



SAPIENZA  
UNIVERSITÀ DI ROMA

## Transizione di fase nei problemi K-SAT

Facoltà di Scienze Matematiche Fisiche e Naturali  
Corso di Laurea in Fisica

Candidato

Lorenzo Gregoris

Matricola 1867373

Relatore

Prof. Federico Ricci Tersenghi

Correlatore

Dr. Raffaele Marino

Anno Accademico 2020/2021

Tesi non ancora discussa

---

**Transizione di fase nei problemi K-SAT**

Tesi di Laurea. Sapienza – Università di Roma

© 2021 Lorenzo Gregoris. Tutti i diritti riservati

Questa tesi è stata composta con L<sup>A</sup>T<sub>E</sub>X e la classe Sapthesis.

Email dell'autore: [lorenzo.gregoris@gmail.com](mailto:lorenzo.gregoris@gmail.com)

# Indice

<b>Introduzione</b>	<b>3</b>
<b>1 Meccanica statistica</b>	<b>6</b>
1.1 Il limite termodinamico . . . . .	6
1.2 Transizioni di fase . . . . .	7
1.3 Il modello di Ising . . . . .	10
<b>2 Random K-SAT</b>	<b>12</b>
2.1 Teoria della complessità . . . . .	12
2.2 La transizione SAT-UNSAT . . . . .	14
2.3 Relazione con gli spin-glasses . . . . .	14
2.4 L'hamiltoniana del K-SAT . . . . .	15
2.5 Upper bound . . . . .	16
2.6 Random 2-SAT . . . . .	17
<b>3 Simulazioni numeriche</b>	<b>21</b>
3.1 L'algoritmo . . . . .	21
3.2 Finite-size scaling . . . . .	23
3.3 Analisi dati . . . . .	24
<b>4 Conclusioni</b>	<b>30</b>

# Introduzione

L'informatica e la fisica sono strettamente legate sin dalla nascita del computer moderno. L'architettura originale di John Von Neumann [1] degli anni '40, motivata da applicazione di idrodinamica e balistica, è quella su cui è basato l'hardware moderno.

Durante il progetto Manhattan, dove collaborarono scienziati come Fermi, Von Neumann, Teller, Ulam, Metropolis, vennero sviluppati i metodi Monte Carlo, e i primi computer digitali cominciarono a simulare processi naturali sfruttando i principi della fisica statistica. E' innegabile riconoscere il contributo dei computer nello sviluppo della fisica (si pensi all'esperimento numerico Fermi-Pasta-Ulam [2]).

Solo vent'anni dopo cominciò a svilupparsi una comprensione teorica sull'efficienza delle computazioni. Nel 1965 Hartmanis e Stearns definirono il concetto di complessità computazionale [3] caratterizzando gli algoritmi in base a quanto rapidamente crescevano la memoria e il tempo impiegato in funzione della dimensione dell'input.

Un problema combinatorio appartiene alla classe di complessità P (*polynomial time*) se esiste un algoritmo che garantisce una soluzione in un tempo computazionale (il numero di step dell'algoritmo) che cresce in maniera polinomiale rispetto alla dimensione dell'input. Questo tipo di problemi sono considerati computazionalmente trattabili. Un esempio classico è ordinare una lista di  $n$  numeri: un algoritmo semplice come il *bubble sort* (ogni coppia di elementi adiacenti viene comparata, ed i due elementi vengono invertiti di posizione se sono nell'ordine sbagliato) è in grado di restituire la lista ordinata in un numero di passi che cresce come  $\mathcal{O}(n^2)$ . Questo problema appartiene alla classe P.

Un problema appartiene alla classe NP (*non-deterministic polynomial time*) se è solo possibile verificare in tempo polinomiale la correttezza di una data soluzione. Il problema del cammino hamiltoniano, dove si richiede l'esistenza di un cammino che passi per ogni vertice di un dato grafo esattamente una volta, appartiene alla classe NP: dato un cammino che suppongo sia hamiltoniano, un algoritmo lineare può percorrerlo e verificare se rispetti le condizioni del problema. A differenza dei problemi in P, questa classe di problemi è considerata *intrattabile* computazionalmente,

i migliori algoritmi sono infatti esponenziali rispetto alla dimensione dell'input. Se pensiamo al problema del cammino hamiltoniano, il numero di percorsi possibili in un grafo ad  $n$  vertici sono  $n!$ , e già per grafi non troppo grandi lo spazio di ricerca del problema diventa enorme.

Ulteriori passi in avanti nella classificazione della complessità furono compiuti agli inizi degli anni '70 da Cook [4] e Karp [5], che introdussero le classi NP-HARD e NP-COMPLETE. Un problema è NP-HARD se ogni problema in NP può essere ricondotto ad esso; un problema è NP-COMPLETE se è NP-HARD ed appartiene alla classe NP. Quindi, i problemi NP-COMPLETE sono i *più difficili* in NP.

Il risultato principale di Cook fu dimostrare che il problema della soddisfacibilità di formule booleane (SAT), ovvero decidere se una data formula proposizionale possa essere soddisfatta, è NP-COMPLETE. Karp mostrò come molti problemi combinatori come la  $k$ -colorabilità di un grafo, il traveling salesman, il percorso hamiltoniano sono in NP-COMPLETE dimostrando come possono essere ridotti al problema SAT. Da allora migliaia di problemi dai campi più disparati, gestione di risorse nelle industrie, il *protein-folding*[6], si sono rivelati NP-COMPLETE.

Se si trovasse un algoritmo polinomiale per risolvere uno di questi problemi, potrebbe essere usato per risolvere ogni problema in NP, e pertanto risolvere uno dei problemi del millennio dell'istituto Clay<sup>1</sup>,  $P = NP$ . Questa rimane una delle più fondamentali domande aperte dell'informatica. La complessità di un algoritmo è garantita per ogni istanza del problema, ovvero è definita sul caso peggiore, una particolare istanza per cui l'algoritmo impiega un numero di step massimo rispetto tutte le altre (ad esempio l'algoritmo bubble sort per ordinare in maniera crescente una lista di  $n$  numeri impiega  $n^2$  step solo quando la lista in input è ordinata in maniera decrescente). Tuttavia è possibile che la maggior parte delle istanze sia più semplice, e solo un sottoinsieme ristretto sia complesso. Qui entra in gioco l'approccio della fisica statistica: studiare proprietà macroscopiche del problema, evitando di specificare completamente ogni istanza. Considerare problemi computazionali che hanno come input strutture combinatorie casuali: per esempio la soddisfacibilità di una formula booleana random, o la  $k$ -colorabilità di un particolare ensemble di grafi random. Kirkpatrick e al. [7] notarono che al crescere della dimensione dei problemi di ottimizzazione, l'analisi del caso peggiore diventa irrilevante, e il caso medio diventa dominante nell'analisi della performance di un algoritmo. Citando Kirkpatrick: "*This large number limit is the domain of statistical mechanics*".

Nel 1991 Cheeseman, Kanefsky e Taylor pubblicarono un articolo intitolato "*Where the really hard problems are*" [8], dove mostrarono empiricamente delle proprietà di istanze random di alcuni problemi, tra cui il problema della soddisfacibilità booleana. In particolare notarono, in problemi NP-COMPLETE, delle "*sharp*

---

<sup>1</sup><https://www.claymath.org/millennium-problems>

*phase boundary*" data un'appropriata parametrizzazione, un problema appartenente ad una delle due fasi era facile da risolvere, mentre era in prossimità di un boundary che vi erano i problemi difficili. Da notare che per problemi difficili non si intende problemi che non hanno una soluzione, ma semplicemente problemi per i quali la soluzione esiste ma è molto difficile da trovare. Inoltre notarono come per problemi in P, non vi era tale regime *easy-hard-easy*.

Ad un fisico statistico, tutto ciò appare simile ad una transizione di fase: come un particolare parametro macroscopico supera una soglia critica, il sistema va incontro ad un rapido cambiamento. L'esempio classico è il congelamento dell'acqua: quando la temperatura scende sotto il punto di congelamento, il comportamento globale cambia drasticamente, nonostante le interazioni locali tra le molecole rimangano lo stesse.

In questa dissertazione verrà analizzato il problema 2-SAT, caso particolare dei problemi K-SAT, in cui si verifica la soddisfacibilità di formule booleane in "*conjunctive normal form*", ovvero congiunzioni di clausole, in cui ogni clausola è una disgiunzione di esattamente K letterali (variabili o negazioni di variabili). In particolare affronteremo il problema con l'approccio della fisica statistica: genereremo istanze random e studieremo la regolarità che appare aumentando il numero di variabili in gioco e il numero di clausole del problema mantenendo costante il loro rapporto, in perfetta analogia col limite termodinamico. Studieremo il 2-SAT perchè a differenza dei problemi K-SAT con  $K \geq 3$  appartiene alla classe P, consentendo di studiare formule con un numero molto grande di variabili.

Nel primo capitolo richiameremo alcuni concetti fondamentali di meccanica statistica, in particolare il limite termodinamico, le transizioni di fase e la loro caratterizzazione, ed il modello di Ising. Nel secondo capitolo introdurremo il problema 2-SAT e il random 2-SAT, argomenteremo ulteriormente la distinzione tra problemi polinomiali ed esponenziali, e mostreremo i legami con la fisica; infine dimostreremo matematicamente una transizione di fase legata alla soddisfacibilità di una formula del random 2-SAT. Nell'ultimo capitolo effettueremo delle analisi numeriche sul random 2-SAT, spiegheremo l'algoritmo implementato per risolvere il problema e introdurremo la *finite-size scaling* per calcolare l'esponente critico associato.

# Capitolo 1

## Meccanica statistica

Nella prima parte di questo capitolo definiremo il limite termodinamico e mostriamo alcune importanti conseguenze, mentre nella seconda parte parleremo brevemente delle transizioni di fase, soffermandoci in particolare sulla caratterizzazione di quelle del secondo ordine tramite gli esponenti critici. Alla fine del capitolo ci sarà una breve sezione dedicata al modello di Ising, base fondamentale dello studio dei fenomeni critici.

### 1.1 Il limite termodinamico

Il merito della meccanica statistica classica fu spiegare la termodinamica a partire dalla meccanica, basandosi sulla teoria atomistica. Il fatto che la materia sia formata da un numero enorme di costituenti (ad esempio una mole di materia contiene un numero di atomi pari al numero di Avogadro  $N_A \sim 10^{23}$ ), rende impossibile applicare direttamente la meccanica classica, ma al tempo stesso fa sì che le quantità macroscopiche trattate dalla termodinamica classica siano praticamente fissate: consideriamo l'energia interna di un sistema termodinamico pari alla somma delle energie di tutte le particelle che lo compongono. Assumendo che le energie di ogni particella  $e_i$  siano variabili aleatorie indipendenti (l'energia di una particella non influenza le altre) e con identica distribuzione di probabilità con i primi due momenti finiti pari a  $\mathbb{E}[e_i] = \mu$  e  $\text{Var}[e_i] = \sigma^2$ . Se consideriamo come osservabile macroscopica la somma delle singole energie di particella  $U = \sum_i^N e_i$  per la linearità della varianza e del valore atteso

$$\mathbb{E}[U] = \sum_i^N \mathbb{E}[e_i] = N\mu \quad \text{Var}[U] = \sum_i^N \text{Var}[e_i] = N\sigma^2 \quad (1.1)$$

le fluttuazioni relative tendono a zero nel limite  $N \rightarrow \infty$ :

$$\lim_{N \rightarrow \infty} \frac{\text{Var}[U]^{1/2}}{\mathbb{E}[U]} = \lim_{N \rightarrow \infty} \frac{N^{1/2}\sigma}{N\mu} = \frac{\sigma}{\mu} \lim_{N \rightarrow \infty} \frac{1}{N^{1/2}} = 0 \quad (1.2)$$

e dal teorema del limite centrale sappiamo che  $U$  è distribuita gaussianamente.

