



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	8
2.4.1	Irriducibilità e periodicità	8
2.4.2	Distribuzione stazionaria	8
2.4.3	Catena di Markov Monte Carlo	8
2.5	Algoritmi di approssimazione	8
3	Logit Dynamics	9
3.1	Definizione	9
3.2	Proprietà	9
3.2.1	Ergodicità	9
3.2.2	Logit dynamics e Glauber dynamics	9
3.3	Movitazioni	9
3.4	Alcuni Esperimenti	9
4	Il lavoro di Jerrum e Sinclair	10
4.1	Spins world e Subgraphs world	10
4.2	Stima della Partition Function	10
4.3	Analisi del subgraphs-world process	10
5	Miglioramenti apportati	11
5.1	Stato dell'arte	11
5.1.1	Miglioramenti Rinaldi	11

<i>INDICE</i>	iii
5.1.2 Esperimenti	11
5.2 Stima della Partition Function	11
5.2.1 Numero di steps	11
6 Mean Magnetic Moment	12
6.1 Lemma 8, Teorema 9	12
6.2 Approssimazione della funzione odd(X)	12
6.2.1 logm Subgraphs	12
6.2.2 Algoritmo L	12
7 Implementazione e testing	13
7.1 Implementazione	13
7.2 Testing	13
7.3 DLib Python Wrapper	13
8 Conclusioni e sviluppi futuri	14
Bibliografia	14

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*.

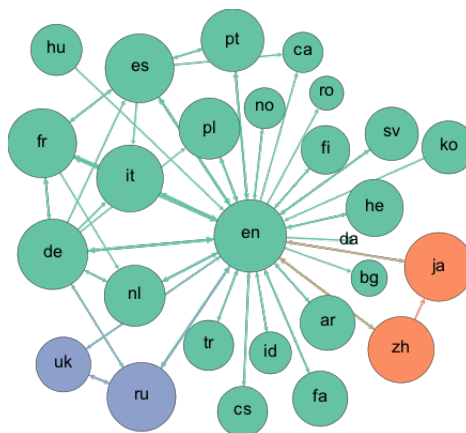


Figura 2.1: Wikipedia Multilingual Network Graph (July 2013)

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.

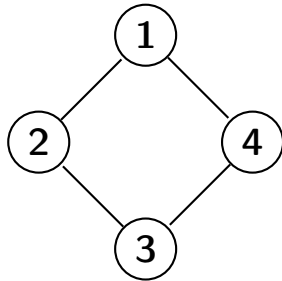


Figura 2.2: Grafo indiretto

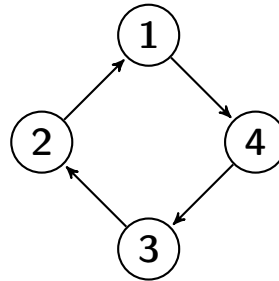


Figura 2.3: Grafo diretto

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

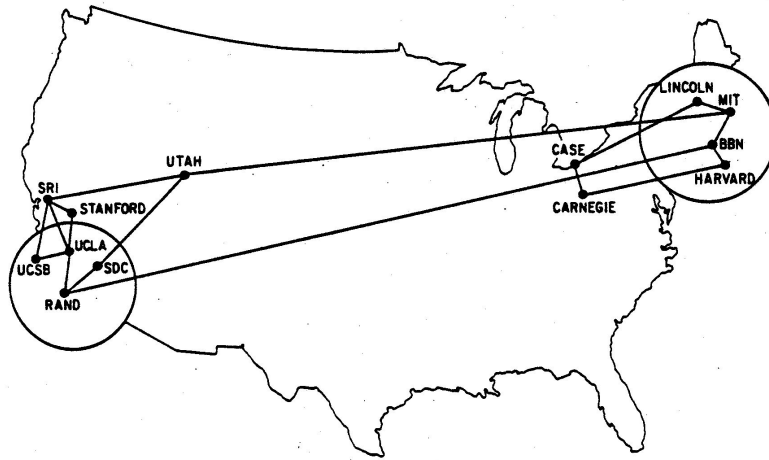


Figura 2.4: ARPANET nel Dicembre 1970

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.

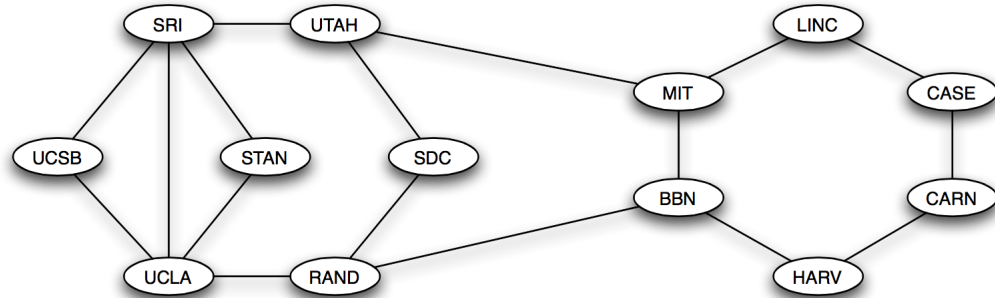


Figura 2.5: Grafo di ARPANET nel Dicembre 1970

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:

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

2.2.1 Partition Function

2.3 Cenni di probabilità e statistica

2.4 Processi Markoviani

2.4.1 Irriducibilità e periodicità

2.4.2 Distribuzione stazionaria

2.4.3 Catena di Markov Monte Carlo

2.5 Algoritmi di approssimazione

Capitolo 3

Logit Dynamics

3.1 Definizione

3.2 Proprietà

3.2.1 Ergodicità

3.2.2 Logit dynamics e Glauber dynamics

3.3 Movitazioni

3.4 Alcuni Esperimenti

Capitolo 4

Il lavoro di Jerrum e Sinclair

4.1 Spins world e Subgraphs world

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.