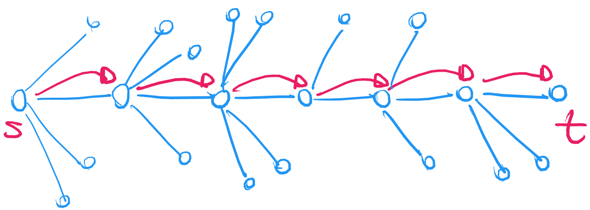
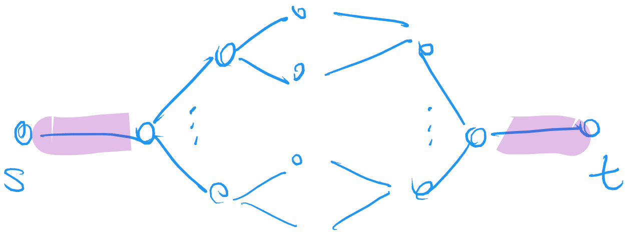
Algoritmos e Estruturas de Dados no Big Data
MLP - ESBD1 - Aula 5

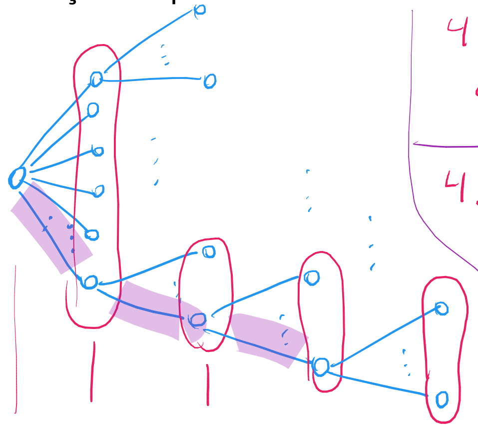
Importância dos breaks a cada 50 min / 1 hora

Overview:

- Analisar a Atividade 2 — motivada pelos 6 graus de separação
- Pilhas e recursão
- Recapitular definição de grafos
- Busca em profundidade: o que é alcançado a partir de um ramo?







$1 \times 1000 \times 1000 \times 1000 \times 1000$

$1 \times 1000 \times 100 \times 100 \times 100$

4 saltos c/ 1000 amigos $= 1000^4 = 1T$

4 saltos c/ 100 amigos distintos $\Rightarrow 100^4 = 10^8$

Analisando a Atividade 2:
https://colab.research.google.com/drive/1CzzCaPp-IrL2s87ovmeULyiqmOqnEUG-#scrollTo=Sy4Z7P9mdhxs

Calculando o # de arestas em cada teste

→ tam. grafo crescimento quadrático

	# Médio de Arestas por Vértice || # Total de Arestas		
# Vértices (n)	5 || 5n	raiz(n) || n * raiz(n)	n / 5 || n^2 / 5
100	5 × 100 = 500	10 * 100 = 1000	20 * 100 = 2000
1000	5 * 1000 = 5000	31,623 × 1000 = 31.623	200 * 1000 = 200000
10000	5 * 10000 = 50000	100 * 10000 = 1000000	2000 × 10000 = 20.000.000
100000	5 * 100000 = 500000	316.227 * 100000 = 31622700	20000 * 100000 = 2000000000

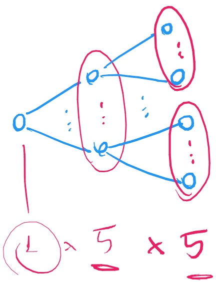
tam. do grafo cresce linearmente no # de vértices crescimento do # médio de arestas reduz o # de saltos

Calculando o grau de separação médio considerando todos os caminhos

	# Médio de Arestas por Vértice || # Total de Arestas		
# Vértices (n)	5 || 5n	raiz(n) || n * raiz(n)	n / 5 || n^2 / 5
100	2.27	1.82	1.56
1000	3.21	1.96	1.62
10000	4.3	2.0	1.63
100000	5.31	2.02	limite de memória

×10 ×10 ×10

+1 +1 +1



crescimento exponencial nos vértices

crescimento linear

crescimento "marginal"

sem crescimento

Análise: Observando a variação dos valores ao longo de uma linha,

- podemos notar que o número de saltos diminui
  - quando cresce o número médio de arestas por vértice.
- Atentando à variação ao longo de uma coluna
  - notamos que, em geral, o número de saltos aumenta
    - quando cresce o número de vértices.