Il passaggio da variabili fissate (posizione e momento) a variabili aleatorie per descrivere lo stato meccanico del sistema, è dovuto all'impossibilità di una conoscenza totale dello stato del sistema. Anche se fossimo in grado di determinare le condizioni iniziali di tutte le molecole di un gas, sarebbe impossibile risolvere tutte le equazioni del moto. Una descrizione probabilistica è ciò di cui dobbiamo accontentarci. Tuttavia la termodinamica è una teoria fisica deterministica, senza traccia di probabilità. Abbiamo fatto vedere come nel limite di infinite particelle, del tutto ragionevole considerando i sistemi macroscopici trattati in termodinamica, le fluttuazioni sull'energia interna (o di una generica osservabile somma) tendano a zero. Le due teorie coincidono nel così detto *limite termodinamico*, definito nel seguente modo:

$$N, E, V \rightarrow \infty \quad \frac{V}{N} \rightarrow v \quad \frac{E}{N} \rightarrow e \quad (1.3)$$

con  $e, v$  rispettivamente l'energia e il volume per particella costanti finite diverse da zero.

## 1.2 Transizioni di fase

Una conseguenza sorprendente delle interazioni tra particelle è la comparsa di comportamenti collettivi emergenti, stati della materia, che si discostano molto dal comportamento di poche particelle.

Lo stato (di equilibrio) di un corpo omogeneo è caratterizzato da due grandezze termodinamiche qualsiasi, ad esempio volume  $V$  ed energia  $E$ . Tuttavia non c'è ragione per cui ad una coppia  $(E, V)$  corrisponda un solo stato omogeneo del corpo. Può risultare che per un volume ed un'energia fissata il corpo non sia omogeneo all'equilibrio termico ma si separi in due parti omogenee a contatto che si trovano in stati diversi. Chiamiamo tali stati della materia, che possono esistere contemporaneamente in mutuo equilibrio e a contatto *fasi della materia*.

Per *transizione di fase* di un sistema fisico s'intende il passaggio da una fase ad un'altra fase del sistema. Ciò comporta una variazione delle proprietà fisiche come ad esempio la densità, la conducibilità elettrica, la magnetizzazione, la struttura cristallina e così via.

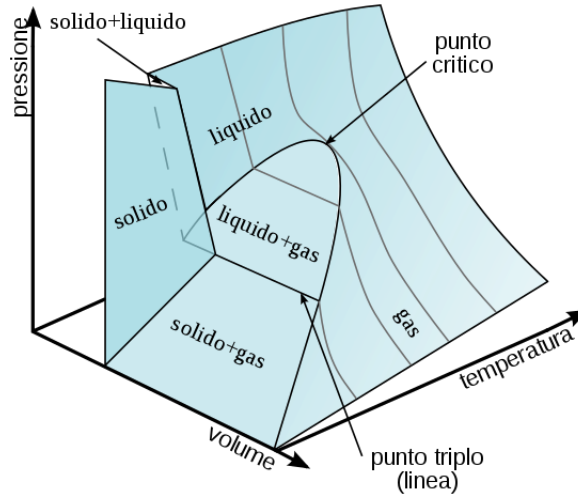
Se descriviamo una fase tramite i parametri di temperatura ( $T$ ), pressione ( $P$ ), e potenziale chimico ( $\mu$ ), affinché la prima fase  $(T_1, P_1, \mu_1)$  sia in equilibrio con la seconda fase  $(T_2, P_2, \mu_2)$  è necessario che  $T_1 = T_2$ ,  $P_1 = P_2$  e  $\mu_1 = \mu_2$ .

Esprimendo i potenziali chimici come funzioni di pressione e temperatura otteniamo l'equazione:

$$\mu_1(P, T) = \mu_2(P, T)$$



che mostra come la pressione e temperatura delle fasi in equilibrio siano una funzione dell'altra, definendo una curva nel piano (P,T), chiamata *curva di equilibrio delle fasi* o *curva di coesistenza*. Riportiamo in figura 1.1 il diagramma di stato dell'acqua.



**Figura 1.1.** Diagramma di fase dell'acqua.

Per distinguere le due fasi quando si è lungo la curva di coesistenza, si introduce un *parametro d'ordine*, tipicamente una variabile termodinamica estensiva che è differente nelle due fasi. La curva di equilibrio delle fasi può arrestarsi: nel punto di arresto, detto *punto critico*, si perde la distinzione tra la fase liquida e quella gassosa[9]. In generale si ha un punto critico in una transizione del secondo ordine[9].

Avvicinandoci al punto critico troviamo sperimentalmente che proprietà termodinamiche come la compressibilità isoterma ed il calore specifico divergono, suggerendo la comparsa di fluttuazioni<sup>1</sup> nel sistema che si estendono ben oltre la scala atomica[10].

Teoricamente, tutte le proprietà macroscopiche possono essere dedotte da un potenziale termodinamico (es. dall'energia libera di Helmholtz  $F(T, V, N)$ ), quindi ad ogni cambiamento repentino corrisponderà una singolarità nel potenziale o nelle sue derivate. Tuttavia, la funzione di ripartizione canonica è sempre una funzione analitica se consideriamo un numero finito di particelle. Solo nel limite termodinamico possono apparire delle non analiticità [11].

Lo studio delle transizioni di fase è perciò la ricerca di queste singolarità e la loro caratterizzazione.

<sup>1</sup>Siccome le funzioni di risposta (suscettibilità) sono legate alle fluttuazioni all'equilibrio, ciò implica che anche le fluttuazioni diventino grandi.

Il primo tentativo di classificare le transizioni di fase è dovuto a Paul Ehrenfest. Una transizione è di ordine  $n$  se è la derivata ennesima dell'energia libera la prima ad essere discontinua.

Le varie transizioni solido/liquido/gas sono classificate al primo ordine in quanto la densità, che è legata all'inverso del volume, e il quale è la derivata prima dell'energia libera rispetto alla pressione, presenta una discontinuità lungo la transizione.

Un esempio di transizione di fase del secondo ordine è la transizione paraferromagnetica<sup>2</sup> [10]: infatti la magnetizzazione, che è la derivata prima dell'energia libera rispetto al campo magnetico applicato, cresce dallo zero con continuità non appena la temperatura è al di sotto di quella di Curie, mentre la suscettività magnetica, che è la derivata seconda dell'energia libera rispetto al campo, cambia con discontinuità.

Tuttavia è uno schema di classificazione non adeguato, basato sui risultati ottenuti da teorie di campo medio. Nell'intorno del punto critico non si hanno discontinuità, ma divergenze: nei ferromagneti reali la capacità termica diverge.

Nel moderno schema di classificazione, le transizioni di fase sono divise in due categorie[12]:

- Le transizioni di fase del primo ordine sono quelle che coinvolgono un calore latente. Durante tali transizioni, un sistema termodinamico assorbe o rilascia una quantità di energia fissa, essendo discontinua l'entropia tra le due fasi. Poiché l'energia non può essere istantaneamente trasferita dal sistema all'ambiente circostante, le transizioni del primo ordine sono associate a "regimi di fase mista" in cui alcune parti del sistema hanno completato la transizione, mentre altre ancora no (si pensi all'ebollizione dell'acqua).
- La seconda classe è detta delle transizioni di fase continue, anche chiamate transizioni di fase del secondo ordine: queste non sono associate ad alcun calore latente essendo l'entropia continua. Esempi di transizioni di fase del secondo ordine sono la transizione ferromagnetica, la transizione  $\lambda$  nell'elio liquido[9].

Il comportamento singolare nelle vicinanze del punto critico nelle transizioni del secondo ordine è caratterizzato da un insieme di *esponenti critici*, che descrivono la non analiticità della parte singolare delle variabili termodinamiche[9]. Gli esponenti critici risultano essere fondamentali poichè essi rappresentano un invariante per

---

<sup>2</sup>Ricordiamo che un ferromagnete è un materiale in grado di magnetizzarsi sotto l'azione di un campo magnetico esterno, e di restare a lungo magnetizzati; un paramagnete invece non si magnetizza permanentemente.

classi di universalità[13]. Riportiamo in tabella 1.1 la definizione di alcuni esponenti critici.

quantità termodinamica	esponente critico
Capacità termica	$C_T \sim  t ^{-\alpha}$
Magnetizzazione media	$M \sim  t ^\beta$
Suscettibilità isoterma	$\chi_T \sim  t ^{-\gamma}$
Campo sull'isoterma critica	$H \sim  M ^\delta \text{sign } M$
Lunghezza di correlazione	$\xi =  t ^{-\nu}$

**Tabella 1.1.** Definizioni di alcuni esponenti critici.

dove è stata introdotta la temperatura ridotta

$$t \equiv \frac{T - T_c}{T_c}$$

Il simbolo " $\sim$ " sta ad indicare:

$$F(t) \sim t^\lambda \equiv \lim_{t \rightarrow 0} \frac{\ln |F(t)|}{\ln |t|} = \lambda$$

L'ultimo esponente critico della tabella 1.1 descrive il comportamento di una quantità interessante, la *lunghezza di correlazione* indicata con  $\xi$ . Come già accennato tra le righe, un'altra caratteristica dei fenomeni critici è la comparsa di fluttuazioni a lungo raggio legate alla funzione di risposta (suscettibilità).

Sperimentalmente si osserva che  $\xi$  diverge al punto critico di una transizione del secondo ordine. Questo fatto ha l'interessante conseguenza che i sistemi al punto critico godono di un'*invarianza di scala* (sono statisticamente invarianti per cambiamenti di scala[10]).

### 1.3 Il modello di Ising

Il modello reticolare di Ising si propone di spiegare il fenomeno del ferromagnetismo della materia. Sperimentalmente si nota che la temperatura di Curie, dove al di sotto di essa avviene magnetizzazione spontanea, è ben inferiore alla temperatura di congelamento. Pertanto è appropriato descrivere il ferromagnetismo in termini di un modello di spin fissati nei siti di un reticolo cristallino.

Il modello consiste in un reticolo di siti  $i$  ciascuno con un grado di libertà a due stati,  $S_i = \pm 1$  chiamato "spin". L'hamiltoniana del sistema è la seguente:

$$\mathcal{H} = - \sum_{\langle ij \rangle} J_{ij} S_i S_j - H \sum_i S_i$$

dove la sommatoria su  $\langle ij \rangle$  è intesa su tutti i vicini,  $J$  è la costante di accoppiamento tra spin, e nell'ultima somma si tiene conto dell'influenza di un campo esterno  $H$ . La principale quantità macroscopica d'interesse è il vettore di magnetizzazione media  $\mathbf{m} = 1/N \sum_i S_i$ , parametro d'ordine della transizione ferromagnetica.

