Università degli Studi di Padova

DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"

Corso di Laurea Triennale in Matematica

Il Teorema di Perron-Frobenius e le Sue Applicazioni

The Perron-Frobenius Theorem and its Applications

Relatrice:

Prof.ssa Detomi Eloisa Michela

Laureando:

Giulia Visentin
Matricola:1162042

Anno Accademico 2023/2024
data ufficiale di laurea 20/09/2024

Indice

Capitolo 1

Introduzione

Il Teorema di Perron-Frobenius ricopre un ruolo fondamentale nell'ambito dell'algebra lineare applicata. Questo Teorema permette una comprensione profonda delle proprietà strutturali delle matrici positive o non-negative irriducibili ed ha applicazioni in molti ambiti scientifici e nella tecnologia. In questa tesi analizzeremo il teorema stesso e una delle principali applicazioni ai sistemi dinamici discreti.

Nel secondo capitolo ci concentreremo quindi sulla comprensione dei due teoremi (di Perron e di Frobenius) e ne daremo una dimostrazione.

Successivamente nel terzo capitolo introdurremo le catene di Markov con le loro proprietà per comprendere l'importanza di questo risultato, evidenziando come esso garantisca l'esistenza di una distribuzione stazionaria sotto alcune ipotesi. Introdurremo inoltre l'algoritmo di PageRank, processo che ha reso il motore di ricerca Google famoso, che permette di stabilire un'ordine tra le pagine Web nella rete. Esso è un'applicazione delle catene di Markov e in quanto tale converge grazie al Teorema di Perron-Frobenius. Cercheremo quindi di comprendere il processo alla base di una piattaforma come Google, che al giorno d'oggi utilizziamo tutti frequentemente, e il legame che esso ha con il Teorema in questione.

Lo scopo di questo testo è di evidenziare come concetti matematici astratti, come i teoremi dell'algebra lineare, in realtà siano il fulcro di processi, come il motore di ricerca Google, di utilizzo comune e a portata di tutti. La connessione tra teoria e pratica, oltre ad evidenziare la potenza del Teorema di Perron-Frobenius, dimostra l'importanza della matematica come disciplina che, attraverso la modellazione e l'analisi, contribuisce alla risoluzione di problemi nel mondo reale.

Capitolo 2

Il Teorema di Perron-Frobenius

2.1 Introduzione

Il Teorema di Perron-Frobenius è un risultato fondamentale nell'algebra lineare e nella teoria delle matrici, con implicazioni estese in molteplici campi della matematica applicata. Esso descrive le proprietà spettrali delle matrici non negative e positive, fornendo informazioni cruciali sugli autovalori e gli autovettori di tali matrici. Vedremo inizialmente il teorema nella sua versione originale proposta da Perron e applicata alle matrici positive, successivamente vedremo l'estensione di questo teorema alle matrici non-negative e irriducibili proposta da Frobenius.

2.2 I Teoremi

Nel 1907 Oscar Perron pubblica l'omonimo teorema. Iniziamo con la definizione di raggio spettrale, fondamentale per il resto del trattato.

Definizione 2.1. *Data A matrice si chiama **Raggio Spettrale** il massimo tra i moduli degli autovalori della matrice A e lo indicheremo come $\boldsymbol{r(A)}$.*

Enunciamo di seguito alcuni risultati che si riveleranno necessari per la dimostrazione.

Lemma 2.1. *[6] Siano A e B due matrici non-negative tali che $A - B$ è una matrice non-negativa. Allora $r(A) \geq r(B)$. In particolare, se A è una matrice non-negativa, $r(A) \geq 0$.*

Lemma 2.2. *Sia $A = (a_{ij})$ una matrice $n \times n$ positiva e sia $\lambda \in \mathbb{C}$ un autovalore di modulo massimo di A con relativo autovettore $x = (x_1, \dots, x_n)^T \in \mathbb{C}^n$. Risulta allora:*

- *il vettore $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ è positivo e $Ax = r(A)x$;*
- *esiste un numero complesso $u \in \mathbb{C}$ di modulo 1 tale che $|x| = ux$ (cioè le coordinate di x hanno tutte il medesimo argomento). [6]*

Omettiamo le dimostrazioni di questi due lemmi e passiamo direttamente all'enunciato e alla dimostrazione del teorema, fulcro di questa tesi.

Teorema 2.1. ***(Perron)*** *Sia A una matrice quadrata positiva di ordine n. Allora:*

a) $r(A)$ è positivo ed è un autovalore di A dotato di autovettore positivo;

b) $r(A)$ è strettamente dominante (ovvero $r(A) > |\lambda|$ per ogni altro autovalore λ di A);

c) $r(A)$ ha molteplicità algebrica 1. [6]

Dimostrazione. Il punto a) del teorema segue direttamente dai lemmi 2.1 e 2.2. Infatti, la non-negatività della matrice A implica per il lemma 2.1 la positività del raggio spettrale. Invece, il lemma 2.2 garantisce l'esistenza dell'autovettore relativo ad entrate positive.

Per dimostrare il punto b) invece vogliamo provare che il raggio spettrale è strettamente dominante quindi che $r(A) > \lambda\ \forall\lambda$ autovalore di A. Sia quindi λ_{max} un autovalore di A tale che $|\lambda_{max}| \geq |\lambda|\ \forall\lambda$ autovalore di A, quindi di modulo massimo, e sia x un suo autovettore associato. Poiché sono soddisfatte le ipotesi del lemma 2.2, segue che esiste un numero complesso u di modulo 1 tale che $|x| = ux$, allora $Ax = \lambda_{max}x$ può essere scritta nella seguente forma: $Aux = \lambda_{max}ux$ e in particolare $A|x| = \lambda_{max}|x|$. Ma per il primo punto del Lemma 2.2 allora si può concludere che $r(A) = \lambda_{max} \geq \lambda\ \forall\lambda$ autovalore di A e quindi $r(A)$ è strettamente dominante.
Il punto c) afferma che $r(A)$ ha molteplicità algebrica 1. Dimostriamo inizialmente che la molteplicità geometrica è uguale a 1 e successivamente utilizzeremo questo fatto per dimostrare la tesi sulla molteplicità algebrica. Siano y e z due autovettori di A associati a $r(A)$. Per il lemma 2.2, y e z hanno tutte le coordinate non nulle. Sia quindi $0 \neq c \in \mathbb{C}$ tale che cy e z abbiano la prima coordinata uguale. In tal caso si avrebbe quindi che $cy - z \neq 0$ sarebbe un autovettore di A relativo a $r(A)$ ma con la prima coordinata nulla. Questo contraddice il Lemma 2.2, quindi necessariamente concludo che $cy = z$. Se quindi z è un multiplo di y, appartiene al sottospazio generato da esso, la molteplicità geometrica di $r(A)$ è uguale a 1.
Vogliamo vedere ora che anche la molteplicità algebrica di $r(A)$ è pari a 1. Si tratta quindi di dimostrare che $r = r(A)$ è radice del polinomio caratteristico $p_A(X)$, ma non della sua derivata $p'_A(X)$. Poiché abbiamo provato che la molteplicità geometrica di r è 1 allora si ha che $\dim N(A - rI) = 1$ (ove N rappresenta il nucleo della matrice tra parentesi). Quindi la matrice $A - rI$ ha rango $n - 1$. Questo implica che $A - rI$ ha un minore di ordine $n - 1$ invertibile e quindi che la matrice aggiunta di $A - rI$ è non nulla. Tenendo conto che quindi la matrice $B = A - rI$ è singolare e ha determinante nullo si ha quindi che:

