



Università degli Studi di Salerno
Facoltà di Scienze Matematiche Fisiche e Naturali

Tesi di Laurea Magistrale in
Informatica

Titolo

Relatori

Prof. Vincenzo Auletta
Dott. Diodato Ferraioli

Candidato

Francesco Farina
Matricola 0522500282

Anno Accademico 2015-2016

Dediche e ringraziamenti

Indice

1	Introduzione	1
2	Alcuni concetti base	2
2.1	Teoria dei Grafi	2
2.1.1	Grafo come modello della realtà	4
2.2	Modello di Ising	6
2.2.1	Partition Function	8
2.3	Cenni di probabilità e statistica	8
2.4	Processi Markoviani	10
2.4.1	Irriducibilità e periodicità	11
2.4.2	Distribuzione stazionaria	11
2.4.3	Catena di Markov Monte Carlo	12
2.5	Algoritmi di approssimazione	13
3	Logit Dynamics	14
3.1	Definizione	14
3.2	Proprietà	15
3.2.1	Ergodicità	15
3.2.2	Logit dynamics e Glauber dynamics	16
3.3	Movitazioni	16
3.4	Alcuni Esperimenti	17
4	Il lavoro di Jerrum e Sinclair	19
4.1	Spins world e Subgraphs world	20
4.2	Stima della Partition Function	22
4.3	Analisi del subgraphs-world process	22
5	Miglioramenti apportati	23
5.1	Stato dell'arte	23
5.1.1	Miglioramenti Rinaldi	23

<i>INDICE</i>	iii
5.1.2 Esperimenti	23
5.2 Stima della Partition Function	23
5.2.1 Numero di steps	23
6 Mean Magnetic Moment	24
6.1 Lemma 8, Teorema 9	24
6.2 Approssimazione della funzione odd(X)	24
6.2.1 logm Subgraphs	24
6.2.2 Algoritmo L	24
7 Implementazione e testing	25
7.1 Implementazione	25
7.2 Testing	25
7.3 DLib Python Wrapper	25
8 Conclusioni e sviluppi futuri	26
Bibliografia	26

Capitolo 1

Introduzione

Capitolo 2

Alcuni concetti base

2.1 Teoria dei Grafi

La teoria dei grafi è una branca della matematica, nata nel 1700 con Eulero, che consente di descrivere le relazioni che intercorrono tra un insieme di oggetti.

Il grafo è lo strumento attraverso il quale tali relazioni possono essere espresse ed organizzate. Infatti, il grafo, consiste di oggetti chiamati *nodi* e relazioni tra coppie di questi oggetti detti *archi*; nodi connessi tra loro da un arco sono detti *vicini* o *adiacenti*.

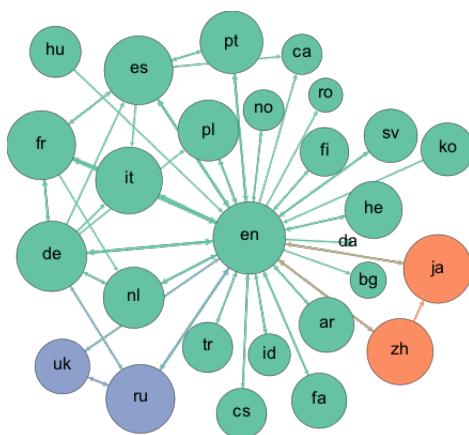


Figura 2.1: Wikipedia Multilingual Network Graph (July 2013)

La relazione tra una coppia di nodi può essere di due tipi:

- Simmetrica: l'arco connette i nodi con un collegamento bidirezionale ed è detto *indiretto*. Un grafo costituito di soli archi indiretti è anch'esso detto indiretto.
- Asimmetrica: l'arco connette i nodi con un collegamento unidirezionale ed è detto *diretto*. Un grafo costituito di soli archi diretti è anch'esso detto diretto.

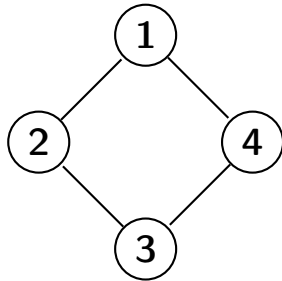


Figura 2.2: Grafo indiretto

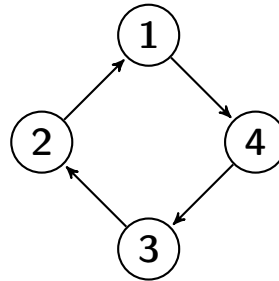


Figura 2.3: Grafo diretto

Un grafo può essere formalmente descritto come una coppia di insiemi $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, dove V è l'insieme dei nodi ed E è l'insieme degli archi. Un arco $e \in E$ è rappresentato come un sottoinsieme di due elementi di V , $e = \{u, v\}$ per $u, v \in V$.

Le rappresentazioni atte a descrivere un grafo sono molteplici:

- *Rappresentazione grafica*: ad ogni nodo corrisponde una figura circolare sul piano e ad ogni arco (i, j) corrisponde una linea che collega il nodo i al nodo j .
- *Matrice di adiacenza*: matrice di dimensione $n \times n$, dove n è il numero di nodi, il cui elemento (i, j) assume valore 1 se esiste l'arco tra il nodo i ed il nodo j , 0 altrimenti.
- *Lista di adiacenza*: ad ogni vertice v è associata la lista dei nodi ad esso vicini.

Negli anni, gli studi sulla teoria dei grafi hanno prodotto una quantità enorme di definizioni e teoremi, per cui, di seguito vengono descritti solamente i concetti necessari alla comprensione di questo lavoro di tesi.

Sottografo. Un grafo H si dice sottografo di un grafo G se i vertici di H sono un sottoinsieme dei vertici di G e gli archi di H sono un sottoinsieme degli archi di G . Siano $G = (V, E)$ ed $H = (V_1, E_1)$ due grafi. H è un sottografo di G se e solo se $V_1 \subseteq V$ ed $E_1 \subseteq E$. Un concetto particolarmente utile alla comprensione di questo lavoro è lo *spanning subgraph*: uno spanning subgraph H di un grafo G è un sottografo che contiene tutti i vertici di G , cioè $V_1 = V$.

