

Laboratório Semanal 05

Prof. Arnaldo Moura e Prof. Lehilton Pedrosa

Prazo para entrega: 22/05/2016 às 22:00:00

1 ANÁLISE DE DADOS

Vivemos em uma era em que uma grande quantidade de dados está sendo gerada, a partir de diferentes fontes de dados (imagens de satélites, GPS de Smartphones, etc), e armazenada a todo momento. Nós seres humanos temos dificuldade em processar e analisar essa grande quantidade de dados. Portanto, utilizamos técnicas computacionais para auxiliar-nos no processamento e entendimento destes dados. Técnicas para análise de dados estão sendo utilizadas em diversos domínios do conhecimento (medicina, análise de redes sociais, bussiness, etc) para, por exemplo, apoiar a tomada de decisões ou permitir a descoberta de novos padrões dos dados estudados. Por exemplo, na área de bussiness, uma grande empresa utilizou técnicas computacionais para analisar os dados de suas transações de vendas e aumentou as vendas de fraldas descartáveis simplesmente colocando-as próximas de cervejas.

Após ler sobre as diferentes aplicações da análise de dados, você deseja utilizar seu conhecimento computacional para analisar os dados de uma turma e ajudar os estudantes a deixarem o posto de “Forever Alone”, encontrando uma “sintonia” entre pares de pessoas. Você está interessado em saber qual a pessoa que tem mais sintonia com a pessoa mais popular da sala. O primeiro passo que você fez foi aplicar um questionário na turma em que as pessoas definiam o grau de afinidade entre ela e seus colegas atribuindo uma nota. Esta nota é um valor inteiro que varia de 0 (menor afinidade) a 10 (maior afinidade). O grau de afinidade entre uma pessoa e ela mesma é o valor máximo. Você usará a afinidade para definir a popularidade de uma pessoa.

Para chegar em seu objetivo final, você irá analisar os dados deste questionário para tentar descobrir uma “sintonia” entre as pessoas da turma com a pessoa mais popular. Para isto você deverá realizar os três passos seguintes: 1- organizar os dados do formulário em uma estrutura que seja melhor para realizar a análise deste tipo de dado; 2- encontrar a pessoa mais popular, ou seja, a pessoa que possui a maior afinidade com os demais colegas; 3- Calcular a sintonia entre as pessoas utilizando a numerologia.

ETAPA 01: ORGANIZAÇÃO DOS DADOS

Uma maneira de organizar os dados, para posteriormente analisá-los, é utilizando a estrutura de grafos. Um grafo $G(V, A)$ é definido pelo par de conjuntos V e A , onde V é um conjunto não vazio conhecidos como os vértices do grafo e A é conjunto de pares não ordenados $a = (v, w)$, v e $w \in V$ conhecidos como as arestas

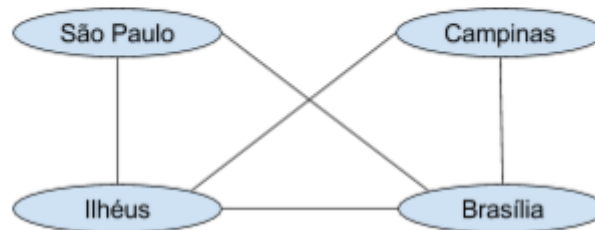
do grafo. Uma aresta sempre conecta dois vértices. A Figura 1.1 abaixo ilustra um grafo onde os vértices são as cidades com aeroportos e as arestas indicam se há linha aérea ligando essas cidades.

Figura 1.1: Exemplo de Grafo

$G(V,A)$

$V = \{v \mid v \text{ é uma cidade com aeroporto}\}$

$A = \{(v,w) \mid \text{<há linha aérea ligando } v \text{ a } w\}$



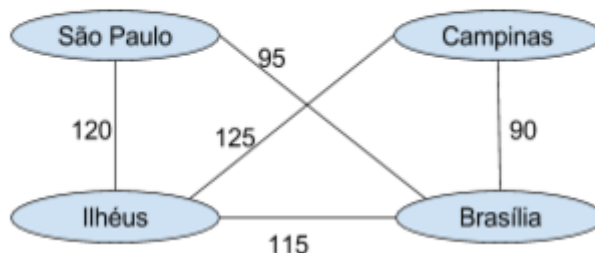
As arestas podem ter um peso (número) associado que indica alguma relação entre os vértices conectados pela aquela aresta. Por exemplo, a Figura 1.2 ilustra um grafo em que o peso das arestas indicam o tempo de voo, em minutos, entre as cidades.