Vale notar que, mesmo quando o número médio de arestas é constante,

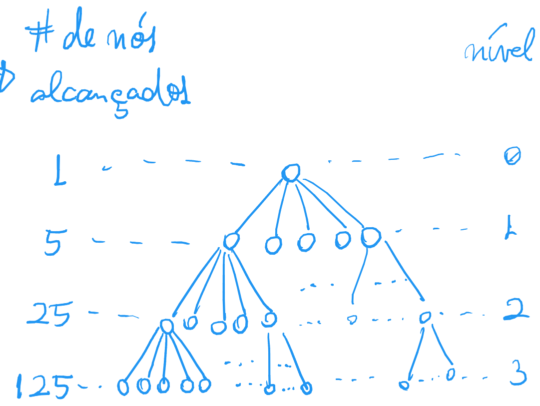
- o crescimento no número de saltos é
  - muito pequeno em relação ao crescimento do número de vértices.

crescimento $O(\lg n)$
sendo $n = |V|$

- Para entender o motivo, vale considerar
  - o crescimento exponencial do número de nós alcançados.

Ademais, quando o número médio de arestas cresce

- em função do número de vértices
  - o crescimento no número médio de saltos é ainda menor,
    - ou mesmo inexistente.



Calculando o grau de separação médio considerando apenas caminhos alternantes

32/2 = 16 arestas úteis em média

	# Médio de Arestas por Vértice || # Total de Arestas		
# Vértices (n)	5 || 5n	raiz(n) || n * raiz(n)	n / 5 || n^2 / 5
100	2.9	2.42	2.1
1000	4.24	2.54	2.06
10000	5.6	2.69	2.15
100000	7.03	2.49	limite de memória

$\frac{2000}{2} = 1K$ arestas úteis em média

Análise: Observando somente a segunda tabela,

- podemos fazer as mesmas observações feitas n a primeira.

- No entanto, comparando as tabelas percebemos que
  - o número de saltos médio aumenta quando
    - vamos da primeira para a segunda tabela

5 || 5n	raiz(n) || n * raiz(n)	n / 5 || n^2 / 5
2.27	1.82	1.56
3.21	1.96	1.62
4.3	2.0	1.63
5.31	2.02	limite de memória

5 arestas em média

100 arestas em média

Uma maneira de interpretar a restrição dos caminhos alternantes

- é pensar que, em média, cada vértice tem
  - metade dos vizinhos que teria sem a restrição.
- Uma consequência disso é que, fixada uma célula
  - da primeira coluna na primeira tabela,
- o valor dela é maior que o da célula
  - à direita da correspondente dela na segunda tabela.

No entanto, fixada uma célula da segunda coluna na primeira tabela,

- o valor dela é menor que o da célula à direita
  - da correspondente dela na segunda tabela. Por que?

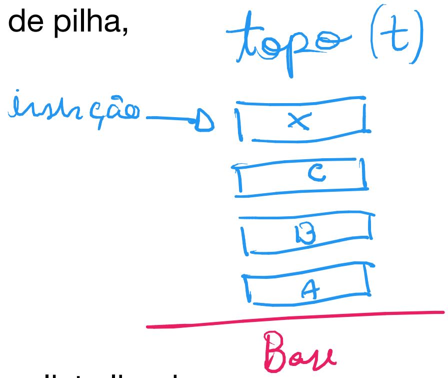
Uma explicação para essas últimas observações é que,

  - embora o maior número de arestas encurte os caminhos,
- a regra do caminho alternante torna inviáveis
  - caminhos de comprimento 1
    - quando os extremos têm o mesmo gênero
  - e caminhos de comprimento 2
    - quando os extremos têm gêneros distintos.
- Por isso, quando o número médio de saltos é próximo ou inferior a 2,
  - é muito difícil que o cenário com mais arestas
    - e caminho alternante "vença".

