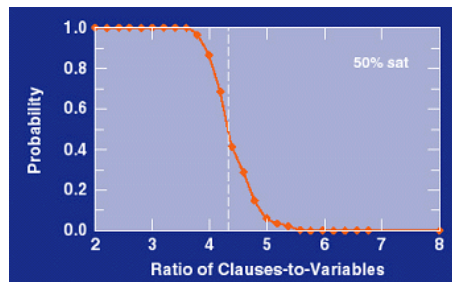


UNIVERSITÀ DEGLI STUDI ROMA TRE  
FACOLTÀ DI SCIENZE M.F.N.  
Corso di Laurea Magistrale in Matematica

Sintesi della Tesi di Laurea  
presentata da  
Luca Avena



# On the threshold of the random k-Sat

Relatore  
Prof. Fabio Martinelli

Correlatore  
Dott. Pietro Caputo

ANNO ACCADEMICO 2005 - 2006  
Luglio 2006

AMS Classification : 82B26, 82B44, 90C27, 68W20.  
Keywords: random combinatorial optimization, satisfiability, random algorithms, phase transition.



## 0.1 Introduzione

Il problema k-Sat (k-Satisfiability) consiste nel dire se è possibile soddisfare contemporaneamente un certo numero di vincoli su  $N$  variabili booleane  $\sigma_i \in \{0, 1\}$ , dove ciascun vincolo è rappresentato da un'espressione logica, una *clausola*, costruita con  $k$  differenti variabili (o le loro negazioni) legate tra loro dall'operatore logico OR. Denotando la negazione di una variabile come  $\overline{\sigma}_i$ , chiameremo *letterale*  $l_i$  una variabile affermata  $\sigma_i$  o negata  $\overline{\sigma}_i$  nell'espressione. Prendiamo ad esempio una clausola  $c$  di 3 variabili  $c = (\sigma_1 \vee \sigma_2 \vee \overline{\sigma}_3)$ , gli assegnamenti di verità che soddisfano tale espressione in  $\{0, 1\}^3$ , sono tutte e solo quelle terne tali che almeno uno dei letterali sia soddisfatto; dunque in questo caso è semplice verificare che l'unico assegnamento tra i  $2^3$  possibili che falsifica la clausola è  $(0, 0, 1)$ . Considerando lo spazio degli assegnamenti di verità  $\Omega := \{0, 1\}^N$ , un'istanza  $F$  del k-Sat con  $M$  vincoli, detta *Formula Congiuntiva Normale* (CNF), sarà quindi un'espressione della forma:

$$F = \bigwedge_{j=1}^M c_j = \bigwedge_{j=1}^M \bigvee_{i=1}^k l_{ji}$$

Notiamo che da  $N$  variabili, si possono costruire  $2^k \binom{N}{k}$  clausole differenti contenenti  $k$  variabili.

**Esempio 0.1.** Dato lo spazio degli assegnamenti di verità  $\Omega := \{0, 1\}^3$ , consideriamo la CNF  $F = (\sigma_1 \vee \overline{\sigma}_2) \wedge (\sigma_2 \vee \sigma_3) \wedge (\overline{\sigma}_1 \vee \overline{\sigma}_3)$ . È facile verificare che  $S(F) := \{(0, 0, 1), (1, 1, 0)\}$  è il sottoinsieme delle sue soluzioni in  $\Omega$

Il modello k-Sat riveste un ruolo fondamentale tra i problemi di ottimizzazione combinatoria e negli ultimi trent'anni è stato oggetto di studio in Matematica (in particolare in Teoria della Probabilità), Informatica Teorica e Meccanica Statistica con risultati estremamente rilevanti, alcuni dei quali apparsi recentemente su *Nature* [ANP]. Lo scopo di questo lavoro è di dare una trattazione chiara e unitaria degli sviluppi di queste ricerche concentrando l'attenzione sui risultati rigorosi e le questioni aperte. Da un punto di vista applicativo, tantissimi problemi di robotica, intelligenza artificiale, computer design, possono essere formalizzati in istanze del k-Sat. Da un punto di vista teorico, nell'ambito della Teoria della Complessità, il problema k-Sat, per  $k \geq 3$ , è stato il primo ad essere dimostrato appartenere alla classe dei problemi NP-completi (1971 [C]). Ciò significa che se trovassimo un algoritmo polinomiale per risolverlo, ogni altro problema NP (i.e. di cui si conosce una verifica polinomiale per una soluzione) potrebbe essere risolto in tempo polinomiale. Purtroppo, l'unico algoritmo in grado di risolvere una qualsiasi istanza del k-Sat sembra essere la procedura di ricerca globale nello