Figura 1.2: Exemplo de Grafo Ponderado

$G(V,A)$

$V = \{v \mid v \text{ é uma cidade com aeroporto}\}$

$A = \{(v,w,t) \mid \text{<há linha aérea ligando } v \text{ a } w, \text{ sendo } t \text{ o tempo de voo}\}$



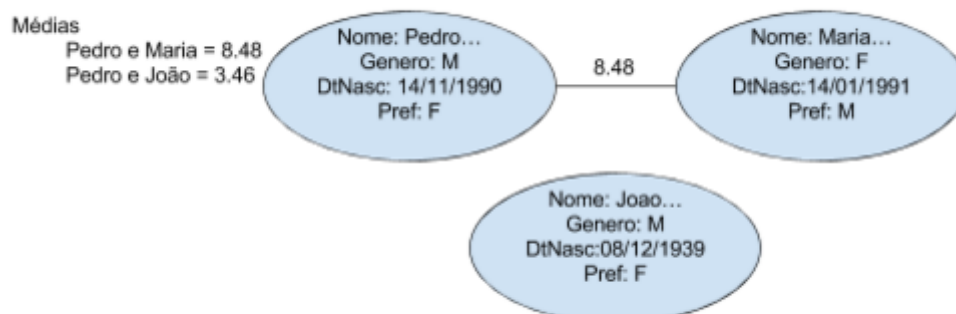
Portanto, sua tarefa nesta etapa é organizar os dados do formulário no formato de grafos. O seu formulário possui os dados da pessoa que responderá o formulário e uma lista das pessoas da turma com um campo para atribuição da nota de afinidade. A Figura 1.3 abaixo ilustra um formulário para uma turma de 4 pessoas.

Figura 1.3: Formulário

PARTE 01	
Nome Completo: _____	
Data Nascimento: ____/____/____	
Gênero: (<input type="checkbox"/>)M (<input type="checkbox"/>)F	
Preferência: (<input type="checkbox"/>)M (<input type="checkbox"/>)F // Marque apenas 1 opção	
// indica se você tem interesse em pessoas do sexo	
//masculino ou feminino.	
PARTE 02	
Pessoa	Afinidade [0-10]
Pedro Ferreira	_____
Maria Santos	_____
Micalateia Silva	_____
João Sousa	_____
Ana Clara Costa	_____

Seu grafo **obrigatoriamente** deverá ser representado por duas estruturas: Pessoa e Aresta. A estrutura Pessoa representa os vértices do grafo e contém as informações contidas na Parte 01 do formulário (ver Figura 1.3). A data de nascimento deve ser uma nova estrutura contendo dia, mês e ano. O nome completo da pessoa possui apenas letras (sem acentos) e espaços. A estrutura Aresta representa as arestas do grafo com um **ponteiro** para cada pessoa e o peso de afinidade entre essas pessoas (**é obrigatório o uso de ponteiros nesta estrutura**). Uma aresta só existirá se a média geométrica das notas dadas por duas pessoas for maior ou igual a 5. Por exemplo, sabendo que Pedro atribuiu 8 para sua afinidade com Maria e 3 para sua afinidade com João; Maria e João atribuíram 9 e 4, respectivamente, para suas afinidades com Pedro, a figura 4 ilustra o grafo que deverá ser montado para estes dados:

Figura 1.4: Exemplo criação de grafo



ATENÇÃO: Uma vez criado o seu grafo, você não deverá utilizar as informações do formulário. Trabalhe diretamente com o grafo, não seguir essa regra terá penalidades.

DICA: Você poderá utilizar um vetor para armazenar os vértices e um vetor para armazenar as arestas de

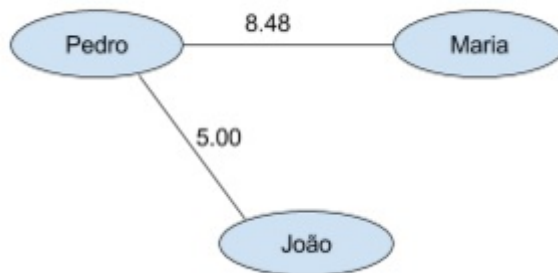
seu grafo;

ETAPA 2 - PESSOA MAIS POPULAR

Agora que você organizou seus dados em forma de grafo, você quer descobrir a pessoa mais popular da turma, ou seja, aquela que possui mais afinidade com a turma. Em grafos, a centralidade de um vértice indica a sua importância no grafo. Nesta etapa, você deve utilizar a centralidade de grau para determinar a pessoa mais popular da sala. **A Centralidade de Grau é definida como o número de ligações incidentes sobre um nó.** A centralidade de grau de um vértice, para um dado grafo $G(V, A)$ com V vértices e A arestas, é definido como $C_D(v) = \deg(v)$. A figura 1.5 ilustra um grafo em que a pessoa mais popular é Pedro. Por questões de visualização, apenas o nome do aluno será exibido nos vértices do grafo.