Pilhas: Comportamento e exemplo de pilha,

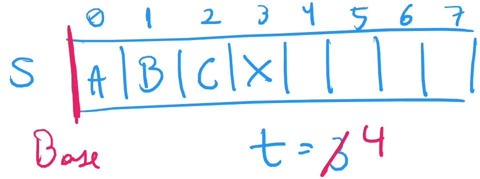
Stack

Last
In
First
Out



- Implementação em vetor e em lista ligada.

– em vetor :

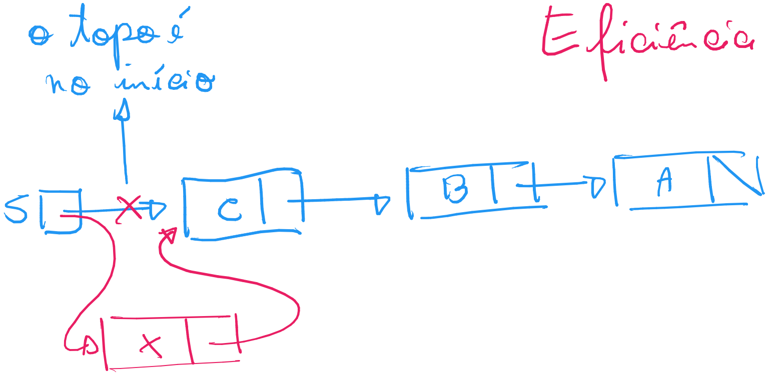


$S[t] = X$

$t += 1$

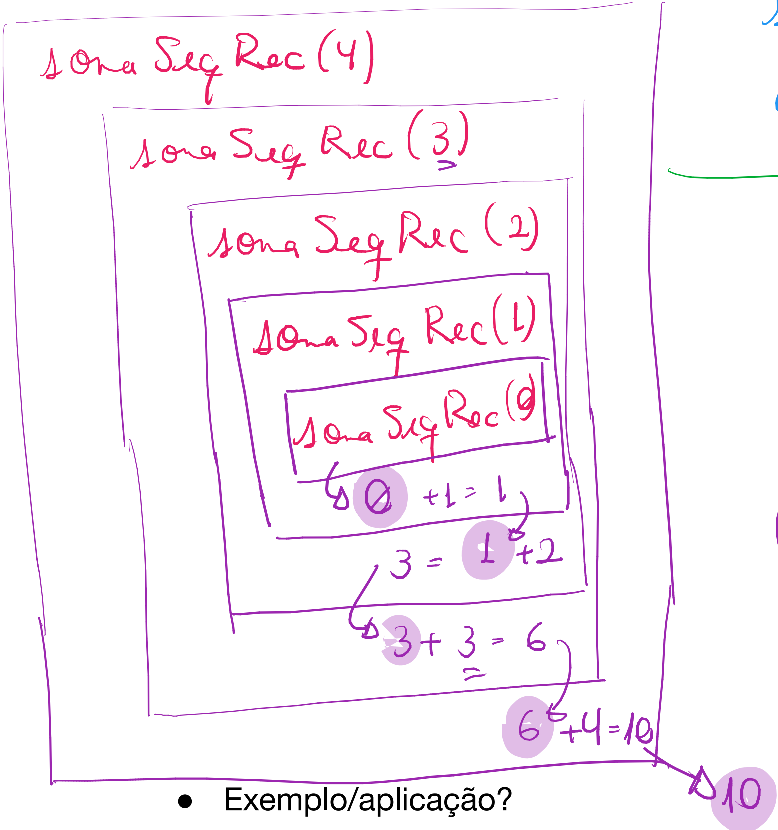
Eficiência: $O(1)$

– em lista ligada :



- Relação com recursão.