$$B \cdot \mathrm{Agg}(B) = \det(B) \cdot I_n = 0$$

Quest'uguaglianza mostra quindi che ogni colonna non nulla della matrice aggiunta di B è un autovettore di A relativo all'autovalore r. Poiché la molteplicità geometrica di r è 1 allora si ha che ogni colonna non nulla di $\mathrm{Agg}(B)$ è un multiplo dell'autovettore positivo di A associato a r quindi ha tutte le coordinate dello stesso segno. Usando lo stesso argomento sulle righe della matrice trasposta si ottiene

quindi che la matrice Agg(B) o è positiva o è negativa.
Deriviamo ora rispetto a X la seguente uguaglianza:

$$\mathrm{Agg}(A - XI_n) \cdot (A - XI_n) = p_A(X)I_n$$

Quindi:

$$\frac{d}{dX}(\mathrm{Agg}(A - XI_n)) \cdot (A - XI_n) - \mathrm{Agg}(A - XI_n)I_n = p'_A(X)I_n$$

Sostituendo r a X e moltiplicando ambo i membri a destra per l'autovettore x di A si ottiene:

$$-\mathrm{Agg}(A - rI_n)x = p'_A(r)x$$

Infatti, sostituendo r a X, il primo elemento del membro destro dell'equazione si annulla. Inoltre, per le considerazioni precedenti, si ha che la matrice al membro sinistro dell'equazione ha tutti i coefficienti dello stesso segno il che implica che il prodotto a sinistra dell'uguaglianza non si annulla, ne segue che neanche il polinomio a destra può annullarsi. Questo conclude la dimostrazione del Teorema. □

Vediamo ora il teorema proposto da Frobenius nel 1912 che permette di applicare parzialmente il teorema di Perron al caso delle matrici non negative. Questo risultato estende ulteriormente l'ambito di applicazione del teorema di Perron in quanto, come vedremo nei capitoli successivi, molti problemi di matematica applicata sono basati su matrici non-negative (matrici di transizione). Elenchiamo di seguito, prima di vedere il teorema e la relativa dimostrazione, alcuni risultati utili.

Corollario 2.1. ***(Teorema di Perron):*** *Sia $A = (a_{ij})$ una matrice non negativa $n \times n$. Allora il raggio spettrale di A, $r(A)$, è autovalore di A ed ha un autovettore associato non negativo. [6]*

Dimostrazione. Sia $\epsilon > 0$ un numero positivo arbitrario e sia $A_\epsilon = (a_{ij} + \epsilon)$. Per il Teorema di Perron, $r_\epsilon = r(A_\epsilon)$ è autovalore di A_ϵ con relativo autovettore x_ϵ (quest'ultimo è quello che viene definito autovettore di Perron, ovvero l'autovettore positivo relativo al raggio spettrale di modulo 1). La dimostrazione del corollario si conclude con il passaggio al limite per $\epsilon \to 0$. Infatti $r = \lim_{\epsilon \to 0} r_\epsilon$ è autovalore di $A = \lim_{\epsilon \to 0} A_\epsilon$ poiché gli autovalori di una matrice sono funzioni continue dei suoi coefficienti. Inoltre, poiché gli autovettori x_ϵ stanno in un sottoinsieme compatto di $\mathbb{R}^n$, esiste allora una sottosuccessione x_{ϵ_k} convergente ad un vettore non negativo x che è autovettore di A associato a r ($x \neq 0$ poiché $\|x\|_1 = 1$). Essendo quindi r_ϵ strettamente dominante per ogni ϵ, allora r risulta dominante in quanto autovalore di A, quindi $r = r(A)$. □

Proposizione 2.1. *Per una matrice non-negativa $A = a_{ij}$ valgono le seguenti disuguaglianze:*

$$\min_i \sum_j a_{ij} \leq r(A) \leq \max_i \sum_j a_{ij}$$

$$\min_j \sum_i a_{ij} \leq r(A) \leq \max_j \sum_i a_{ij}$$

[6]

Proposizione 2.2. *Sia $A = (a_{ij})$ una matrice non negativa $n \times n$. A è irriducibile se e solo se la matrice $(I_n + A)^{n-1}$ risulta positiva. [6]*

Omettiamo la dimostrazione di questi risultati e vediamo direttamente il Teorema di Frobenius e la dimostrazione.

Teorema 2.2. ***(Frobenius)*** *Sia A una matrice non-negativa irriducibile. Allora:*

a) $r(A)$ è positivo ed è un autovalore di A dotato di un autovettore positivo;

b) $r(A)$ ha molteplicità algebrica 1. [6]

Dimostrazione. Il primo punto del teorema discende direttamente dal Corollario del Teorema di Perron, il quale garantisce che $r(A)$ è un autovalore di A con relativo autovettore x non negativo. Dalle disuguaglianze della Proposizione 2.1, poichè una matrice non negativa irriducibile non ha righe nulle, si deduce inoltre che $r(A) > 0$. Per completare la dimostrazione del punto 1) resta solo da dimostrare che l'autovettore x è positivo. Da $Ax = r(A)x$ segue che $(I_n + A)x = (1 + r(A))x$ e quindi che $(I_n + A)^{n-1}x = (1 + r(A))^{n-1}x$. Ma per la Proposizione 2.2, poiché A è irriducibile per ipotesi, si ha che la matrice $(I_n + A)^{n-1}$ è positiva, pertanto a tale matrice si può applicare il teorema di Perron e dedurne che x è positivo.
Per dimostrare il secondo punto del teorema è necessario osservare che $r(A)$ è autovalore per A con la stessa molteplicità algebrica m per cui $1+r(A)$ è autovalore per $I_n + A$. Inoltre, m è minore o uguale alla molteplicità algebrica per cui $(1 + r(A))^{n-1}$ è autovalore di $(I_n+A)^{n-1}$. Poiché a quest'ultima matrice si può applicare il Teorema di Perron, concludo quindi che $m = 1$. □

Unificando i due teoremi esposti in questo capitolo si ottiene il famoso Teorema di Perron-Frobenius.