spazio degli assegnamenti, l'algoritmo DPLL; poiché la cardinalità di  $\Omega$  è  $2^N$ , tale procedura può richiedere un tempo esponenziale in  $N$ . Negli anni '80, per capire le origini di questa difficoltà, usando il DPLL, sono state fatte molte simulazioni al fine di distinguere da un punto di vista computazionale istanze difficili da istanze facili. Dalle analisi di queste simulazioni è emerso che un buon parametro caratterizzante per questa distinzione è il rapporto tra il numero di vincoli  $M$  e il numero di variabili in gioco  $N$ :

$$\alpha = \alpha(k) := \frac{M}{N}$$

Per avere un'idea chiara di questo scenario consideriamo istanze random generate come segue:

**Definizione 0.1. Formule Random**

Dato lo spazio degli assegnamenti di verità  $\Omega = \{0, 1\}^N$ , indichiamo con  $F_N(k, \alpha)$  una formula random generata dalla congiunzione di  $M = \alpha N$  clausole scelte uniformemente, indipendentemente e con rimpiazzo tra le  $2^k \binom{N}{k}$  possibili.

Tramite simulazioni numeriche, si osserva che per valori relativamente bassi di  $\alpha$  (il problema è sotto-vincolato) l'algoritmo DPLL trova facilmente una soluzione. I valori alti di  $\alpha$  corrispondono ad istanze sopra-vincolate, e l'algoritmo in tempo rapido ci dice che non esistono soluzioni. I casi più difficili (che impiegano molto tempo) sono quelli con valori di  $\alpha$  intermedi. In termini probabilistici, se si provasse a graficare in funzione di  $\alpha$  la probabilità empirica (intesa come frequenza) di soddisfare una formula random, si osserverebbe che tale probabilità è monotona decrescente e passa da valori prossimi a 1 a valori vicini a 0. Il fenomeno più interessante è che al crescere di  $N$ , il passaggio di tale probabilità da valori prossimi a 1 a valori vicini a 0 avviene in un intervallo di  $\alpha$  sempre più piccolo: siamo di fronte a un tipico esempio di *transizione di fase*.

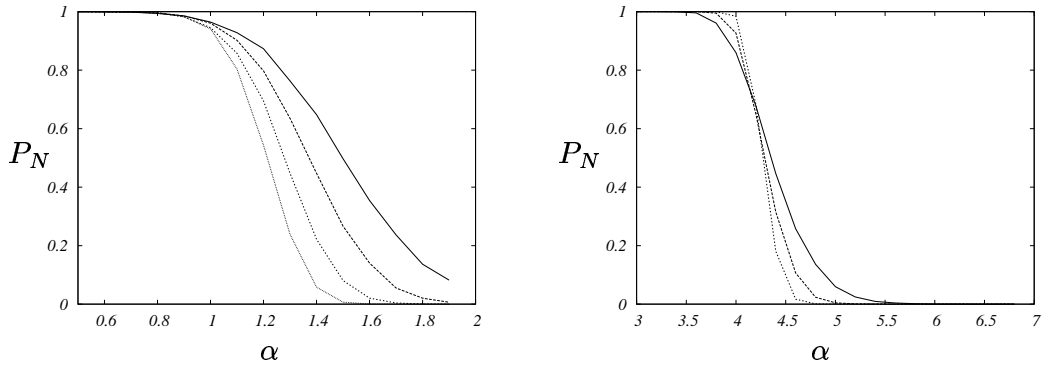


Figura 1: *Probabilità di soddisfare una formula random del  $k$ -Sat in funzione di  $\alpha$ . Ogni punto è stato ottenuto mediando su  $10^4$  formule random. Le curve per  $k=2$  (sinistra) corrispondono a formule su  $N = 50, 100, 200, 400$  variabili (da destra a sinistra). Per  $k=3$  (destra) le curve corrispondono ad istanze su  $N = 50, 100, 200$  variabili, rispettivamente: linea continua, tratteggiata e a tratti più leggeri. La transizione tra formule soddisfacenti e non, diviene più marcata al crescere di  $N$ .*

Viene spontaneo chiedersi quale sia il comportamento di questa probabilità nel limite per  $N \rightarrow \infty$ . In altre parole, ci si chiede se per ogni  $k$  fissato, esiste un valore critico di soglia  $\alpha_c(k)$  (*threshold value*) tale che:

$$\lim_{N \rightarrow \infty} \mathbb{P}(F_N(k, \alpha) \text{ è soddisfacibile}) = \begin{cases} 1 & \text{if } \alpha < \alpha_c(k), \\ 0 & \text{if } \alpha > \alpha_c(k) \end{cases}$$

Questa domanda, posta indipendentemente da diversi ricercatori, è nota come *Satisfiability Threshold Conjecture*. Nel 1992 Chvatal and Reed [CR] dimostrarono la congettura per il 2-Sat, (che è un problema polinomiale), stabilendo come soglia  $\alpha_c(2) = 1$ . Per  $k \geq 3$ , il problema è ancora aperto ma si possono ottenere delle ottime stime del supposto valore critico  $\alpha_c(k)$  con metodi probabilistici. L'analisi e l'applicazione di questi metodi rappresenta l'oggetto principale di questa tesi. Prima di affrontarli, diamo un'altra utile formulazione della congettura.