L'energia d'interazione tra due spin è  $-J$  se sono paralleli,  $J$  se antiparalleli. Quindi se  $J > 0$ , il modello favorisce lo stato parallelo: l'interazione è ferromagnetica. Quando la temperatura tende a zero, in assenza di campo magnetico esterno, tutti gli spin diventano paralleli, e la magnetizzazione media diventa  $\mathbf{m} = \pm 1$ . Se  $J < 0$  l'interazione è antiferromagnetica; ad alte temperature ci aspettiamo che sia l'entropia a dominare: gli spin fluttueranno rapidamente, e la magnetizzazione media sarà praticamente nulla (fase paramagnetica).

Lo studio del problema K-SAT e il modello di Ising sono interconnessi. In effetti è molto intuitivo associare alle variabili di spin  $S_i = \pm 1$  una variabile booleana. L'analogia prosegue introducendo i vetri di spin che verranno trattati nel capitolo 3.

## Capitolo 2

# Random K-SAT

I problemi K-SAT sono una famiglia di problemi di decisione, ovvero problemi dove la risposta è binaria: Vero o Falso. Un'istanza del problema K-SAT è quindi un insieme di  $N$  variabili booleane  $\{x_1, \dots, x_N\}$  che possono assumere i valori Vero o Falso (indicati convenzionalmente con 1 e 0), e un insieme di  $M$  vincoli sotto forma di clausole, che consistono in un **OR** logico di esattamente  $K$  letterali (ovvero variabili o negazioni di variabili). La formula  $\varphi$  sarà una congiunzione (**AND** logico) di clausole. Questa particolare forma, una congiunzione di disgiunzioni, è detta forma normale congiuntiva (CNF).

$$\varphi = C_1 \wedge \dots \wedge C_M \quad C_i = (l_1 \vee \dots \vee l_k) \quad l_i = x_i, \neg x_i$$

Una formula  $\varphi$  si dice soddisfacibile (SAT) se esiste un particolare assegnamento dei valori di verità di ciascuna variabile che rende logicamente vera l'intera formula. Per formule CNF ciò significa che l'assegnamento deve soddisfare tutte le clausole.

In questo capitolo studieremo il problema K-SAT. Nella prima parte introdurremo alcune classi di complessità e mostreremo perchè è importante lo studio del problema K-SAT per l'informatica, enunciando anche il noto teorema di Cook. Proseguiremo spiegando il problema del random K-SAT e la sua relazione con i vetri di spin, costruendo l'hamiltoniana. Nella quinta parte del capitolo calcoleremo un upper bound su quella che chiameremo *densità di clausola critica* con il metodo del primo momento. Il capitolo si conclude con uno studio approfondito del random 2-SAT.

### 2.1 Teoria della complessità

La teoria della complessità computazionale classifica problemi basandosi sulle risorse del computer necessarie per risolverli. Le risorse necessarie aumentano in funzione della dimensione dell'input (nel caso del K-SAT, del numero delle variabili

delle formula). Per questo motivo la teoria della complessità classifica gli algoritmi basandosi sul tasso di crescita delle risorse in funzione della dimensione dell'input  $N$ .

Sia  $t(N)$  un generico tempo computazionale impiegato dall'algoritmo. L'algoritmo si dice lineare o quadratico se nel limite  $N \rightarrow \infty$ , si ha che  $t(N) = \mathcal{O}(N)$  o  $t(N) = \mathcal{O}(N^2)$  rispettivamente<sup>1</sup>.

Dal punto di vista pratico, la classificazione più utile è quella che separa gli algoritmi nelle seguenti due classi [14]:

- Polinomiali: il tempo di esecuzione cresce come  $t(N) = \mathcal{O}(N^{\theta_1})$ , e l'algoritmo risolve problemi appartenenti alla classe P;
- Esponenziali: il tempo di esecuzione cresce come  $t(N) = \mathcal{O}(e^{\theta_2 N})$ , e l'algoritmo risolve problemi anche appartenente alla classe NP.

Come già detto nell'introduzione i problemi Polinomiali vengono considerati trattabili, mentre quelli esponenziali no. Facciamo un esempio concreto:

Consideriamo il problema 2-SAT, che appartiene alla classe P. In particolare, esiste un algoritmo in grado di decidere se una formula booleana 2-CNF sia decidibile o meno in tempo lineare 3.1. L'algoritmo più semplice e naïf consiste nel cercare tra tutte le combinazioni dei valori di verità delle variabili della formula e controllare se mettono la formula a Vero. Se ciò avviene, allora la formula è SAT, altrimenti è UNSAT. Il problema di questo approccio è che il numero totale degli assegnamenti è  $2^N$ , pertanto l'algoritmo nel caso peggiore cresce come  $\mathcal{O}(2^N)$ . Assumendo che ogni operazione impieghi 1 ns (ragionevole per le moderne CPU, che operano nell'ordine dei GHz) allora per una formula con  $N = 10^4$  variabili, nel caso peggiore il programma impiegherebbe  $6.3 \cdot 10^{2993}$  anni<sup>2</sup>.

Da notare però, che l'analisi teorica di un algoritmo non può soffermarsi solo allo studio del caso peggiore; nella pratica è anche importante studiare le prestazioni medie di un algoritmo. Infatti, per problemi che il K-SAT per  $K > 2$ , che appartengono alla classe NP, gli algoritmi migliori sono esponenziali, ma mediamente performano molto meglio dell'algoritmo naïf. Pertanto, l'analisi teorica di un algoritmo, per quanto importante, potrebbe non catturarne le performance tipiche. Quest'ultime possono essere valutate empiricamente misurando i tempi di esecuzione.

---

<sup>1</sup>Diciamo che  $t(N) = \mathcal{O}(N^\theta)$  se  $\lim_{N \rightarrow \infty} t(N)/N^\theta$  esiste e tende ad una costante  $c$  diversa da zero ( $0 < c < \infty$ ).

<sup>2</sup>L'età dell'universo è circa  $10^{10}$  anni.

## 2.2 La transizione SAT-UNSAT

Un aspetto interessante di questi problemi, già menzionato nell'introduzione è stato notato nel 1991 da Cheeseman, Kanefsky e Taylor nel loro articolo *"Where the really hard problems are"* [8], che mostra empiricamente come la probabilità di essere SAT di istanze random controllate dal parametro  $\alpha = M/N$ , ovvero il rapporto tra clausole e variabili, nel limite  $N \rightarrow \infty$  presenta un andamento simile ad una transizione di fase per un certo valore critico del parametro  $\alpha$ .

Gli elementi dell'ensemble  $SAT(N, M, K)$  sono formule CNF con  $M$  clausole da  $K$  letterali, ciascuna selezionata casualmente dalle  $\binom{N}{K} 2^K$  clausole possibili.

**Congettura 2.2.1** (Satisfiability threshold). *Per ogni  $K$  esiste una soglia critica  $\alpha_c(K)$  tale che per ogni  $\varphi \in SAT(N, \alpha N, K)$ :*

$$\lim_{N \rightarrow \infty} P(\varphi \text{ è SAT}) = \begin{cases} 1 & \text{se } \alpha < \alpha_c(K) \\ 0 & \text{se } \alpha > \alpha_c(K) \end{cases}$$

Quindi, nel limite termodinamico, per  $\alpha < \alpha_c(K)$  ogni formula è soddisfacibile mentre per  $\alpha > \alpha_c(K)$  ogni formula è insoddisfacibile. Chiamiamo queste due fasi SAT e UNSAT rispettivamente.

Questa transizione è stata analiticamente dimostrata solo per il 1-SAT e 2-SAT, con rispettive soglie critiche  $\alpha_c(1) = 0$  e  $\alpha_c(2) = 1$ . La dimostrazione per 1-SAT è triviale, mentre quella del 2-SAT è decisamente non banale, e richiederà l'uso della teoria dei grafi random.

## 2.3 Relazione con gli spin-glasses

Spin glasses (vetri di spin) sono sistemi fisici con disordine magnetico, chiamati in analogia coi vetri e solidi amorfi, che possiedono disordine. A differenza del modello di Ising, dove la costante di accoppiamento è fissata (una  $J > 0$  per materiali ferromagnetici,  $J < 0$  per antiferromagnetici), nei sistemi vetrosi le  $J_{ij}$  diventano variabili aleatorie indipendenti.

Nel 1975 David Sherrington e Scott Kirkpatrick introdussero un modello per gli spin glass (SK model) [15], dove le costanti di accoppiamento tra spin sono distribuite gaussianamente.

$$\mathcal{H} = - \sum_{i,j} J_{ij} S_i S_j - H \sum_i S_i$$

Dove la prima somma è su tutte le  $N(N-1)/2$  coppie distinte.

A differenza del modello di Ising, nel modello SK non c'è nozione di distanza: tutte le coppie di spin possono interagire, non si richiede che siano vicine. Per tale ragione è un modello ad interazioni a raggio infinito, un modello a campo medio[16].

L'hamiltoniana e l'energia libera diventano quindi variabili aleatorie: per studiare il caso tipico si dovrà mediare sulle realizzazioni delle  $J_{ij}$ .

$$f_J = -\frac{1}{\beta N} \ln Z_J = -\frac{1}{\beta N} \ln \text{Tr}(e^{-\beta \mathcal{H}_J(\{\underline{S}\})}) \quad (2.1)$$

dove  $\beta = 1/k_B T$ ,  $\underline{S} \in \{-1, +1\}^N$  è lo stato del sistema, e  $f_J$ ,  $Z_J$  indicano l'energia libera e la funzione di partizione per una data realizzazione delle  $\{J_{ij}\}$ . Per studiare il comportamento tipico occorre mediare sul disordine, ovvero sulle diverse realizzazioni. Indicheremo l'energia libera mediata sul disordine con  $\bar{f}$ . Esistono due modi per effettuare questa media, che corrispondono a due casi fisici diversi [17]: *quenched* dove la dinamica del disordine è molto più lenta rispetto quella degli spin, quindi sono supposte costanti: la media andrà fatta dopo il calcolo dell'energia libera; *annealed* in cui vale l'opposto, perciò la media sul disordine va calcolata prima dell'energia libera. Calcolare la media di un logaritmo è un compito arduo che richiede l'uso di metodi come quello delle repliche (Edwards [18]) o delle cavità (Parisi, Mezard e Virasoro [19]).