Figura 1.5: Exemplo pessoa popular

$C_D(\text{Pedro}) = 2$
 $C_D(\text{Maria}) = 1$
 $C_D(\text{João}) = 1$



ETAPA 3: CALCULAR A SINTONIA ENTRE PARES DE PESSOAS

Dado que você conhece a pessoa mais popular da turma, agora deve descobrir o seu grau de sintonia com cada outra pessoa que se adequa à sua preferência sexual e que possuem afinidade ≥ 5.0 entre si. Seja A a pessoa mais popular e suponha que sua preferência seja pelo sexo masculino, você deverá calcular a sintonia da pessoa A com todos os homens da sala de aula. O cálculo da sintonia é baseado em numerologia aplicada ao primeiro nome, data de nascimento e idade de cada pessoa envolvida. **O cálculo se inicia pela codificação do primeiro nome de cada pessoa gerada pelo algoritmo Soundex.** Esse algoritmo foi utilizado primeiramente na língua inglesa para indexar nomes através de sua pronúncia.

Etapa 3. a) Algoritmo Soundex

O Soundex codifica um nome por uma letra seguida de 3 dígitos (Xddd). A letra (X) é a primeira letra do nome e os dígitos (d) correspondem a codificação das consoantes segundo uma tabela. Nesta tabela, as consoantes com pronúncia semelhante recebem a mesma codificação. Para esta análise, usaremos as seguintes regras e tabelas para a codificação:

1. **Retenha a primeira letra do nome (corresponde ao X em Xddd);**
2. **Desconsidere todas as ocorrências de: a, e, i, o, u, y, h, w;**
3. **Atribua valores para as consoantes de acordo com a tabela abaixo:**
 - 1 -> g, j;
 - 2 -> s, z, x;
 - 3 -> c, k, q;
 - 4 -> b, p;
 - 5 -> m, n;
 - 6 -> d, t;
 - 7 -> f, v;
 - 8 -> l;

9 -> r;

4. Se, após a regra 2, existirem duas ou mais consoantes adjacentes com a mesma codificação, codifique apenas a primeira e desconsidere as demais. Esta regra também se aplica para a primeira letra.
5. Após obter a codificação Xddd para as 3 primeiras consoantes do nome, as demais serão ignoradas.
6. Para nomes curtos, adicione zeros para completar a codificação em 3 dígitos (Xddd).

Exemplo de codificação:

Robert => R496

R segundo a regra 1;

string "brt" segundo a regra 2;

496 segundo a codificação da regra 3;

Jessica => J230

J segundo a regra 1;

string "ssc" segundo a regra 2;

223 segundo a regra 3;

23 segundo a regra 4;

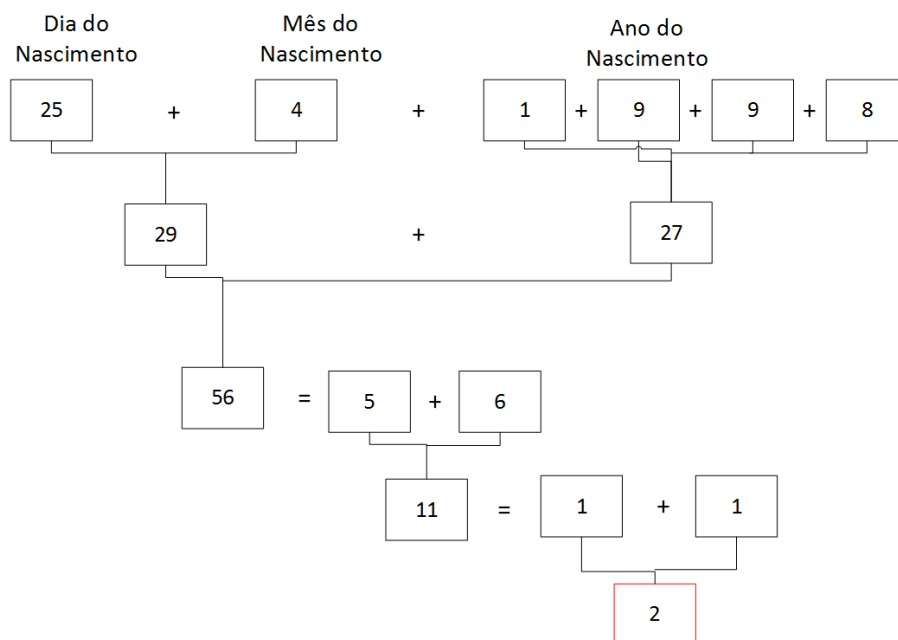
230 segundo a regra 6;

