Aula 11

Roteiro

- Autovalores, autovetores, decomposição
- Convergência para estacionaridade
- Tempo de mistura
- Spectral gap
- Tempo de mistura de passeios aleatórios

Estacionaridade

- Seja *P* a matriz de transição de estados de uma CM
- π é uma distribuição estacionária sse

$$\pi P = \pi \qquad \pi_s \geq 0 \qquad \sum_{s \in S} \pi_s = 1$$

- Seja $\pi(0)$ a distribuição inicial da CM. A distribuição no tempo *t* é dada por

$$\pi(t) = \pi(t-1) P = \pi(0) P^t$$

- Para qualquer CM aperiódica e irredutível, temos

$$\lim_{t \to \infty} d_{TV}(\pi(t), \pi) = 0$$

- Convergência para π é única e independe de $\pi(0)$

Convergência

- Mas o quão rápido é esta convergência?
- Lembrando da distância de variação total entre dois vetores de probabilidade

$$d_{TV}(\alpha,\beta)=\frac{1}{2}\sum_k |\alpha_k-\beta_k|$$ ← d_{TV} tem valor entre 0 e 1

- Como que $d_{TV}(\pi(t),\pi)$ vai a zero com t ?

$$d_{TV}(\pi(t),\pi)=\theta(e^{-at})\,?$$ ← Muito rápido (exponencial)

$$d_{TV}(\pi(t),\pi)=\theta(t^{-b})\,?$$ ← Meio lento (lei de potência)

$$d_{TV}(\pi(t),\pi)=\theta\left((\log t)^{-c}\right)?$$ ← Muito lento

- Depende de $\pi(0)$? Depende de P?

Autovalores e Autovetores

- Dada uma matriz *P*, *v* é chamado de autovetor associado ao autovalor λ, se

$$Pv = \lambda v$$

← Multiplicar *v* por *P* é igual a escalar *v* por λ

- *P* possui até *n* autovetores linearmente independentes, cada qual associado a um autovalor
- *u* é chamado de autovetor a esquerda se

$$uP = \lambda u$$

← Multiplicar *u* a esquerda de *P*

- Se *u* é autovetor a esquerda, então existe autovetor *v* tq

$$P' v = \lambda u$$

← *P'* é a transposta da matriz *P*

- Autovalores são os mesmos (esquerda e direita), relação entre autovetores obtida pela transposta da matriz

Autovetores e Matriz *P*

- π é uma distribuição estacionária da CM com matriz *P* sse

$$\pi P = \pi$$

- Ou seja, π é o autovetor à esquerda de *P* associado ao autovalor λ = 1
- Precisamos ainda $\sum_{s \in S} \pi_s = 1$
- Solução: normalizar o autovetor para garantir soma 1
- **Teorema:** Se *P* é uma matriz estocástica, temos $|\lambda|$ <= 1 para todo autovalor, e apenas um autovalor λ = 1
    - matriz estocástica = matriz de transição de probabilidade

Decomposição em Autovetores

- Uma matriz *P* pode ser escrita através de seus autovetores e autovalores

$$P = Q L Q^{-1}$$

- Onde *Q* é matriz com autovetores de *P* como colunas
- *L* é matriz diagonal, com L_{ii} autovalor associado ao autovetor *i* (*i*-ésima coluna de *Q*)
- Q^{-1} é a inversa da matriz *Q*

Exemplo

$$P=\begin{pmatrix} .3 & .2 & .5 \\ .4 & .5 & .1 \\ .7 & .2 & 1 \end{pmatrix}$$

- Autovalores: 1, 0.3, -0.4 (de acordo com nosso teorema)
- Autovetores associados: (1, 1, 1), (2, -5, 2), (-43, 13, 55)

$$P=Q\,L\,Q^{-1}$$

Autovetor à esquerda associado ao λ=1

$$Q=\begin{pmatrix} 1 & 2 & -43 \\ 1 & -5 & 13 \\ 1 & 2 & 55 \end{pmatrix} \quad L=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & -0.4 \end{pmatrix} \quad Q^{-1}=\begin{pmatrix} \boxed{43/98 \quad 2/7 \quad 27/98} \\ 3/49 \quad -1/7 \quad 4/49 \\ -1/98 \quad 0 \quad 1/98 \end{pmatrix}$$

Autovetores de *P* (um em cada coluna)

Autovalores de *P* (correspondentes)

Inversa de *Q*

Q^{-1}: autovetores à esquerda de *P*, um em cada linha!