## 2.4 L'hamiltoniana del K-SAT

Istanze del problema K-SAT possono essere ricondotte ad hamiltoniane di forma simile ad un vetro di spin. Il primo passo consiste nella banale sostituzione  $S_i = 2(x_i - 1/2)$ , che dalle variabili proposizionali  $x_i = 0, 1$  ci riconduce ad uno spin. Ogni clausola è ricondotta ad un'interazione:  $J_{ij} = +1$  se la clausola  $j$  contiene la variabile  $x_i$ ,  $J_{ij} = -1$  se contiene la variabile  $\neg x_i$ ,  $J_{ij} = 0$  altrimenti. Possiamo definire la seguente funzione indice:

$$V_j = \frac{1}{2^K} \prod_{i=1}^N (1 - J_{ij} S_i)$$

che ha la proprietà di essere nulla se la clausola  $j$  è soddisfatta, maggiore di zero nel caso non lo sia, in modo tale che lo stato fondamentale sia l'assegnamento che soddisfi ogni clausola della formula. Sommando su tutte le funzioni indice delle clausole otteniamo un'hamiltoniana del K-SAT, la cui energia corrisponde al numero di clausole violate:

$$\mathcal{H}_{\text{K-SAT}} = \sum_{j=1}^M V_j = \frac{1}{2^K} \sum_{j=1}^M \prod_{i=1}^N (1 - J_{ij} S_i)$$

---

<sup>3</sup>D'ora in avanti porremo la costante di Boltzmann  $k_B = 1$  come si usa tipicamente in meccanica statistica.



quindi il problema K-SAT è ricondotto al problema di trovare lo stato fondamentale di una particolare hamiltoniana. Diverse istanze dei parametri  $J_{ij}$  corrispondono a diverse formule.

Grazie a questa corrispondenza, trovata da Zecchina e Monasson [20] i fisici hanno potuto applicare metodi creati per studiare i sistemi di vetri di spin per analizzare il comportamento critico del K-SAT.

Per calcolare l'energia dello stato fondamentale introduciamo una temperatura fittizia che alla fine mandiamo a zero  $\beta \rightarrow \infty$ . La funzione di partizione e l'energia libera sono:

$$Z = \text{Tr}(e^{-\beta \mathcal{H}(\underline{s})}) \quad F(\beta) = -\frac{1}{\beta} \ln Z$$

dove  $\underline{s} \in \{0, 1\}^N$  è un particolare assegnamento, lo stato del sistema. Tramite la funzione di partizione è possibile calcolare l'energia media, da cui è possibile calcolare l'energia dello stato fondamentale nel limite  $\beta \rightarrow \infty$ .

$$E = -\frac{\partial \ln Z}{\partial \beta} \quad E_{GS} = \lim_{\beta \rightarrow \infty} E$$

Queste quantità riguardano una particolare istanza del problema K-SAT, ovvero una particolare matrice  $J_{ij}$ . Per studiare la transizione di fase in vicinanza della soglia critica, siamo interessati al comportamento di un ensemble di formule regolate dal parametro  $\alpha$ . Chiamando  $\overline{E_{GS}}(\alpha)$  l'energia dello stato fondamentale mediata sull'ensemble per un fissato  $\alpha$ , ci aspettiamo che nel limite termodinamico  $\overline{E_{GS}}(\alpha) = 0$  se ci troviamo al di sotto della soglia critica (fase SAT),  $\overline{E_{GS}}(\alpha) > 0$  se siamo oltre (fase UNSAT).

## 2.5 Upper bound

Calcoliamo un upper bound per i valori critici  $\alpha_c(K)$  utilizzando il metodo del primo momento.

Se  $X$  è una variabile aleatoria non negativa, intera e il cui valore atteso esiste, segue dalla disuguaglianza di Markov che

$$P(X \neq 0) \leq \mathbb{E}[X] \quad (2.2)$$

Scegliamo  $X(\varphi)$  in maniera che

$$X(\varphi) = \begin{cases} 0 & \text{se } \varphi \text{ UNSAT} \\ \geq 1 & \text{altrimenti} \end{cases} \quad (2.3)$$

quindi per ogni formula  $\varphi$  abbiamo

$$P(\varphi \text{ SAT}) \leq \mathbb{E}[X(\varphi)]$$

Conviene scegliere una  $X(\varphi)$  di cui si sappia calcolare il valore atteso. Una scelta naturale è il numero di assegnamenti che rendono vera  $\varphi$ . Il valore atteso di  $X(\varphi)$  sarà il numero di assegnamenti,  $2^N$ , per la probabilità che un assegnamento sia SAT (che non dipende dal particolare assegnamento). La probabilità che un particolare assegnamento soddisfi la formula è pari al prodotto delle probabilità che soddisfi ogni clausola. Ogni clausola di  $K$  letterali è falsa se tutte i letterali che la compongono sono messi a falso, ciò avviene con probabilità  $2^{-K}$ . Perciò la probabilità che un dato assegnamento soddisfi una clausola è  $(1 - 2^{-K})$  [16]. Quindi otteniamo:

$$\mathbb{E}[X(\varphi)] = 2^N (1 - 2^{-K})^M = \exp N(\ln 2 + \alpha \ln(1 - 2^{-K}))$$

Sappiamo che per  $N \rightarrow \infty$ ,  $\mathbb{E}[X(\varphi)] \rightarrow 0$ , il che richiede che l'argomento dell'esponenziale sia negativo, condizione che avviene quando  $\alpha > \alpha_{UB}$

$$\alpha_{UB}(K) = -\frac{\ln 2}{\ln(1 - 2^{-K})} \quad (2.4)$$

Questo risultato è dovuto a Franco [21]. La conseguenza importante è che la soglia critica  $\alpha_C(2)$  dovrà per forza essere inferiore a  $\alpha_{UB}(2) \approx 2.409$ .

## 2.6 Random 2-SAT

Per studiare il problema 2-SAT ricorriamo ad una rappresentazione equivalente del problema: a partire da una formula  $\varphi$  generiamo un grafo diretto  $\vec{G}(\varphi)$ <sup>4</sup>[22]. L'insieme di vertici del nostro grafo diretto saranno tutte le variabili proposizionali che compaiono nella formula  $\varphi$ , e la loro negazione. Per definire gli archi del grafo sfruttiamo le seguenti identità:

$$x_1 \vee x_2 \equiv \neg x_1 \rightarrow x_2 \equiv \neg x_2 \rightarrow x_1$$

Da ogni clausola generiamo due implicazioni logicamente equivalenti dalle quali, in maniera molto naturale, definiamo gli archi del grafo diretto (da qui il nome *implication graph*). È bene notare che questa costruzione del grafo a partire da una formula è realizzabile in tempo lineare.

L'insieme dei vertici  $V$  del grafo d'implicazione  $\vec{G}(\varphi)$  saranno quindi i  $2N$  letterali della formula  $L = \{x_1, \neg x_1, \dots, x_n, \neg x_n\}$ , mentre ogni clausola  $(l_i \vee l_j)$  genera i due archi  $(\neg l_i \rightarrow l_j)$  e  $(\neg l_j \rightarrow l_i)$ . Inoltre chiamiamo  $E$  l'insieme degli archi.

Un'interessante proprietà di dualità del grafo d'implicazione è la seguente:  $\vec{G}(\varphi)$  è isomorfo al suo *grafo complementare*  $\vec{G}(\varphi)^c$ , definito invertendo i versi degli archi e

---

<sup>4</sup>Naturalmente  $\vec{G}$  non è un vettore, la notazione ricorda come il verso degli archi sia rilevante in un grafo diretto.

scambiando i letterali con la loro negazione. Ciò è una conseguenza dell'equivalenza  $x_i \rightarrow x_j \equiv \neg x_j \rightarrow \neg x_i$ , ovvero per contrapposizione.

Il motivo per cui definiamo i grafi d'implicazione è il seguente teorema, che mette in luce come lo studio della soddisfacibilità di una formula sia legato allo studio della topologia del suo grado d'implicazione. Ricordiamo che una componente fortemente connessa di un grafo  $\vec{G}$  è un sottografo massimale<sup>5</sup> di  $\vec{G}$  in cui esiste un cammino orientato tra ogni coppia di nodi ad esso appartenenti.

**Teorema 2.6.1.** *Una formula  $\varphi$  2-CNF è soddisfacibile se e solo se il suo grafo d'implicazione  $\vec{G}(\varphi)$  non ha una componente fortemente connessa in cui compare una variabile e la sua negazione (chiamiamo questi sottografi componenti contraddittorie).*

*Dimostrazione.* Un verso è molto semplice. Infatti, possiamo procedere per assurdo e supporre che il grafo ammetta una componente fortemente connessa in cui compaia una variabile  $x$  e la sua negazione. Ma le connessioni fra vertici nel grafo d'implicazione corrispondono a implicazioni logiche (per definizione di grafo d'implicazione). Dato che  $x$  e  $\neg x$  appartengono allo stessa componente fortemente connessa, otteniamo quindi

$$x \rightarrow \neg x \text{ e } \neg x \rightarrow x$$

per tanto otteniamo che la formula  $x \leftrightarrow \neg x$  è messa a Vero, ma questa formula è ovviamente insoddisfacibile. Questo assurdo prova la prima implicazione.

Per dimostrare l'altro verso si procede per induzione sulle variabili di  $\varphi$ . Il caso  $N = 1$  è triviale. Supponiamo che il grafo diretto  $\vec{G}(\varphi)$  non abbia componenti contraddittorie. Dobbiamo dimostrare che la formula  $\varphi$  è SAT. Chiamiamo  $SC(x)$  la componente fortemente connessa in cui compare  $x$ . Definiamo un ordine parziale sull'insieme delle componenti fortemente connesse del grafo nella seguente maniera:

$$SC(x) \prec SC(y) \iff x \rightsquigarrow y \text{ e } \forall x' \in SC(x) \forall y' \in SC(y) \quad x' \rightsquigarrow y' \quad (2.5)$$

dove con  $x \rightsquigarrow y$  intendiamo che esiste un cammino tra  $x$  e  $y$ , ovvero un sottoinsieme ordinato  $P = \{e_1, \dots, e_n\}$  con  $e_i \in E \forall i \in [1, n]$  tale che  $x \in e_1$  e  $y \in e_n$ .

Questo ordine parziale ha un significato preciso: se poniamo una delle variabili nella  $SC(x)$  a vero, lo saranno anche quelle di  $SC(y)$ . Ponendo  $SC(x)$  a falso non abbiamo restrizioni (*ex falso quod libet*). Sia  $SC_0$  l'elemento minimale dell'ordine parziale<sup>6</sup>  $\prec$  ( $SC_0$  non contiene archi  $x \rightsquigarrow y$  con  $x \notin SC_0$  e  $y \in SC_0$ ). Allora, il grafo complementare  $SC_0^c$  è l'elemento massimale di  $\prec$ . Inoltre, siccome per ipotesi il grafo non contiene componenti contraddittorie,  $SC_0 \cap SC_0^c = \emptyset$ . Poniamo tutti i

<sup>5</sup>Un sottografo  $G'$  è massimale per una proprietà  $\rho$  se ogni sottografo contenente  $G'$  non ha  $\rho$

<sup>6</sup>Ricordiamo che  $m$  si dice elemento minimale (massimale) dell'insieme ordinato  $(A, \leq)$  se  $a \leq m(a \geq m) \implies a = m, a \in A$ .

letterali di  $SC_0$  a falso (tutti i letterali di  $SC_0^c$  a vero). Allora tutte le clausole che contengono almeno un letterale in  $SC_0 \cup SC_0^c$  sono soddisfatte. Questo processo rimuove tutte le variabili corrispondenti ai letterali in  $SC_0$  e  $SC_0^c$ , e tutte le clausole di  $\varphi$  corrispondenti, generando una nuova formula  $\varphi'$ . Essendo  $\vec{G}(\varphi')$  sotto grafo di  $\vec{G}(\varphi)$ , anche questo è privo di componenti contraddittorie. Grazie all'ipotesi induttiva, otteniamo un assegnamento che soddisfa  $\varphi$ .  $\square$

Consideriamo l'insieme di tutte le formule  $\varphi_{N,M}$  2-CNF di un linguaggio a  $N$  variabili e con  $M$  clausole. Scegliendo casualmente una formula in questo insieme (ogni formula è equiprobabile) qual è la probabilità che sia soddisfacibile? Dimostreremo come associato a questo problema vi è una transizione di fase controllata dal seguente parametro:  $\alpha = \frac{M}{N}$ . Un  $\alpha$  piccolo significa molti gradi di libertà (le variabili) rispetto ai vincoli (le clausole): ci aspettiamo che tali formule siano in generale più soddisfacibili rispetto ad un  $\alpha$  alto.