Etapa 3. b) Numerologia usando data de nascimento e idade

Segundo a numerologia, a soma dos números da data de nascimento pode dizer muito sobre uma pessoa. O número final representa o caminho da vida, que indica traços positivos e negativos da personalidade de uma pessoa. Para realizar este cálculo, você deve somar o dia, mês, e cada dígito, separadamente, do ano do nascimento. A partir deste valor, deve-se realizar a soma de cada dígito deste resultado até que se obtenha um único dígito.

A Figura 1.6 abaixo ilustra um exemplo para uma pessoa que nasceu em 25/04/1998:

Figura 1.6: Codificação da Data de Nascimento



Após este processo, componha uma codificação de 3 dígitos adicionando a sua idade no final do dígito encontrado. Se sua idade possui apenas um dígito, adicione zeros no final da codificação; se sua idade possui mais de 2 dígitos, desconsidere os dígitos finais.

Desta forma, a pessoa que nasceu em 25 de abril de 1998, terá a seguinte codificação: 218.

Etapa 3. c) Sintonia entre 2 pessoas usando as codificações das etapas a e b

Esta etapa calcula a similaridade entre as codificações anteriores. Esta similaridade representa o grau de sintonia entre duas pessoas. Quanto maior este grau, maior chance de ter encontrado a pessoa certa para você! Para calcular a similaridade/sintonia entre duas pessoas utilize a seguinte fórmula, onde a cada elemento que compõe a fórmula possui um peso específico:

$$Sintonia = \frac{3 * Sim(CodNome1, CodNome2) + 5 * Sim(CodNasc1, CodNasc2) + 2 * \left(\frac{PesoAresta(Pessoa1, Pessoa2)}{10} \right)}{10}$$

CodNome é a codificação gerada pela etapa 3.a; CodNasc é a codificação gerada pela etapa 3.b; e Sim(A, B) calcula o valor da similaridade (variando de 0 a 1) entre duas cadeias A e B de mesmo tamanho. Sim (A, B) verifica a quantidade de caracteres iguais na mesma posição em A e B, retornando a taxa de similaridade. Por exemplo, para as codificações de nomes R123 e R321 retorna o valor 0.5 (50%), visto que os caracteres R e 2 estão na mesma posição em ambas as cadeias. Por fim, PesoAresta(Pessoa1, Pessoa2) é o peso da aresta, calculado na etapa 1, entre as pessoas envolvidas.

Terminado as três etapas você deverá imprimir a frase: “NomePessoaPopular combina com NomePessoaSintonia com X de sintonia s2”, onde NomePessoaPopular é o nome da pessoa mais popular; NomePessoaSintonia é o nome da pessoa que possui mais sintonia com a pessoa mais popular; e X é a taxa de sintonia entre elas.

ENTRADA:

A entrada consiste de:

- Uma linha contendo a data (no formato dd/mm/aaaa) em que foi realizado o questionário;
- Uma linha contendo um inteiro $N \leq 40$ que representa o número de alunos do turma;

Para cada aluno ordenado de 1 a N de acordo com a lista aplicada no formulário:

- Uma linha com o nome do aluno que respondeu o formulário. O nome do aluno tem no máximo 200 caracteres;
- Uma linha com os demais dados do aluno que respondeu o formulário separados por espaço. A data de nascimento possui o formato dd/mm/aaaa;
- Uma linha com N-1 respostas do questionário contendo as notas de afinidades atribuídas para cada colega.

SAÍDA:

A frase: “NomePessoaPopular combina com NomePessoaSintonia com X de sintonia s2”, onde NomePessoaPopular é o nome completo da pessoa mais popular; NomePessoaSintonia é o nome completo da pessoa que possui mais sintonia com a pessoa mais popular; e X é a taxa de sintonia entre elas com duas casas decimais. Utilize toda a precisão de sua variável de sintonia para realizar os cálculos necessários e utilize duas casas decimais apenas na impressão.

EXEMPLO:

Considerando o formulário mostrado na Figura 1.3, os primeiros dados da entrada serão o de Pedro que informa as notas para os alunos Maria, Micalateia, João e Ana, nesta ordem. Os dados seguintes serão de Maria que dá notas para Pedro, Micalateia, João e Ana, nesta ordem e assim sucessivamente até acabar a lista do formulário.