Capitolo 3

Le Catene di Markov e il Teorema di Perron-Frobenius

3.1 Introduzione

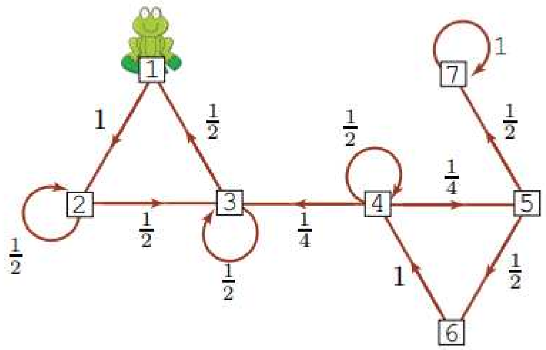
Le catene di Markov rappresentano uno degli strumenti più potenti della matematica applicata. Introdotte da Andrey Markov all'inizio del XX secolo, le catene di stati offrono un modello per descrivere sistemi che evolvono nel tempo in modo probabilistico, dove la transizione di un oggetto da uno stato all'altro può dipendere da vari fattori come ad esempio il suo stato attuale, dal tempo, da alcuni degli stati precedenti dell'oggetto e dagli stati attuali e/o precedenti degli altri oggetti nel sistema. Ci concentreremo principalmente sul caso in cui la probabilità che un oggetto passi da uno stato ad un altro dipende solo dallo stato corrente. [5]

In questo capitolo, esploreremo le fondamenta teoriche delle catene di Markov, definendo i concetti chiave e discutendo le loro principali proprietà. Successivamente, ci concentreremo sul Teorema di Perron-Frobenius che gioca un ruolo fondamentale nello studio delle catene di Markov, garantendo l'esistenza e l'unicità di distribuzioni stazionarie per matrici stocastiche irriducibili.

L'obiettivo di questo capitolo è fornire una comprensione approfondita delle catene di Markov e delle condizioni sotto le quali esse convergono a una distribuzione stazionaria. Tale comprensione è essenziale per poter apprezzare appieno l'efficacia di algoritmi come il PageRank, che vedremo alla fine di questo capitolo e che sfruttano le proprietà delle catene di Markov per ordinare in modo efficiente grandi insiemi di dati, come le pagine web nel contesto del World Wide Web.

Per comprendere al meglio la funzionalità delle catene di Markov forniremo un esempio molto semplice. Chiaramente le applicazioni sono molto più ampie e più complesse di quelle riportate nell'esempio seguente.

Esempio 3.1. *[5] Supponiamo che una rana possa saltare sopra a 7 foglie differenti come in figura. I numeri vicino alle frecce indicano la probabilità con la quale la rana, al salto successivo, si muova in direzione della freccia.*



Il sistema è quindi composto da 7 stati. Vediamo come costruire la matrice di transizione di tale catena. Ad ogni riga e ad ogni colonna facciamo corrispondere uno stato, in ogni entrata della matrice (i, j) *corrisponderà la probabilità che la rana partendo dallo stato i si muova verso lo stato j al passo successivo. Quindi riassumendo i valori nel grafico in figura si ottiene la matrice:*

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

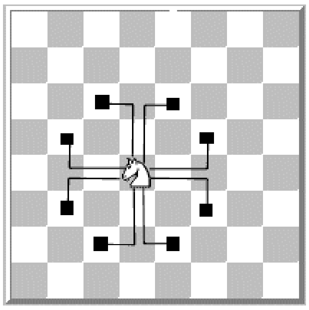
Le domande che quindi ci poniamo sono:

1. *Qual è la probabilità che partendo dallo stato 1 dopo tre salti la rana si trovi nello stato 3?*

2. *Qual è la probabilità che partendo dallo stato 1 si raggiunga lo stato 7 in un numero di iterazioni arbitrario?*

Le risposte alle seguenti domande sono semplici. Per rispondere alla prima domanda è sufficiente reiterare la matrice tante volte quanti sono i salti da eseguire e nell'entrata $(1, 3)$ *si potrà vedere la risposta. Per la seconda domanda invece è necessario valutare le potenze della matrice finchè il termine di posto* $(1, 7)$ *non diventa diverso da zero.*

Esempio 3.2. *[5] Supponiamo di muoverci con un cavallo su una scacchiera in maniera casuale partendo da una casella vicina al centro. Supponiamo che vi sia una distribuzione di probabilità uniforme nella scelta della posizione successiva. Quindi dalla sua posizione di partenza il cavallo ha la probabilità uguale a* $\frac{1}{8}$ *di muoversi verso una qualsiasi delle 8 caselle possibili. In questo esempio lo spazio degli stati ha 64 elementi.*



Quante mosse sono necessarie affinchè il cavallo torni nella sua posizione di partenza compiendo mosse casuali con identica probabilità? Lo studio delle catene di Markov è utile anche per rispondere a domande come questa.

3.2 Le Basi Teoriche

Definiamo innanzitutto l'ambiente in cui si studiano le Catene di Markov con una serie di definizioni generali e successivamente diamo la definizione di Catene di Markov. Lavoriamo nello spazio di probabilità $(\Omega, \mathcal{F}, P)$ ove Ω è un insieme di esiti possibili, $\mathcal{F}$ è una sigma-algebra di sottoinsiemi di Ω e per $A \in \mathcal{F}$, $P(A)$ è la probabilità di A.

In questo capitolo esaminiamo le sequenze di variabili aleatorie $\{X_n\}_{n\geq 0}$, ove ciascuna variabile X_n assume valori in uno spazio discreto di stati I. Gli stati X_n descrivono l'evoluzione di un sistema nel tempo.

Definizione 3.1. *Sia I un insieme numerabile* $I = \{i, j, k, \ldots\}$ *ogni* $i \in I$ *è chiamato* ***Stato*** *(o configurazione) e I è lo* ***Spazio degli Stati****. [5]*

Le variabili $X_0, X_1, \ldots$ assumono valori in I, e quindi sono delle funzioni $X : \Omega \longrightarrow I$.

Definizione 3.2. *Un vettore riga* $(\lambda_i : i \in I)$ *è chiamato* ***Misura*** *se* $\lambda_i \geq 0$ *per ogni i. Se* $\sum_i \lambda_i = 1$ *allora* λ *è chiamato* ***Misura di probabilità*** *(o Distribuzione). [5]*

Consideriamo una distribuzione iniziale λ^0 su I: $\{\lambda_i^0 : i \in I\}$ tale che $0 \leq \lambda_i \leq 1$ per ogni i e tale che $\sum_i \lambda_i^0 = 1$. Ogni componente di questo vettore rappresenta la probabilità che il sistema si trovi in una data configurazione iniziale i al tempo $t = 0$. Come definito nell'esempio 3.1 allo scopo di studiare l'evoluzione delle configurazioni del sistema nel tempo è necessaria anche una matrice che descriva le probabilità di passare da una configurazione alla successiva.

