

Monte Carlo e Las Vegas

- Dois tipos de algoritmos que diferem no local de sua incerteza

→ Monte Carlo : Na correteude.

PODE DAR **SOLUÇÕES ERRADAS**,
MAS GASTA **TEMPO DETERMINÍSTICO**

→ Las Vegas : NO TEMPO DE EXECUÇÃO

DAÍ APENAS **SOLUÇÕES CORRETOS**
MAS GASTA **TEMPO ALEATÓRIO**

MONTE CARLO

- DECISÃO DOIS TIPOS DE INSTÂNCIA : SIM E NÃO

- ERRO UNILATERAL

1. BASEADO NO SIM : NUNCA ERRA QUANDO DIZ SIM

2. BASEADO NO NÃO : NUNCA ERRA QUANDO DIZ NÃO

- ERRO BILATERAL : PODE ERRAR EM AMBAS RESPOSTAS

EX: IDENTIDADE DE POLINÔMIOS :

DADOS $F(x) = (x - a_1)(x - a_2) \dots (x - a_d)$

e $G(x) = b_d x^d + b_{d-1} x^{d-1} + \dots + b_0$

$a_i, b_i \in \mathbb{R}$

DECIDIR SE $F = G$

~> PODEMOS EXPANDIR F E COMPARAR OS COEFICIENTES : $O(d^d)$

~> PODEMOS TESTAR AS RAÍZES DE F EM G $O(d^2)$

EX: IDENTIDADE DE POLINÔMIOS:

DADOS $F(x) = (x - a_1)(x - a_2) \dots (x - a_d)$
E $G(x) = b_d x^d + b_{d-1} x^{d-1} + \dots + b_0$

DECIDIR SE $F = G$

~> PODEMOS EXPANDIR F E COMPARAR OS COEFICIENTES: $O(d^d)$

~> PODEMOS TESTAR AS RAÍZES DE F EM G : $O(d^2)$

~> SORTEAR UM INTEIRO r ENTRE L E $100d$ E TESTAR SE $F(r) = G(r)$ $O(d)$

SE $F(r) \neq G(r)$ ENTÃO r É UM CERTIFICADO DE QUE $F \neq G$.