Entrada:

4/5/2016

5

Pedro Ferreira

14/11/1990 M F

7 10 9 8

Maria Santos

11/1/1991 F M

10 10 8 7

Micalateia Silva

30/12/1968 F M

3 10 7 6

Joao Sousa

8/12/1939 M F

10 6 3 5

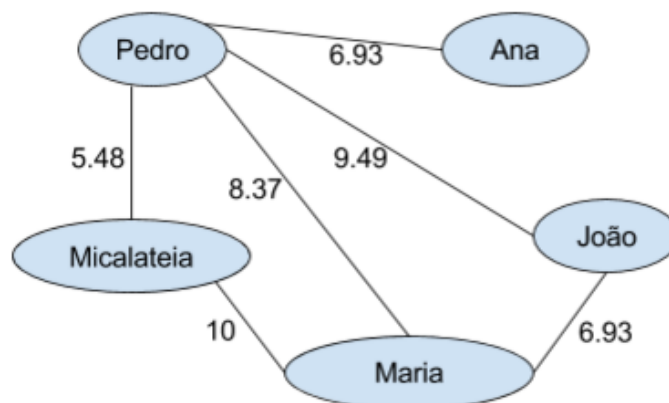
Ana Clara Costa

9/10/1999 F M

6 3 2 2

ETAPA 01

Figura 1.7: Exemplo Etapa 01



As demais arestas não foram desenhadas, pois possuem peso menor que 5.0. Os valores representados nesta figura estão arredondados e com duas casas decimais para uma melhor visualização.

ETAPA 02: De acordo com a centralidade dos vértices do grafo da Etapa 01, podemos concluir que a pessoa mais popular é Pedro com Grau 4. Enquanto que Maria tem grau 3, Micalateia e João têm grau 2 e Ana tem grau 1.

ETAPA 03: Nesta etapa, a codificação e similaridade são calculadas apenas para Pedro e as pessoas de sua preferência de gênero, neste caso, as mulheres.

Etapa 03.A:

Maria = M900

Micalateia = M386

Ana = A500

Pedro = P690

Etapa03.B:

Numerologia com data de nascimento:

Maria = $11+1+1+9+9+1=32 = 3+2 = 5$

Micalateia = $30+12+1+9+6+8 = 66 = 6+6 = 12 = 1+2 = 3$

Ana = $9+10+1+9+9+9 = 47 = 4+7 = 11 = 1+1 = 2$

Pedro $14+11+1+9+9+0 = 44 = 4+4 = 8$

Concatenando com as idades

Maria = 525

Micalateia = 347

Ana = 216

Pedro = 825

Etapa03.C:

Sintonia(Pedro, Maria) = $(3*0,25+5*0,67+((2*8,37)/10))/10 = 0.58$

Sintonia(Pedro, Micalateia) = $(3*0+5*0+((2*5,48)/10))/10 = 0.11$

Sintonia(Pedro, Ana) = $(3*0,25 + 5*0+((2*6,93)/10))/10 = 0.21$

Obs: Os valores aqui foram arredondados por questão de visualização. Utilize toda a precisão de sua variável e apenas na impressão utilize duas casas decimais.

Saída

“Pedro Ferreira combina com Maria Santos com 0.58 de sintonia s2”

• Observações

- É **obrigatório** o uso de ponteiros e estruturas para representar o seu grafo. Não atender a esta regra resultará em nota **0 (zero)**;
- É **proibido** utilizar variáveis globais;
- Em caso de empate, para decidir a pessoa mais popular ou a pessoa com maior sintonia considere sempre a primeira pessoa que aparece primeiro no questionário;
- Sempre há pelo menos uma pessoa cujo gênero é o mesmo preferido pela pessoa mais popular;
- Utilize funções e comentários para um melhor entendimento do seu código.
- **ATENÇÃO: Uma vez criado o seu grafo você não deve utilizar as informações do formulário. Trabalhe diretamente com o grafo, não seguir essa regra terá penalidades.**
- **DICA: Você poderá utilizar um vetor para armazenar os vértices e um vetor para armazenar as arestas de seu grafo;**
- **DICA: Divida o problema em subproblemas e utilize funções para resolver esses subproblemas.**
- O programa deve ser submetido em C (labSemanal05.c)
- Inclua um comentário no início do programa com o seu nome completo, turma e RA
- Faça comentários e indentação do seu código
- O aluno pode assumir que todas as linhas da entrada terminam com o fim-de-linha
- Todas as linhas da saída devem terminar com o fim-de-linha
- O número máximo de submissões é 15
- O comando de compilação utilizado será:

gcc -std=c99 -pedantic -Wall -lm labSemanal05.c -o labSemanal05

– Para executar seu programa faça: ./nome_executável