somaSeqRec(int n):
se n = 0 devolva 0
devolva n + somaSeqRec(n-1)



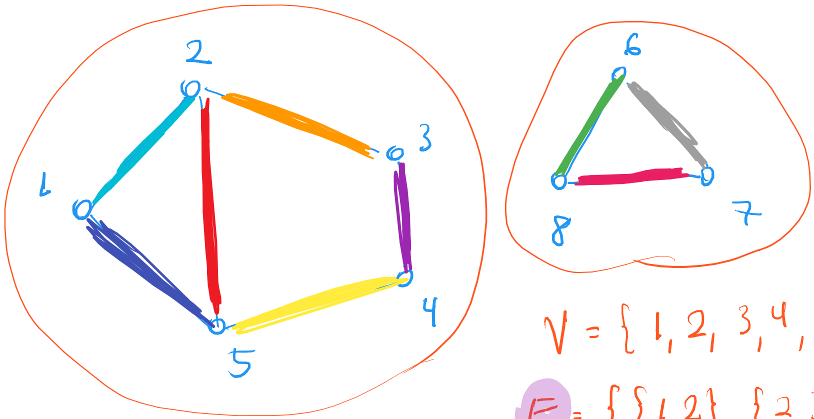
- Exemplo/aplicação?

http://bit.ly/EduMolinaCodigosAula5

Recapitulando definição e exemplos de Grafos: não orientados

o nós

— arestas



$G = (V, E)$

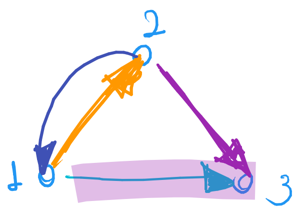
$V = \{1, 2, 3, 4, 5, 6, 7, 8\}$

$E = \{\{1,2\}, \{2,3\}, \{2,5\}, \{1,5\}, \{5,4\}, \{4,3\}, \{6,8\}, \{8,7\}, \{6,7\}\}$

- orientados ou dirigidos

o nós

→ arcos



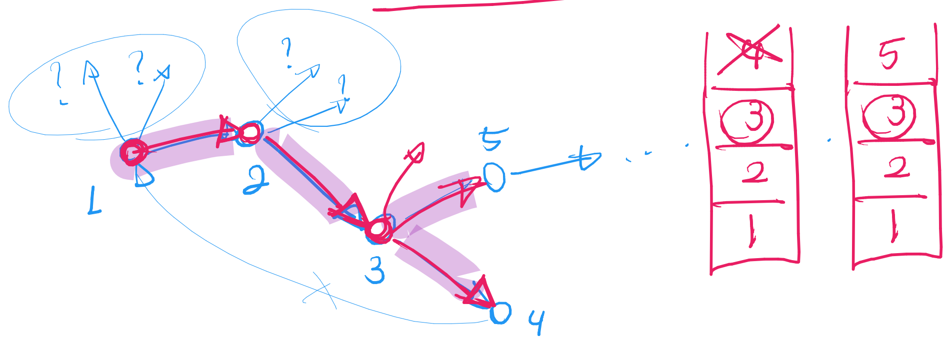
$D = (V, A)$

$V = \{1, 2, 3\}$

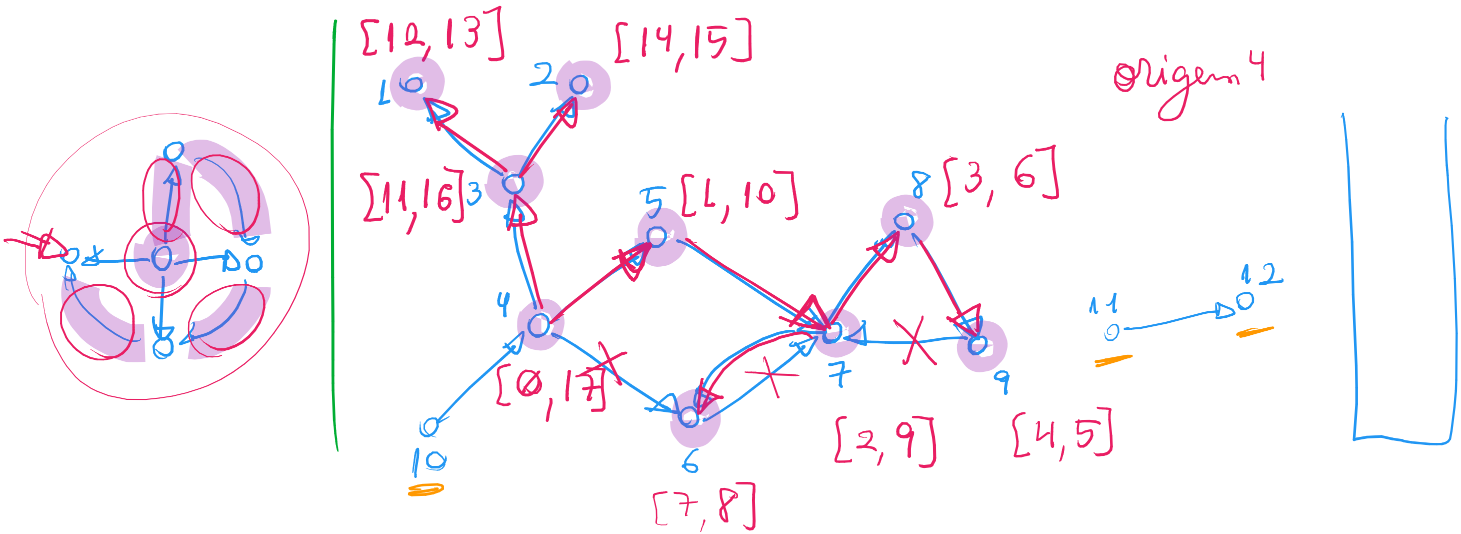
$A = \{(1,2), (2,3), (1,3), (2,1)\}$

Busca em profundidade: Ideia de visitar vértices ao longo de um caminho,

Depth
First
Search



- exemplo com cálculo de tempos de abertura e fechamento.



  - Quais vértices são garantidamente alcançados a partir de um ramo?

Pseudocódigo da busca em profundidade:

```
buscaProfRec (G=(V,E), origem v):
    visitado[v] = 1
    abertura[v] = t ; t += 1
    p/ cada aresta (v,w) ∈ E:
        se w ñ foi visitado:
            buscaProfRec(G, w)
    fechamento[v] = t ; t += 1
```

```