Figueiredo 2021

Distribuição no Tempo *t*

- Seja $\pi(0)$ a distribuição inicial da CM. A distribuição no tempo *t* é dada por

$$\pi(t) = \pi(t-1)P = \pi(0)P^t$$

- Mas sabemos que $P = QLQ^{-1}$, então temos

$$P^t = PP\ldots P = QLQ^{-1}QLQ^{-1}\ldots QLQ^{-1}$$
$$= QLILI\ldots LQ^{-1} = QLL\ldots LQ^{-1} = QL^tQ^{-1}$$

Matriz diagonal, com cada elemento elevado a potência *t*

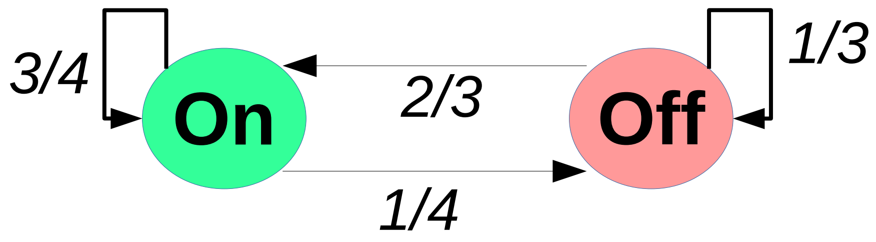
- Logo, temos que

$$\pi(t) = \pi(0)QL^tQ^{-1}$$

- Para onde vai L^t com *t* crescente ?
- $\lambda = 1$ fica no mesmo lugar, todos os outros valores vão a zero, pois $|\lambda| < 1$

Figueiredo 2021

Exemplo



$$P=\begin{pmatrix} .75 & .25 \\ .67 & .33 \end{pmatrix}$$

$$\pi=\begin{pmatrix} 8/11 & 3/11 \end{pmatrix}$$

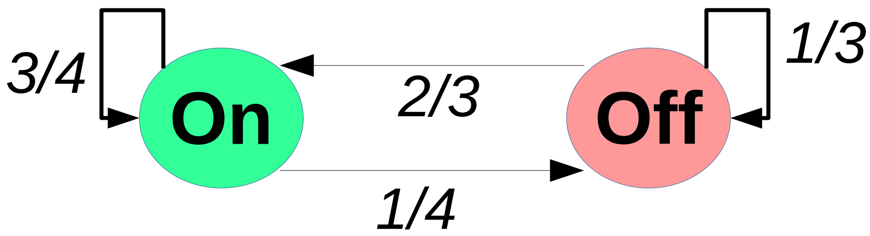
- Autovalores de P: $\lambda_1 = 1$, $\lambda_2 = 1/12$
- Autovetores a esquerda de P: $v_1 = (8/3\ 1)$, $v_2 = (1\ -1)$
- Temos então $\pi=\left(\frac{8/3}{8/3+1}\ \ \frac{1}{8/3+1}\right)=\begin{pmatrix} 8/11 & 3/11 \end{pmatrix}$
- Supor $\pi(0) = (1\ 0)$. Podemos escrever

$$\pi(0)=(1\,0)=\pi+3/11\,v_2$$

- Temos então

$$\pi(t)=\pi(0)P^t=(\pi+3/11\,v_2)P^t=\pi P^t+3/11\,v_2 P^t$$
$$=\pi+3/11\lambda_2^t v_2$$
$$=\left(8/11+3/11(1/12)^t \quad 3/11-3/11(1/12)^t\right)$$

