

Exercício de Programação 1: Processando dados da pesquisa Origem Destino

29 de Março de 2020

1 Introdução Téorica

Na disciplina de AED1 estudamos uma série de estruturas de dados como pilhas, filas e listas ligadas. Vimos que a escolha da estrutura de dados pode mudar muito o tempo de processamento de uma tarefa computacional podendo inclusive tornar viável uma solução que usando outras estruturas eram inviáveis. Tendo visto em IAA como estimar a eficiência de um algoritmo em termos assintótico, pudemos aplicar esta técnica em operações de diferentes estruturas. Nesta disciplina estudaremos outras estruturas de dados, mas para este primeiro exercício usaremos as estruturas que já conhecemos.

2 Introdução ao Problema

2.1 COVID19

Os dados de quinta-feira indicam 505 mil casos confirmados de contágio e 23 mil mortes por COVID19 em todo o mundo. A COVID19 é causada pelo novo coronavírus e é transmitida por meio de gotículas de saliva de pessoa para pessoa. Assim, a probabilidade de uma pessoa passar o vírus para outra é muito maior se elas se encontraram presencialmente. Para diminuir a taxa de contaminação da doença, a Organização Mundial da Saúde tem indicado que as pessoas no mundo todo se esforcem para diminuir deslocamentos e evitar locais de aglomeração, uma estratégia chamada de “distanciamento

social”. Foi por este motivo que o CRUESP decidiu que as universidades paulistas deveriam suspender suas aulas, o que nos afetou diretamente.

Aparentemente a epidemia de COVID19 segue um modelo de transmissão chamado de SIR (susceptible, infected, recovered). Nesse modelo as pessoas se dividem em três categorias: suscetíveis, infectadas e recuperadas. Essa última categoria é das pessoas que pegaram o vírus, se recuperaram e estão imunizadas. Um gráfico dinâmico publicado pelo Washington Post ilustra esse processo de contágio mostrando a evolução exponencial no número de infectados (<https://www.washingtonpost.com/graphics/2020/world/corona-simulator>). O gráfico mostra a eficácia da estratégia do distanciamento social para achatar a curva do número de contaminado e, assim, diminuir o impacto de um possível colapso no sistema de saúde.

A simulação publicada no Washington Post representa pessoas se deslocando aleatoriamente em um espaço bidimensional finito. Um modelo mais realista deveria considerar o deslocamento real das pessoas pela cidade.

2.2 Pesquisa Origem Destino de São Paulo

A cada 10 anos é realizado uma pesquisa na cidade de São Paulo para avaliar as rotas que os paulistanos fazem pela cidade. A última pesquisa Origem/Destino foi realizada em 2017 e entrevistou pessoas em 32 mil domicílios. Além de coletar dados demográficos das pessoas entrevistadas (idade, faixa de renda etc.), a pesquisa coleta dados sobre suas viagens frequentes: local de origem, local de destino, pontos de transferência, motivo da viagem, forma de deslocamento etc. Os dados públicos dessa pesquisa estão disponíveis em <http://www.metro.sp.gov.br/pesquisa-od/>.

Com os dados da pesquisa Origem/Destino temos uma melhor aproximação do padrão de encontro de pessoas que moram na região metropolitana de São Paulo. Por um lado, é bom lembrar que estamos usando dados amostrais (eles representam menos de 1% da população). O número de encontros em nossa simulação será muito menor do que os encontros reais, caso tivéssemos os dados de toda a população. Por outro lado, vamos considerar que pessoas que frequentam os mesmos lugares se encontrem mesmo que elas frequentem em horários diferentes ou andares diferentes de um prédio. Essas são limitações metodológicas graves, mas julgamos que os dados servem como um bom exercício de programação.

3 Tarefa

A primeira tarefa de vocês será processar os dados da pesquisa Origem/ Destino. O arquivo está em formato DBF que é uma linguagem para representar dados em tabelas. Cada registro na tabela possui dados sobre uma viagem: dados pessoais do entrevistado (idade, renda etc.), motivo da viagem, coordenadas da origem, coordenadas do destino, coordenadas dos pontos de transferência etc. Além de folhear o material disponível online sobre os dados disponíveis da pesquisa, vocês deverão estudar como processar dados DBF usando a linguagem de programação da preferência de vocês.

A primeira tarefa consiste em organizar os dados para os fins que desejamos. Para isso, criem uma classe ou estrutura `Local` com os seguintes atributos: `coordenada_x`, `coordenada_y`, `frequentadores`. Os primeiros são inteiros que indicam as coordenadas do lugar, o segundo é um arranjo de inteiros que indica os ids das pessoas que passam por aquele lugar. Usem alguma das estruturas que vocês estudaram em AED1 (arranjo, lista ligada, lista duplamente ligada etc.) para listar esses objetos e atualizar a lista de frequentadores.

A primeira tarefa é construir um histograma do número de pessoas que frequentam cada lugar mencionado na pesquisa. Ou seja, um gráfico cujo eixo x é a quantidade de lugares e o eixo y o número de pessoas que frequentam (quantos lugares são frequentados por um único entrevistado, quantos são frequentados por 2, por 3 e assim por diante).

Quanto tempo demorou para processar tudo? Vocês conseguem estimar assintoticamente o tempo de processamento do programa que vocês escreveram em termos do número de entrevistados e e do número de lugares mencionados l ? Mesmo que não implementem, conseguem pensar uma solução assintoticamente mais eficiente? Como seria uma solução mais eficiente? Qual seria o tempo de processamento dessa solução?

4 Entrega

O código de vocês deve estar disponível em um servidor de controle de versão público (como o Github). Pelo TIDIA-AE vocês devem entregar um relatório em PDF(!) de uma ou duas páginas descrevendo o que fizeram, os resultados (o histograma, o tempo real de processamento total, a estimativa do tempo de processamento assintótico e sugestões de algoritmos e estruturas de dados

que fariam o programa mais eficiente e porque).

A entrega não tem prazo e é facultativa. Aqueles que entregarem serão recompensados com pontos na média.