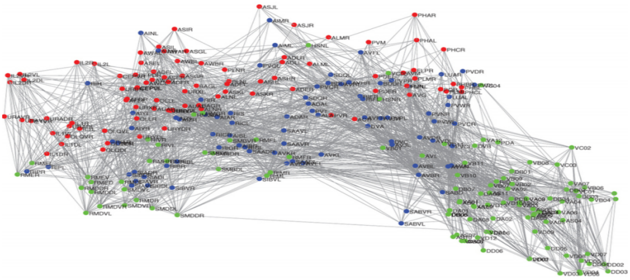
Catene di Markov

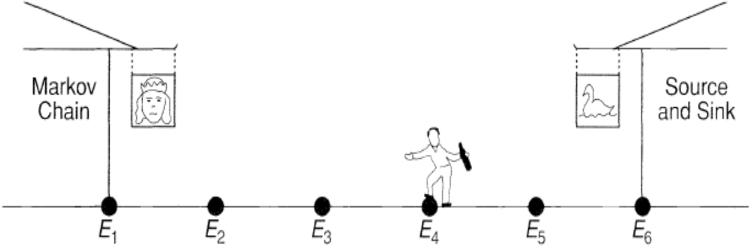
Sara Giuli

Università degli studi "Roma Tre"



Piano della presentazione

Introduzione



Mi chiedo, data la distanza tra i due pub:

1. Quale pub è privilegiato?
2. Quanto tempo ci metto ad arrivarci?

Posizione

Una posizione è descritta attraverso una variabile aleatoria $X = (0, 0, ...0, 1, 0, ..0)$

Matrice di Transizione

$p_{ij} = \mathbf{P}(\text{Mi muovo da } E_i \text{ ad } E_j \text{ in 1 passo})$

Catena di Markov

Matrice stocastica

P è una matrice stocastica se

1. $\sum_{y \in \Omega} P(x,y) = 1 \quad \forall x \in \Omega$
2. $0 \leq P(x,y) \leq 1$

Catena Di Markov

Una sequenza di variabili aleatorie $X = (X_0, X_1, \ldots)$ è una Catena di Markov con spazio degli stati Ω e matrice di incidenza P se $\forall x, y \in \Omega$

$$\mathbf{P}(X_{t+1} = y \mid X_t = x) = \mathbf{P}(x,y)$$

Dato $X = (0, ..., 0, 1, 0, ...0)$ posizione iniziale, allora XP^k è la posizione dopo k passi

Rappresentazione di una Catena

Possiamo rappresentare i passi di una catena tramite un grafo orientato in cui:

1. i vertici sono gli stati della catena
2. i lati indicano se si può passare da uno stato ad un altro in un passo della catena

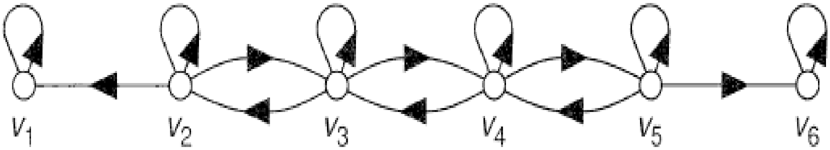


Figure: Grafo orientato associato

Un modo alternativo di rappresentare la catena di Markov è attraverso la matrice di adiacenza

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

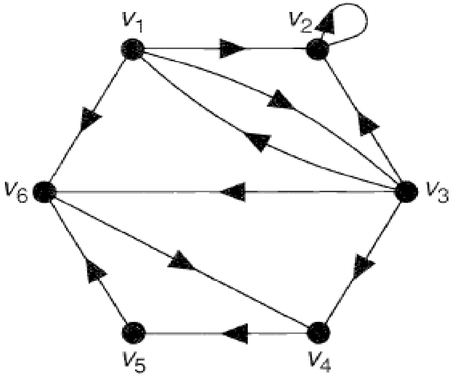


Figure: Matrice di adiacenza associata

Proprietà di una catena di Markov

Irriducibilità

Definizione

Una Catena di Markov è irriducibile se posso raggiungere qualunque stato della catena da qualunque stato iniziale

Osserviamo allora che alcune proprietà dei grafi si deduce che:

1. posso andare dallo stato E_i allo stato E_j $\Leftrightarrow$ esiste un cammino da v_i a v_j;
2. La lunghezza di tale cammino è il più piccolo tempo che impiega la catena per spostarsi tra i due stati
3. La Catena è irriducibile se e solo se il grafo orientato alla catena è strettamente connesso

Proprietà di una catena di Markov

Aperiodicità

Definizione

1. Sia $\tau(x) = \{t \geq 1 : P^t(x,x) \geq 0\}$. Allora il **periodo di una catena** è il massimo comun divisore di $\tau(x)$
2. Una catena è detta aperiodica se tutti gli stati hanno periodo 1.