Grado di un nodo. Il grado di un nodo v è il numero di nodi ad esso adiacenti ed è indicato con $\deg(v)$.

In un grafo diretto, si distinguono due tipi di grado:

- *in-deg*(v), il grado in ingresso del nodo v , dato dal numero di archi in cui v compare come nodo destinazione;
- *out-deg*(v), il grado in uscita del nodo v , dato dal numero di archi in cui v compare come nodo sorgente.

Cammino. Un cammino è una sequenza di nodi, in cui ogni coppia consecutiva della sequenza sia connessa da un arco. Formalmente, un cammino è una sequenza di vertici $v_0, v_1, \dots, v_n \in V$ tale che $\{v_{i-1}, v_i\} \in E, \forall 1 \leq i \leq n$. Un cammino con almeno tre vertici distinti, i cui vertici di inizio e fine coincidono, è detto *ciclo*.

Grafo connesso. Un grafo è connesso se, per ogni coppia distinta di vertici (i, j) , esiste un cammino da i a j .

2.1.1 Grafo come modello della realtà

I grafi hanno una grande utilità, in quanto consentono di astrarre le relazioni che intercorrono tra più oggetti, e di rappresentare tali relazioni in strutture su cui è possibile applicare modelli matematici. In [1] viene proposto un esempio reale: la Figura 2.4 rappresenta la struttura della rete Internet nel Dicembre del 1970, noto come ARPANET allora, composto solo da 13 macchine. I nodi rappresentano gli host, e vi è un arco tra due host se esiste una comunicazione diretta tra di essi. Come è possibile intuire, la posizione geografica dei nodi non ha molta importanza, ma quel che conta è il come ogni nodo sia connesso agli altri. Infatti la figura 2.5 mostra lo stesso grafo di ARPANET, attraverso una rappresentazione logica. Il grafo di ARPANET mostrato in precedenza è un esempio di *communication network*, i cui

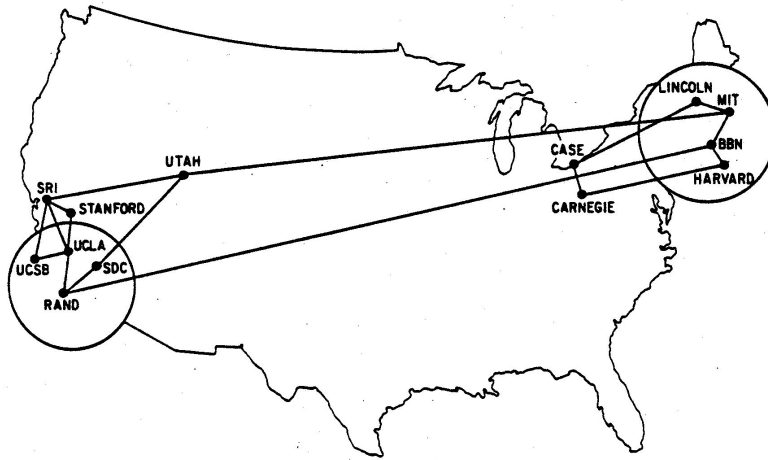


Figura 2.4: ARPANET nel Dicembre 1970

nodi sono computer o altri dispositivi capaci di inviare messaggi mentre gli archi rappresentano i collegamenti diretti lungo i quali tali messaggi possono viaggiare. Ma questo è solamente uno dei tipi di rete che possiamo avere.

Le **social network** i cui nodi sono persone o gruppi di persone, ed i cui archi rappresentano un tipo di interazione (amicizia, inimicizia, ecc.), sono reti massive che al giorno d'oggi comprendono gran parte della popolazione mondiale, come ad esempio le reti di Facebook e Twitter.

Le **information network** sono reti che rappresentano il mondo dell'informazione, i cui nodi sono le fonti di informazione e gli archi rappresentano collegamenti logici come riferimenti, citazioni identificati da hyperlink. A tale categoria appartiene il grafo del Web, così come la rete di documenti di Wikipedia.

Le **dependency network**, che descrivono le dipendenze esistenti in una collezione di oggetti. Ne sono esempi reti di dipendenze tra task, per cui sono stati sviluppati molti lavori, applicati anche in altri campi come lo studio del sistema immunitario e delle reti semantiche.

I grafi mostrano la loro grande utilità anche nelle **transportation network**, reti i cui nodi sono luoghi geografici ed i cui archi sono le linee stradali, ferroviarie o aeree che li collegano. Questo tipo di rete è stata di grande importanza nello sviluppo di concetti ed algoritmi su grafi, infatti molta della terminologia deriva dal mondo dei trasporti.

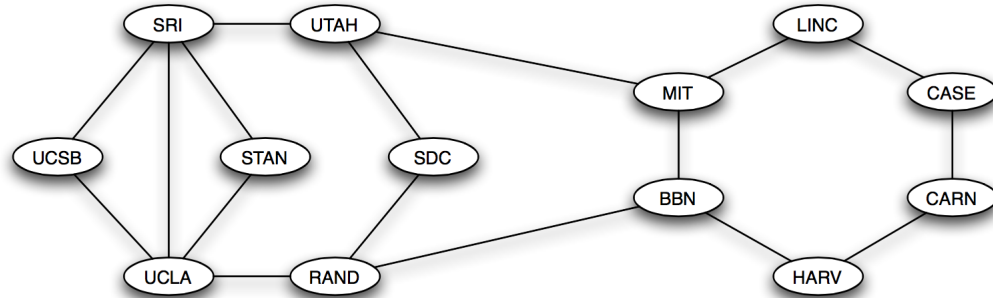


Figura 2.5: Grafo di ARPANET nel Dicembre 1970

Quindi, come si evince dai vari tipi di network descritte finora, i grafi sono uno strumento potente che trova applicazione nelle realtà più disparate, partendo dalla matematica ed informatica, fino ad arrivare agli studi della sociologia, biologia, linguistica, chimica e fisica. Ed è proprio l'ultimo ambito ad interessare questo lavoro di tesi: la *fisica statistica* è una branca della fisica che utilizza metodi della teoria delle probabilità e statistici, ed in particolare gli strumenti matematici usati per gestire grandi popolazioni ed approssimazioni, per la risoluzione di problemi fisici. Può descrivere una grande varietà di campi dotati di una natura stocastica intrinseca. Le sue applicazioni coinvolgono problemi nel campo della fisica, biologia, chimica, neurologia e scienze sociali.