Inicialização
p/ todo v ∈ V:
    visitado[v] = 0
t = 0
// inicializa abertura e fechamento
```

variáveis globais ou passadas por referência

- análise de eficiência. — $O(n+m)$, sendo $n = |V|$ e $m = |E|$

Códigos do Edu - https://colab.research.google.com/drive/1QdOsCbcWFOLkdal1X2S30a7hZ359UwTb?usp=sharing (http://bit.ly/EduMolinaCodigosAula5)MLP ESBD1 Aula 5 Códigos

Material complementar

Pilhas, recursão e busca em profundidade:

- [PDF] Pilha implementada em vetor, aplicação com parênteses e colchetes, pilha de execução, relação de pilha com recursão - http://bit.ly/MarioSanFeliceCompPilhaVetorPDF
- [PDF] Pilhas em listas encadeadas (com e sem nó cabeça) - http://bit.ly/MarioSanFeliceCompPilhaListaPDF
- [Playlist] Busca em profundidade, conectividade - http://bit.ly/MarioSanFeliceBuscaProfVideo
- [PDF] Busca em profundidade, conectividade - http://bit.ly/MarioSanFeliceCompBuscaProfPDF

Ordenação Topológica e DAGs

- [Playlist] Ordenação topológica, DFS, DAGs aleatórios - http://bit.ly/MarioSanFeliceOrdTopVideo
- [PDF] Ordenação topológica, DFS, DAGs aleatórios - http://bit.ly/MarioSanFeliceCompOrdTopPDF