DO CONTRÁRIO, RETORNAMOS SIM

→ O POLINÔMIO $H(x) = F(x) - G(x)$ POSSUI NO MÁXIMO d RAÍZES EM $[1, 100d]$

A PROBABILIDADE DE r SER UMA DESSAS RAÍZES É $d/100d = 1/100$

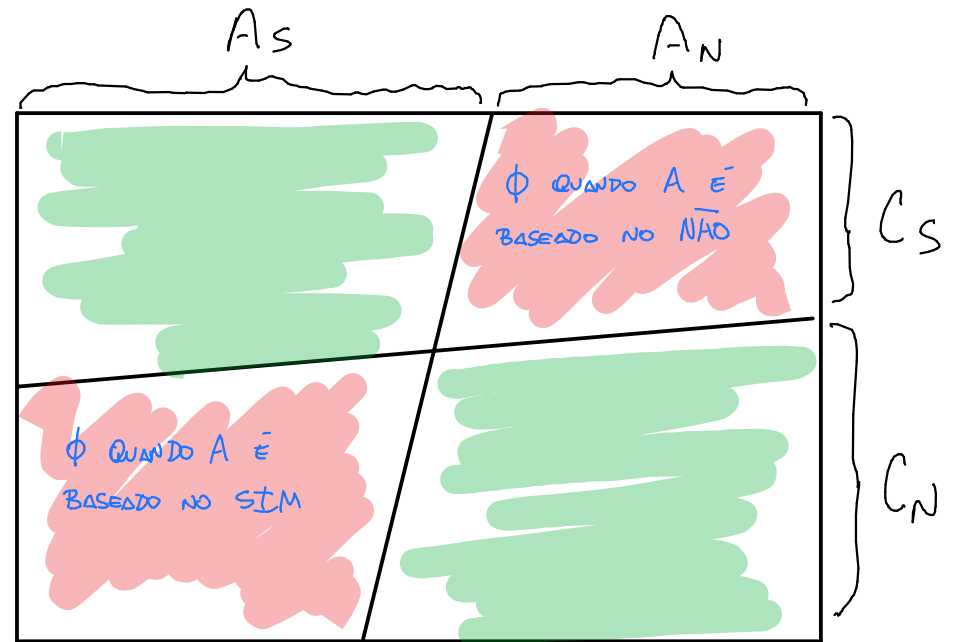
CUSTO $O(d)$

PROBABILIDADES MONTE CARLO UNILATERAL

EVENTOS

- 1 A_S - O ALGORITMO RESPONDE SIM
- 2 A_N - O ALGORITMO RESPONDE NÃO

- 3 C_S - A RESPOSTA CORRETA É SIM
- 4 C_N - A RESPOSTA CORRETA É NÃO



GOSTARIAMOS DE ESTUDAR AS PROBABILIDADES

$$\Pr(A_S | C_S), \Pr(A_N | C_S), \Pr(A_S | C_N), \Pr(A_N | C_N)$$

- SE O ALG É BASEADO EM SIM, ENTÃO $A_S \subseteq C_S$

$$\Pr(C_S | A_S) = 1$$

$$\Pr(C_N | A_S) = 0$$

- SE O ALG É BASEADO EM NÃO, ENTÃO $A_N \subseteq C_N$

$$\Pr(C_N | A_N) = 1$$

$$\Pr(C_S | A_N) = 0$$

EX: ID. DE POLINÔMIOS.

BASEADO NO NÃO, LOGO $A_N \subseteq C_N$

$\Pr(A_S | C_N)$?

→ UM ALG BASEADO NO NÃO RESPONDE CORRETAMENTE SIM PARA INSTÂNCIAS EM C_S

OU SEJA $C_S \subseteq A_S$.

↳ ALGUMAS EM C_N

→ DEVEMOS NOS PREOCUPAR COM A PROB. DE ENCONTRAR UM CERTIFICADO.

C_S	C_N
A_S	A_N

SE $I \in C_S$ E $A(I) = N$

$A_N = C_N$?

$I \in A_N \subseteq C_N$

$A_S = C_S$

$A_S \subseteq C_N \cup C_S$

1. BASEADO NO SIM : NUNCA ERRA QUANDO DIZ SIM

$A_S \subseteq C_S$

2. BASEADO NO NÃO : NUNCA ERRA QUANDO DIZ NÃO

$A_N \subseteq C_N$

• CONSIDERE UM ALG. BASEADO NO NÃO E QUE ERRA COM PROB NO MÁX ϵ_1 .

$$\rightarrow \Pr(A_S | C_N) \leq \epsilon_1$$

↓
INSTÃNCIAS
EM C_N

→ EXECUTE-O t VEZES $\rightsquigarrow A^t$

→ Qual a prob. global ϵ_t DE ERRO DE A^t ?

Ω ESP. DE UMA EXECUÇÃO	$\Omega \times \Omega \times \dots \times \Omega$ ESP DE PROB. DE VÁRIAS EXECUÇÕES
--	--

→ CERTAMENTE UMA INSTÃNCIA EM C_N
PARA A QUAL RESPONDEMOS SIM t VEZES (A^1, A^2, \dots, A^t)

$$\begin{aligned} \text{Logo, } \epsilon_t &= \Pr(C_N \wedge \bigwedge_{i=1}^t A_S^i) = \Pr(C_N) \cdot \Pr\left(\bigwedge_{i=1}^t A_S^i \mid C_N\right) \\ &= \Pr(C_N) \cdot \Pr(A_S^t \mid C_N) \cdot \Pr\left(\bigwedge_{i=1}^{t-1} A_S^i \mid C_N, A_S^t\right) \\ &= \Pr(C_N) \prod_{k=1}^t \Pr(A_S^k \mid C_N, \bigwedge_{i=1}^{k-1} A_S^i) \end{aligned}$$

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

$$\Pr(A|B) = \Pr(A)$$

$$\begin{aligned} &\stackrel{\text{IND DAS EXECUÇÕES}}{\rightarrow} \underbrace{\Pr(C_N)}_{\leq 1} \prod_{k=1}^t \underbrace{\Pr(A_S^k \mid C_N)}_{\epsilon_1} \leq \epsilon_1^t \end{aligned}$$

LAS VEGAS

- O TEMPO É UMA VAR ALÉATORIA
- JULGAMOS PELA ESPERANÇA, VARIÂNCIA, ...