Lo scopo finale per cui nasce la fisica statistica, in particolare la meccanica statistica, è quello di fornire strutture e strumenti per mettere in relazione proprietà microscopiche di atomi e molecole individuali rispetto a proprietà del materiale che possono essere osservate ad occhio nudo. Spiegando ad esempio la termodinamica come il risultato naturale di statistica, meccanica classica e meccanica quantistica al livello microscopico: il grafo riesce a modellare le particelle del sistema su cui agisce la dinamica del processo fisico.

2.2 Modello di Ising

Il modello di Ising è un classico problema combinatoriale della fisica statistica, studiato per la prima volta da Ernest Ising nel 1920. Tale modello è interessato alla fisica delle transizioni di fase, che occorrono quando un

piccolo cambiamento nei parametri, come temperatura o pressione, causa un grande cambiamento qualitativo nello stato del sistema. Le transizioni di fase sono comuni nella fisica, come ad esempio nel fenomeno del ferromagnetismo.

Uno degli scopi principali del modello di Ising è spiegare come interazioni di breve raggio tra le componenti del sistema siano in grado di dare luogo a comportamenti correlati di lungo raggio e, quindi, di predire in un certo senso il potenziale per una transizione di fase. Tale modello trova applicazione in molti ambiti, in generale in tutti quegli ambiti dove si studia il comportamento cooperativo di grandi sistemi: tali applicazioni sono possibili perché il modello di Ising può essere formulato come un problema matematico.

Una semplice definizione matematica del modello di Ising è data in [2] e in [3]: consideriamo una collezione di siti $[n] = \{0, 1, \dots, n-1\}$, in cui ogni coppia i, j ha associata un'energia di interazione V_{ij} . Nella maggior parte dei casi di interesse fisico, l'insieme E di coppie con energia di interazione non nulla forma un *regular lattice graph* $([n], E)$: detto anche *mesh graph* o *grid graph*, è un grafo la cui rappresentazione grafica sul piano forma delle tassellature regolari.

Il primo passo del modello è quello di assegnare ad ogni sito i una variabile σ_i , detta **spin**, con $i = 1, \dots, n$. Le variabili σ_i possono assumere solo due valori, $\sigma_i = \pm 1$, che si possono definire come gli *stati* dei siti.

Una **configurazione** σ del sistema è un assegnamento di spin positivi ($\sigma_i = 1$) o negativi ($\sigma_i = -1$) ad ogni sito $i \in [n]$.

L'*energia* di una configurazione è data dall'Hamiltoniana del sistema: nella fisica matematica, tale quantità governa le dinamiche del sistema. Per il modello di Ising, l'Hamiltoniana è definita sotto un'assunzione ben precisa: si assume che solo le interazioni di breve raggio e con i siti più vicini e le interazioni dei siti con un "campo esterno" contribuiscono al livello di energia del sistema. Pertanto, per ogni configurazione $\sigma = (\sigma_1, \dots, \sigma_n)$ si ha:

$$H(\sigma) = - \sum_{\{i,j\} \in E} V_{ij} \sigma_i \sigma_j - B \sum_{k \in [n]} \sigma_k \quad (2.1)$$

in cui la prima sommatoria varia su tutte le coppie di vicini nel lattice graph mentre la seconda varia su tutti i siti del sistema. V_{ij} e B sono parametri associati rispettivamente alle interazioni dei vicini ed alle interazioni con il campo esterno.

Nel caso in cui tutte le energie di interazione siano non negative, tale sistema modella il comportamento di un *ferromagnete*: una configurazione "magnetizzata" (con la maggior parte di coppie di siti vicini aventi $\sigma_i = \sigma_j$), ha un

livello di energia più basso rispetto ad una configurazione non magnetizzata. Il parametro B corrisponde alla presenza di un “campo magnetico esterno”.

2.2.1 Partition Function

La funzione di partizione è senza alcun dubbio l’oggetto di maggior rilievo introdotto dalla meccanica statistica. La si ottiene dall’esponenziazione dell’Hamiltoniana e sommando su tutte le possibili configurazioni σ , cioè su 2^n possibili assegnamenti di valori ± 1 alle n variabili $(\sigma_1, \dots, \sigma_n)$:

$$Z = Z(V_{ij}, B, \beta) = \sum_{\sigma} \exp(-\beta H(\sigma)). \quad (2.2)$$

Nell’ambito della meccanica statistica, solitamente $\beta = 1/kT$, dove k è la costante di Boltzmann e T è la temperatura ($\beta > 0$).

La partition function è utilizzata come “denominatore” nel calcolo delle probabilità. In particolare, la probabilità che il sistema si trovi in una specifica configurazione σ è data dalla formula

$$P(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z} \quad (2.3)$$

Il segno negativo apporta una maggiore probabilità agli stati con energia più bassa. Valori bassi di β , corrispondenti ad alta temperatura, tendono ad “appiattire” la distribuzione, rendendo così tutte le configurazioni equiprobabili o quasi. Alti valori di β , corrispondenti a bassa temperatura, tendono ad apportare maggiore probabilità agli stati con energia minore.

È possibile ottenere, a partire dalla partition function Z , due quantità altrettanto importanti, la *mean energy* $\epsilon = -\partial(\ln Z)/\partial\beta$ ed il *mean magnetic moment* $\mathcal{M} = \beta^{-1}\partial(\ln Z)/\partial\beta$.

2.3 Cenni di probabilità e statistica

Di seguito vengono riportati concetti e definizioni utili alla comprensione di questo lavoro di tesi (per maggiori dettagli consultare [4]).

Spazio delle probabilità. Tripla (Ω, \mathcal{F}, P) , dove Ω è lo *spazio campionario*, cioè l’insieme dei possibili esiti (o *outcome*), \mathcal{F} è l’insieme degli eventi, mentre $P : \mathcal{F} \rightarrow [0, 1]$ è una funzione che assegna probabilità ad eventi.