Dimostreremo che tale transizione SAT-UNSAT avviene per un valore critico  $\alpha_c(2) = 1$ , sfruttando la rappresentazione del grafo diretto. Transizioni di fase simili in ensemble di grafi erano già state notate da Erdős–Rényi [23].

**Teorema 2.6.2** (Problema di soglia). *Esiste una soglia critica per il problema 2-SAT tale che:*

1. se  $\alpha < 1$  la probabilità che  $\varphi$  sia SAT tende a 1 quando  $N \rightarrow \infty$ ;
2. se  $\alpha > 1$  la probabilità che  $\varphi$  sia SAT tende a 0 quando  $N \rightarrow \infty$ .

*Dimostrazione.* Cominciamo provando il punto 1. Seguiremo la dimostrazione combinatoria di Chvátal e Reed [24]. Un *biciclo* di lunghezza  $s$  è definito nel seguente modo: un cammino  $(u, w_1, \dots, w_s, v)$  dove  $w_i$  sono letterali di  $s$  variabili distinte e  $u, v \in \{w_1, \dots, w_s, \neg w_1, \dots, \neg w_s\}$ . Dal teorema 2.6.1, se  $\varphi$  è UNSAT allora contiene una componente contraddittoria, ovvero un cammino  $l_0, l_1, \dots, l_t$  tale che  $l_0 = \neg l_t$  e  $l_t = \neg l_0$  per un qualche  $t$ . Scegliendo un cammino che minimizza  $t$ , troviamo che tutti i letterali  $l_1, \dots, l_t$  sono distinti e nessuno è la negazione di un altro. La probabilità che  $\varphi$  sia UNSAT è uguale alla probabilità che abbia una componente contraddittoria, che è minore uguale alla probabilità che abbia un biciclo. Quest'ultima è ovviamente maggiorata numero di bicicli aspettati (vale la disuguaglianza 2.2):

$$\mathbb{P}(\varphi \text{ UNSAT}) \leq \mathbb{P}(\vec{G}(\varphi) \text{ ha un biciclo}) \leq \sum_{s=2}^N N^s 2^s (2s)^2 M^{s+1} \left( \frac{1}{4 \binom{N}{2}} \right)^{s+1} = \frac{2M}{N(N-1)} \sum_{s=2}^N s^2 \left( \frac{M}{N-1} \right)^s$$

$N^s$  è un upper bound sulle possibili scelte di variabili che compaiono in un biciclo di lunghezza  $s$ ;  $2^s$  sono le scelte dei letterali date le variabili;  $(2s)^2$  sono il numero di

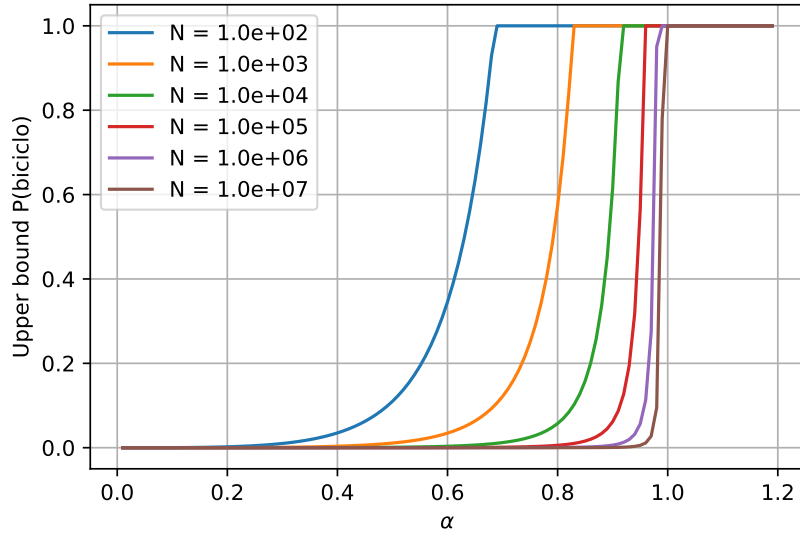
scelte per  $u, v$ ;  $M^{s+1}$  è un upper bound sul numero di possibili clausole del biciclo; l'ultimo termine è la probabilità che ogni clausola del biciclo appartenga al grafo.

Lo stesso conto risulta più agevole nell'ensemble con numero di (archi) clausole variabili[25]. In questo modello la probabilità che una formula contenga una delle  $4\binom{N}{2}$  possibili clausole distinte è  $p$ . Con la giusta scelta di  $p$  i due modelli  $G(N, M)$  e  $G(N, p)$  sono equivalenti per proprietà monotone e quasi-certe scegliendo:

$$p = \frac{M}{4\binom{N}{2}} = \frac{M}{2N^2}(1 + \mathcal{O}(1/N))$$

ovvero imponendo che il valore atteso di clausole sia  $M$  (essendo eventi indipendenti ripetuti la distribuzione è binomiale). Nel limite  $N$  grande il numero di clausole sarà sempre più vicino al numero  $M$ . Per una dimostrazione si veda [26].

$$\mathbb{P}(\varphi \text{ UNSAT}) \leq \mathbb{P}(\vec{G}(\varphi) \text{ ha biciclo}) \leq \sum_{s=2}^N N^s 2^s (2s)^2 p^{s+1} = \frac{2\alpha}{N} \sum_{s=2}^N s^2 \alpha^s \quad (2.6)$$



**Figura 2.1.** L'andamento dell'espressione 2.6 in funzione della densità di clausola  $\alpha$  per un numero di variabili  $N$  da  $10^2$  a  $10^7$ . Il grafico mostra come nel limite termodinamico, per valori di  $\alpha < 1$  la probabilità che compaia un biciclo nel grafo d'implicazione tende a zero.

Che mostra chiaramente che per  $\alpha < 1$  nel limite  $N \rightarrow \infty$  l'upper bound sul numero di bicicli del grafo d'implicazione tende a zero (si veda la figura 2.1) di conseguenza anche la probabilità di essere UNSAT tende a zero.

Per dimostrare il secondo punto, sfruttiamo l'upper bound 2.4, che ci assicura  $\alpha_c(2) \leq \alpha_{UB}(2) \approx 2.41$ . □

## Capitolo 3

# Simulazioni numeriche

In questo capitolo verificheremo quanto detto precedentemente sul random 2-SAT, mostrando la transizione di fase SAT-UNSAT. Faremo ciò generando istanze random per studiare la percentuale di formule soddisfacibili e annotando il numero di step impiegati dall'algoritmo, al variare del parametro  $\alpha = M/N$  detto *densità di clausola*.

Nella prima parte spiegheremo l'algoritmo che utilizzeremo per risolvere il problema 2-SAT e studiare il rapporto di formule soddisfacibili. Nella seconda parte introdurremo brevemente il *finite-size scaling* utilizzato per studiare sistemi finiti nel punto critico, e come verrà applicato nello studio del random 2-SAT per calcolare un particolare esponente critico. L'ultima parte è l'analisi dei dati in cui verificheremo le predizioni teoriche e le performance dell'algoritmo implementato.

### 3.1 L'algoritmo

Il 2-SAT appartiene alla classe di complessità polinomiale P ed è pertanto considerato un problema *facile*. Un algoritmo lineare fu trovato da Aspvall Plass e Tarjan nel 1978 [27]. Ciò consente di poter studiare formule con un grande numero di variabili, prerogativa per lo studio del limite termodinamico.

Il teorema 2.6.1 è di fatto una procedura per verificare che una formula 2-CNF sia SAT o UNSAT: una volta trovate le componenti fortemente connesse del grafo d'implicazione della formula, è sufficiente controllare se siano contraddittorie, ovvero contengano una letterale e la sua negazione. Tale controllo è ovviamente lineare per il numero di vertici del grafo ( $2N$ ). Fortunatamente anche trovare le componenti fortemente connesse di un grafo è un problema risolvibile in tempo lineare da algoritmi come quello di Kosaraju [28] o Tarjan [29].

Un approccio più diretto che non richiede di passare esplicitamente per il grafo d'implicazione è l'algoritmo proposto da Mézard e Montanari [16]. Se iniziamo da

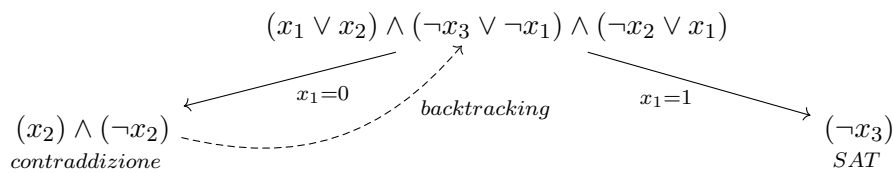
una clausola  $C = l_1 \vee l_2$  e fissiamo il valore di verità del letterale  $l_1$ , possono accadere due cose:

- Se fissiamo  $l_1 = 1$ , la clausola è soddisfatta e può essere rimossa dalla formula.
- Se fissiamo  $l_1 = 0$ , la clausola si trasforma in una clausola con un solo letterale, detta *unit clause*, che implica  $l_2 = 1$

Questa procedura di semplificazione chiamata *unit clause propagation* (UCP) è legittimata dalle seguenti equivalenze:  $1 \wedge x_i \equiv x_i$  e  $0 \vee x_i \equiv x_i$ , dove  $x_i$  può essere una clausola o un letterale (o una variabile booleana).

Data una formula 2-CNF, una istanza del 2-SAT, iniziamo scegliendo una variabile  $x_i \in \{x_1, \dots, x_N\}$  e fissiamo  $x_i = 0$ . Procediamo applicando la regola di riduzione (UCP) alle clausole che contengono  $x_i$  o  $\neg x_i$ . Si prosegue ricorsivamente fissando il valore di verità dei letterali che appaiono in unit clause. La procedura può arrestarsi per una delle seguenti ragioni: la formula non contiene unit clause; la formula contiene la contraddizione  $l_i \wedge \neg l_i$ , che a causa della regola di riduzione comporta la comparsa di una clausola vuota.