Definizione 3.3. *Chiamiamo* ***Matrice di Transizione*** *la matrice* $P = (p_{ij} : i, j \in I)$ *con* $p_{ij} \geq 0$ *per ogni* i, j *ove ogni entrata descrive la probabilità di passare dallo stato i allo stato j con* $i, j \in I$*. [5]*

Definizione 3.4. *Una matrice P si dice* ***Matrice Stocastica*** *se* $p_{ij} \geq 0$ *per ogni* $i, j \in I$ *e* $\sum_{j\in I} p_{ij} = 1$ *(ovvero se ogni riga di P è una misura di probabilità su I) [5]*

Ora abbiamo quindi il contesto e gli strumenti necessari per definire le **Catene di Markov**:

Definizione 3.5. *Chiamiamo* $(X_n)_{n\geq 0}$ ***Catena di Markov*** *con distribuzione iniziale* λ^0 *e matrice di transizione* P *se per ogni* $n \geq 0$ *e per ogni* $i_0, \ldots, i_{n+1} \in I$,

a) $P(X_0 = i_0) = \lambda^0_{i_0}$;

b) $P(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n) = p_{i_n i_{n+1}}$

Per abbreviare possiamo quindi scrivere che $(X_n)_{n\geq 0}$ *è* ***Markov***(λ^0, P). *[5]*

La prima caratteristica nella definizione è coerente con la definizione di distribuzione iniziale, infatti afferma che la probabilità che il sistema si trovi nello stato iniziale i_0 è esattamente la prima componente del vettore λ^0. La seconda caratteristica delle catene di Markov può essere interpretata come indipendenza della transizione tra stati dagli stati precedenti e da altri elementi. Sostanzialmente la probabilità che al passo $n+1$ il sistema si trovi nella configurazione i_{n+1} dipende solo dallo stato in cui si trovava il sistema al passo n, non dal tempo, nè dagli stati precedenti. Diamo in seguito l'enunciato e la dimostrazione di un teorema che caraterizza le catene di Markov. É quindi sufficiente che una successione di stati $(X_n)_{n\geq 0}$ soddisfi la condizione necessaria e sufficiente per concludere che è $Markov(\lambda, P)$.

Teorema 3.1. $(X_n)_{n\geq 0}$ *è* $Markov(\lambda, P)$ *se e solo se per ogni* $n \geq 0$ *e ogni* $i_0, i_1, \ldots, i_n \in I$, *vale*

$$P(X_0 = i_0, \ldots, X_n = i_n) = \lambda^0_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n} \tag{3.1}$$

ove $\lambda^0 = (\lambda^0_{i_0}, \ldots, \lambda^0_{i_n})$ *è la distribuzione iniziale. [5]*

Dimostrazione. Supponiamo inizialmente che $(X_n)_{n\geq 0}$ sia $Markov(\lambda, P)$. Allora

$$P(X_0 = i_0, \ldots, X_n = i_n) =$$
$$P(X_n = i_n | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) P(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) =$$

Per la definizione di probabilità condizionata.

$$= P(X_0 = i_0) P(X_1 = i_1 | X_0 = i_0) \ldots P(X_n = i_n | X_0 = i_1, \ldots, X_{n-1} = i_{n-1})$$

$$= \lambda^0_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n}$$

Ove quest'ultima uguaglianza vale per la seconda proprietà di catena di Markov
.
Viceversa partendo dall'ipotesi che l'uguaglianza (3.1) sia valida per tutte le sequenze di stati $i_0, \ldots, i_n$, si deve dimostrare che quest'uguaglianza implica che $(X_n)_{n\geq 0}$ soddisfa le proprietà delle catene di Markov. Per verificare la prima condizione, sommiamo l'uguaglianza $P(X_0 = i, \ldots, X_n = i_n) = \lambda^0_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n}$ su tutti i possibili stati $i_0, \ldots, i_n$ iniziali.

$$P(X_1 = i_1, \ldots, X_n = i_n) = \sum_i \lambda^0_i p_{i_0 i_1} \ldots p_{i_{n-1} i_n} =$$

ove λ_i^0 sono tutte le componenti del vettore λ^0. Poichè $p_{i_0 i_1} \dots p_{i_{n-1} i_n}$ non dipendono da i :

$$p_{i_0 i_1} \dots p_{i_{n-1} i_n} \sum_i \lambda_i^0 = p_{i_0 i_1} \dots p_{i_{n-1} i_n}$$

Confrontiamo la probabilità con cui $X_1 = i_1, \dots, X_n = i_n$ ottenuta partendo da qualsiasi stato iniziale X_0 con la 3.1 dimostrando quindi che la probabilità che $X_0 = i_0$ è proprio la componente $\lambda_{i_0}^0$.

Dobbiamo ora dimostrare che anche la condizione b) è verificata , quindi si osserva che la somma:

$$\sum_i \lambda_i^0 p_{i_0 i_1} = P(X_1 = i_1)$$

Questo è il risultato della somma totale su tutti gli stati iniziali i_0. Quindi si ottiene che

$$P(X_1 = i_1, \dots, X_n = i_n) = P(X_1 = i_1) p_{i_1 i_2} \dots p_{i_{n-1} i_n}$$

Analogamente continuando a sommare su tutti i possibili $i_1, i_2, \dots$ si ha che :

$$P(X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p_{i_{n-1} i_n}$$

Questo quindi implica che la probabilità condizionata $P(X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0)$ dipende solo dallo stato immediatamente precedente X_{n-1} e non dagli altri precedenti. Questa è esattamente la condizione b) della definizione.

□

Un'altra proprietà fondamentale delle catene di Markov è descritta nel lemma seguente.

Lemma 3.1. *Sia* $(X_n)_{n \geq 0}$ $Markov(\lambda, P)$ *allora per ogni k intero positivo si ha che:*

$$\lambda^k = \lambda^{k-1} P = \lambda^0 P^k$$

ove λ^0 *è il vettore riga che rappresenta la distribuzione di probabilità iniziale e* λ^k *è il vettore riga la cui n-esima entrata è la misura di probabilità che il processo si trovi nello stato* i_n *dopo k iterazioni. [1]*

Dimostrazione. Sia k un intero positivo. Dato che la probabilità che il sistema si muova da una configurazione i ad una configurazione i_n dipende solo dalle configurazioni e non da altri elementi, segue che la probabilità che il processo si trovi in uno stato i_m, dopo $k - 1$ iterazioni, e che da tale stato passi allo stato successivo i_n , è

$$\lambda^k = \lambda^{k-1} P.$$

La conclusione discende direttamente dalla definizione di catene di Markov e in particolare dalla seconda proprietà nella definizione. Infatti la probabilità che il sistema si trovi in una configurazione i_n dopo k passi è il prodotto delle probabilità precedenti:

$$\lambda^k = \lambda^0 P^k$$

Per dimostrare il teorema si può seguire anche il ragionamento inverso. La distribuzione di probabilità che il sistema assuma una certa configurazione i_1 dopo il primo passo è la distribuzione di probabilità della variabile X_1 ovvero : $\lambda^1 = \lambda^0 P$. Procedendo fino al passo k si ha che la distribuzione di probabilità al passo k è data da:

$$\lambda^k = \lambda^{k-1} P = \lambda^{k-2} PP = \ldots = \lambda^0 P^k$$

□

Nello studio delle catene di Markov uno degli interessi fondamentali è quello di riuscire a calcolare le distribuzioni di probabilità dopo k passi, il che non è impossibile manualmente, come afferma il teorema precedente, ma può essere molto dispendioso senza il supporto di un calcolatore. Quello che invece risulta spesso più semplice è il calcolo delle probabilità invarianti.

Definizione 3.6. *Sia $(X_n)_{n\geq 0}$ una $Markov(\lambda, P)$. Una distribuzione di probabilità $\pi = (\pi_1, \ldots, \pi_n)$ è detta* ***invariante o stazionaria*** *per la catena di Markov se soddisfa la seguente equazione:*

$$\pi P = \pi$$

Esplicitamente la condizione può essere espressa anche come:

$$\lambda_j = \sum_{i=1}^{n} \lambda_i p_{ij}$$

ove λ_j è la probabilità che la catena si trovi nello stato j e p_{ij} è la probabilità di transizione dallo stato i allo stato j.