Variabile aleatoria. Definita su uno spazio di probabilità, esprime il valore numerico di un outcome di un evento. Una variabile aleatoria X è una funzione definita sullo spazio degli outcome Ω , ed associa ad ogni elemento di Ω un valore in \mathcal{R} . Sia X una variabile casuale che rappresenta il valore degli outcome di un certo evento e si assuma che tale evento possa avere solo un numero finito di possibili outcome. Sia Ω lo spazio campionario dell'esperimento. Una *funzione di distribuzione* per X è una funzione p di valori reali il cui dominio è Ω e che soddisfa $p(\omega) \geq 0$ e $\sum p(\omega) = 1$, con $\omega \in \Omega$. Per ogni sottoinsieme E di Ω , definiamo la *probabilità* di E come

$$P(E) = \sum_{\omega \in E} p(\omega). \quad (2.4)$$

La funzione $p(\omega)$ è la **distribuzione di probabilità** della variabile aleatoria X .

Valore atteso. Sia X una variabile casuale discreta con spazio campionario Ω e funzione di distribuzione $p(x)$. Il *valore atteso* $E(X)$ è definito come

$$E(X) = \sum_{x \in \Omega} xp(x) \quad (2.5)$$

a patto che tale sommatoria converga (altrimenti X non avrebbe valore atteso). Di solito, il valore atteso è anche chiamato *media*, ed è indicato con il simbolo μ . Alcune delle principali proprietà di cui gode il valore atteso sono riportate di seguito:

- Siano X ed Y due variabili aleatorie con valori attesi finiti, allora $E(X + Y) = E(X) + E(Y)$
- Sia X una variabile aleatoria con valore atteso finito e sia c una costante, allora $E(cX) = cE(X)$
- Siano X e Y due variabili aleatorie indipendenti, allora $E(X \cdot Y) = E(X)E(Y)$.

Varianza. Sia X una variabile casuale con valore atteso $\mu = E(X)$. Allora la *varianza* di X , indicata con $V(X)$, o con σ^2 , è pari a

$$V(X) = E((X - \mu)^2). \quad (2.6)$$

La varianza può essere anche scritta come

$$V(X) = \sum_x (x - \mu^2)p(x). \quad (2.7)$$

Una utile alternativa al calcolo della varianza è data dal seguente teorema: *sia X una variabile casuale con $E(X) = \mu^2$, allora $V(X) = E(X^2) - \mu^2$* . Per maggiori dettagli riguardanti la prova di tale teorema, consultare pagina 258 di [4].

Alcune importanti proprietà della varianza:

- Sia X una variabile casuale e c una costante, allora $V(cX) = c^2V(X)$
- Siano X ed Y due variabili casuali indipendenti, allora $V(X + Y) = V(X) + V(Y)$.

Disuguaglianza di Chebyshev. Sia X una variabile casuale con valore atteso finito μ e varianza non negativa finita σ^2 . Allora per qualsiasi numeri reale $k > 0$ si ha che

$$Pr(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}. \quad (2.8)$$

Tale disuguaglianza garantisce che in qualsiasi distribuzione di probabilità, quasi tutti i valori siano vicini alla media.

2.4 Processi Markoviani

Di seguito sono introdotti i concetti alla base delle catene di Markov (consultare [5] per maggiori dettagli).

Una catena di Markov finita è un processo che si muove tra gli elementi di un insieme finito Ω nel seguente modo: se si suppone di trovarsi in $x \in \Omega$, allora la prossima posizione è scelta in accordo ad una distribuzione di probabilità fissata $P(x, \cdot)$. Più precisamente, una sequenza di variabili casuali (X_0, X_1, \dots) è una **catena di Markov con spazio degli stati Ω e matrice di transizione \mathbf{P}** se per tutti gli $x, y \in \Omega$, per tutti i $t \geq 1$, e per tutti gli eventi $H_{t-1} = \bigcap_{s=0}^{t-1} X_s = x_s$ che soddisfano $Pr(H_{t-1} \cap X_t = x) > 0$, si ha

$$Pr(X_{t+1} = y | H_{t-1} \cap X_t = x) = Pr(X_{t+1} = y | X_t = x) = Pr(x, y). \quad (2.9)$$

L'equazione 2.9, nota come *proprietà di Markov*, sta a significare che la probabilità condizionale di procedere dallo stato x allo stato y è la stessa, non importa quale sia la lunghezza degli stati x_0, \dots, x_{t-1} che precedono lo stato corrente x .

2.4.1 Irriducibilità e periodicità

Una catena di Markov P è detta **irriducibile** se, per qualsiasi coppia di stati $x, y \in \Omega$, esiste un intero t tale che $P^t(x, y) > 0$. L'affermazione sta ad indicare che è possibile andare da uno stato in un qualsiasi altro stato utilizzando solo transizioni di probabilità positive.

Sia $\mathcal{T}(x) := \{t \geq 1 : P^t(x, x) > 0\}$ l'insieme di volte in cui è possibile per la catena ritornare nella posizione iniziale x . Il **periodo** di x è definito come il *massimo comune divisore* (\gcd) di $\mathcal{T}(x)$.

Lemma. *Se P è irriducibile, allora $\gcd(\mathcal{T}(x)) = \gcd(\mathcal{T}(y)) \forall x, y \in \Omega$.*

Data una catena irriducibile, il periodo della catena è definito come il periodo che è comune a tutti gli stati. La catena sarà detta *aperiodica* se tutti gli stati hanno periodo 1, altrimenti *periodica*.

Proposizione. *Se P è aperiodica ed irriducibile, allora vi è un intero r t.c. $P^r(x, y) > 0 \forall x, y \in \Omega$.*

Una catena di Markov finita (con spazio degli stati Ω finito), *irriducibile* ed *aperiodica* è detta **ergodica**.