EX: BUSCA EM LISTA COM REPETIÇÃO.

- LISTA COM n ENTRADAS
- NÚMEROS DE 0 A 9
- CADA NÚMERO APARECE $1/10$ DAS VEZES.

$$\Omega = \{H, TH, TTH, T^3H, \dots\}$$

$$n \left(\frac{1}{n}\right)^{10} = \frac{1}{n^9}$$

→ VARRER A LISTA DO COMEÇO AO FIM $\Theta(n)$

→ ÍMPARES E DEPOIS PARES $\Theta(n)$

→ TESTAR ALÉAT. UMA POSIÇÃO. REPETIR ATÉ ACHAR.

→ DISTRIB. GEOMÉTRICA COM PROB. DE SUCESSO $p = 1/10$

⇒ ESPERANÇA : $1/p = 10 = O(1) = o(n)$

}

TRANSFORMAR LAS VEGAS (QUE BUSCA CERT. P/ SIM) EM MONTE CARLO:

RODAR O LV UM NÚMERO LIMITADO DE VEZES (t). SE NÃO ENCONTRAR, RESPONDE NÃO.

PROB. DE ERRO DO MC = PROB. DO LV PRECISAR DE MAIS TEMPO

SEJA X A VAR. DE TEMPO DO LV.

SE $t = k\mu$, ONDE $\mu = E(X) = 1/p$, TEMOS

$$k = \frac{t}{\mu} = t \cdot p$$

$$E_{MC} \leq \Pr(X \geq k\mu) \stackrel{\text{MARKOV}}{\leq} \frac{E(X)}{k\mu} = 1/k$$

TRANSFORMAR MC EM LV

$$\frac{1}{100} \quad \frac{99}{100}$$

→ MENOS EFICAZ. O LV PODE RODAR ? / SEMPRE

$$\frac{100}{99} \leftarrow 2$$

EX: SE A É BASEADO NO NÃO E REPETIMOS ELE ATÉ ENCONTRAR CERTIFICADO

O QUE ACONTECE COM INSTÂNCIAS SIM?

→ EM ALGUNS CASOS PODEMOS FAZER COM QUE AS EXECUÇÕES DE MC **NÃO** SEJAM IND. E AÍ O ALG. PODE PARAR DEPOIS DE FAZER TODAS AS ESCOLHAS.

→ ISSO SERIA UM CERTIFICADO

→ SE HÁ DOIS ALGS Alg_S E Alg_N BASEADOS NO SIM E NÃO RESP. EXECUTAMOS ALTERNADAMENTE

→ A PROB. DE SUCESSO EM CADA EXEC É $\geq p = \min\{P_S, P_N\}$

→ LOGO, O TEMPO ESPERADO É $1/p$.

CLASSES DE COMPLEXIDADE

- RP (RANDOMIZED POLYNOMIAL TIME): Probs PARA OS QUAIS EXISTEM ALGORITMOS QUE RESPONDEM SIM COM PROB $\geq \frac{1}{2}$ SE \mathcal{IEC}_S E NÃO COM PROB. \perp SE \mathcal{IEC}_N

\rightarrow Probs. PARA OS QUAIS EXISTE Alg DE MONTE CARLO BASEADO NO SIM.

- Co-RP: Probs. t.q. \exists Alg MONTE CARLO BASEADO NO NÃO.

- ZPP: (ZERO-ERROUR PROBABILISTIC POLYNOMIAL TIME)
Probs. PARA OS QUAIS EXISTE Alg. DE LAS VEGAS DE TEMPO ESPERADO POLINOMIAL.