Sostanzialmente la probabilità invariante rappresenta una situazione di "equilibrio" per la catena di Markov. Infatti se il processo ha distribuzione iniziale invariante, ad ogni passo k il sistema presenterà sempre la stessa probabilità di trovarsi in una data configurazione. In questo caso infatti il processo stocastico $(X_n)_{n\geq 0}$ è *stazionario in senso stretto* .

Chiameremo in seguito $\lambda^{(t)} = (\lambda^{(t)}_{i_0}, \ldots, \lambda^{(t)}_{i_n})$ la distribuzione di probabilità del sistema al passo t. La componente k-esima, $\lambda^{(t)}_{i_k}$, di questo vettore quindi rappresenta la probabilità che al passo t il sistema assuma la configurazione $i_k \in I$.

Il principale interesse per le probabilità invarianti deriva dallo studio del limite di una generica distribuzione $\lambda^{(t)}$ al tendere di t all'infinito. Supponeniamo quindi che il limite esista per ogni componente ed indichiamolo con π_{i_j}.

$$\pi_{i_j} = lim_{t \longrightarrow \infty} \lambda^{(t)}_{i_j}$$

Poichè per il Lemma 3.1 si ha che $\lambda^{(t)} = \lambda^{(0)} P^t$ calcolando il limite per $t \to \infty$ si ottiene che $\pi = \pi P$. Quindi π è una distribuzione di probabilità invariante. Abbiamo quindi dimostrato il seguente lemma:

Lemma 3.2. *Se la probabilità $\lambda^{(t)}$ al passo t-esimo converge ad un vettore di probabilità π, allora π è un vettore di probabilità invariante .*

Prima di studiare i teoremi sulla convergenza delle catene di Markov ad una distribuzione stazionaria, vediamo una classificazione generale degli stati del sistema e delle catene di Markov. Questo aiuterà a chiarire gli argomenti della prossima sezione.

3.2.1 Stati del sistema e Classificazione delle Catene di Markov

Poichè il nostro scopo è quello di indagare le applicazioni del teorema di Perron-Frobenius alla teoria delle catene di Markov saranno utili alcune definizioni e caratterizzazioni degli stati del sistema.

Definizione 3.7. *Dati due stati del sistema $i, j \in I$, si dice che lo stato i comunica con lo stato j se esiste un $k > 0$ tale che $p_{ij}^{(k)} > 0$. Questo si indica con $i \to j$. Se i non comunica con j si scrive invece $i \nrightarrow j$. Due stati i, j si dicono comunicanti se i comunica con j e j comunica con i. [5]*

Sostanzialmente due stati si dicono comunicanti se vi è una possibilità con un numero k di passi che il sistema passi da uno stato all'altro.

Questa relazione induce una relazione di equivalenza nello spazio degli stati I. Indicando quindi con I lo spazio degli stati di una catena, si può dare la seguente definizione.

Definizione 3.8. *Un sottoinsieme $S \subset I$ è detta una* ***classe chiusa*** *se gli stati di S non comunicano con quelli del complementare S^c . Una classe S si dice inoltre* ***irriducibile*** *se tutti i suoi stati comunicano tra loro. [5]*

Definizione 3.9. *Uno stato $i \in I$ si dice* ***transitorio*** *se la catena che parte dallo stato i ha probabilità nulla di tornare in quello stato. Si dice invece che uno stato $i \in I$ è* ***ricorrente*** *se la probabilità, che partendo dallo stato i si ritorni in esso, vale 1. [5]*

Diamo in seguito alcune proprietà degli stati di cui omettiamo le dimostrazioni.

Proposizione 3.1. *Se per uno stato vale la seguente condizione:*

$$\exists j \text{ tale che } i \to j \text{ ma } j \nrightarrow i$$

allora lo stato i è transitorio. [5]

Proposizione 3.2. *Una catena finita ha almeno uno stato ricorrente [5]*

Proposizione 3.3. *Se uno stato i è ricorrente e $i \to j$ allora anche lo stato j è ricorrente. [5]*

Proposizione 3.4. *Se una catena è irriducibile allora gli stati o sono tutti ricorrenti o tutti transitori. [5]*

Proposizione 3.5. *Se una catena è finita e irriducibile allora tutti i suoi stati sono ricorrenti. [5]*

Alle proprietà degli stati corrispondono delle proprietà omonime per le classi di stati definite dalla relazione di equivalenza.

Definizione 3.10. *Si dice che una classe $S_1 \subset I$ **ha accesso** ad una classe $S_2 \subset I$ se $s_{1_i} \rightarrow s_{2_j}$ per qualche $s_{1_i} \in S_1$ e qualche $s_{2_j} \in S_2$. Una classe S si chiama **finale** se non ha accesso ad altre classi. Se la classe finale contiene un solo stato allora lo stato si chiama **assorbente**. [5]*

Definizione 3.11. *Una classe S si chiama **transitoria** se un suo stato è transitorio, e si chiama **ricorrente** altrimenti. [5]*

Proposizione 3.6. *Una classe è ricorrente se e solo se è finale. [5]*

Chiaramente vi sono delle proprietà corrispondenti anche per le catene di Markov. Vediamone alcune.

Definizione 3.12. *Una catena di Markov si dice **ricorrente** se consiste in un'unica classe ricorrente di stati. Si chiama **regolare** se per un fissato passo k ogni stato della catena si può muovere ad un altro stato della catena con probabilità positiva, e si dice **periodica** se è ricorrente ma non regolare. [5]*

Definizione 3.13. *Una catena di Markov nella quale ogni classe ricorrente è composta da un solo elemento assorbente si chiama catena **assorbente**. [1]*

Teorema 3.2. *Sia P la matrice di transizione di una catena di Markov finita . Allora la catena è:*

1. ***ricorrente*** *se e solo se P è irriducibile.*
2. ***regolare*** *se e solo se P è primitiva.*
3. ***periodica*** *se e solo se P è irriducibile e ciclica. [1]*

Dimostrazione.

1. Poichè trattiamo catene finite, allora per il Teorema 3.5 tutti gli stati della catena sono ricorrenti, quindi la catena è ricorrente.
2. Per la caratterizzazione delle matrici primitive , ovvero le matrici tali che $\exists k$ tale che $A^k > 0$, allora è chiaro che se una matrice è primitiva allora la sua potenza A^k avrà un grafo fortemente connesso. Poichè i nodi del grafo sono gli stati del sistema, allora $\exists k$ tale per cui gli stati del sistema dopo il passo k sono tutti connessi e quindi la classe è regolare. Viceversa analogo.
3. Ricordiamo che una matrice, di ordine n, si dice ciclica se ha h autovalori di modulo $\rho(A)$, ove $\rho(A)$ è il raggio spettrale di A. In questo caso il grafo associato contiene un ciclo di lunghezza h, il che implica che, partendo da un elemento $i \in I$, è possibile ritornare in i esattamente dopo h passi. Poichè la matrice di transizione P è irriducibile, la classe degli stati della catena è ricorrente. Questo significa che il grafo associato alla catena di Markov è costituito da un unico ciclo. Di conseguenza la catena è periodica con periodo h. Viceversa analogo.