Nel primo caso, possiamo scegliere arbitrariamente un'altra variabile non ancora fissata e continuare ad applicare la regola di riduzione. Nel secondo, si effettua un *backtracking*, ovvero si cambia la scelta originale sul valore di verità di  $x_i$ . Se anche con questa scelta si arriva ad una contraddizione, la formula è UNSAT. Se la procedura non incontra nessuna contraddizione e fissa il valore di verità di tutte le variabili rimuovendo ogni clausola, allora un assegnamento è stato trovato e la formula è SAT. Vediamo l'algoritmo sopra descritto applicato ad una particolare formula:



La cosa non banale è che in caso di contraddizioni occorre effettuare un *backtracking* di solo un passo, cosa in generale non valida per i problemi K-SAT (si veda l'algoritmo DPLL[30]). Nel 2-SAT infatti, ogni assegnamento parziale (che dopo la riduzione non produce una formula contraddittoria) è estendibile ad un assegnamento completo.

L'algoritmo così definito è efficiente, dato che fisserà il valore delle variabili al massimo  $2N$  volte.

Riportiamo di seguito lo pseudocodice dell'algoritmo sopra descritto.

**Algorithm 1** 2satSolver(Formula  $\varphi$  2CNF)

---

```

 $x = \text{ChoseVariable}(\varphi)$  ▷ Viene scelta una variabile
 $\varphi_p = \text{UCP}(\varphi, x)$  ▷ Pone  $x = 0$  e semplifica ricorsivamente la formula
if  $\varphi_p$  è vuota then
    return SAT
else if  $\varphi_p$  contiene una clausola vuota then
     $\varphi_n = \text{UCP}(\varphi, \neg x)$  ▷ in caso di contraddizioni prova  $x = 1$ 
    if  $\varphi_n$  è vuota then
        return SAT
    else if  $\varphi_n$  contiene una clausola vuota then
        return UNSAT
    else
        return 2satSolver( $\varphi_n$ )
    end if
else
    return 2satSolver( $\varphi_p$ )
end if

```

---

### 3.2 Finite-size scaling

Le simulazioni numeriche sono limitate dagli errori statistici e dalla dimensione finita dei sistemi simulati. L'ultima limitazione è la più problematica per lo studio delle transizioni di fase, per quanto già detto nella sezione 1.2 solo per sistemi infiniti il potenziale termodinamico può diventare non analitico e il sistema presentare una transizione di fase.

La lunghezza di correlazione  $\xi$  diverge nel punto critico: finché è piccola rispetto alle dimensioni del sistema simulato  $L$  la simulazione è indistinguibile da un sistema reale, tuttavia quando ci si avvicina al punto critico le fluttuazioni superano le dimensioni del sistema, ne consegue che le singolarità sono smussate 3.2, invece di divergere assumono un valore finito. Nel punto critico del sistema finito si ha:

$$\xi \approx L \tag{3.1}$$

$$C_T \propto L^{\underline{\alpha}/\nu} \tag{3.2}$$

dove  $\underline{\alpha}$  e  $\nu$  sono gli esponenti critici della capacità termica e della lunghezza di correlazioni definiti nel capitolo 1.2 tabella 1.1.

Studiare come le proprietà del sistema finito *scalano* simulando sistemi con dimensioni  $L$  diverse consente di studiare il caso limite, trovando il rapporto  $\underline{\alpha}/\nu$ . Questa tecnica è nota come *finite-size scaling* [31].



Applicheremo questa tecnica per lo studio della *scaling window* (o regione critica) del 2-SAT, definita come la quantità  $w(N)$  con cui si deve incrementare la densità di clausole  $\alpha$  per abbassare la probabilità di essere SAT da  $1 - \epsilon$  ad  $\epsilon$ . Ci aspettiamo che al crescere di  $N$  questa regione critica si restringa rendendo la transizione sempre più netta. Assumiamo che la dimensione della regione critica sia inversamente proporzionale alla derivata della probabilità di essere SAT nel punto critico[32]:

$$w(N)^{-1} \propto \left. \frac{\partial P_{SAT}(\alpha)}{\partial \alpha} \right|_{\alpha=\alpha_c} \quad (3.3)$$

in maniera tale che per  $N \rightarrow \infty$  la regione diventi nulla. Postuliamo che l'andamento di  $w(N)$  sia regolato da una legge di potenza

$$w(N) \propto N^{-1/\nu} \quad (3.4)$$

che definisce l'esponente critico  $\nu$ , in analogia con la 3.2, nel nostro caso le dimensioni del sistema  $L$  è il numero di variabili della formula  $N$ [33].

Non è ovvio a priori che  $\nu$  sia ben definito, in principio potrebbe essere una funzione debolmente oscillante di  $N$  o dipendere da  $\epsilon$ . Solo per il 2-SAT è stato provato rigorosamente che  $\nu = 3$  da Bollobas et. al [26], mentre per  $K \geq 3$  si sa solo che se  $\nu$  è ben definito, vale  $\nu \geq 2$ , risultato ottenuto da Wilson [34].

Dalle relazioni 3.3 e 3.4 segue che nell'intorno del valore critico

$$P_{SAT}(N, \alpha) \propto N^{1/\nu}(\alpha - \alpha_c) = f(y) \quad (3.5)$$

dove il membro di destra è una funzione di scala universale, che dipende dalla quantità adimensionale  $y \equiv N^{1/\nu}(\alpha - \alpha_c)$ .

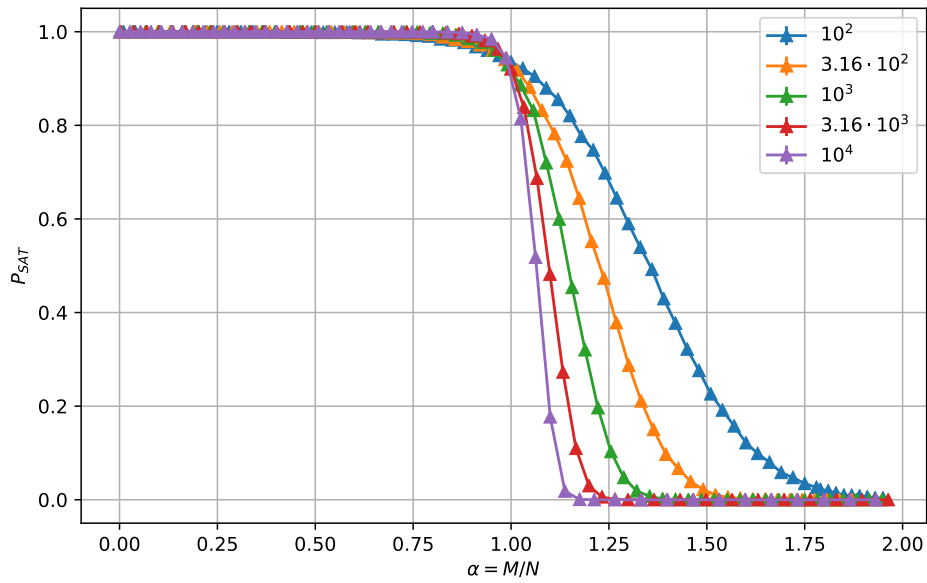
### 3.3 Analisi dati

Generiamo  $10^4$  istanze del random 2-SAT con un numero di variabili  $N = 10^2, 3.16 \cdot 10^2, 10^3, 3.16 \cdot 10^3$  e  $10^4$ , facendo variare il parametro  $\alpha$  nell'intervallo  $[0, 2]$  per stimare la funzione  $P_{SAT}(\alpha, N)$ , calcolando il rapporto di formule SAT tramite un'implementazione Python dell'algoritmo esposto nella sezione 3.1<sup>1</sup> Stimiamo la complessità (si veda 2.1) annotando il numero medio di chiamate ricorsive dell'algoritmo 3.1 per ciascun valore di  $\alpha$ . Inoltre valutiamo le performance dell'algoritmo usato studiando il tempo impiegato dall'implementazione in funzione della dimensione del numero di variabili  $N$  delle formule.

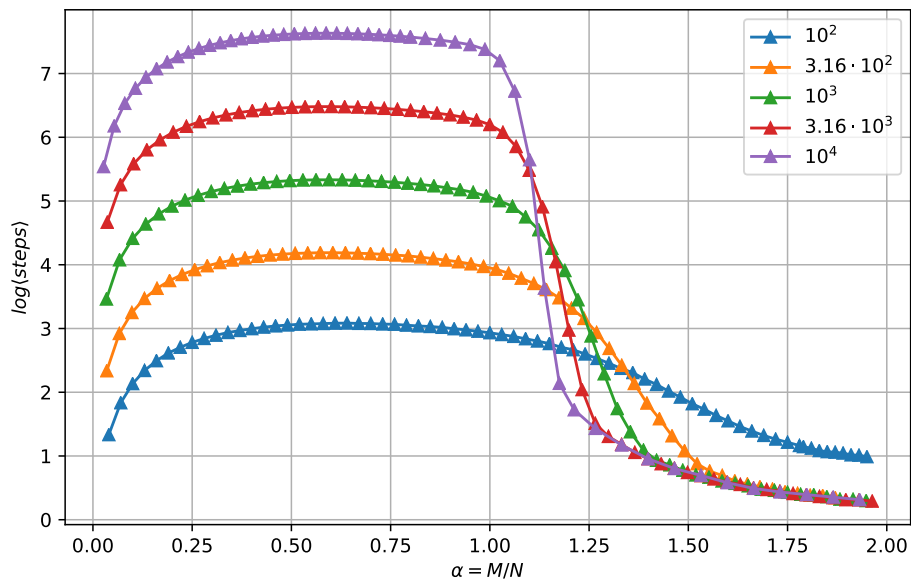
Riportiamo di seguito i grafici della  $P_{SAT}(\alpha, N)$  ottenuti numericamente, e gli step medi impiegati dall'algoritmo rispettivamente nelle figure 3.1 e 3.2.

---

<sup>1</sup>Specifiche hardware: CPU AMD Ryzen 5 2400G 3.6 GHz, RAM 16 GB 3000 MHz, OS Windows 10 64 bit.



**Figura 3.1.** Probabilità che un'istanza del 2-SAT sia soddisfacibile, per  $N = 10^2, 3.16 \cdot 10^2, 10^3, 3.16 \cdot 10^3$  e  $10^4$ . Ogni punto è generato mediando su  $10^4$  formule random.