EXISTE SORTEIO COM PROB. DE SUCESSO p T.q. $\frac{1}{p}$ é Poly em n

$$p = \frac{1}{f(n)}$$

CLASSES DE COMPLEXIDADE

• BPP (Bounded-error Probabilistic Polynomial Time)

Probs. Para as quais existe Alg de Monte Carlo bilateral onde tanto a prob de resp. SIM com IECs, quanto responder NÃO com IECN SÃO $> 3/4$.
 $\epsilon_1^t < \left(\frac{1}{4}\right)^t \rightarrow 0$

• PP (Probabilistic Polynomial Time)

Probs. Para as quais existe Alg de Monte Carlo bilateral onde tanto a prob de resp. SIM com IECs, quanto responder NÃO com IECN SÃO $> 1/2$.
 $\epsilon_1^t < \left(\frac{1}{2}\right)^t \rightarrow 0$

OBS: $BPP \subseteq PP$. A diferença é o número de repetições do Alg para

que o erro seja $< \epsilon$.

~* BPP = Polinomial

PP = Exponencial

PARADIGMAS COMBINATÓRIOS

I. BOLAS E LATAS

- Há m BOLAS IDÊNTICAS
- Há n LATAS
- AS BOLAS SÃO ASSOCIADAS ÀS LATAS ALEATORIAMENTE (COM PROB $1/n$)
- UMA LATA PODE RECEBER MAIS DE UMA BOLA
- QUANTAS LATAS FICAM VAZIAS?
- Qual o número esperado de latas com mais do que k bolas?
- Quantas bolas precisamos? / que existam mais latas com bolas do que vazias?
- Quantas bolas há na lata mais cheia?

EX: m PESSOAS, HÁ DUAS PESSOAS QUE FAZEM ANIVERSÁRIO NO MESMO DIA?

SEJA A_i O EVENTO NO QUAL A BOLA i CAI NUMA LATA DESOCUPADA.

$$p = \Pr\left(\bigwedge A_i\right) = \prod_{i=1}^m \Pr\left(A_i \mid \bigwedge_{j < i} A_j\right)$$

NOTE QUE $\Pr\left(A_i \mid \bigwedge_{j < i} A_j\right)$ É A PROBABILIDADE DE i CAIR EM UMA LATA, DADO QUE HÁ $i-1$ LATAS OCUPADAS, OU SEJA $\frac{m-i+1}{m}$
LOGO

$$p = \prod_{i=1}^m \left(\frac{m-i+1}{m}\right) = \frac{\binom{m}{m} m!}{m^m} < \frac{1}{2}$$

- Qual o valor de m para que seja mais provável haver duas pessoas com mesmo aniversário?

↳ DEVEMOS RESOLVER A EQUAÇÃO $p < \frac{1}{2}$, $m \approx \sqrt{2m \ln 2} = O(\sqrt{m \log m})$

COLECIONADOR DE CUPONS

- Há um conjunto de cupons $D = \{d_1, \dots, d_m\}$
- Gostaríamos de ter um de cada

~> seq. de var. X_1, \dots cada uma indica o índice que saiu.

~> cada índice tem prob. $1/m$ de sair.

DEFININDO $W_{m,k} = \#$ DE SORTEIOS REALIZADOS ATÉ QUE SE TENHA
OBTIDO k ITENS DISTINTOS, DOS m EXISTENTES

~> ESTAMOS INTERESSADOS EM $E[W_{m,m}]$.

~> PARECE COM BOLAS x LATAS

↳ SORTEIOS

↳ ITENS EXISTENTES

m LATAS, INFINITAS BOLAS

DEFINA $W_{m,k} = \#$ DE SORTEIOS REALIZADOS ATÉ QUE SE TENHA
OBTIDO k ITENS DISTINTOS, DOS m EXISTENTES

DEFINA $Z_i =$ QUANTIDADE DE SORTEIOS PARA QUE O $\#$ DE CUPONS
DISTINTOS VÁ DE $i-1$ P/ i .

~> QUANDO O COLECIONADOR POSSUI $i-1$ CUPONS DISTINTOS

A PROB. DE SORTEARMOS UM CUPOM DIFERENTE É $p_i = \frac{m - (i-1)}{m}$

~> LOGO, CADA Z_i É UMA V.A. GEOMÉTRICA COM PROB. DE SUCESSO p_i

$$\Rightarrow E(Z_i) = \frac{1}{p_i} = \frac{m}{m - (i-1)}$$

$$\Rightarrow W_{m,k} = Z_1 + \dots + Z_k \Rightarrow E(W_{m,k}) = \sum_{i=1}^k E(Z_i)$$

