4\binom{n}{2}$ clauses.
  $c < 1$ Proba SAT $\to 1$, $c > 1$ Proba SAT $\to 0$.
  Underlying combinatorial structures : directed graphs.

$$\texttt{Write} \quad x \vee y \quad \texttt{as} \quad \begin{cases} \neg x = 1 \Longrightarrow y = 1 \\ \neg y = 1 \Longrightarrow x = 1 \end{cases}$$

Characterization : SAT iff no directed path between x and $\neg x$ (and vice-versa).
Proof. First and second moments method [Goerdt 92, De la Vega 92, Chvàtal, Reed 92].

2-XOR-SAT and the Boole Project

Main motivations and results

- What can be the contributions of **ENUMERATIVE/ANALYTIC COMBINATORICS** to SAT/CSP-like problems ?
- MONASSON (2007) inferred that (statistical physics) :

$$\lim_{n \to +\infty} n^{\textbf{critical exponent}} \times \mathtt{Proba}\left[\text{2XORSAT}(n, \frac{n}{2}) \right] = O(1),$$

  where "critical exponent" $= \mathbf{1/12}$.

- We have shown that "critical exponent" is in fact $= \mathbf{1/12}$ and we made explicit the hidden constant behind the $O(1)$.

Remark :

This was the beginning of a whole line of research within the Boole Project cf. [Vonjy Rasendrahasina 2012 and Elie de Panafieu's theses $\sim$ 2014].

2-XOR-SAT and the Boole Project

Main motivations and results

- What can be the contributions of **ENUMERATIVE/ANALYTIC COMBINATORICS** to SAT/CSP-like problems ?
- MONASSON (2007) inferred that (statistical physics) :

$$\lim_{n\to+\infty} n^{\textbf{critical exponent}} \times \texttt{Proba}\left[2\mathrm{XORSAT}(n, \frac{n}{2})\right] = O(1),$$

  where “critical exponent” $=$ **1/12** .

- We have shown that “critical exponent” is in fact $=$ **1/12** and we made explicit the hidden constant behind the $O(1)$.

Remark :

This was the beginning of a whole line of research within the Boole Project cf. [Vonjy Rasendrahasina 2012 and Elie de Panafieu’s theses $\sim$ 2014].

Random 2-XOR-SAT

- Ex :
  $x_1 \oplus x_2 = 1, x_2 \oplus x_3 = 0, x_1 \oplus x_3 = 0, x_3 \oplus x_4 = 1, \cdots$.
- General form : $AX = C$ where A has m rows and 2 columns and C is a m-dimensional $0/1$ vector.
- Distribution : uniform. We pick m clauses of the form $x_i \oplus x_j = \varepsilon \in \{0, 1\}$ from the set of $n(n-1)$ clauses.
- Underlying structures : **graphs with weighted edges**
  $x \oplus y = \varepsilon \iff$ edges of weight $\varepsilon \in \{0, 1\}$.

Characterisation :

SAT iff no elementary cycle of odd weight.

Main ideas of our approach

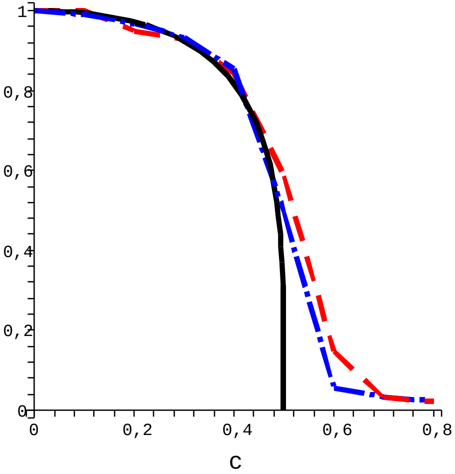
A basic scheme

1. Enumeration of "SAT"-graphs (graphs without cycles of odd weight) by means of generating functions.

2. Use the obtained results with analytic combinatorics to compute :

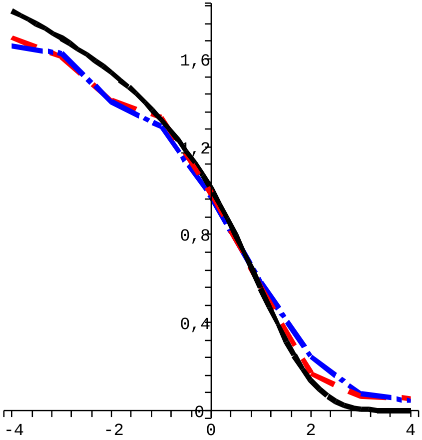
$$\textbf{Prob. SAT} = \frac{\textbf{Nbr of configurations without cycles of odd weight}}{\textbf{Nbr total of configurations}}.$$