2.4.2 Distribuzione stazionaria

Sia π una distribuzione di probabilità. Se tale distribuzione soddisfa

$$\pi = \pi P \quad (2.10)$$

allora π è detta *distribuzione stazionaria* della catena di Markov. tale distribuzione può essere riscritta anche per i singoli elementi della catena:

$$\pi(x) = \sum_{y \in \Omega} \pi(y) P(y, x) \forall x, y \in \Omega. \quad (2.11)$$

Il risultato ben noto, che si ottiene dalla teoria dei processi Markoviani è che se la catena in questione è ergodica, allora la distribuzione stazionaria è *unica*.

Una catena di Markov è detta *reversibile* se

$$\pi(x) P(x, y) = \pi(y) P(y, x) \forall x, y \in \Omega. \quad (2.12)$$

Quindi, data una catena (X_t) che soddisfa 2.12 ed ha una distribuzione stazionaria iniziale, allora la distribuzione di (X_0, X_1, \dots, X_n) è la stessa di

$(X_n, X_{n-1}, \dots, X_0)$.

Un risultato di grande importanza è il *teorema della convergenza*: tale teorema prova che una catena di Markov irriducibile ed aperiodica converge alla sua distribuzione stazionaria.

Prima di enunciare il problema, però, è necessaria la definizione di *total variation distance*: siano π e μ due distribuzioni di probabilità sullo stesso spazio campionario Ω , allora la total variation distance è definita come

$$\|\pi - \mu\| = \max_{A \subseteq \Omega} |\pi(A) - \mu(A)| = \frac{1}{2} \sum_{x \in \Omega} |\pi(x) - \mu(x)|. \quad (2.13)$$

Si enuncia ora, il **teorema della convergenza**:

Teorema. *Si supponga che P sia irriducibile ed aperiodica, con distribuzione stazionaria π . Allora esistono due costanti $\alpha \in (0, 1)$ e $C > 0$ t.c.*

$$\max_{x \in \Omega} \|P^t(x, \cdot) - \pi\| \leq C\alpha^t. \quad (2.14)$$

Per la dimostrazione e maggiori dettagli consultare [5] (pag. 52).

2.4.3 Catena di Markov Monte Carlo

Data una matrice di transizione irriducibile P , vi è un'unica distribuzione stazionaria π che soddisfa $\pi = \pi P$, come affermato nella sezione precedente. Si consideri ora il problema inverso: data una distribuzione di probabilità π su Ω , è possibile trovare una matrice di transizione P per cui π è la distribuzione stazionaria? [5] fornisce un esempio pratico per capire come risolvere tale problema.

Il *campione casuale* da un insieme finito Ω è definito come una selezione uniformemente casuale da Ω ; una selezione tale che ogni elemento ha la stessa probabilità $1/|\Omega|$ di essere scelto.

Sia $1, 2, \dots, q$ un insieme di *colori*. Un *proper q -coloring* di un grafo $G = (V, E)$ è un assegnamento di colori ai vertici di V , con il vincolo che nodi vicini non debbano avere lo stesso colore. Per alcune tipologie di grafo, come ad esempio gli alberi, esistono semplici metodi ricorsivi che generano una colorazione casuale; per altri grafi, invece, risulta più difficile. un approccio consiste nell'utilizzare le catene di Markov per campionare: supponiamo che (X_t) sia una catena con spazio degli stati Ω e con una distribuzione stazionaria uniforme su Ω . Grazie al teorema della convergenza enunciato in 2.14, X_t è approssimativamente uniformemente distribuita quando t è grande.

tale metodo di campionamento a partire da una data distribuzione di probabilità è chiamato **Catena di Markov Monte Carlo**. Si supponga che π sia una distribuzione di probabilità su Ω . se + possibile costruire una catena di Markov (X_t) con distribuzione stazionaria π allora, per t grande abbastanza si ha che la distribuzione di X_t è vicina a π .

2.5 Algoritmi di approssimazione

Un **α -approximation algorithm** per un dato problema è un algoritmo con tempo di esecuzione polinomiale che, per tutte le istanze del problema, produce una soluzione il cui valore è entro un fattore α del valore della soluzione ottima.

Dato un α -approximation algorithm, α è la *performance* garantita dell'algoritmo. In letteratura è spesso anche chiamato *rapporto di approssimazione* o *fattore di approssimazione*.

Un **polynomial time approximation scheme** è una famiglia di algoritmi A_ϵ , in cui vi è un algoritmo per ogni $\epsilon > 0$ tale che A_ϵ è un $(1 + \epsilon)$ -approximation algorithm (per problemi di minimizzazione) o un $(1 - \epsilon)$ -approximation algorithm (per problemi di massimizzazione).

Oltre a queste due definizioni riportate in [6], è interessante fornire le definizioni tratte dal paper di riferimento di questo lavoro di tesi [3]:

dati numeri reali non negativi a, \tilde{a}, ϵ , si dice che \tilde{a} approssima a in un range $(1 + \epsilon)$ se

$$a(1 + \epsilon)^{-1} \leq \tilde{a} \leq a(1 + \epsilon). \quad (2.15)$$

Sia f una qualsiasi funzione che mappa istanze del problema in numeri reali. Un **randomised approximation scheme** per f è un algoritmo probabilistico che, quando presentato con un'istanza x ed un numero reale $\epsilon \in (0, 1]$, restituisce un numero che, con alta probabilità, approssima $f(x)$ in un range $(1 + \epsilon)$. Il risultato di tale algoritmo, però, non deve essere ottenuto solo con *alta affidabilità*, ma anche *efficientemente*. In accordo a ciò, si definisce un approximation scheme **fully polynomial** se il suo tempo di esecuzione è polinomiale in ϵ^{-1} e nella taglia dell'istanza x del problema.

Capitolo 3

Logit Dynamics

In questo capitolo si descrive la *Logit dynamics*, introdotta per la prima volta da Blume in [7]. Si evidenziano le sue proprietà e le sue motivazioni. Maggiori dettagli in [8].

3.1 Definizione

Si consideri un gioco strategico $\mathcal{G} = ([n], S_1, \dots, S_n, u_1, \dots, u_n)$, in cui $[n] = 1, \dots, n$ è un insieme finito di giocatori, S_i è l'insieme finito di strategie per il giocatore i , $S = S_1 \times \dots \times S_n$ è l'insieme dei profili di strategie ed $u_i : S \rightarrow \mathcal{R}$ è la funzione utilità del giocatore $i \in [n]$.