$$E[W_{m,k}] = \sum_{i=1}^k \frac{m}{m - i + 1}$$

$$\Rightarrow E[W_{m,m}] = \sum_{i=1}^m \frac{m}{m - i + 1} = m \sum_{i=1}^m \frac{1}{i}$$

$$= m \ln m + \Theta(m)$$

EX: MENSAGEM EM PACOTES CLIENTE → SERVIDOR

- CAMINHOS LONGOS (MUITOS ROTEADORES)
- UM PACOTE PODE GUARDAR A INFO DE NO MAX UM ROTEADOR
↳ O QUE OCUPARIA MUITO ESPAÇO/BANDA
- SE HÁ m ROTEADORES, CADA PACOTE GUARDA A INFO DE UM ROTEADOR COM PROB $1/m$.

⇒ O NÚMERO DE PACOTES QUE PRECISAM SER RECEBIDOS PARA TERMOIS INFO DE TODOS OS ROTEADORES É

$$m \log m + \theta(m)$$

- E SE O PACOTE NÃO SABE QUANTOS ROTEADORES TEM NO CAM.

COMO FAZER PARA QUE CADA ROTEADOR APAREÇA COM PROB. $1/m$?

ESTRATÉGIA:

- 1) GUARDA INFO DO PRIMEIRO ROTEADOR
- 2) NO k -ÉSIMO ROTEADOR, TROCA A INFO GUARDADA PELA INFO DO ROTEADOR ATUAL COM PROB $1/k$.

→ Para que a info do k -ésimo roteador chegue no servidor, precisamos trocar no passo z (Prob. $1/k$) e não trocar depois (Prob. $\prod_{i=k+1}^m \frac{i-1}{i}$)

$$\Rightarrow \Pr(\text{GUARDAR a } k\text{-INFO}) = \frac{1}{k} \cdot \frac{k}{k+1} \cdot \frac{k+1}{k+2} \cdots \frac{m-1}{m} = \frac{1}{m}$$

Análise Probabilística

- ALGORITMOS DETERMINÍSTICOS EXECUTAM DA MESMA FORMA PARA A MESMA INSTÂNCIA. POR ISSO OS ANALIZAMOS PELO PIOR CASO.
 - NA PRÁTICA PODE SER QUE O CASO TÍPICO SEJA MUITO MELHOR QUE O PIOR CASO. OU SEJA, QUE A PROBABILIDADE DE OCORRER O PIOR CASO SEJA MUITO BAIXA.
- ANALIZAR O ALG LEVANDO EM CONSIDERAÇÃO A DISTRIBUIÇÃO DAS INSTÂNCIAS É CHAMADO DE ANÁLISE PROBABILÍSTICA.

- DE ALGS DETERMINÍSTICOS E RANDOMIZADOS!

EX: QUICKSORT. PIOR CASO: $O(n^2)$
CASO MÉDIO: $O(n \log n)$

OBS: PODEMOS RANDOMIZAR O QUICKSORT!

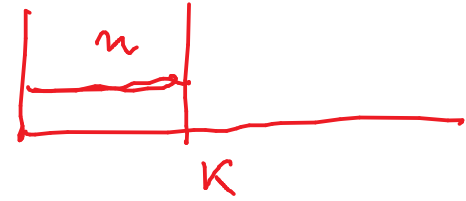
QUICKSORT - RANDOMIZADO (S)

1. REORDENA S DE FORMA ALEATÓRIA
2. QUICKSORT (S)

- DIFERENTES EXECUÇÕES DE ALGS. RANDOMIZADOS POSSUEM TEMPOS DIFERENTES P/ A MESMA INSTÂNCIA. POR ISSO ANALIZAMOS O CASO MÉDIO.

BUCKET SORT

- RODA EM TEMPO **LINEAR!**
- Qual o modelo para isso?

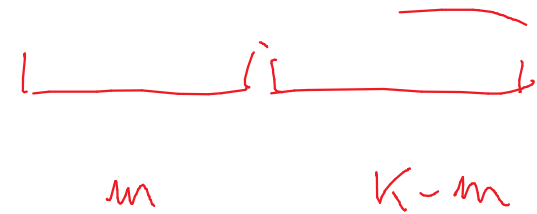


→ Uma ENTRADA COM $n = 2^m$ NÚMEROS

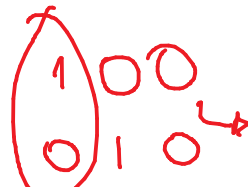
→ ESCOLHIDO DE FORMA ALEATÓRIA E UNIFORME NO INTERVALO $[0, 2^k[$
COM $k > m$.