### Satisfiability Threshold Conjecture

Diremo che una successione di eventi  $E_N$  si verifica w.h.p. (with high probability) se  $\lim_{N \rightarrow \infty} \mathbb{P}(E_N) = 1$ .

Per ogni  $k \geq 3$ , siano

$$\overline{\alpha(k)} := \inf\{\alpha \mid F_N(k, \alpha) \text{ non è soddisfacibile w.h.p.}\} \quad (1)$$

$$\underline{\alpha(k)} := \sup\{\alpha \mid F_N(k, \alpha) \text{ è soddisfacibile w.h.p.}\} \quad (2)$$

allora

$$\underline{\alpha(k)} = \overline{\alpha(k)}.$$

## 0.2 Metodi del Momento Primo e Secondo

Vediamo come fornire delle stime dall'alto di  $\overline{\alpha(k)}$  e dal basso per  $\underline{\alpha(k)}$ .

### Metodo del momento Primo

Consideriamo la variabile aleatoria che conta il numero di assegnamenti  $\underline{\sigma} \in \{0, 1\}^N$  che soddisfano una formula random  $F = F_N(k, \alpha)$ :

$$X = X(F) := \sum_{\underline{\sigma} \in \Omega} \mathbb{I}_{\{\underline{\sigma} \in S(F)\}} = |S(F)| \quad (3)$$

dove  $S(F)$  rappresenta l'insieme delle soluzioni della formula random e  $\mathbb{I}_{\{\cdot\}}$  è la funzione indicatrice o caratteristica. Osservando che

$$X(F) = \begin{cases} 0 & \text{se } F_N(k, \alpha) \text{ non è soddisfacibile} \\ \geq 1 & \text{altrimenti} \end{cases}$$

dalla disuguaglianza di Markov si ha che:

$$\mathbb{P}(F_N(k, \alpha) \text{ è soddisfacibile}) = \mathbb{P}(X(F) \geq 1) \leq \mathbb{E}[X(F)] \quad (4)$$

Nel 1983 Franco e Paull [FP] utilizzarono questa variabile aleatoria per mostrare con un semplice calcolo che

$$\text{se } \alpha \geq 2^k \log 2 \Rightarrow \mathbb{E}[X] = o(1)$$

Ricordando la definizione in (1), ciò equivale a dire che  $\overline{\alpha(k)} \leq 2^k \log 2$ . Questa stima dall'alto della soglia di non-soddisfacibilità  $\overline{\alpha(k)}$  è la più semplice da ottenere ma non è la migliore. Nella fase di soddisfacibilità "Sat", la variabile  $X(F)$  può fluttuare moltissimo visto che potrebbero esserci un numero esponenziale di soluzioni. Quindi, per migliorare la stima, si può considerare una nuova variabile aleatoria che conti la cardinalità di un adatto sottoinsieme di  $S(F)$  e tale che valga (4). Negli ultimi anni sono stati fatti diversi tentativi in questa direzione, uno tra i migliori tentativi è stato fatto considerando in  $S(F)$  il sottoinsieme degli assegnamenti tali che cambiando un singolo 0 in 1, il nuovo assegnamento di verità non soddisfa più alla formula [KKK]. La stima che si ottiene non ha però un'espressione esplicita in  $k$ , è rappresentata dall'unica soluzione positiva della seguente equazione :

$$(1 - 2^{-k})^\alpha (2 - e^{-\frac{k\alpha}{2^k-1}}) = 1 \quad (5)$$

Dunque, fissando il  $k$  d'interesse, con metodi numerici si cerca lo zero dell'equazione (5).