La *Logit dynamics* per un gioco \mathcal{G} procede come segue, ad ogni step:

1. si sceglie un giocatore $i \in [n]$ a caso;
2. si aggiorna la strategia del giocatore i in accordo alla *logit update rule* con parametro $\beta \geq 0$ sull'insieme S_i delle sue strategie. Cioè. la dinamica sceglie una strategia $s \in S_i$ con probabilità

$$\sigma_i(s|x) = e^{\beta u_i(s, x_{-i})} / Z_i(x), \quad (3.1)$$

in cui $x = (x_1, \dots, x_n) \in S$ è il profilo di strategie corrente, $\beta \geq 0$ e

$$Z_i(x) = \sum_{z \in S_i} e^{\beta u_i(z, x_{-i})} \quad (3.2)$$

è il fattore di normalizzazione.

Il parametro β indica il *livello di razionalità* del sistema; pertanto, quando $\beta = 0$ il giocatore i sceglie la sua strategia in maniera casuale, mentre

con $\beta > 0$, la probabilità è influenzata dalle strategie che permettono un *payoff* maggiore, e per $\beta \rightarrow \infty$ il giocatore sceglie la sua *best response*. Inoltre, bisogna notare che la probabilità $\sigma_i(s|x)$ non dipende dalla strategia x_i attualmente adottata dal giocatore i . Formalmente, la dinamica può essere definita anche nel seguente modo.

Definizione 3.1.1 Sia $\mathcal{G} = ([n], \mathcal{S}, \mathcal{U})$ un gioco strategico e sia $\beta \geq 0$. La Logit dynamics per \mathcal{G} è la catena di Markov $\mathcal{M}_\beta = (X_{t \in \mathbb{N}}, S, P)$ dove $S = S_1 \times \dots \times S_n$ e

$$P(x, y) = \frac{1}{n} \cdot \begin{cases} \sigma_i(s_i|x), & \text{se } y_{-i} = x_{-i} \text{ ed } y_i \neq x_i \\ \sum_{i=1}^n \sigma_i(s_i|x), & \text{se } x = y \\ 0, & \text{altrimenti} \end{cases} \quad (3.3)$$

con $\sigma_i(s_i|x)$ definita in 3.1. Come sarà mostrato nella sezione 3.2.1, la catena di Markov descritta in 3.3 è ergodica. Pertanto, indipendentemente dallo stato di partenza, la distribuzione della catena al tempo t convergerà alla distribuzione stazionaria π man mano che t tende ad infinito.

Si consideri ora un *potential game*. Per tale gioco, la distribuzione stazionaria è la ben nota *Gibbs measure*.

Teorema 3.1.1 Se $\mathcal{G} = ([n], \mathcal{S}, \mathcal{U})$ è un *potential game* con funzione potenziale Φ , allora la catena di Markov descritta in 3.3 è reversibile e la sua distribuzione stazionaria è data dalla Gibbs measure

$$\pi(x) = \frac{1}{Z} e^{\beta \Phi(x)}, \quad (3.4)$$

con $Z = \sum_{y \in S} e^{\beta \Phi(y)}$ è la costante di normalizzazione.

3.2 Proprietà

Si analizzano ora alcune proprietà della *Logit dynamics* utili alla comprensione di questo lavoro di tesi. Per ulteriori proprietà e dettagli consultare [8].

3.2.1 Ergodicità

È semplice vedere come la catena di Markov descritta in 3.3 sia ergodica. Siano $x = (x_1, \dots, x_n)$ ed $y = (y_1, \dots, y_n)$ due profili e sia $z = (z^0, \dots, z^n)$ un cammino di profili dove $z^0 = x$, $z^n = y$ e $z^i = (y_1, \dots, y_i, x_{i+1}, \dots, x_n)$

per $i = 1, \dots, n-1$. La probabilità che la catena, iniziando in x , si trovi in y dopo n passi è

$$P^n(x, y) = P^n(z^0, z^n) \geq P^{n-1}(z^0, z^{n-1})P(z^{n-1}, z^n), \quad (3.5)$$

e, ricorsivamente

$$P^n(x, y) \geq \prod_{i=1}^n P(z^{i-1}, z^i) > 0, \quad (3.6)$$

dove l'ultima disuguaglianza deriva da 3.3.

3.2.2 Logit dynamics e Glauber dynamics

In un *potential game*, la *Logit dynamics* è equivalente alla ben nota *Glauber dynamics*.

Sia $S = S_1 \times \dots \times S_n$ uno spazio degli stati e sia μ una distribuzione di probabilità su S , allora la *Glauber dynamics* per μ procede come segue:

da un profilo $x \in S$ si sceglie un giocatore $i \in [n]$ in maniera casuale e si modifica la sua strategia in $y \in S_i$ con probabilità μ condizionata dagli altri giocatori che sono in x_{-i}

$$\mu(y|x_{-i}) = \frac{\mu(x_{-i}, y)}{\sum_{z \in S_i} \mu(x_{-i}, z)}. \quad (3.7)$$

È facile vedere come la catena di Markov definita dalla *Glauber dynamics* sia irriducibile, aperiodica e reversibile, con distribuzione stazionaria μ . Quando $\mathcal{G} = ([n], \mathcal{S}, \mathcal{U})$ è un *potential game* con funzione potenziale Φ , la *Logit dynamics* definisce la stessa catena di Markov che la *Glauber dynamics* definisce per la *Gibbs distribution* π in 3.4. In accordo a quanto appena detto, la *Logit dynamics* per un *potential game* e la *Glauber dynamics* per la *Gibbs distribution* sono due modi di guardare la stessa catena di Markov. Grazie a tale analogia, ci si può riferire alla terminologia utilizzata dai fisici per indicare le quantità coinvolte, in particolare, il parametro β è detto *inverse temperature* ed il fattore di normalizzazione Z della *Gibbs distribution* 3.4 è detto *partition function*.