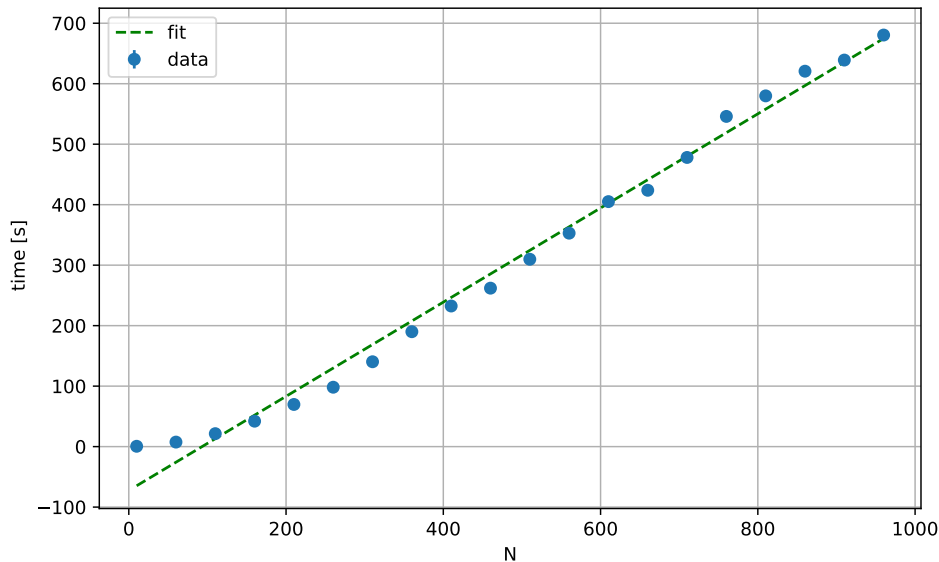
**Figura 3.2.** Numero medio di step impiegati dall'algoritmo, calcolato contando le chiamate ricorsive. Ogni punto è generato mediando su  $10^4$  formule random.

Per valori di  $N$  grandi, gli effetti di taglia finita diminuiscono e il comportamento limite della  $P_{SAT}(\alpha, N)$  diventa evidente: per  $\alpha < 1$  la formula è quasi certamente soddisfacibile, per  $\alpha > 1$  quasi certamente UNSAT, come si nota dalla figura 3.1. Infatti per  $N = 10^2$  la probabilità di essere SAT raggiunge lo zero solo per valori della densità di clausola  $\alpha \approx 2$ , mentre per  $N = 10^4$  la formula è quasi certamente insoddisfacibile già per  $\alpha \approx 1.1$ .

Notiamo inoltre come il numero di step medi impiegati dall'algoritmo, riportati in figura 3.2, che rispecchia la complessità delle istanze, mostri l'andamento notato da Kanefsky Cheeseman P. e W. M. Taylor [8] per problemi P: non vi è un picco nel valore critico della densità di clausola, come invece avviene per problemi NP come il 3-SAT. Il numero di chiamate dell'algoritmo è praticamente costante nella fase SAT, ed i massimi delle curve per i diversi valori di  $N$  scalando linearmente come ci si aspettava ( $\mathcal{O}(2N)$ ). Notiamo come nella fase UNSAT il numero di step necessari per risolvere il problema si riduca notevolmente, spiegabile dal fatto che l'algoritmo trova subito una contraddizione (un ciclo contraddittorio).

Per valutare le performance dell'implementazione dell'algoritmo generiamo  $10^3$  formule con un valore di densità di clausola  $\alpha = 0.75$ , ovvero nella fase SAT (come si vede in figura 3.1). Annotiamo poi il tempo necessario per decidere la soddisfacibilità delle  $10^3$  formule per un numero di variabili  $N$  fissato.

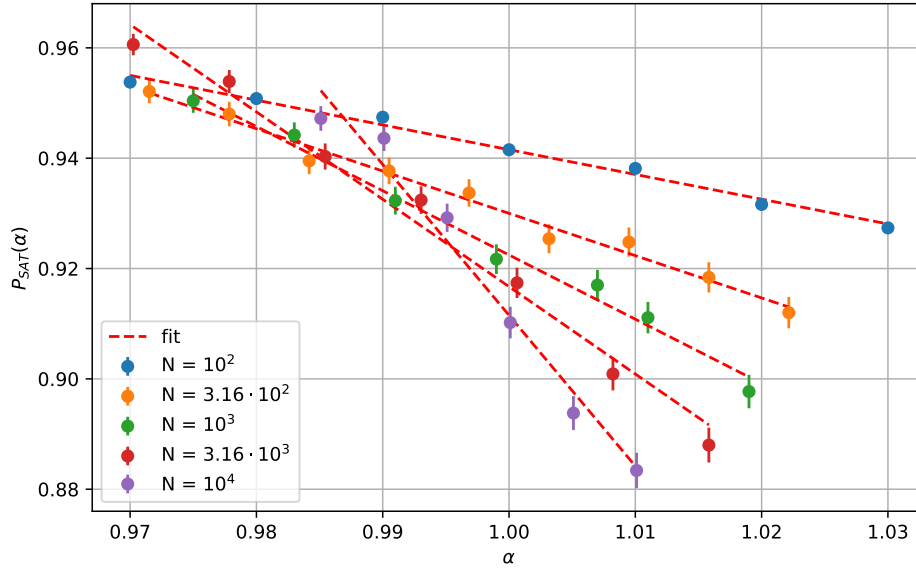
Riportiamo nella seguente figura 3.3 il grafico del tempo impiegato  $t(N)$  in funzione del numero di variabili  $N$  delle formule.



**Figura 3.3.** Tempo impiegato dall'algoritmo 3.1 per decidere la soddisfacibilità di  $10^3$  formule per il 2-SAT.

L'andamento è lineare come ci si aspettava. Un fit lineare eseguito su questi dati rivela un coefficiente angolare di  $0.778 \pm 0.001$  ed un'intercetta di  $-72.53 \pm 0.01$ .

Usiamo i dati nell'intorno del punto critico per stimare la derivata  $P_{SAT}(\alpha, N)$  rispetto ad  $\alpha$  tramite un fit lineare (figura 3.4).



**Figura 3.4.** Stima della derivata di  $P_{SAT}(\alpha, N)$  calcolata nel punto critico per  $N = 10^2, 3.16 \cdot 10^2, 10^3, 3.16 \cdot 10^3$  e  $10^4$ .

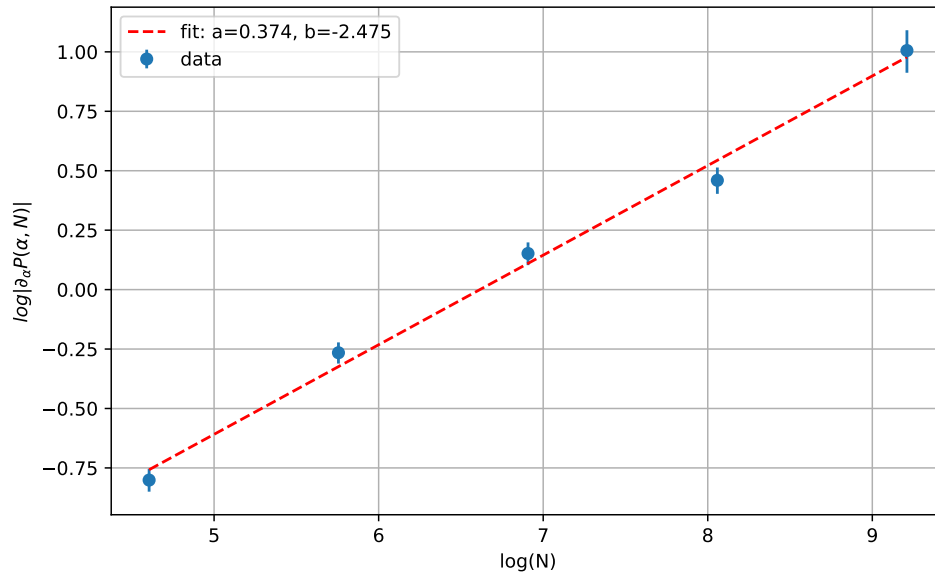
N	coefficiente angolare
$10^2$	$-0.449 \pm 0.021$
$3.16 \cdot 10^2$	$-0.767 \pm 0.034$
$10^3$	$-1.114 \pm 0.061$
$3.16 \cdot 10^3$	$-1.584 \pm 0.088$
$10^4$	$-2.68 \pm 0.20$

**Tabella 3.1.** Coefficienti angolari delle rette in figura 3.4, stimati tramite un fit lineare.

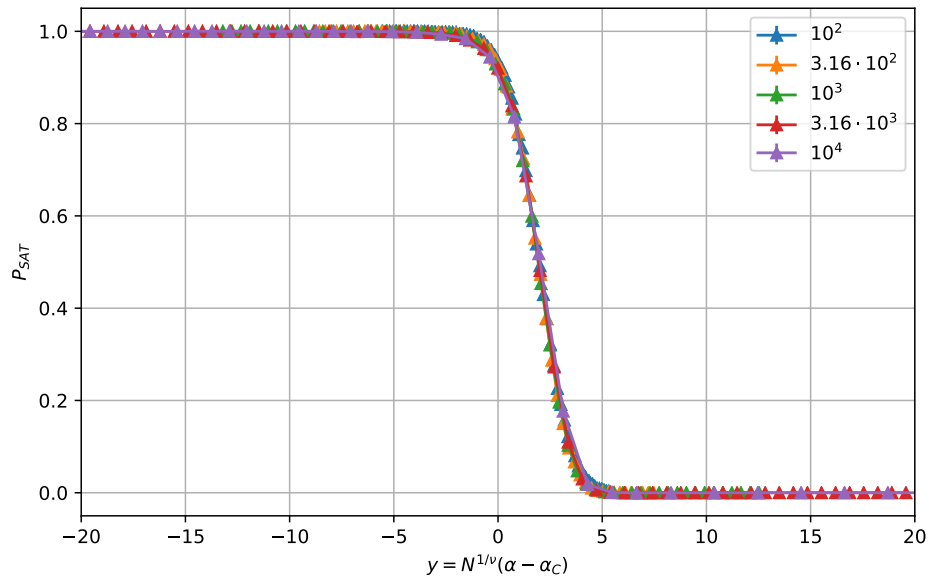
Possiamo considerare le intersezioni delle varie curve per stimare il valore critico  $\alpha_c$ . Mediando i punti d'intersezione delle rette in figura 3.4 otteniamo

$$\alpha_c = 0.99 \pm 0.01$$

compatibile con il valore teorico  $\alpha_c = 1$ . Calcoliamo l'esponente critico  $\nu$  sfruttando le relazioni 3.3 e 3.4, ed effettuando un altro fit lineare in scala  $\log - \log$  (figura 3.5).



**Figura 3.5.** Stima dell'esponente critico  $\nu$  tramite fit lineare in scala log-log dei valori della derivata di  $P_{SAT}$  nel punto critico in funzione di  $N$ .



**Figura 3.6.** Curva universale ottenuta riscalando i valori di  $\alpha$  in  $N^{1/\nu}(\alpha - 1)$ .

In figura 3.6 è riportata la curva universale 3.5 ottenuta riscaland  $\alpha$  con la variabile  $y \equiv N^{1/\nu}(\alpha - 1)$ . Tutte le curve con un numero diverso di variabili collassano

in una funzione di scala universale. La stima dell'esponente critico  $\nu$  tramite il fit risulta essere:

$$\boxed{\nu = 2.68 \pm 0.13}$$

In accordo con con la prima stima di  $\nu = 2.6 \pm 0.2$  dovuta a Kirkpatrick e Selman [35], ma distante dal valore teorico  $\nu = 3$ . Il regime asintotico dove  $\nu = 3$  si raggiunge quando il numero di variabili supera 4096, mostrato empiricamente da Wilson<sup>2</sup>. Kirkpatrick e Selman non superarono  $N = 500$ , mentre nelle nostre simulazioni abbiamo raggiunto  $N = 10000$ , questo spiega come la stima da noi ottenuta sia leggermente superiore.