### Metodo del momento Secondo

Una stima della soglia di soddisfacibilità  $\underline{\alpha(k)}$  (vedi (2)) è molto più complicata ed è stata possibile solo grazie a un importante lavoro di E.Friedgut [FB] in teoria dei grafi random. In questo lavoro si dimostra una sorta di legge 0 – 1 che stabilisce l'esistenza di un *valore di soglia non uniforme* :

#### **Teorema 0.2.1. *Non-uniform Sharp Threshold***

*Per ogni  $k \geq 2$ , esiste una successione  $\alpha_N(k)$  tale che per ogni  $\epsilon > 0$ ,*

$$\lim_{N \rightarrow \infty} \mathbb{P}(F_N(k, \alpha_N - \epsilon) \text{ è soddisfacibile}) = 1$$

*e*

$$\lim_{N \rightarrow \infty} \mathbb{P}(F_N(k, \alpha_N + \epsilon) \text{ non è soddisfacibile}) = 0$$

Per mostrare l'utilità di questo teorema, introduciamo l'idea alla base del metodo del momento secondo.

#### **Lemma 0.2.2.**

*Sia  $X$  una variabile aleatoria non negativa. Allora:*

$$\mathbb{P}(X > 0) \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \quad (6)$$

A partire da questo lemma, possiamo considerare una variabile aleatoria  $X(F)$  di cui si sappia calcolare momento primo e secondo e tale che

$$X = X(F) = \begin{cases} 0 & \text{se } F_N(k, \alpha) \text{ non è soddisfacibile} \\ > 0 & \text{altrimenti} \end{cases} \quad (7)$$

In tal modo da (6) e (7) si ha che:

$$\mathbb{P}(F_N(k, \alpha) \text{ è soddisfacibile}) = \mathbb{P}(X(F) > 0) \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \quad (8)$$

Ora, trovando un  $\alpha^* = \alpha^*(k)$  tale che:

$$\mathbb{E}[X^2] = O(\mathbb{E}[X]^2) \text{ i.e. } \exists C > 0 \text{ t.c. } \mathbb{E}[X^2] \leq C\mathbb{E}[X]^2 \quad (9)$$

da (8) segue che:

$$\liminf_{N \rightarrow \infty} \mathbb{P}(F_N(k, \alpha^*) \text{ è soddisfacibile}) = \liminf_{N \rightarrow \infty} \mathbb{P}(X > 0) > \frac{1}{C} \quad (10)$$

Finalmente, grazie al teorema 0.2.1, (10) implica che:

$$\lim_{N \rightarrow \infty} \mathbb{P}(F_N(k, \alpha^*) \text{ è soddisfacibile}) = 1$$

che equivale a dire

$$\underline{\alpha(k)} \geq \alpha^*(k)$$

Questo metodo adottato da D.Achlioptas e Y.Peres nel 2004 [AP], ha portato al seguente teorema:

**Teorema 0.2.3.**

*Esiste una successione  $\delta_k \rightarrow 0$  tale che  $\forall k \geq 3$*

$$\underline{\alpha(k)} \geq 2^k \log 2 - (k+1) \frac{\log 2}{2} - 1 - \delta_k$$

Il teorema non prova l'esistenza del valore di soglia ma rafforza la congettura. Infatti insieme alla precedente stima su  $\underline{\alpha(k)}$ , ci dice che la transizione di fase ha luogo per  $\alpha(k)$  vicino al valore  $2^k \log 2$ . Per capirne l'importanza, si pensi che attraverso le analisi di algoritmi risolutivi per il k-Sat, il miglior risultato ottenuto forniva una stima pessima rispetto a quella del teorema 0.2.3:

$$\underline{\alpha(k)} \geq \frac{s_k 2^k}{k} \text{ dove } \lim_{k \rightarrow \infty} s_k = 1.817\dots$$



stima che ha resistito per circa 10 anni [FS]. La difficoltà del metodo del momento secondo sta nel trovare un'adeguata variabile aleatoria  $X = X(F)$  e il miglior  $\alpha^*$  per i quali  $\mathbb{E}[X^2] = O(\mathbb{E}[X]^2)$ . Abbiamo detto che per il metodo del momento primo è sufficiente considerare la variabile aleatoria che conti il numero di assegnamenti che soddisfano una formula random (3). Purtroppo per la stessa variabile aleatoria si ha che per ogni  $\alpha$  esiste un  $\beta = \beta(\alpha) > 0$  tale che

$$\mathbb{E}[X^2] > (1 + \beta)^N \mathbb{E}[X]^2$$

e il metodo del momento secondo fallisce. Vedendo la dimostrazione di questo fatto, si evince che la causa del fallimento è dovuta essenzialmente alle correlazioni tra gli assegnamenti di verità nello spazio  $\{0, 1\}^N$ . Per averne un'idea definiamo la *distanza di Hamming* tra  $\underline{\sigma}, \underline{\tau} \in \{0, 1\}^N$  come :

$$d_{\underline{\sigma}, \underline{\tau}} := \sum_{i=1}^N \mathbb{I}_{\{\sigma_i \neq \tau_i\}}$$