- Há $n = 2^m$ LATA BUCKETS ROTULADOS DE 0 a $n-1$ EM BINÁRIO
- CADA NÚMERO É COLOCADO NO BUCKET QUE REPRESENTA SEUS PRIMEIROS m DÍGITOS BINÁRIOS.

→ FEITO EM TEMPO $O(m)$



$n = 1, k = 3$



$$1 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0 = 4$$

$$0 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 = 2$$

- SEJA X_J O # DE ELTS NO BUCKET J .

→ BOLAS e LATAS } X_J É V.A. BINOM. $B(m, 1/m)$
↓ ↓
ENTRADA BUCKETS

- ORDENAMOS CADA BUCKET INDIVIDUALMENTE (USANDO, POR EX, O QUICKSORT)

→ TEMPO QUADRÁTICO EM $X_J : c X_J^2$

- O VALOR ESPERADO PARA O TEMPO TOTAL $X = \sum c X_J^2$ É

$$E[X] = E\left[\sum_J c X_J^2\right] \stackrel{lin}{=} c \cdot n \cdot E[X_1^2] < 2 \cdot c m$$

→ LEMBREMOS QUE COMO X_i É V.A. BIN. $B(m, p)$, TEMOS

$$E[X_i^2] = m(m-1)p^2 + mp$$

$$\begin{cases} X = Y_1 + \dots + Y_m \\ X^2 = \sum_{i=1}^m Y_i^2 + \sum_{i < j} Y_i Y_j \\ E(X_i^2) = m E(Y_i) + m(m-1) p^2 \end{cases}$$

→ NESTE CASO, $p = 1/m$ E, PORTANTO, $E[X_i^2] = 2 - 1/m < 2$.

⇒ BUCKET SORT RODA EM TEMPO ESPERADO $O(n)$

O MÉTODO PROBABILÍSTICO

- PROVAS DE EXISTÊNCIA

- É POSSÍVEL DE-RANDOMIZAR ALGS. RANDOMIZADOS.

EX: É POSSÍVEL COLORIR AS ARESTAS DO K_{40} COM DUAS CORES SEM CLIQUE MONOCROMÁTICA DE TAMANHO 8?

$\rightarrow \binom{40}{2} = \frac{40 \cdot 39}{2} = 780$
- Há 780 arestas. Logo 2^{780} colorações

\rightarrow SEJAM $C_1, \dots, C_{\binom{40}{8}}$ AS CLIQUES DE TAMANHO 8 EM $G = K_{40}$.

\rightarrow SEJA M_i O EVENTO "TODAS AS ARESTAS DE C_i SÃO AZUIS OU TODAS AS ARESTAS DE C_i SÃO VERMELHAS"

$$\rightarrow \Pr(M_i) = \frac{1}{2^{\binom{8}{2}-1}} = 2^{-27}$$

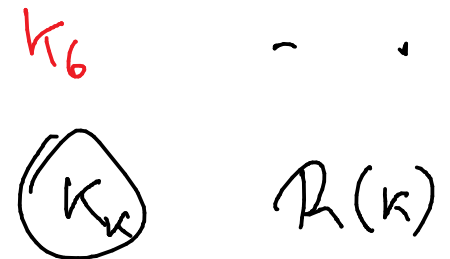
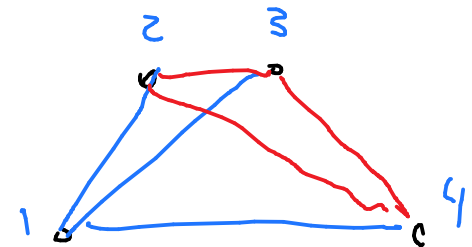
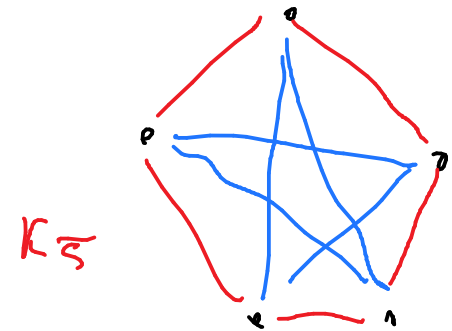
Queremos calcular $\Pr(\bigcap \bar{M}_i) = 1 - \Pr(\bigcup M_i)$