□

Ora che abbiamo analizzato e classificato le catene di markov e gli stati, possiamo indagare il legame tra il teoorema di Perron-Frobenius e le catene di Markov.

3.3 Le applicazioni del Teorema

L'interesse principale dello studio di una catena di Markov risiede nella comprensione del comportamento asintotico della catena, ovvero se essa tende a raggiungere uno stato di equilibrio a lungo termine. Questo stato di equilibrio, se esiste, è rappresentato da una distribuzione stazionaria. Determinare se una catena di Markov converge a questa distribuzione stazionaria consente di prevedere la distribuzione degli stati della catena nel tempo, indipendentemente dalla configurazione iniziale, garantendo una visione a lungo termine e a tempo indefinito delle possibili configurazioni degli stati. Questo è l'ambito in cui entra in gioco il teorema di nostro interesse.

3.3.1 Esistenza di una distribuzione stazionaria

Teorema 3.3. *Tutte le catene finite di Markov hanno un vettore di distribuzione di probabilità stazionaria. [1]*

Dimostrazione. Sia P la matrice di transizione associata alla catena di Markov $(X_n)_{n\geq 0}$. La matrice P ha per definizione alcune proprietà:

1. Essere una matrice non negativa, dato che le entrate della matrice sono tutte comprese tra 0 e 1.

2. Essere una matrice stocastica

Supponiamo che la matrice P sia una matrice irriducibile, allora per il teorema di Perron-Frobenius esiste un autovalore $\lambda = 1$ per P che è reale positivo e semplice. Inoltre il teorema garantisce anche che a questo autovalore sia associato un autovettore unico con tutte le componenti non negative che può essere normalizzato in modo da sommare a 1, costituendo così un vettore di probabilità π. L'autovettore normalizzato è una distribuzione stazionaria per definizione per tutta la catena in quanto è autovettore di autovalore 1 e quindi tale che $\pi P = \pi$.
Se la matrice P non fosse riducibile lo stesso processo può essere applicato ad ogni blocco di P irriducibile, ottenendo comunque lo stesso risultato, ovvero che esiste un autovettore di autovalore 1 che se normalizzato costituisce un vettore di distribuzione di probabilità. □

Da questa dimostrazione si comprende l'importanza del teorema in quanto garantisce l'esistenza dell' "equilibrio asintotico" a cui tende la catena di Markov. Questo teorema garantisce che la distribuzione stazionaria funge da attrattore asintotico, verso cui la distribuzione della catena converge nel lungo periodo.

3.3.2 Unicità della distribuzione stazionaria

Vedremo, nel seguente teorema, come sia sempre il Teorema di Perron-Frobenius a garantire l'unicità della distribuzione stazionaria.

Teorema 3.4. *Una matrice stocastica P è la matrice di transizione di una catena di Markov ricorrente se e solo se :*

1. *1 è autovalore semplice di P*
2. *Esiste un vettore riga $x \gg 0$ a entrate positive unico, a meno di moltiplicazione per scalari, tale che :*

$$xP = x$$

[1]

Dimostrazione. Supponiamo inizialmente che P sia la matrice di transizione di una catena di Markov ricorrente. Dobbiamo dimostrare i punti 1 e 2 del teorema. La prima parte della dimostrazione è banalmente garantita dal teorema sull'esistenza della distribuzione stazionaria 3.3. Resta da dimostrare solo che il vettore riga x è unico a meno di moltiplicazione per scalari. L'unicità di questo autovettore è garantita dal teorema di Perron-Frobenius in quanto se il raggio spettrale è autovalore semplice, quindi di molteplicità geometrica 1, è garantita l'unicità dell'autovettore relativo. Viceversa invece se la matrice P rispetta le condizioni 1 e 2 allora il vettore x rappresenta una distribuzione di probabilità stazionaria della catena di Markov. Affermare che esiste una distribuzione stazionaria per la catena di Markov, significa che ad ogni stato della catena è assegnata una probabilità positiva, questo è possibile solo se tutti gli stati della catena sono comunicanti tra loro. Quindi il grafo associato alla catena di Markov è fortemente connesso e la matrice P è irriducibile. Allora, sapendo che P è irriducibile e stocastica, per il teorema 3.2 posso concludere che la catena di Markov è ricorrente. □

Corollario 3.1. *Una catena di Markov è* ***ricorrente*** *se e solo se ha un unico vettore positivo di distribuzione di probabilità stazionaria. [1]*

Riassumiamo quindi nel teorema successivo le conclusioni che il teorema di Perron-Frobenius permette di trarre nello studio delle catene di Markov. Poichè come sottolineato in precedenza uno degli aspetti di maggior interesse dello studio di tali processi è l'andamento a lungo termine degli stessi, il teorema ricopre un ruolo fondamentale. Il fatto che garantisca, in alcune circonstanze, la presenza di una situazione di equilibrio, permette di conoscerne l'andamento, anche senza uno studio approfondito della transizione tra i primi stati del sistema.Poichè la dimostrazione è contenuta delle dimostrazioni dei teoremi precedenti verrà quindi omessa.

Teorema 3.5. *Sia $(X_n)_{n\geq 0}$ una catena di Markov finita e irriducibile . Allora $(X_n)_{n\geq 0}$ possiede una distribuzione stazionaria unica π e, indipendentemente dalla distribuzione iniziale, le distribuzioni degli stati al tempo n convergono a π quando n tende all'infinito. In simboli:*

$$lim_{n\to\infty} P_{ij}^{(n)} = \pi_j$$

Nel caso in cui la distribuzione iniziale sia una distribuzione stazionaria, allora la catena di Markov sarà una catena stazionaria. [1]

3.4 L'algoritmo di PageRank

Il PageRank è un algoritmo, quindi un processo informatico, che consente di definire un ordine, attraverso l'assegnazione di un peso numerico, ad un insieme di documenti connessi tramite dei collegamenti ipertestuali. La maggior applicazione di questo algoritmo si ha nell'insieme delle pagine nel World Wide Web, con l'obbiettivo di definire un'ordine di importanza tra le stesse. Questo processo è stato brevettato dalla Stanford University, ma deve il suo nome a Larry Page, uno dei due fondatori di Google. Attualmente il PageRank continua ad essere utilizzato nelle valutazioni per il posizionamento dei siti web, ma è diventato parte di un sistema molto più complesso e avanzato. Vediamo quindi inizialmente di comprendere il funzionamento del processo, senza approfondire l'aspetto informatico, al fine di comprendere il legame con le catene di Markov e successivamente come il teorema di Perron-Frobenius ne possa garantire l'efficienza.

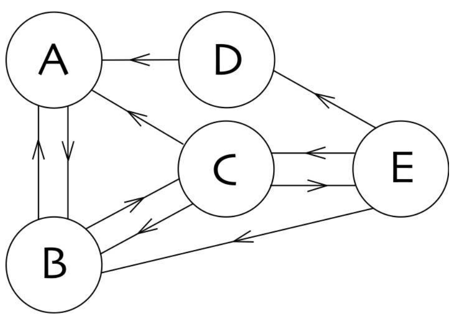
3.4.1 Un'idea dell'algoritmo e il legame con le catene di Markov

Iniziamo con una panoramica sul meccanismo alla base della classificazione dei documenti testuali da parte di un calcolatore, mettendoci inizialmente in una situazione molto semplificata.