Sia  $f_k(x) := \mathbb{E}_x[\mathbb{I}_{\{\underline{\sigma} \in S(c)\}} \mathbb{I}_{\{\underline{\tau} \in S(c)\}}] = \mathbb{P}_x(\underline{\sigma}, \underline{\tau} \in S(c))$  la probabilità che due assegnamenti a distanza di Hamming  $d_{\underline{\sigma}, \underline{\tau}} = Nx$  ( $Nx$  intero) soddisfano entrambi la clausola random  $c$ . Si può far vedere che:

$$f_k(x) = 1 - 2^{1-k} + 2^{-k}(1-x)^k$$

$$\mathbb{E}[X^2] \leq (2 \max_{x \in [0,1]} [\mathcal{E}(x) f_k(x)^\alpha])^N O(1) =: (\max_{x \in [0,1]} \Lambda(x, k, \alpha))^N O(1) \quad (11)$$

dove  $X$  è come in (3) e  $\mathcal{E}(x) := 1/x^x(1-x)^{1-x}$ ,

$$f_k(1/2) = \mathbb{P}_{1/2}(\underline{\sigma}, \underline{\tau} \in S(c)) = (1 - 2^{-k})^2 = \mathbb{P}(\underline{\sigma} \in S(c))^2$$

i.e. assegnamenti a distanza di Hamming  $N/2$  sono scorrelati e

$$\mathbb{E}[X]^2 = \Lambda(1/2, k, \alpha)^N \quad (12)$$

Quindi, da (11) e (12) segue che:

$$\frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \geq \frac{\Lambda(1/2, k, \alpha)^N}{(\max_{x \in [0,1]} \Lambda(x, k, \alpha))^N O(1)}$$

Ora se il massimo della funzione  $\Lambda(x, k, \alpha)$  fosse raggiunto per  $x = 1/2$ , tale rapporto rimarrebbe strettamente positivo nel limite per  $N \rightarrow \infty$  e

il metodo del momento secondo avrebbe successo. Come detto questo non è il caso per la variabile aleatoria definita in (3) per la quale il massimo globale della funzione  $\Lambda$  per ogni  $\alpha$  è raggiunto per qualche  $x < 1/2$ . Di conseguenza bisogna considerare una nuova variabile aleatoria  $Y = Y(F)$  che pesi in maniera differente gli assegnamenti di verità in modo che il contributo principale al momento secondo derivi dalle coppie di assegnamenti scorrelati (i.e. a distanza di Hamming  $N/2$ ). La dimostrazione del teorema 0.2.3 è stata fatta utilizzando una variabile aleatoria che verifica le suddette condizioni della forma:

$$Y = Y(F) = \sum_{\underline{\sigma} \in S(F)} \gamma^{H(\underline{\sigma}, F)} \mathbb{I}_{\{H(\underline{\sigma}, F) \geq 0\}}$$

dove  $H(\underline{\sigma}, F)$  rappresenta il numero di letterali in  $F_N(k, \alpha)$  soddisfatti da  $\underline{\sigma}$  meno il numero di quelli non soddisfatti tra tutti i  $kM$  letterali in  $F_N(k, \alpha)$  e  $\gamma \in (0, 1]$ . Nella seguente tabella riportiamo per qualche valore di  $k$  le stime “sull’ upper e il lower bound”, rispettivamente  $\overline{\alpha}(k)$  e  $\underline{\alpha}(k)$  :

$k$	3	4	7	10	20	21
Upper Bound	4.51	10.23	87.88	708.94	726, 817	1, 453, 635
Lower Bound	2.68	7.91	84.82	704.94	726, 809	1, 453, 626
Algorithmic LB	3.42	5.54	33.23	172.65	95, 263	181, 453

Tabella 1: *Nelle prime due righe sono riportati i migliori risultati ottenuti con i metodi dei momenti per bassi valori di  $k$ . Nella terza riga i lower bound ottenuti dall’analisi delle prestazioni degli algoritmi. Nonostante il fatto che per  $k=3$  il “lower bound algoritmico” è il migliore, a partire da  $k=4$ , la stima “probabilistica” migliora incredibilmente*

### 0.3 Clusterizzazione e Meccanica Statistica

Sistemi costituiti da un gran numero di componenti interagenti tra loro (gas, materiali ferromagnetici, vetri di spin, modelli neurali, etc...) costituiscono l’oggetto di studio della Meccanica Statistica contemporanea. In quest’ambito sono stati sviluppati metodi euristici per studiare il comportamento di sistemi disordinati, cioè sistemi caratterizzati da un *Hamiltoniana* random [MPV]. E’ stato osservato recentemente che gli stessi metodi si possono usare per analizzare problemi di ottimizzazione combinatoria. Ciascuno di tali problemi può essere infatti formulato in termini di una funzione costo o d’en-

ergia da minimizzare: un Hamiltoniana. Vediamo il caso del k-Sat:

**Definizione 0.2. Hamiltoniana del k-Sat**

Denotiamo il valore logico FALSO con -1 invece di 0. Consideriamo lo spazio degli assegnamenti  $\underline{\sigma} \in \{-1, 1\}^N$ . Data un'istanza  $F$  del k-Sat :

$$F = \bigwedge_{j=1}^M c_j = \bigwedge_{j=1}^M \bigvee_{i=1}^k l_{ji}$$

per ogni clausola  $c_j$  definiamo la funzione:

$$E_j(\underline{\sigma}) := \prod_{i=1}^k (1 + J_j^i \sigma_{j_i}) \quad \text{con } J_j^i := \begin{cases} -1 & \text{se } l_{ji} = \sigma_{j_i}, \\ 1 & \text{se } l_{ji} = \overline{\sigma_{j_i}} \end{cases}$$

E' facile verificare che:

$$E_j(\underline{\sigma}) = 0 \Leftrightarrow \underline{\sigma} \text{ soddisfa } c_j$$

Prendendo in considerazione la funzione costo:

$$H(\underline{\sigma}) := \sum_{j=1}^M E_j(\underline{\sigma}) = \sum_{j=1}^M \prod_{i=1}^k (1 + J_j^i \sigma_{j_i}) \tag{13}$$

chiedersi se la formula  $F$  ammette soluzioni, equivale a chiedere l'esistenza di un assegnamento  $\underline{\sigma}^* \in \{-1, 1\}^N$  tale che  $H(\underline{\sigma}^*) = 0$

In tal modo si ha un hamiltoniana (13) caratterizzata da interazioni di k variabili che in Meccanica Statistica si chiama un problema *k-spin*. A partire da questa connessione, i fisici hanno studiato il modello k-Sat con metodi non rigorosi (*teoria delle repliche, metodo della cavità*) che a)mostrano che la transizione di fase "Sat-Unsat" ha luogo per valori di  $\alpha(k)$  vicino a  $2^k \log 2$  e che b) permettono la costruzione di un algoritmo polinomiale (*survey propagation*) che riesce a risolvere singole istanze con valori di  $\alpha$  vicini a  $2^k \log 2$  (lì dove gli altri algoritmi falliscono). Alla base di questo algoritmo ci sono però delle assunzioni riguardanti la geometria dello spazio delle soluzioni di una formula, che per quanto plausibili non sono dimostrate. Nel 2006 ci sono stati diversi risultati rigorosi che in qualche modo sembrano confermare una delle principali assunzioni di questo algoritmo: l'esistenza di una *fase di clusterizzazione*.

**Definizione 0.3. Fase di Clusterizzazione**

Per ogni  $k \geq 3$  esiste un secondo valore di soglia  $\alpha_d = \alpha_d(k) < \alpha_c(k)$  tale che:

1. Per  $\alpha < \alpha_d$ , un'istanza tipica del k-Sat ammette molte soluzioni che tendono, nel limite per  $N \rightarrow \infty$ , a formare un'unica grande componente connessa: *un cluster*. Ciò significa che in questo cluster esistono cammini brevi (costituiti da passi a distanza di Hamming finita) per passare da una soluzione ad un'altra. Questa regione è detta "Easy-Sat" perchè anche gli algoritmi basati su ricerche locali nello spazio degli assegnamenti, possono trovare facilmente una soluzione.
2. Per  $\alpha_d < \alpha < \alpha_c$ , lo spazio delle soluzioni è invece costituito dall'unione di un numero esponenziale di regioni connesse distanti tra loro. Questa fase è chiamata "Hard-Sat phase" perchè costituita dalle istanze difficili da risolvere.