Pelo limite da união, temos

$$\Pr(\bigcup M_i) \leq \sum \Pr(M_i) = \frac{\binom{40}{8}}{2^{27}} < 0.573$$

Logo, $\Pr(\bigcap \bar{M}_i) > 0.427 > 0$

EX: TODO GRAFO CONTÉM UM GRAFO
BIP. COM METADE DAS ARESTAS.



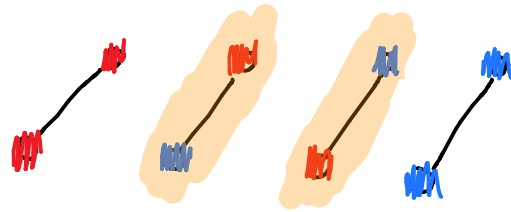
PROP: TODO GRAFO G CONTÉM UM SUBGRAFO BIPARTIDO H T.q $|E(H)| \geq \frac{|E(G)|}{2}$.

PROVA: CONSIDERE UMA 2-COLORAÇÃO ALEATÓRIA DE G

→ CADA VÉRTICE É COLORIDO COM VERMELHO OU AZUL COM PROB. $\frac{1}{2}$

OBS: CADA COLORAÇÃO ESPECÍFICA TEM (IGUAL) PROBABILIDADE $2^{-n} = \frac{1}{2^n}$

CADA ARESTA PODE SER COLORIDA DE QUATRO FORMAS:



→ COM IGUAL PROB.

O GRAFO H INDUZIDO PELAS ARESTAS BICOLORIDAS É BIPARTIDO.

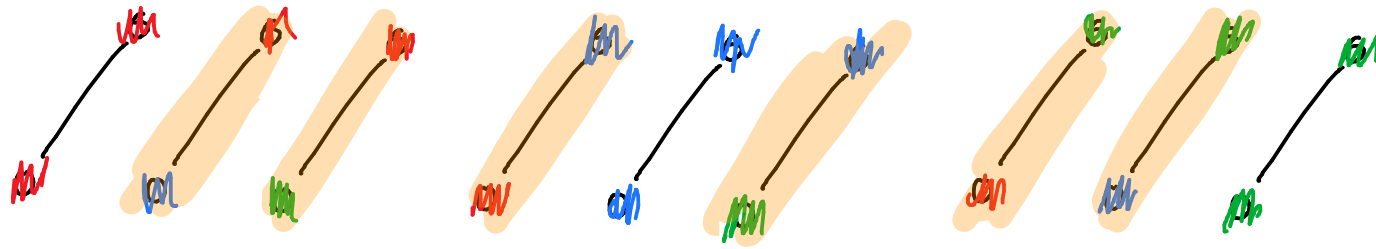
$X = \#$ ARESTAS BICOLORIDAS $= \sum_{e \in E(G)} I_e$

→ VARIÁVEL INDICADORA QUE INDICA SE $e \in E(H)$ SE e É BICOLORIDA

$$E(X) = \sum_{e \in E(G)} E(I_e) = \sum_{e \in E(G)} \Pr(e \text{ é Bic}) = \frac{1}{2} |E(G)| \Pr(e \text{ é Bicolorida}) = \frac{1}{2} |E(G)|$$

LOGO, HÁ 2-COLORAÇÃO C T.q. $X(C) \geq E(X) = \frac{1}{2} |E(G)|$. ▣

EX : SUBGRAFO TRIPARTIDO



$$Pr(e \text{ é bicolorida}) = \frac{6}{9} = \frac{2}{3}$$

EXISTE SUBGRAFO TRIPARTIDO COM $\frac{2}{3} |E(G)|$ ARESTAS

EXISTE SUBGRAFO K-PARTIDO COM $\frac{(K-1)}{K} |E(G)|$ ARESTAS. \square