

MC202 — Estruturas de Dados

Laboratório 4 - O Jogo do Bacon

Primeiro semestre de 2017 - Turmas B e C
Professor: Emilio Franceschini
franceschini@ic.unicamp.br

1 O problema

Existe uma conjectura chamada “6 graus de separação” que afirma que qualquer um está a no máximo 6 conhecidos de distância de qualquer outra pessoa no mundo¹. Na verdade, Watts and Strogatz mostraram que o custo do caminho médio entre dois vértices de um grafo onde as conexões são feitas aleatoriamente e cada aresta tem custo 1 é igual a $\frac{\ln N}{\ln K}$, onde N é o número total de vértices e K é o número de arestas incidentes em cada vértice. Portanto, se $N = 208.000.000$ (população do Brasil em 2015) e $K = 30$ então o caminho médio tem custo $\frac{19,2}{3,4} = 5,6$. Contudo se tomarmos $N = 7.200.000.000$ (estimativa da população mundial em 2013) e $K = 30$, temos que o caminho médio entre dois vértices é de $\frac{22,5}{3,4} = 6,7$.

Esta ideia acabou sendo levada para outros lugares. Por exemplo, matemáticos costumam medir colaborações entre eles medindo a distância até um dos matemáticos mais prolíficos: Paul Erdős². Neste caso dois pesquisadores estão ligados no grafo caso tenham sido co-autores de um artigo. Começou-se inclusive a calcular a probabilidade de se ter um Nobel baseada na distância até Erdős.

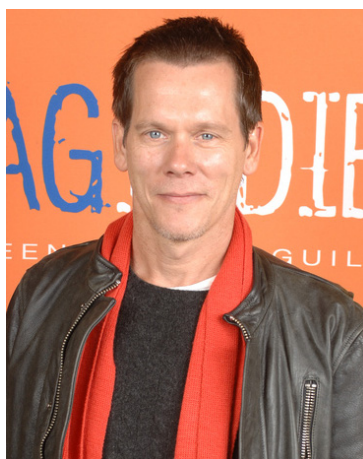


Figura 1: Kevin Bacon mal pode esperar para jogar.

¹https://en.wikipedia.org/wiki/Six_degrees_of_separation

²https://en.wikipedia.org/wiki/Erdős_number

Não demorou muito até levarem essa ideia para Hollywood, usando como marco zero o famoso Kevin Bacon³! Sites dedicados a brincadeira surgiram, apesar do Kevin Bacon não ter gostado muito⁴...

Apesar de ser apenas uma brincadeira, o problema de buscar caminhos em grafos se apresenta frequentemente em computação. Ele é usado desde GPSs até bolsas de valores. Neste laboratório faremos algo semelhante ao site “Oracle of Bacon”⁵.

Ao contrário dos laboratórios anteriores, não há restrições nem tampouco sugestões a respeito de como você deverá organizar o seu código. Você deve levar em consideração todas as estruturas de dados que você viu ao longo do semestre para escolher as estruturas mais apropriadas e criar um código que seja correto e eficiente. Lembre-se: vocês podem ver as estruturas que estudamos ao longo do semestre como modelos, como pontos iniciais da sua implementação. Certamente você não conseguirá utilizar as estruturas exatamente como foram dadas, adapte-as para resolver o seu problema!

2 Entradas e saídas

O seu programa deverá receber dois parâmetros. O primeiro será o nome do arquivo contendo o banco de dados e o segundo o nome do arquivo contendo o nome dos atores a serem pesquisados. A saída deve ser escrita para a `stdout` utilizando funções como `printf`.

2.1 Arquivo de banco de dados

O formato do arquivo de banco de dados é o seguinte. Cada linha trará na primeira coluna o nome do ator e na segunda coluna o nome do filme que ele participou. As colunas são separadas pelo caractere `tab` (equivalente ao `'\t'`, caractere número 9 da tabela ASCII). Para processar cada linha e separar as colunas facilmente veja a função `strtok` contida no `string.h`. Para isto, basta fazer:

```
char *nomeAtor = strtok(linha, "\t");
char *nomeFilme = strtok(NULL, "\t");
```

para cada uma das linhas lidas a partir do arquivo do banco de dados. O formato é semelhante ao mostrado abaixo:

³https://en.wikipedia.org/wiki/Six_Degrees_of_Kevin_Bacon#Bacon_numbers

⁴<http://edition.cnn.com/2014/03/08/tech/web/kevin-bacon-six-degrees-sxsw/>

⁵<http://oracleofbacon.org/>

Sergio Mallandro	Lua de Cristal
Sergio Mallandro	Menino do Rio
Sergio Mallandro	Muita Calma Nessa Hora
Xuxa	Lua de Cristal
Xuxa	O Fuscão Preto
Xuxa	O Mistério de Robin Hood
Xuxa	Os Trapalhões e o Mágico de Oróz
Letícia Spiller	Flordelis: Basta Uma Palavra Para Mudar
Letícia Spiller	Lua de Cristal
Letícia Spiller	O Casamento de Gorete
Letícia Spiller	O Inventor de Sonhos
Matheus Nachtergaele	Castelo Rá-Tim-Bum, O Filme
Matheus Nachtergaele	O Auto da Compadecida
Matheus Nachtergaele	O Inventor de Sonhos
Zezé Motta	A Comédia Divina
Zezé Motta	A Força do Xangô
Zezé Motta	Águia na Cabeça

2.2 Arquivo de atores a serem pesquisados

O segundo parâmetro será o nome de um arquivo que conterá um nome por linha. Você pode assumir que todos os atores incluídos no arquivo existem no banco de dados. Contudo, é possível que não haja caminhos entre dois atores.

Sergio Mallandro
 Letícia Spiller
 Xuxa
 Matheus Nachtergaele
 Sergio Mallandro
 Zezé Motta

2.3 Saída esperada

O segundo parâmetro trará o nome de um ator por linha. O seu programa deve ler este arquivo e imprimir o caminho (incluindo o nome dos filmes) do ator da linha 1 até o ator na linha 2, do ator da linha 3 ao ator da linha 4, e assim sucessivamente para todas as linhas do arquivo. Quando não houver um caminho entre dois seu programa deverá imprimir CAMINHO

NAO ENCONTRADO. Lembre-se estamos interessados em um **CAMINHO MINIMO** e não em um caminho qualquer. Em alguns casos existem vários caminhos de comprimento mínimo, neste caso você pode imprimir qualquer um deles.

Utilizando como exemplo os arquivos de banco de dados e atores a serem pesquisados dados nas Seções 2.1 e 2.2 a saída esperada é:

```
Sergio Mallandro
trabalhou em "Lua de Cristal"com
Leticia Spiller
Xuxa
trabalhou em "Lua de Cristal"com
Leticia Spiller
trabalhou em "O Inventor de Sonhos"com
Matheus Nachtergaele
CAMINHO NAO ENCONTRADO
```

Os arquivos de teste descritos acima pode ser baixados aqui: [BD Exemplo](#), [Consulta Exemplo](#), [Saída Exemplo](#).

Bônus: Além disto, preparei um script que baixa o banco de dados completo do IMDB e o processa para colocá-lo no formato descrito acima. Assim você vai poder brincar com o seu *Jogo do Bacon* de verdade. O script gera três arquivos de interesse: (i) `atores.completo` a base de dados completa do IMDB, `atores.2010` a base de dados com todos os filmes desde 2010 e finalmente `atores.2015` a base com os dados de filmes desde 2015. O script pode ser baixado [aqui](#).

Atenção, o banco de dados do IMDB é imenso, na minha máquina leva por volta de 10 minutos para baixar e gerar os arquivos acima. Os arquivos também não são pequenos: `atores.completo` (~ 300MB), `atores.2010` (~ 100MB) e `atores.2015` (~ 35MB). Contudo, haverá **pontos de bonificação extras** para quem conseguir processar o arquivo gerado pelo script (`atores.2010`) e **pontos super-extras** para aqueles programas que forem capazes de processar o arquivo completo (`atores.completo`).

Ainda que o uso desses arquivos seja estritamente educativo, a licença de uso exige que o uso da base de dados seja estritamente pessoal e não permite que versões modificadas dos arquivos sejam distribuídas online. Por isso, legalmente, os arquivos pré-processados não serão colocados online, mas você pode rodar o script na sua própria máquina ou, caso você não use Linux, você pode pedir para um colega rodá-lo e enviar os arquivos gerados diretamente para você.

3 Entrega e avaliação

Este laboratório poderá ser feito em duplas. A entrega deverá ser feita pelo Susy **apenas** por um dos integrantes da dupla. O arquivo enviado deve conter um cabeçalho contendo o nome completo e RA de cada um. Verifique a página da disciplina (<http://www.ic.unicamp.br/~francesquini/mc202/>) para informações sobre as datas de entrega e acesso ao Susy.

Todos os trabalhos entregues passarão por uma correção automática no Susy. A nota será proporcional ao número de testes que passarem com sucesso. Contudo, os monitores e o professor poderão, ao seu próprio critério, alterar as notas para cima ou para baixo. Passar na metade do testes não significa que você tirou nota 5 automaticamente, indica apenas a provável nota que será atribuída ao seu trabalho.

ATENÇÃO: Plágios serão severamente punidos com a reprovação na disciplina.

Bom trabalho!