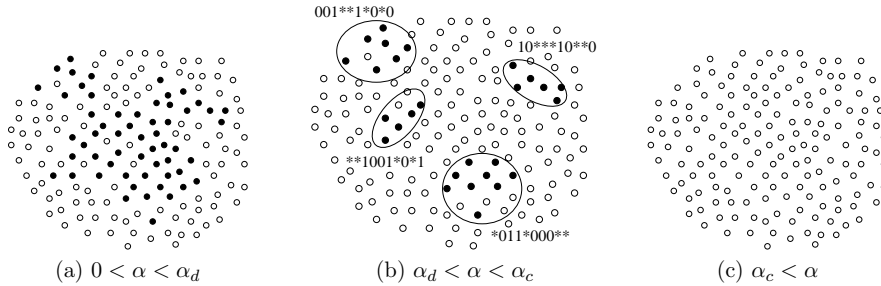


Figure 2.3: *I punti rappresentano assegnamenti di verità soddisfacenti e non, rispettivamente in nero e bianco. (a) Per valori piccoli di  $\alpha$ ,  $S(F)$  è un insieme connesso. (b) Esiste un valore  $\alpha_d$  tale che in una regione di  $\alpha_d < \alpha < \alpha_c$ ,  $S(F)$  si divide in un numero esponenziale di clusters. (c) Sopra di  $\alpha_c$  w.h.p. non esistono assegnamenti soddisfacenti.*

Il risultato principale riguardante la fase di clusterizzazione è rappresentato dal seguente teorema [AR]:

**Teorema 0.3.1. Clusters**

Dati due sottoinsiemi  $X, Y \subseteq \{0, 1\}^N$  definiamo la loro distanza come:

$$dist(X, Y) := \min_{(\underline{x}, \underline{y}) \in X \times Y} d_{\underline{x}, \underline{y}}$$

Data una componente connessa dello spazio delle soluzioni di una formula, un cluster  $C$ , definiamo il diametro del cluster come:

$$\text{diam}(C) := \max_{\underline{\sigma}, \underline{\tau} \in C} d_{\underline{\sigma}, \underline{\tau}}$$

Per ogni  $k \geq 8$  esistono  $x_1(k) < x_2(k) < 1/2$  and  $\alpha < \underline{\alpha}(k)$  tali che w.h.p. per una formula random del k-Sat  $F_N(k, \alpha)$ :

1. Per ogni coppia di soluzioni  $(\underline{\sigma}, \underline{\tau})$ :  $d_{\underline{\sigma}, \underline{\tau}} \in [0, Nx_1(k)) \cup (Nx_2(k), N]$
2. Esistono  $(1 + \xi)^N$  clusters  $C_i$  tali che  $\text{dist}(C_i, C_j) \geq Nx_2(k) \forall i \neq j$ , con  $\xi > 0$
3. Per ogni cluster  $C$  :  $\text{diam}(C) \leq Nx_1(k)$

Si noti che il risultato si tiene per  $k \geq 8$ , dunque l'assunzione della fase di clusterizzazione per valori di  $k$  inferiori a 8 rimane una questione aperta.

## 0.4 Struttura della Tesi

- Il capitolo 1 è dedicato alla definizione del modello k-Sat e alla sua versione random introducendo la classe dei problemi di ottimizzazione combinatoria e vedendone le relazioni con la classe dei problemi NP.
- Nel capitolo 2 presentiamo i tre approcci differenti al fenomeno di soglia (*threshold phenomenon*): l'approccio algoritmico, fisico e probabilistico.
- Nel capitolo 3 discutiamo in dettaglio i risultati riguardanti la transizione di fase ottenuti tramite i metodi del momento primo e del momento secondo.
- Il capitolo 4 è dedicato alla fase di clusterizzazione. Presentiamo due approcci diversi a tale problematica ([AR],[MMZ2],[MMZ3]), uno dei quali sintetizzato dal teorema 0.3.1.
- Concludiamo mettendo in evidenza le questioni che rimangono aperte.



# Bibliografia

- [AM] D.ACHLIOPTAS and C.MOORE, “The asymptotic order of the k-Sat threshold,” *In Proc. 43th Annual Symposium on Foundations of Computer Science* (2002), pp.126-127.
- [ANP] D.ACHLIOPTAS, A.NAOR and Y.PERES, “Rigorous location and phase transitions in hard optimization problems,” *Nature*, 435 (2005), pp.759-764.
- [AP] D.ACHLIOPTAS and Y.PERES, “The threshold for random k-Sat is  $2^k \log 2 - O(k)$ ,” *Journal of the American Mathematical Society*, 17 (2004), pp.947-973.
- [AR] D.ACHLIOPTAS and F.RICCI-TERSENGHI, “On the solution-space geometry of random constraint satisfaction problems,” *In Proc. 38th Annual ACM Symposium on Theory of Computing* (2006), pp.130-139.
- [BCM] G.BIROLI, S.COCCO and R.MONASSON, “Phase transition and complexity in computer science: an overview of the statistical physics approach to the random satisfiability problem,” *Physica A* 306 (2002), pp.381-394.
- [BMZ] A.BRAUNSTEIN, M.MEZARD and R.ZECCHINA, “Survey propagation: an algorithm for satisfiability,” *Random Structures and Algorithms* 27 (2005), pp.201-226.
- [C] S.COOK, “Complexity of theorem proving procedures,” *In Proc. 3th Annual ACM Symposium on Theory of Computing* (1971), pp.151-158.
- [CF] M.T.CHAO and J.FRANCO, “Probabilistic analysis of a generalization of the unit-clause literal selection heuristics for the k-satisfiability problem,” *Inform. Sci.*,51(3) (1990) pp.289-314.