Esempio 3.3. *[4] Consideriamo un sistema di cinque pagine Web, che chiamiamo A, B, C , D, E, connesse tra loro (attraverso collegamenti ipertestuali che conosciamo come Link) . Costruiamo il grafo che descrive i collegamenti tra i cinque testi. Sappiamo che nel grafo un nodo X è connesso a un secondo nodo Y se nel testo X vi è un collegamento ad Y. Supponiamo di avere un sistema con le condizioni seguenti:*

- *A è connesso solo a B.*
- *B è connesso a A e C.*
- *C è connesso a E, B e A.*
- *D è connesso solo ad A.*
- *E è connesso a B, C e D.*

Rappresentiamo un grafo che descrive le condizioni indicate e otteniamo il seguente:



L'idea è che la pagina più importante sia corrispondente al nodo del grafo con più lati entranti, quindi la pagina Web alla quale vi sono più riferimenti alle altre pagine. Nell'esempio si equivalgono le pagine A e B, con due lati entranti ciascuna. Supponiamo, quindi, di seguire un percorso casuale nel grafo orientato. Partiamo da una pagina generica e scegliamo un link passando alla successiva. Per esempio supponiamo di partire dalla pagina B, possiamo quindi procedere verso la pagina C o la pagina A con probabilità $\frac{1}{2}$ *per ognuna. Se partiamo, invece, dalla pagina D, dobbiamo necessariamente andare in A con probabilità 1. Mettiamo in una matrice le probabilità e costruiamo la matrice come segue:*

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Quindi partendo da A ho probabilità 1 di muovermi verso la pagina B, come da D ho probabilità 1 di spostarmi sulla pagina A. Invece da E vado nelle pagine B,C,D con probabilità $\frac{1}{3}$ *e così via.*

Vedendo la matrice è chiaro che rappresenta la matrice di transizione di una catena di Markov ove gli stati sono A,B,C,D,E. Possiamo quindi applicare le catene di Markov a questo contesto. Sia $(X_n)_{n\geq 0}$ *la successione di variabili casuali che assumono valori nello spazio degli stati* $I = \{A, B, C, D, E\}$ *.*

Vogliamo valutare la probabilità di trovarci su una delle 5 pagine dopo n passi. Come affermato nel paragrafo precedente, poichè la catena di Markov così generata è una catena finita e ricorrente, possiamo calcolarne la probabilità stazionaria. Per questo esempio specifico si otterrà un vettore come segue:

$$\pi = \begin{pmatrix} 0.293 & 0.390 & 0.220 & 0.024 & 0.073 \end{pmatrix}$$

Ove le entrate del vettore sono state approssimate a 3 cifre decimali. L'ordine di importanza, che verrà quindi assegnato alle pagine dell'esempio, corrisponde quindi a B, A, C, E, D in ordine crescente rispetto alla probabilità di trovarsi dopo n passi sulla pagina.

3.4.2 Il funzionamento dell'algoritmo

Vediamo di approfondire meglio dal punto di vista tecnico il funzionamento di questo processo. Il modello originale dell'algoritmo sfrutta quindi la struttura con hyperlink per costruire una catena di Markov tra le pagine Web con una matrice di transizione irriducibile. L'irriducibilità della matrice garantisce la esistenza di un vettore di probabilità stazionaria a cui tende il processo a lungo termine. Un'idea dell'algoritmo originario è già data nell'esempio precedente. Ma cosa succederebbe alla matrice di transizione se una pagina non avesse collegamenti ad altre? Si avrebbe una matrice P che ha una riga di zeri, il che implicherebbe che la matrice non fosse stocastica e non rappresenterebbe quindi una matrice di transizione. Diamo quindi una definizione iniziale di ranking semplificato per valutare poi alcune modifiche da effettuare alla matrice per comprendere la vera definizione di

PageRank.

Definizione 3.14. *Dato un sistema di pagine Web connesse tra loro il ranking semplificato è il vettore-riga R tale che:*

$$R = cPR$$

[2]

Ove c è un fattore di normalizzazione ed è sicuramente positivo e minore di uno e R è un autovettore di P con autovalore 1.

Come spiegato in precedenza questo calcolo del ranking non è sempre applicabile, in quanto vi sono situazioni in cui la matrice potrebbe non risultare stocastica. L'algoritmo richiede quindi alcuni accorgimenti.
Supponiamo di avere una pagina in cui non vi siano iper-collegamenti ad altre pagine. In questo caso per ottenere una matrice stocastica si impone che da una pagina priva di collegamenti vi sia uguale probabilità (maggiore di zero) di giungere ad una qualsiasi delle altre pagine. In questo caso si costruisce quindi la matrice inserendo una riga ove ogni entrata vale $\frac{1}{|I|}$. Chiamiamo $\bar{P}$ la matrice con questa modifica, è sempre stocastica? Rispetta le ipotesi per cui esiste sempre una distribuzione stazionaria? Sappiamo che una catena di Markov tende ad una data distribuzione stazionaria se e solo se la catena è finita e ricorrente. Ciò implica che la matrice P sia irriducibile. Allora per rendere $\bar{P}$ irriducibile, servono ulteriori modifiche. Siano quindi $n = |I|$ e $e^T = \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}$. Si definisce la matrice seguente:

$$\bar{\bar{P}} = c\bar{P} + (1-c)\frac{1}{n}ee^T$$

La dimostrazione che la matrice $\bar{\bar{P}}$ è stocastica e irriducibile si basa sul fatto che la combinazione lineare di una matrice $\bar{P}$ e di una matrice stocastica di perturbazione è irriducibile, la vedremo in seguito. Ora quindi si ha che ogni nodo del grafo è direttamente connesso ad ogni altro nodo, creando quindi una catena irriducibile per definizione. Queste modifiche alla matrice di transizione assicurano anche che questa sia primitiva, il che soddisfa le ipotesi di convergenza alla distribuzione stazionaria π .
A questa modifica classica i fondatori dell'algoritmo propongono una modifica alla *matrice di teletrasporto E*. Invece di utilizzare $\frac{1}{n}ee^T$, si utilizza ev^T ove v è un vettore di probabilità con elementi positivi chiamato *vettore di teletrasporto o di personalizzazione*. Questo vettore permette di aumentare la probabilità di un comune utente di visitare una pagina di suo interesse, rendendo quindi la ricerca personalizzata secondo la classe dell'utente. Questo metodo di modifica della matrice, dalla forma P alla forma

$$\bar{\bar{P}} = c\bar{P} + (1-c)ev^T$$

è detto *metodo di irriducibilità massima.* Diamo quindi la definizione generale di PageRanking.