Osservazione: Una catena irriducibile e aperiodica ha buone proprietà

$\Rightarrow$ si può sostituire la catena di partenza con una *"Lazy"*

Random Walks su grafi

Definizione

Dato $G = (V, E)$, siano $x \sim y$ se $\exists e = (x, y) \in E$. Allora x è vicino ad y.

RW su un grafo

Un random Walks su un grafo $G = (V, E)$ è definito dalla seguente matrice di transizione:

$$P(x, y) = \begin{cases} \frac{1}{d(x)}, & \text{se } x \sim y, \\ 0, & \text{altrimenti.} \end{cases}$$

Misura stazionaria

Data una catena di Markov (P, Ω), π è una distribuzione stazionaria su Ω per la catena se

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

1. Nell'esempio del Random Walk sul grafo avremo che

$$\sum_{x \in V(G)} d(x) P(x, y) = \sum_{x \sim y} \frac{d(x)}{d(x)} = d(y)$$

   Quindi, per ottenere una probabilità, normalizziamo per $\sum_{y \in V(G)} d(y) = 2\,|E|$ (Grazie all'Handshaking Lemma)

2. Per grafi regolari, ossia tali che $d(v) = d$, $\pi(y) = \frac{1}{|V|}$

Tempo di Mixing

1. Quanto tempo ci impiega una catena di Markov (P, Ω) ad avere una distribuzione simile a quella di una misura stazionaria?

Distanza in variazione

Date due distribuzioni di probabilità μ e ν su Ω, allora si definisce la loro distanza in variazione:

$$\|\mu - \nu\|_{TV} = \max_{A \subseteq \Omega} |\ \mu(A) - \nu(A)\ |$$

Tempo di Mixing per una catena

Sia $d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV}$; Allora definiamo il tempo di mixing come

$$t_{mix}(\varepsilon) = \min\{t \mid d(t) \leq \varepsilon\}$$

Stime per il tempo di mixing

Obiettivi principali:

1. Stimare il tempo di mixing dall'alto (ad esempio con le tecniche di *coupling*)
2. Stimare il tempo di mixing dal basso

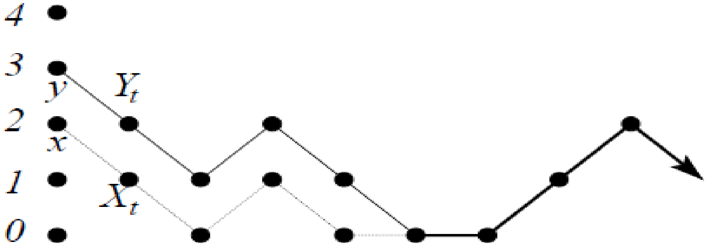


Figure: Coupling random walks su $\{0, 1, 2, 3, 4\}$

Stima con il diametro

Diametro

Data una matrice di transizione P, su Ω, si costruisca un grafo con insieme di vertici su Ω e dove $\{x, y\}$ è un lato se e solo se $P(x, y) + P(y, x) > 0$. Definiamo il diametro di tale grafo come la massima distanza tra vertici distinti.

Data P una matrice di transizione irriducibile e aperiodica su Ω, con diametro L, e siano x_0 e y_0 gli stati alla massima distanza del grafo. Allora

$$\Rightarrow P^{\lfloor (L-1)/2 \rfloor}(x_0, \cdot) \quad \text{e} \quad P^{\lfloor (L-1)/2 \rfloor}(y_0, \cdot)$$

sono positive su insiemi di vertici distinti.