- [CR] V.CHVATAL and B.REED, “Mick gets some(the odds are on his side),” *In Proc. 33th Annual Symposium on Foundations of Computer Science* (1992), pp.620-627.
- [CLR] T.H.CORMEN, C.E.LEISERSON and R.L.RIVEST, “Introduction to algorithms,” *Cambridge: The MIT press* (1990).
- [DLL] M.DAVIS, G.LOGEMANN and D.LOVELAND, “A machine programming for theorem-proving,” *Comm. of ACM*, 5(7) (1962).
- [DP] M.DAVIS and H.PUTNAM, “A computing procedure for quantification theory,” *Journal of the Association for Computing Machinery* (1960) , pp. 201-215.
- [D] N.G.DE BRUIJN, “Asymptotic methods in analysis,” *Dover Publications Inc. New York, 3th edition* (1981).
- [FB] E.FRIEDGUT and J.BOURGAIN, “ Sharp threshold of graph properties and the k-Sat problem,” *Journal of the American Mathematical Society*, 12 (1999), pp.1017-1054.
- [FP] J.FRANCO and M.PAULL, “Probabilistic analysis of the Davis-Putnam procedure for solving the satisfiability problem,” *Discrete Appl. Math*, 5(1)(1983), pp.77-87.
- [FS] A.FRIEZE and S.SUEN, “Analysis of two simple heuristics on a random instance of k-Sat,” *J.Algorithms*, 20(2) (1996), pp.312-355.
- [GJ] M.R.GAREY and D.S.JOHNSON, “Computers and Intractability: A Guide to the Theory of NP-Completeness,” *New York: W.H.Freeman and Co.* (1979).
- [H] B.HAYES, “Can’t get satisfaction,” *Am. Sci.* 85 (2) (1997), pp.108-112.
- [JLR] S.JANSON, T.LUCZAK and A.RUCINSKI, “Random graphs,” *Wiley Interscience* (2000).
- [JT] M.JORDAN and S.TATIKONDA, “Loopy Belief Propagation and Gibbs Measures,” *In Proc. Uncertainty in Artificial Intelligence*, 18 (2002), pp.493-500.
- [KKK] L.KIROUSIS, E.KRANAKIS and D.KRIZANC, “A better upper bound for the unsatisfiability threshold,” *to appear in DIMACS Series in Discrete Mathematics and Theoretical Computer Science* (1996).



- [MMW] E.MANEVA, E.MOSSEL and M.WAINWRIGHT, “A new look at survey propagation and its generalizations,” *In Proceedings of SODA* (2005), pp.1089-1098.
- [MMZ1] S.MERTENS, M.MEZARD and R.ZECCHINA, “Threshold values of random k-Sat from the cavity method,” *Computer Science 0309020* (2003).
- [MM] M.MEZARD and A.MONTANARI, “Constraint satisfaction networks in physics and computation,” *It will be published in 2006 by Oxford University Press*.
- [MMZ2] M.MEZARD, T.MORA and R.ZECCHINA, “Clustering of solutions in random satisfiability problem,” *cond-mat 0504070* (2005).
- [MMZ3] M.MEZARD, T.MORA and R.ZECCHINA, “Pairs of Sat assignments and clustering in random boolean formulae,” *cond-mat 0506053* (2005).
- [MPV] M.MEZARD, G.PARISI and M.A.VIRAROSO, “Spin glass theory and beyond,” *Singapore: World Scientific* (1988).
- [MPZ] M.MEZARD, G.PARISI and R.ZECCHINA, “Analytic and Algorithmic Solution of Random Satisfiability Problems,” *Science*, 297 (2002) pp.812-815.
- [MSL] D.MITCHELL, B.SELMAN and H.LEVESQUE, “Hard and easy distributions of Sat problems,” *In Proc. 10th National Conference on Artificial Intelligence* (1992), pp.459-462.
- [N] H.NISHIMORI, “Statistical physics of spin glasses and information processing ,” *Oxford, U.K.: Oxford University Press* (2001).
- [P] G.PARISI, “A backtracking survey propagation algorithm for k-satisfiability,” *cond-mat 0308510* (2003).
- [S] J.P.SETHNA, “Statistical Mechanics: Entropy, Order Parameters and Complexity” *It will be published in 2006 by Oxford University Press*.