Taste of our results : the whole window



$\mathbf{p}(\mathbf{n}, \mathbf{cn}) \stackrel{\text{def}}{=} \mathtt{Proba}$ [2-XOR with **n** variables , **cn** clauses] is **SAT**
for $n = 1000$, $n = 2000$ and the **theoretical** function : $\mathbf{e^{c/2}(1-2c)^{1/4}}$.

Taste of our results : rescaling the critical window



Rescaling at the point "zero", i.e $c = 1/2$: $n = 1000$, $n = 2000$ and $\lim_{n\to\infty} \underbrace{n^{1/12}\times}\ p(n, n/2 + \mu n^{2/3})$ as a **function of** μ**.**

Enumerating graphs of 2-XORSAT.

We will **enumerate** the `connected graphs without cycles of odd weight` according to two parameters : **number of vertices n** and **number of edges n** $+ \ell$. $\ell \overset{\text{def}}{=}$ **excess.**

Let

$$C_\ell(z) = \sum_{n>0} c_{n,n+\ell} \frac{z^n}{n!} .$$

What are the series C_ℓ ?

Th.

$$C_\ell(z) = \frac{1}{2} W_\ell(2z)$$

with W_ℓ = Exponential generating functions of connected graphs WRIGHT (1977).

Enumerating graphs of 2-XORSAT.

We will **enumerate** the `connected graphs without cycles of odd weight` according to two parameters : **number of vertices n** and **number of edges $\mathbf{n}+\ell$**. $\ell \stackrel{\text{def}}{=}$ **excess.**
Let

$$C_\ell(z) = \sum_{n>0} c_{n,n+\ell} \frac{z^n}{n!} .$$

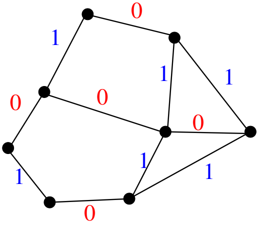
What are the series C_ℓ ?

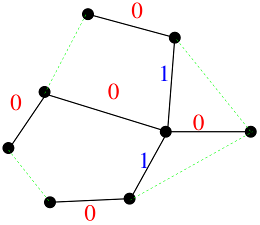
Th.

$$C_\ell(z) = \frac{1}{2} W_\ell(2z)$$

with W_ℓ = Exponential generating functions of connected graphs WRIGHT (1977).

Main ideas behind the enumerations



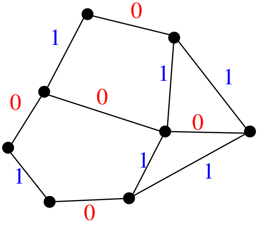


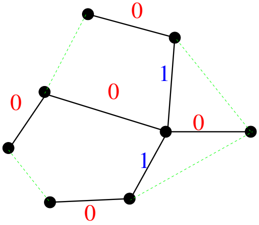
On a connected “SAT”-graph with n vertices and $n + \ell$ edges, the edges of a spanning tree can be colored in 2^{n-1} ways. The colors of the other edges are “determined”.

Remark :

As we obtained the generating functions, the whole machinery to study random graphs apply to random 2-XOR-SAT to get the results.

Main ideas behind the enumerations





On a connected “SAT”-graph with n vertices and $n+\ell$ edges, the edges of a spanning tree can be colored in 2^{n-1} ways. The colors of the other edges are “determined”.

Remark :

As we obtained the generating functions, the whole machinery to study random graphs apply to random 2-XOR-SAT to get the results.