3.3 Movitazioni

Nel lavoro introduttivo alla *Logit dynamics* [7], Blume sottolinea che essa evidenzia due concetti chiave del comportamento strategico: *lock-in* e *bounded rationality*.

La prima proprietà stabilisce che, una volta che un giocatore effettua una scelta, si lega ad essa per qualche tempo: assumendo che tale regola valga nella realtà, giustifica una regola per la scelta di una strategia che tiene conto solamente delle strategie attualmente giocate dagli altri partecipanti e non le strategie che sono state precedentemente scelte. Questo è esattamente ciò che la *Logit update rule* fa in 3.1.

La seconda proprietà interviene nella dinamica in due modi: nel comportamento *miopico* dei giocatori che considerano solo il guadagno attuale e non una prospettiva di guadagni futuri e l'*informazione limitata* che tali giocatori hanno. È evidente che la *Logit update rule* in 3.1 considera tali aspetti. Inoltre, poiché la dinamica descrive una catena di Markov, l'evoluzione del sistema viene trattata in maniera chiara e semplice, permettendo analisi avanzate attraverso gli strumenti forniti dalla teoria dei processi markoviani (alcuni descritti in 2.4). Non essendoci restrizioni circa la struttura dei giochi o la loro funzione utilità, tali dinamiche possono essere applicate ad ogni sistema di interesse.

Un'altra caratteristica, che rende tali dinamiche appetibili, è la forte presenza di *probabilità* e *casualità*. L'approccio probabilistico della *Logit dynamics* è motivato anche dalla volontà di modellare sistemi complessi, intrinsecamente casuali e, pertanto, descrivibili solo attraverso modelli probabilistici. La *Logit dynamics*, quindi, soddisfa una serie di proprietà interessanti che motivano il loro impiego come modello utilizzato per studiare l'evoluzione dei giochi.

3.4 Alcuni Esperimenti

In letteratura sono presenti diversi esperimenti atti a paragonare le predizioni della *Logit dynamics* con dati reali.

McKelvey e Palfrey, in [9], si focalizzano su esperimenti che coinvolgono giochi con due persone in cui vi è un unico equilibrio Nash. I dati reali sono stati raccolti da esperimenti che sono stati eseguiti in più di trent'anni.

Per ogni esperimento, è stata calcolata la stima a *maximum likelihood* del parametro β ed è stato analizzato quanto bene il modello approssimi i dati. I risultati mostrano che la *Logit dynamics* predice deviazioni sistematiche dall'equilibrio Nash. Ciononostante, gli autori fanno notare che, sebbene ci sia consistenza con il parametro β negli esperimenti fatti, vi siano alcuni aspetti dei dati reali che non trovano spiegazione in tale dinamica.

Cameter et al. [10] considerano sette giochi e stimano il parametro β utilizzando il 70% dei soggetti, ed utilizzano tale stima per predire il comporta-

mento del rimanente 30% dei soggetti. Mentre, in un altro test, utilizzano sei giochi per stimare β ed utilizzano tali stime sul settimo gioco. I risultati mostrano che ci sono giochi in cui le predizioni della *Logit dynamics* sono paragonabili ai risultati di dinamiche più forti che, però, sono intrattabili analiticamente e si suppone che non convergeranno mai in qualsiasi gioco.

Capitolo 4

Il lavoro di Jerrum e Sinclair

Questo capitolo focalizza l'attenzione sul lavoro che ha gettato le fondamenta per lo sviluppo del e su cui si basa questa tesi e tanti altri lavori. L'articolo di cui si parla è “*Polynomial-time approximation algorithms for the Ising model*”, scritto da Mark Jerrum ed Alistair Sinclair nel 1993 [3].

In breve, [3] presenta un algoritmo randomizzato che calcola la partition function di un sistema di Ising ferromagnetico qualunque, con grado di accuratezza arbitrario. Il tempo di esecuzione di tale algoritmo è polinomiale nella taglia del sistema (i.e. il numero di siti) e variabile in base ad un parametro che controlla l'accuratezza del risultato. L'algoritmo si basa su una simulazione Monte Carlo di una catena di Markov ergodica opportunamente definita. Si può già anticipare che gli stati della catena non sono, come accade di solito, configurazioni di spin di Ising, ma spanning subgraph del grafo di interazione del sistema. Le performance dell'algoritmo sono garantite da prove rigorose e poggiano sulla proprietà di *rapidly mixing* della catena di Markov: converge alla sua distribuzione d'equilibrio in un numero *polinomiale* di passi.

Si prosegue con l'analisi dettagliata dell'algoritmo proposto da Jerrum e Sinclair, descrivendo prima il modo con cui si trasforma il modello di Ising in un nuovo dominio, in cui le configurazioni sono spanning subgraph del grafo di interazione; prosegue con la costruzione di un *fully polynomial randomised approximation scheme (fpras)* per il calcolo della partition function ed infine analizzando la catena di Markov definita sulle nuove configurazioni.

Nei capitoli successivi, l'algoritmo proposto verrà identificato come “*algoritmo naive*”.

4.1 Spins world e Subgraphs world

L'obiettivo è costruire un algoritmo per il seguente problema: **Istanza:** una matrice reale simmetrica $(V_{ij} : i, j \in [n])$ delle energie di interazione, un numero reale B che rappresenta il campo esterno ed un numero reale positivo β . **Output:** la partition function

$$Z = Z(V_{ij}, B, \beta) = \sum_{\sigma} \exp(-\beta H(\sigma)), \quad (4.1)$$

Dove l'Hamiltoniana $H(\sigma)$ è data da

$$H(\sigma) = - \sum_{\{i,j\} \in E} V_{ij} \sigma_i \sigma_j - B \sum_{k \in [n]} \sigma_k \quad (4.2)$$

ed E è l'insieme di coppie non ordinate i, j con $V_{ij} \neq 0$.

Tale algoritmo tratta il caso *ferromagnetico* del modello di Ising, che è caratterizzato da energie di interazione V_{ij} non negative. È importante tenere a mente che piuttosto di calcolare *esattamente* la partition function, si preferisce approssimarla.