Exemplo



$$P = \begin{pmatrix} .75 & .25 \\ .67 & .33 \end{pmatrix}$$

$$\pi = \begin{pmatrix} 8/11 & 3/11 \end{pmatrix}$$

- Temos então

$$\pi(t) = \left(8/11 + 3/11(1/12)^t \quad 3/11 - 3/11(1/12)^t\right)$$

- Distância de variação total

$$d_{TV}(\pi(t), \pi) = \frac{1}{2}\frac{3}{11}\left(1/12^t + 1/12^t\right) = \frac{3}{11} 1/12^t = \theta\left(12^{-t}\right)$$

Converge exponencialmente rápido!

- Resultado vale para qualquer CM
  - converge exponencialmente rápido em *t*
  - constantes dependem de *P* e $\pi(0)$

Teorema da Convergência

- Considere uma CM aperiódica e irredutível com matriz de probabilidade *P* com distribuição estacionária π
- Existem constantes α em (0, 1) e C > 0, tq

$$max_{\pi(0)} d_{TV}(\pi(t), \pi) \leq C\alpha^t$$

- Distribuição transiente $\pi(t)$ converge exponencialmente rápido em *t* para distribuição estacionária π, independente de *P* e $\pi(0)$

Tempo para Convergência

- Quantos passos até determinar a convergência?
- **Ideia:** definir $\varepsilon > 0$ como distância até equilíbrio
  - calcular t tal $d_{TV}(\pi(t), \pi) = \varepsilon$
- Exemplo anterior

$$d_{TV}(\pi(t), \pi) = \frac{3}{11} 12^{-t} = \epsilon \longrightarrow t = \frac{\log \epsilon + \log(11/3)}{-\log 12}$$

- Se $\varepsilon = 10^{-6}$ então $t = 5$
- Poucas transições são necessárias para chegar próximo do equilíbrio
  - em geral, $t = \theta\left(\log \frac{1}{\epsilon}\right)$
  - constante em θ depende de P e $\pi(0)$

Figueiredo 2021

Tempo de Mistura

- τ_ε : tempo de mistura-ε para constante $\varepsilon > 0$

$$\tau_\epsilon = min\{t \mid max_{\pi(0)} d_{TV}(\pi(t), \pi) \leq \epsilon\}$$

- τ_ε : menor valor de *t* tal que para qualquer distribuição inicial, π(t) está a distância menor que ε da estacionária
- Teorema: Para qualquer CM aperiódica, irredutível, temos $\tau_\epsilon \leq \tau_{1/4} \log 1/\epsilon$
- Depois de estar perto o suficiente, chegar mais perto é mais fácil (um fator log ε^{-1})
- Tempo de mistura depende fracamente em ε
  - ε vai ser tomado como constante

Spectral Gap

- Convergência depende da relação dos autovalores de *P*
  - segundo maior autovalor (em módulo) domina convergência
- *Spectral Gap* (δ): distância entre os dois maiores (em módulo) autovalores de *P* (maior é sempre igual a 1)

$$\delta = 1 - max_{k>1}\{|\lambda_k|\}$$

k-ésimo autovalor de *P*, $\lambda_1 = 1$

- Quanto maior for δ, mais rápido é a convergência
  - base da exponencial que domina a convergência é dada por $|\lambda_2|$ (segundo maior autovalor)
  - todas as outras componentes vão a zero mais rapidamente (menor base)

Exemplo

- Processo de Nascimento e Morte, fila com capacidade K
  - p = probabilidade de chegada, q = probabilidade de saída, r = 1-p-q (não chega nem sai ninguém)

$$P=\begin{pmatrix} q+r & p & 0 & 0 & 0 \\ q & r & p & 0 & 0 \\ 0 & q & r & p & 0 \\ 0 & 0 & q & r & p \\ 0 & 0 & 0 & q & p+r \end{pmatrix}$$