Dalle osservazioni precedenti si deduce che:

Stime del t_{mix} con il diametro

$\left\| P^{\lfloor (L-1)/2 \rfloor}(x_0, \cdot) - P^{\lfloor (L-1)/2 \rfloor}(y_0, \cdot) \right\|_{TV} = 1$ e $\forall \varepsilon < \frac{1}{2}$ segue che:

$$t_{mix}(\varepsilon) \geq \frac{L}{2} \tag{1}$$

Counting bound

Definizione

Sia (X_t) una catena di Markov con matrice di transizione aperiodica e irriducibile P su Ω e supponiamo che la misura stazionaria π è uniforme su Ω. Definito $d_{out} := |\{y : P(x,y) > 0\}|$ il numero di stati accessibili in un passo da x, allora sia

$$\Delta = \max_{x \in \Omega} d_{out}(x)$$

Sia Ω_t^x l'insieme degli stati accessibili da x in esattamente t passi, allora $|\Omega_t^x| \leq \Delta^t$.

Counting bounding

Se $\Delta^t \leq (1-\varepsilon)|\Omega|$ allora per la definizione di distanza in variazione:

$$\left\|P^t(x,\cdot)-\pi\right\|_{TV} \geq P^t(x,\Omega_t^x)-\pi(\Omega_t^x) \geq 1-\frac{\Delta^t}{\Omega_t^x} > \varepsilon$$

Allora, si deduce immediatamente che:

$$t_{mix}(\varepsilon) \geq \frac{\log(\Omega(1-\varepsilon))}{\log\Delta}$$

Applicazione: Random Walk sul grafo

Per il random walk sul grafo d regolare, dato che la misura è uniforme, si ottiene che:

$$t_{mix}(\varepsilon) \geq \frac{\log(|\Omega|\,(1-\varepsilon))}{\log(d-1)}$$

Esempi

Ipercubo

Ipercubo n-dimensionale

l'ipercubo n dimensionale è un grafo i cui vertici sono i multipli binari di $\{0, 1\}^n$. Due vertici sono collegati da un lato se differiscono esattamente per una coordinata

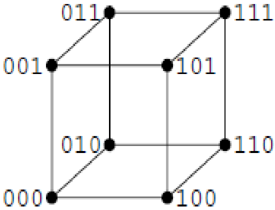


Figure: Ipercubo tridimensionale

Random Walk sull'ipercubo

Il random walk semplice sull'ipercubo si muove da un vertice $(x^1, ..., x^n)$ scegliendo una coordinata $j \in \{1, .., n\}$ uniformemente random e pone il nuovo stato uguale a $(x^1, \ldots x^{j-1}, 1 - x^j, x^{j+1}, \ldots x^n)$.

1. Il random walk sull'ipercubo è periodico
2. Il problema può essere risolto costruendo il **Lazy random walk**, che rimane nella sua posizione corrente con probabilità $\frac{1}{2}$
3. L'ipercubo è un grafo n regolare, quindi la distribuzione uniforme per la catena su tale grafo è quella uniforme su $\{0, 1\}^n$

Stima per il tempo di mixing del random walk per l'ipercupo

$\rightarrow$ Coupling $\quad\Rightarrow t_{mix} = O(n\log(n))$

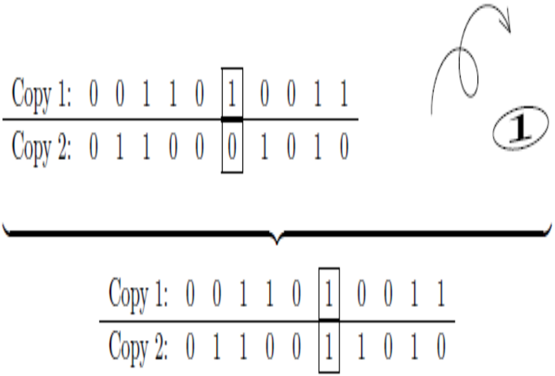


Figure: Coupling per il random walk sull'ipercubo

Expanders

Definizione

Sia $G = (V, E)$ un grafo d regolare, con $|V(G)| = n$. Siano $S, T \subset V(G)$. Sia

$$E(S, T) = \{(u, v) \mid u \in S, v \in T : (u, v) \in E\}$$

Osserviamo allora che $E(S) = \{(u, v) \mid u, v \in S, (u, v) \in E\}$.