Una strategia, rivelatasi molto utile per problemi di questo tipo, prevede la simulazione di una catena di Markov opportuna. Un'applicazione diretta di questa strategia alla partition function di Ising procederebbe come segue: si considerino le configurazioni del sistema di Ising, cioè i 2^n possibili vettori di spin $\sigma \in -1, +1^n$, come gli stati della catena di Markov. Si scelgano poi le probabilità di transizione tra gli stati così da rendere la catena *ergodica* e quindi, nella distribuzione stazionaria, la probabilità di essere nello stato σ è $Z^{-1} \exp(-\beta H(\sigma))$. Un modo ragionevole per raggiungere ciò, spesso utilizzato, è permettere che le transizioni occorranza tra configurazioni di spin che differiscano in una sola componente, e scegliere le probabilità di transizione in accordo alla regola di Metropolis [11]. Se la catena di Markov risultante è *rapidly mixing*, cioè se converge rapidamente alla distribuzione stazionaria indipendentemente dalla scelta dello stato iniziale, allora può essere usata efficacemente per campionare le configurazioni σ da una distribuzione che è vicina alla distribuzione stazionaria. Raccogliendo sufficienti campioni di configurazioni, usando differenti valori di B e β , dovrebbe essere possibile stimare la partition function Z con buona accuratezza.

Purtroppo, la catena di Markov così descritta (lo *spin-world process*) non è *rapidly mixing*. È ben noto che i sistemi ferromagnetici di Ising esibiscono tipicamente una transizione di fase ad un certo valore del parametro β , per valori al di sopra del punto critico, il sistema si stabilizza in uno stato in cui

vi è una preponderanza di spin di uno o l'altro segno. transizioni tra gli stati con maggioranza di +1 e stati con maggioranza di -1 occorrono raramente, semplicemente perché la distribuzione stazionaria assegna un peso totale di basso valore alle configurazioni con spin bilanciati.

Il problema causato dall'assenza di *rapid mixing* nello *spin-world process* può essere aggirato simulando una catena di Markov differente: il *subgraphs-world process*. Sebbene le due catene di Markov siano strutturalmente differenti, ed inoltre, il *subgraphs-world process* non abbia alcun significato fisico, quest'ultimo ha una forte connessione con la *partition function* del modello di Ising e, cosa fondamentale nell'applicazione corrente, è *rapidly mixing*. Tale processo sarà descritto in dettaglio nella sezione 4.3.

Un sottografo si dice *spanning* se include tutti i vertici del grafo “genitore” (in generale gli *spanning subgraph* non sono connessi). Le configurazioni del *subgraph-world* sono *spanning subgraph* del grafo di interazione $([n], E)$. Per semplificare la notazione, siano

$$\lambda_{ij} = \tanh \beta V_{ij} \quad (4.3)$$

$$\mu = \tanh B \beta \quad (4.4)$$

Ad ogni configurazione $X \subseteq E$ è assegnato un *peso*, in accordo alla formula

$$w(X) = \mu^{|\text{odd}(X)|} \prod_{\{i,j\} \in X} \lambda_{ij}, \quad (4.5)$$

in cui la notazione $\text{odd}(X)$ sta ad indicare l'insieme di tutti i vertici che hanno grado dispari nel grafo X .

La partition function per il *subgraphs-world* è

$$Z' = \sum_{X \subseteq E} w(X). \quad (4.6)$$

La formula 4.6 è conosciuta come “*the high temperature expansion*”.

Una relazione interessante è quella tra le partition function Z e Z' : esse sono correlate in maniera semplice. Si definisca

$$A = (2 \cosh \beta B)^n \prod_{\{i,j\} \in E} \cosh \beta V_{i,j}, \quad (4.7)$$

si noti che A è una funzione che può essere facilmente calcolata, poiché è composta dai parametri che specificano il sistema di Ising.

Il seguente risultato classico [12] lega le due partition function:

Teorema 4.1. $Z = AZ'$

4.2 Stima della Partition Function

4.3 Analisi del subgraphs-world process

Capitolo 5

Miglioramenti apportati

5.1 Stato dell'arte

5.1.1 Miglioramenti Rinaldi

5.1.2 Esperimenti

5.2 Stima della Partition Function

5.2.1 Numero di steps

Capitolo 6

Mean Magnetic Moment

6.1 Lemma 8, Teorema 9

6.2 Approssimazione della funzione $\text{odd}(X)$

6.2.1 $\log m$ Subgraphs

6.2.2 Algoritmo L

Capitolo 7

Implementazione e testing

7.1 Implementazione

7.2 Testing

7.3 DLib Python Wrapper

Capitolo 8

Conclusioni e sviluppi futuri

Bibliografia

- [1] D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.
- [2] B. A. Cipra, “An introduction to the ising model,” *American Mathematical Monthly*, vol. 94, no. 10, pp. 937–959, 1987.
- [3] M. Jerrum and A. Sinclair, “Polynomial-time approximation algorithms for the ising model,” *SIAM Journal on computing*, vol. 22, no. 5, pp. 1087–1116, 1993.
- [4] C. M. Grinstead and J. L. Snell, *Introduction to probability*. American Mathematical Soc., 2012.
- [5] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*. American Mathematical Soc., 2009.
- [6] D. P. Williamson and D. B. Shmoys, *The design of approximation algorithms*. Cambridge university press, 2011.
- [7] L. E. Blume, “The statistical mechanics of strategic interaction,” *Games and economic behavior*, vol. 5, no. 3, pp. 387–424, 1993.
- [8] D. Ferraioli, “Logit dynamics for strategic games mixing time and metastability,” 2012.
- [9] R. D. McKelvey and T. R. Palfrey, “Quantal response equilibria for normal form games,” 1993.
- [10] T. H. Ho, C. F. Camerer, and J.-K. Chong, “Self-tuning experience weighted attraction learning in games,” *Journal of Economic Theory*, vol. 133, no. 1, pp. 177–198, 2007.
- [11] C. Gelatt, M. Vecchi, *et al.*, “Optimization by simulated annealing,” *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

- [12] G. F. Newell and E. W. Montroll, "On the theory of the ising model of ferromagnetism," *Reviews of Modern Physics*, vol. 25, no. 2, p. 353, 1953.