- K=4, p=0.3, q=0.4 → λ_2 ~ 0.86
  - δ = 1 – 0.86 = 0.14
- K=4, p=0.1, q=0.8 → λ_2 ~ 0.56
  - δ = 1 – 0.56 = 0.44

- Spectral gap depende dos parâmetros do modelo
  - quanto mais simétrico (*p* e *q* parecidos) menor é o gap, mais lento será a convergência

Spectral Gap e Tempo de Mistura

- Relação entre δ e τ_ε
- Considere CM reversível, irredutível e aperiódica com *spectral gap* δ e $\pi_o = \min_i \pi_i$ (menor valor de probabilidade da distribuição estacionária)
- Temos a seguinte relação

$$\left(\frac{1}{\delta}-1\right)\log 1/(2\epsilon) \leq \tau_\epsilon \leq \frac{\log 1/(\pi_o \epsilon)}{\delta}$$

- Maior δ, menor τ_ε
- Maior π_o, menor τ_ε

Limitante inferior para tempo de mistura

Limitante superior para tempo de mistura

- Usar limitante superior na prática não é fácil, pois precisamos de π_o e δ

Figueiredo 2021

Exemplo

- Processo de Nascimento e Morte, fila com capacidade K
  - CM é reversível, irredutível e aperiódica
- Supondo p < q, podemos calcular π_K, que será o estado menos provável (aula passada)
- K=4, p=0.3, q=0.4 $\rightarrow$ $\lambda_2 \sim 0.86$
  - $\delta = 1 - 0.86 = 0.14$
  - $\pi_K = 0.015 \rightarrow \pi_o = \min_i \pi_i = 0.015$
- Assumindo $\varepsilon = 10^{-6}$, e substituindo temos

$$\left(\frac{1}{0.14}-1\right)\log 1/\left(2*10^{-6}\right) \leq \tau_\epsilon \leq \frac{\log 1/\left(0.015*10^{-6}\right)}{0.14}$$

$$80.6 \leq \tau_\epsilon \leq 128.7$$

- Com t = 129 estaremos 10^{-6} próximo de π, para qualquer distribuição inicial

Figueiredo 2021

Tempo de Mistura em *n*

- Resultados anteriores são para CM de tamanho fixo
- Espaço de estados da CM pode crescer muito com o tamanho do problema
- Seja $n = f(k)$ o número de estados da CM
  - k é um parâmetro do modelo
  - $n = 2^k$, ex. S = número de colorações de k objetos com duas cores
  - $n = k!$, ex. S = permutações de k cartas
- Como τ depende de n e k ?
  - para algum ε fixo
  - convergência em função de n

Passeios Aleatórios

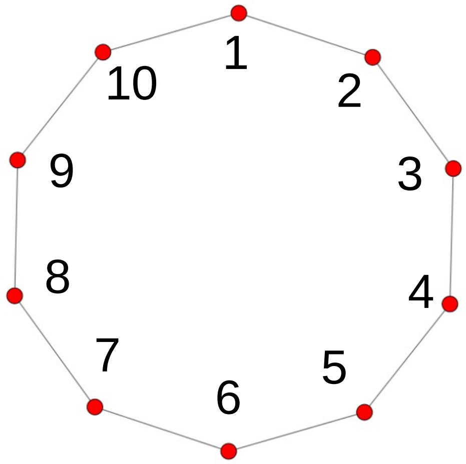
- Tempo de mistura de passeios aleatórios em grafos
  - modelo do grafo parametrizado por n (vértices)
  - grafo completo, grafo em anel, hipercubo, etc
- Passeio aleatório preguiçoso (*lazy random walk*)
  - permanece no vértice atual com prob ½, caso contrário, escolhe vizinho uniformemente
  - implica CM aperiódica e irredutível (e reversível)
- Como que o tempo de mistura depende da estrutura do grafo?
  - τ_n é o tempo de mistura com n vértices para um ε constante