Random MAX-2-XOR-SAT

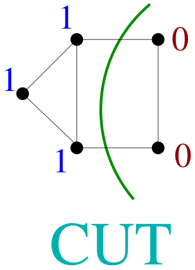
- **MAX-2-XORSAT** is an NP-optimization problem (NPO). The corresponding decision problem is in NP (deciding if the size of the MAX is k ...).

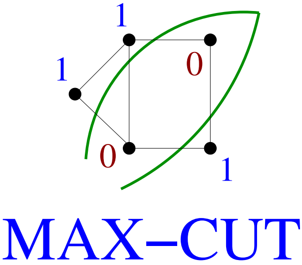
- **MAX/MIN** problems are interesting (and difficult) in randomness context.

- **PREVIOUS WORKS :** [Coppersmith, Gamarnik, Hajiaghayi, Sorkin 04]
  Expectations of the **Maximum** number of satisfiable clauses in MAX-2-SAT and MAX-CUT for the subcritical phases. **Bounds** of these expectations for some cases (namely for the critical and supercritical phases of random graphs) !

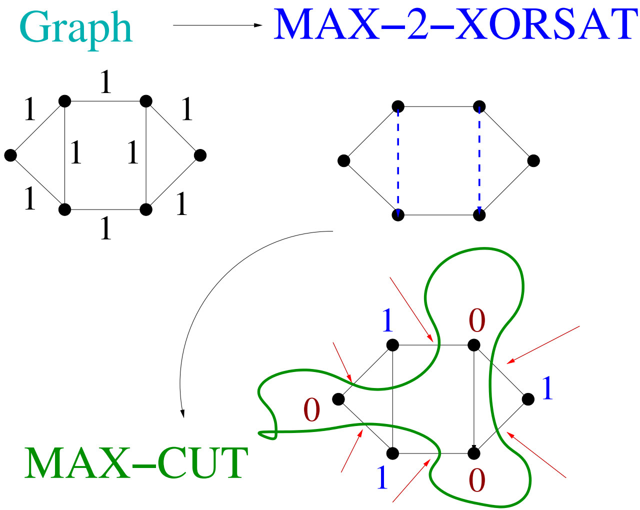
- **OUR WORK :**
  Quantification of the **Minimum** number of clauses to remove in order to get satisfiable formula.

MAX-CUT $\sim$ MAX-2-XORSAT (i)





MAX-CUT ~ MAX-2-XORSAT (ii)



Our results on MAX-2-XOR-SAT

(Daudé, Martinez, Rasendrahasina, R. 2012)

- Precise results such as convergence in distribution to Poisson r.v and then to Gaussian r.v. if the random graph is **`very sparse`**
- "Only" bounds of the size of the MAX-CUT as soon as the giant component of the random graph is born (**truly disappointing** !).
- We have a conjecture about the size of the MAX-CUT in the latter case (confirmed by colleagues empirical data).

Conclusion and perspectives

The **enumerative/analytic approaches** and the random 2-XOR-SAT problem has lead to several directions.

Similar **methods** `worked` on other problems such as

1. bipartiteness (or 2-COL).
2. 2-QXORSAT (quantified formula) and generalizations (cf. [Elie's talk].
3. planarity

`but not completely` on MAX-2-COL, MAX-CUT, MIN-VERTEX-COVER, MIN-BISECTION, MAX-PLANAR

Conclusion and perspectives

The **enumerative/analytic approaches** and the random 2-XOR-SAT problem has lead to several directions.

Similar **methods** **`worked`** on other problems such as

1. bipartiteness (or 2-COL).
2. 2-QXORSAT (quantified formula) and generalizations (cf. [Elie's talk].
3. planarity

`but not completely` on MAX-2-COL, MAX-CUT, MIN-VERTEX-COVER, MIN-BISECTION, MAX-PLANAR

Conclusion and perspectives

The **enumerative/analytic approaches** and the random 2-XOR-SAT problem has lead to several directions.

Similar **methods** **`worked`** on other problems such as

1. bipartiteness (or 2-COL).
2. 2-QXORSAT (quantified formula) and generalizations (cf. [Elie's talk].
3. planarity

`but not completely` on MAX-2-COL, MAX-CUT, MIN-VERTEX-COVER, MIN-BISECTION, MAX-PLANAR