---

<sup>2</sup><http://dbwilson.com/2sat-data/>

## Capitolo 4

# Conclusioni

Questa dissertazione sul K-SAT mostra la versatilità della meccanica statistica, come i metodi matematici sviluppati dai fisici possano essere applicati a campi apparentemente lontani.

Nel primo capitolo dedicato alla meccanica statistica, si è discusso brevemente del limite termodinamico e della sua validità, delle transizioni di fase e la loro caratterizzazione tramite esponenti critici, per finire con il modello di Ising, fondamentale come base per lo studio dei fenomeni critici.

Nel secondo capitolo sono stati definiti i problemi K-SAT importanti per l'informatica teorica, e si è discusso della distinzione fondamentale tra algoritmi polinomiali ed esponenziali. Dopo è stato introdotto il random K-SAT e la sua congettura sulle soglie critiche, dove allo studio del caso medio è stato possibile applicare la meccanica statistica, mostrando come costruire un'hamiltoniana del K-SAT analizzabile con i metodi sviluppati per i vetri di spin, avendo in comune costanti di accoppiamento aleatorie come nel modello SK brevemente esposto. E' stato calcolato un upper bound per le soglie critiche sfruttando il metodo del primo momento, poi è stata esposta la dimostrazione combinatoria della transizione di fase SAT-UNSAT del 2-SAT.

Nell'ultimo capitolo è stata introdotta la finite-size scaling per il calcolo degli esponenti critici, ed è stato studiato numericamente il problema random 2-SAT mostrando la transizione SAT-UNSAT e calcolando l'esponente critico associato.

# Bibliografia

- [1] J. von Neumann. «First draft of a report on the EDVAC». In: *IEEE Annals of the History of Computing* 15.4 (1993), pp. 27–75. DOI: 10.1109/85.238389.
- [2] E. Fermi et al. «STUDIES OF THE NONLINEAR PROBLEMS». In: (mag. 1955). DOI: 10.2172/4376203. URL: <https://www.osti.gov/biblio/4376203>.
- [3] Juris Hartmanis e R. Stearns. «On the Computational Complexity of Algorithms». In: *Transactions of The American Mathematical Society - TRANS AMER MATH SOC* 117 (mag. 1965), pp. 285–285. DOI: 10.2307/1994208.
- [4] Stephen A. Cook. «The complexity of Theorem-proving procedures». In: *Proceedings of the Third Annual ACM Symposium on Theory of Computing*. (1971), pp. 151–158.
- [5] Ronald V. Book. «Richard M. Karp. Reducibility Among Combinatorial Problems. Complexity of Computer Computations, Proceedings of a Symposium on the Complexity of Computer Computations, Held March 20-22, 1972, at the IBM Thomas J. Watson Center, Yorktown Heights, New York, Edited by Raymond E. Miller and James W. Thatcher, Plenum Press, New York and London 1972, Pp. 85–103». In: *Journal of Symbolic Logic* 40.4 (1975), pp. 618–619. DOI: 10.2307/2271828.
- [6] Pierluigi Crescenzi et al. «On the complexity of protein folding». In: *Journal of computational biology* 5.3 (1998), pp. 423–465.
- [7] Mario P. Vecchi Scott Kirkpatrick C. Daniel Gelatt. «Optimization by simulated annealing». In: *American association for the advancement of science* (1983).
- [8] B. Kanefsky Cheeseman P. e W. M. Taylor. «Where the really hard problems are». In: *In Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI '91)* (1991), pp. 331–337.
- [9] K. Huang. *Statistical Mechanics*. Wiley, 1987. ISBN: 9780471815181. URL: <https://books.google.it/books?id=M8PvAAAAAAAJ>.



- [10] James Sethna. *Statistical mechanics: entropy, order parameters, and complexity*. Vol. 14. Oxford University Press, USA, 2021.
- [11] T. D. Lee e C. N. Yang. «Statistical Theory of Equations of State and Phase Transitions. II. Lattice Gas and Ising Model». In: *Phys. Rev.* 87 (3 ago. 1952), pp. 410–419. DOI: 10.1103/PhysRev.87.410. URL: <https://link.aps.org/doi/10.1103/PhysRev.87.410>.
- [12] Nino Zanghì. *Appunti di meccanica statistica*. 2014.
- [13] N. Goldenfeld. *Lectures on phase transitions and the renormalization group*. 1992.
- [14] Luciano M. Barone e Enzo Marinari e Giovanni Organtini e Federico Ricci-Tersenghi. *Scientific programming*. Pearson education, 2013.
- [15] David Sherrington e Scott Kirkpatrick. «Solvable Model of a Spin-Glass». In: *Phys. Rev. Lett.* 35 (26 dic. 1975), pp. 1792–1796. DOI: 10.1103/PhysRevLett.35.1792. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.35.1792>.
- [16] Marc Mezard e Andrea Montanari. *Information, Physics, and Computation*. USA: Oxford University Press, Inc., 2009. ISBN: 019857083X.
- [17] Cirano De Dominicis e Irene Giardina. *Random Fields and Spin Glasses: A Field Theory Approach*. Cambridge University Press, 2006. DOI: 10.1017/CB09780511534836.
- [18] S F Edwards e P W Anderson. «Theory of spin glasses». In: *Journal of Physics F: Metal Physics* 5.5 (mag. 1975), pp. 965–974. DOI: 10.1088/0305-4608/5/5/017. URL: <https://doi.org/10.1088/0305-4608/5/5/017>.
- [19] Marc Mézard Giorgio Parisi Miguel Angel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*. World Scientific Publishing Company, 1987.
- [20] Rémi Monasson e Riccardo Zecchina. «Statistical mechanics of the randomK-satisfiability model». In: *Physical Review E* 56.2 (ago. 1997), pp. 1357–1370. ISSN: 1095-3787. DOI: 10.1103/physreve.56.1357. URL: <http://dx.doi.org/10.1103/PhysRevE.56.1357>.
- [21] John Franco e Marvin Paull. «Probabilistic analysis of the Davis Putnam procedure for solving the satisfiability problem». In: *Discrete Applied Mathematics* 5.1 (1983), pp. 77–87. ISSN: 0166-218X. DOI: [https://doi.org/10.1016/0166-218X\(83\)90017-3](https://doi.org/10.1016/0166-218X(83)90017-3). URL: <https://www.sciencedirect.com/science/article/pii/0166218X83900173>.

- [22] Béla Bollobás. *Graph theory: an introductory course*. Vol. 63. Springer Science & Business Media, 2012.
- [23] P. Erdős e A Rényi. «On the Evolution of Random Graphs». In: *Publication of the mathematical institute of the hungarian academy of sciences*. 1960, pp. 17–61.
- [24] V. Chavatal B. Reed. «Mick Gets Some (the Odds are on its Side)». In: *IEE press, London* (1992), pp. 620–627.
- [25] W. Fernandez de la Vega. «Random 2-SAT: results and problems». In: *Theoretical Computer Science* 265.1 (2001). Phase Transitions in Combinatorial Problems, pp. 131–146. ISSN: 0304-3975. DOI: [https://doi.org/10.1016/S0304-3975\(01\)00156-6](https://doi.org/10.1016/S0304-3975(01)00156-6). URL: <https://www.sciencedirect.com/science/article/pii/S0304397501001566>.
- [26] B. Bollobás C. Borges J. T. Chayes J.H. Kim D.B. Wilson. «The scaling window of the 2-SAT transition». In: *Random Struct. Algorithms, to appear* (2001).
- [27] Bengt ASPVALL Michael F. PLASS Robert Endre TARJAN. «A LINEAR-TIME ALGORITHM FOR TESTING THE TRUTH OF CERTAIN QUANTIFIED BOOLEAN FORMULAS». In: (1978).
- [28] S Rao Kosaraju. «Insertions and deletions in one-sided height-balanced trees». In: *Communications of the ACM* 21.3 (1978), pp. 226–227.
- [29] Robert Tarjan. «Depth-first search and linear graph algorithms». In: *SIAM journal on computing* 1.2 (1972), pp. 146–160.
- [30] Davis M. G. Logemann e D. Loveland. «A machine program for theorem proving». In: *Communications of the ACM vol. 5 n. 7* (1962), pp. 394–397.
- [31] Scott Kirkpatrick e Robert H Swendsen. «Statistical mechanics and disordered systems». In: *Communications of the ACM* 28.4 (1985), pp. 363–373.
- [32] Michele Leone, Federico Ricci-Tersenghi e Riccardo Zecchina. «Phase coexistence and finite-size scaling in random combinatorial problems». In: *Journal of Physics A: Mathematical and General* 34.22 (2001), p. 4615.
- [33] Scott Kirkpatrick e Bart Selman. *Critical Behavior at the k-Satisfiability Threshold*. Citeseer, 1993.
- [34] David B. Wilson. «On the critical exponents of randomk-SAT». In: *Random Structures and Algorithms* 21.2 (ago. 2002), pp. 182–195. ISSN: 1098-2418. DOI: [10.1002/rsa.10050](http://dx.doi.org/10.1002/rsa.10050). URL: <http://dx.doi.org/10.1002/rsa.10050>.

- [35] Scott Kirkpatrick e Bart Selman. «Critical Behavior in the Satisfiability of Random Boolean Expressions». In: *Science (New York, N.Y.)* 264 (giu. 1994), pp. 1297–301. DOI: 10.1126/science.264.5163.1297.

# Ringraziamenti

Alla fine di un lungo percorso non possono non esserci dei lunghi ringraziamenti, perché tutti abbiamo bisogno di qualcuno che ci aiuti nei momenti di difficoltà.

In primis ringrazio Giulio Ricci, non solo per la precedente frase, ma per una serie di ragioni difficili da riassumere in poche righe: l'avermi fatto conoscere un sacco di gente e matematica, in tutti e due i casi con molta pazienza e nei pressi di una lavagna. Inoltre non posso non menzionare il suo aiuto e supporto psicologico durante la stesura di questa dissertazione.

Non posso non ringraziare i miei compagni di corso Youyou Lin, Paolo Mangini, Daniele Lanciotti e Enrico Massara, che oltre a sopportare le mie farneticazioni sulla fisica sono stati ottimi compagni di studio e fornitori di vitto e alloggio a basso costo. Spero che i nostri percorsi accademici si rincontrino in futuro.

Ringrazio il mio relatore il Dr. Raffaele Marino, che con molta pazienza mi ha spinto a migliorare sempre più il mio lavoro, e il professor Federico Ricci-Tersenghi per aver esaudito le mie richieste.

Mi sento obbligato a ringraziare i miei finanziatori: i miei genitori. Grazie, spero che il vostro investimento vada a buon fine!

Infine devo ringraziare la persona senza la quale non avrei potuto risolvere ogni problema burocratico, essendo incapace a scrivere email formali: la mia ragazza Elena Ferrara (ha anche altri meriti meno concreti ma altrettanto indispensabili).

Per ultimo (volutamente) ringrazio Jacopo Masia, per avermi aiutato a scrivere questi ringraziamenti, e per aver sopportato 45 minuti di presentazione sul K-SAT in preparazione alla discussione con i controrelatori. Spero che la nostra futura collaborazione accademica sia produttiva e proficua.