Mistura em Anel e Árvore

- Anel (um ciclo) com *n* vértices

$$c\, n^2 \leq \tau_n \leq n^2$$

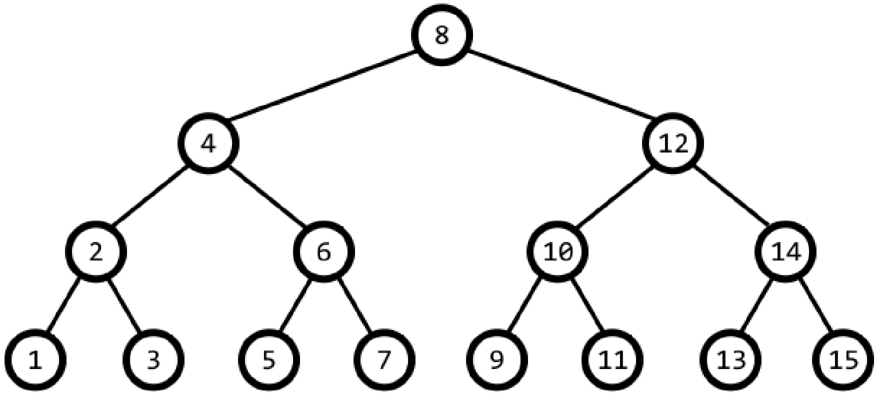
  - quadrático no tamanho do anel



- Árvore binária completa com *n* vértices

$$\tau_n \leq 16\, n$$

  - linear no tamanho da árvore

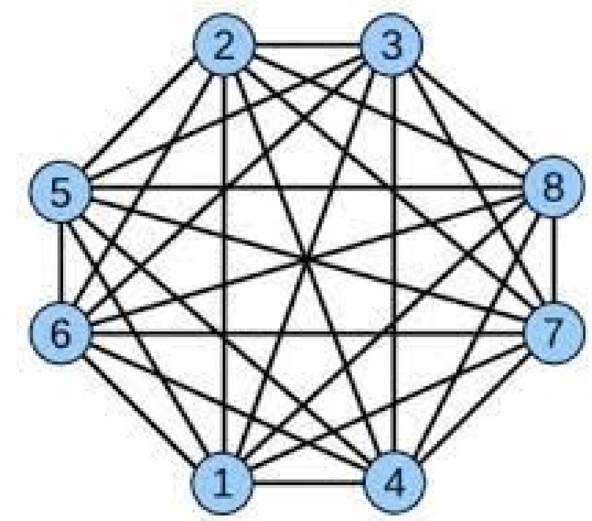


- Passeios misturam muito mais rapidamente em árvores do que anéis!

Figueiredo 2021

Mistura em Outros Grafos

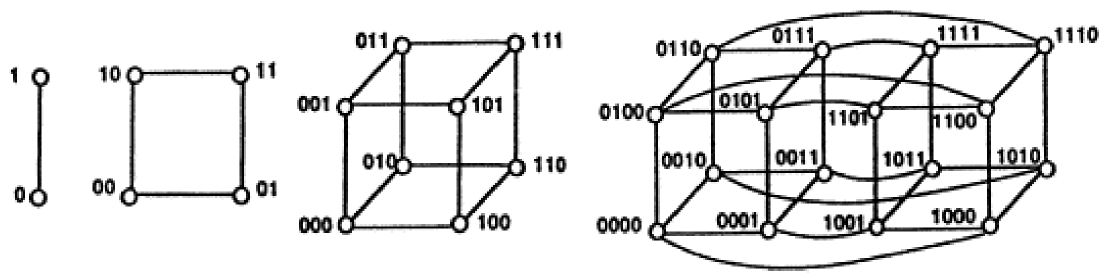
- Grafo completo com *n* vértices (com *loops* ou se *n* é grande)



$$\tau_n = 1$$

  - tempo constante

- Hipercubo com *d* dimensões e $n = 2^d$ vértices



$$\tau_n \le c\, d \log d$$ ← chamado de “*fast mixing*”

  - tempo *polylog* do tamanho do grafo

Tema atual de pesquisa na matemática (e computação)