Definizione 3.15. *Il PageRank R' è l'autovettore della matrice $\bar{\bar{P}}$ di autovalore 1. In simboli*

$$R' = \bar{\bar{P}}R'$$

3.4.3 Le applicazioni del teorema

La panoramica precedente sull'algoritmo in questione rende evidente lo stretto legame presente tra tale processo e le catene di Markov, poichè esso si basa direttamente sulla costruzione di una $Markov(v, P)$. Apportando dirette modifiche alla matrice di transizione, è possibile soddisfare le ipotesi dei teoremi che garantiscono l'esistenza e unicità della distribuzione stazionaria. Vediamo quindi quale teorema garantisce che la matrice $\bar{\bar{P}}$ rispetti le condizioni sotto le quali l'algoritmo converge.

Teorema 3.6. *Dato lo spettro di una matrice $\bar{P}$ stocastica, $\{1, \lambda_2, \lambda_3, \ldots, \lambda_n\}$, lo spettro della matrice stocastica primitiva $\bar{\bar{P}} = c\bar{P}+(1-c)ev^T$ è $\{1, c\lambda_2, c\lambda_3, \ldots, c\lambda_n\}$, ove v^T è un vettore di probabilità. [2]*

Dimostrazione. Sia $e = \begin{pmatrix} 1 & 1 & \ldots & 1 \end{pmatrix}$. Poichè $\bar{P}$ è stocastica per ipotesi, ha autovalore 1 con relativo autovettore e . Sia allora Q la matrice non singolache che abbia l'autovettore e sulla prima colonna

$$Q = \begin{pmatrix} e & X \end{pmatrix}.$$

Sia

$$Q^{-1} = \begin{pmatrix} y^T \\ Y^T \end{pmatrix}$$

tale che

$$Q^{-1}Q = \begin{pmatrix} y^t e & y^T X \\ Y^T e & Y^T X \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0^T & \mathbb{1} \end{pmatrix}$$

Allora posso applicare la seguente trasformazione alla matrice $\bar{P}$:

$$Q^{-1}\bar{P}Q = \begin{pmatrix} y^t e & y^T \bar{P} X \\ Y^T e & Y^T \bar{P} X \end{pmatrix} = \begin{pmatrix} 1 & y^T \bar{P} X \\ 0 & Y^T \bar{P} X \end{pmatrix}.$$

Questo poichè il vettore e è autovettore per la matrice $\bar{P}$ di autovalore 1 . Si deduce quindi che la matrice $Y^T \bar{P} X$ ha come autovalori $\{\lambda_2, \ldots, \lambda_n\}$. Applicando la stessa trasformazione alla matrice $\bar{\bar{P}} = c\bar{P} + (1-c)ev^T$ si ottiene:

$$\begin{aligned} Q^{-1}(c\bar{P} + (1-c)\mathbf{e}\mathbf{v}^T)Q &= cQ^{-1}\bar{P}Q + (1-c)Q^{-1}\mathbf{e}\mathbf{v}^T Q & (3.2) \\ &= \begin{pmatrix} c & cy^T \bar{P} X \\ 0 & cY^T \bar{P} X \end{pmatrix} + (1-c) \begin{pmatrix} y^T \mathbf{e} \\ Y^T \mathbf{e} \end{pmatrix} \begin{pmatrix} \mathbf{v}^T \mathbf{e} & \mathbf{v}^T X \end{pmatrix} & (3.3) \\ &= \begin{pmatrix} c & cy^T \bar{P} X \\ 0 & cY^T \bar{P} X \end{pmatrix} + \begin{pmatrix} (1-c) & (1-c)\mathbf{v}^T X \\ 0 & 0 \end{pmatrix} & (3.4) \\ &= \begin{pmatrix} 1 & cy^T \bar{P} X + (1-c)\mathbf{v}^T X \\ 0 & cY^T \bar{P} X \end{pmatrix} & (3.5) \end{aligned}$$

Poichè abbiamo dimostrato precedentemente che la matrice $Y^T \bar{P} X$ ha come autovalori $\{\lambda_2, \ldots, \lambda_n\}$, allora la matrice $cY^T \bar{P} X$ ha come autovalori $\{c\lambda_2, \ldots, c\lambda_n\}$. □

Questo teorema dimostra che le modifiche effettuate alla matrice $\bar{P}$, non alterano il risultato che si otterrebbe cercando il vettore stazionario partendo da $\bar{P}$. Inoltre, per costruzione, $\bar{P}$ rispetta tutti i criteri necessari affinchè sia garantita l'esistenza di una distribuzione stazionaria. Infatti, poichè $\bar{P}$ è irriducibile (il grafo ad essa associato è connesso), allora avrà esattamente n autovalori distinti, che sono uguali in numero a quelli di $\bar{\bar{P}}$. Anche $\bar{\bar{P}}$ è irriducibile e rispetta le ipotesi del Teorema 3.5. Si ha quindi che grazie al Teorema di Perron-Frobenius, che garantisce l'esistenza di una tale distribuzione, è garantita la convergenza del suddetto algoritmo.

Capitolo 4

Conclusioni

In questa tesi abbiamo esaminato il Teorema di Perron-Frobenius e le sue applicazioni, soprattutto nell'ambito delle catene di Markov e dell'algoritmo di PageRank. Questo teorema è fondamentale nell'algebra lineare e nella teoria delle matrici ed ha molte applicazioni in diversi ambiti. [3]

Abbiamo studiato che, grazie a questo Teorema e alle informazioni che fornisce sullo spettro di una matrice non-negativa, è possibile predire i risultati di sistemi dinamici complessi. L'algoritmo di PageRank, utilizzato da motori di ricerca come Google per ordinare le pagine Web rispetto al numero di collegamenti che esse ricevono, costituisce una particolare applicazione pratica di questo teorema. Questo processo, basato sulla costruzione di una matrice di transizione stocastica, permette di calcolare una distribuzione stazionaria che rappresenta la "popolarità" delle pagine.

Oltre alle catene di Markov, questo Teorema trova applicazioni in molti campi, dalla biologia in cui viene utilizzzato nello studio della crescita di popolazioni, come in economia nello studio di modelli input-output che descrivono le relazioni tra i diversi settori economici. [3]

É stato un mio particolare interesse studiare le applicazioni alle catene di Markov e soprattutto all'algoritmo di PageRank, poichè è la base dei motori di ricerca che sono di utilizzo comune al giorno d'oggi. Sebbene questo breve trattato sia focalizzato principalmente sulle matrici non-negative e irriducibili, la vasta gamma delle applicazioni del teorema offre spunto per ulteriori ricerche e approfondimenti.

In conclusione, il Teorema di Perron-Frobenius si conferma uno strumento potente, con applicazioni in vari ambiti della scienza e della tecnologia. La sua maggiore potenzialità è quella di garantire la convergenza ad un equilibrio stazionario, non solo nell'ambito delle catene di Markov ma anche in molti altri.

Bibliografia

[1] A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1994.

[2] Amy N. Langville and Carl D. Meyer. Deeper Inside PageRank. *Internet Mathematics*, 1(3):335 – 380, 2003.

[3] C. R. MacCluer. The many proofs and applications of perron's theorem. *SIAM Review*, 42(3):487–498, 2000.

[4] Antoine Nectoux. Come funziona google: Catene di markov e autovalori, 2013.

[5] J.R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[6] L. Salce. *Note sulle matrici: Esercizi sulle matrici*. Libreria Progetto, 2019.