Definizione

1. Il bordo dei lati di un insieme S, $\partial S = E(\bar{S}, S)$.
2. La **Costante di Espansione**:

$$h(G) = \min_{\{S \mid |S| \leq \frac{n}{2}\}} \frac{|\partial S|}{|S|}$$

Expanders

Una famiglia di grafi *d*regolari $\{G_n\}_{n\in\mathbb{N}}$ si chiama famiglia di $(d,\varepsilon)-$ Expanders se:

1. $\lim_{n\to\infty} |V(G_n)| = \infty$,
2. G_n è d regolare,
3. $\exists\varepsilon > 0 : h(G_n) \geq \varepsilon, \forall n$

Osservazioni:

1. $h > 0$ se e solo se G_n è connesso.
2. Se $|S| \leq \frac{n}{2}$, allora $|\partial S| \geq h|S| \Rightarrow$ se h non è piccola allora ogni sottoinsieme S con tale proprietà ha molti vicini fuori S (Da qui il nome Expanders)

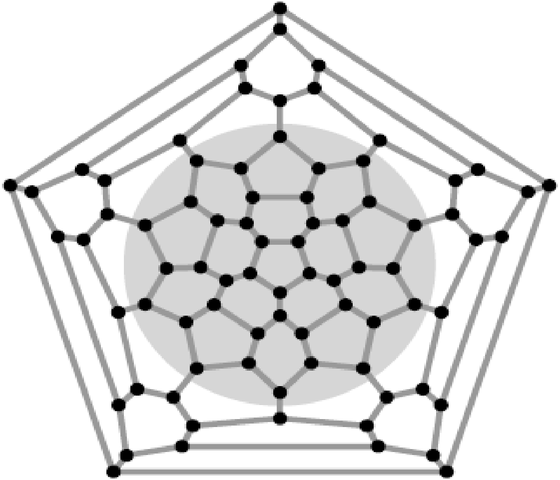


Figure: Ramanujan graph

- *Perchè studiare gli Expanders?*
  1. In matematica pura $\Rightarrow$ Costruzione esplicita di grafi con *girth* molto grande e con un elevato *numero cromatico*
  2. In **computer science** riscontrano le maggiori applicazioni.
     1. costruzione di codici con codifica e decodifica molto efficienti di algoritmi
     2. *"derandomization"* di algoritmi random
     3. disegni espliciti di reti superefficienti

- *Per quale motivo sono molto utilizzati?*

Sfruttando la proprietà che ogni membro della famiglia dei G_n è un grafo d regolare, allora si può stimare il tempo di mixing del *lazy random walk* su G_n:

Proposizione

Quando $\{G_n\}$ è una famiglia di (d, ε) expanders, allora il random walk su $\{G_n\}$ soddisfa $t_{mix} = O(\log |V(G_n)|)$

Esistenza degli expanders

Tenendo conto della stima con il diametro fatta in (1) allora la proposizione precedente ci dice che la famiglia degli expanders esibisce il tempo di mixing più veloce possibile (a meno di costanti) per famiglie di grafi con grado limitato.

Per questo motivo ci poniamo come obbiettivo di dimostrare il seguente risultato:

Proposizione

Esiste una famiglia $(3, 0.004)$ *Expanders.*

- La dimostrazione è costruttiva (Pinsker, 1973)

Dimostrazione

Construiamo inizialmente un multigrafo G_n con $h(G_n) > 0.01$, dove:

1. $V(G_n) = \{a_1, \dots a_n, b_1, \dots b_n\}$;
2. Scegliamo due permutazioni random $\sigma_1, \sigma_2 \in S_n$ tale che l'insieme dei lati è definito da:

$$E(G_n) = \left\{(a_i, b_i), (a_i, b_{\sigma_1(i)}), (a_i, b_{\sigma_2(i)}), 1 \leq i \leq n\right\}$$

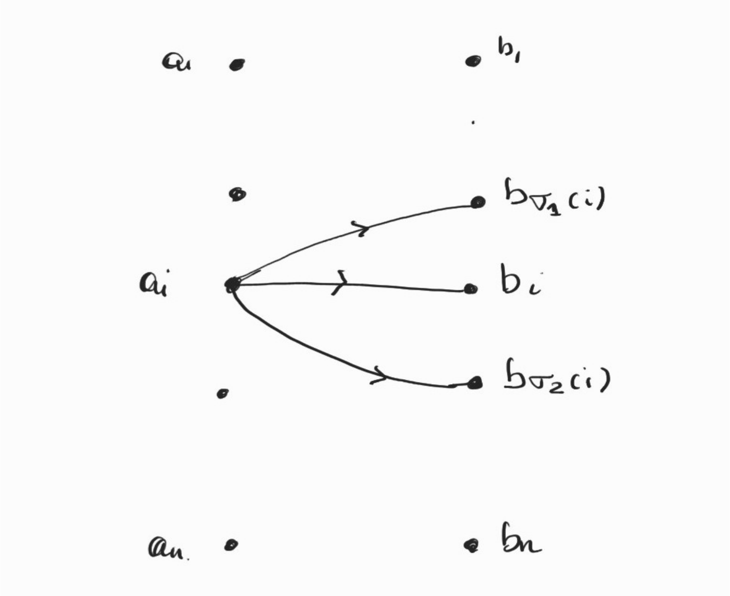


Figure: Multigrafo Random G_n

- Diamo per vero il seguente risultato:

Lemma

Per la famiglia G_n di multigrafi random descritti precedentemente si ha che:

$$P(h(G_n) \geq 0.01) \to 1 \quad n \to +\infty \tag{2}$$

Osservazione : Grazie a (2) G_n non ha tripli edge: infatti, se così non fosse, G_n avrebbe una coppia di vertici disconnessa e $h(\{a_i, b_i\}) = 0$.

$$\Rightarrow \quad P(\text{in } G_n \ \exists \text{ tripli lati}) = 0$$

Quindi il grafo G_n potrebbe ammettere solo doppi lati.

1. Studiamo la seguente probabilità

$$P(G_n \text{ abbia più di 100 doppi lati}) \tag{3}$$

2. Calcoliamo N_{tot}, il numero di doppi lati in G_n:
   1. Sia N1 il numero di doppi lati di G_n dati da (Π_1, σ_2);
   2. Sia N2 il numero di doppi lati di G_n dati da (Π_2, σ_1);
   3. Sia N3 il numero di doppi lati di G_n dati da (σ_1, σ_2);
3. Allora il numero di doppi lati totali di G_n è:

$$\mathrm{N}_{tot} \leq \mathrm{N1} + \mathrm{N2} + \mathrm{N3}$$

1. Calcoliamo $P(\mathrm{N1} > 30)$ utilizzando la disuguaglianza di *Chebychev*:

$$P(|X - \mu| \geq \lambda\sigma) \leq 1 - \frac{1}{\lambda^2}$$

2. Calcoliamo valore atteso e varianza di $N1$:

$$E(N1) = E(\sum_{i=1}^{n} 1_{\{\sigma_1(i)=i\}}) = \sum_{i=1}^{n} P(\sigma_1(i) = i) = 1$$

3. Calcolando in maniera analoga la varianza:

$$Var(\mathrm{N1}) \leq 1$$

4. Allora otteniamo, applicando la disuguaglianza di Chebychev

$$\Rightarrow \quad P(\mathrm{N1} > 30) \leq \frac{1}{29^2}$$

1. In questo modo, otteniamo una stima sulla probabilità totale:

$$\mathbf{P}(N_{tot} > 100) \leq 3\mathbf{P}(N1 > 30)$$

2. Questo ci consente di dire che, per n molto grande, si trovano al più 100 multilati (doppi lati), mentre tutti gli altri sono semplici (Ossia il numero di multilati è un numero **finito**)
3. Possiamo allora "aggiustare" i multiedge che compaiono nel grafo G_n costruendo un nuovo grafo, $\tilde{G}_n$, *semplice*, partendo da G_n e aggiungendo un vertice in presenza di ogni lato doppio

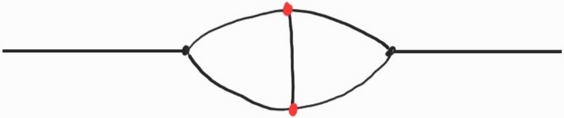


Figure: Costruzione del grafo semplice, $\tilde{G}_n$

Con questo concludiamo che:

$$\mathbf{P}(\tilde{G}_n \text{ è semplice e } h(\tilde{G}_n) > 0.004) > 0 \tag{4}$$

(4) ci permette di concludere che:

$$\Rightarrow \quad \exists \left\{ \tilde{G}_n \right\} \text{grafo semplice}$$

Ossia

$\tilde{G}_n$ è l'n esimo elemento degli expanders.

- Markov Chains and Mixing Times, David A. Levin, Yuval Peres, Second Edition
- Introduction to Graph Theory, Robin J. Wilson
- Expander Graphs and their applications, Shlomo hoory, Nathan Linial and Avi Wigderson
- What is an Expander?, Peter Sarnak, notices of AMS, 2004