

## Prática 1 – Aplicações de Teoria dos Grafos

Grupo de 5 aluno(a)s

**Objetivo:** Modelar e representar graficamente situações-problema reais usando grafos.

### Etapas:

1. Escolher uma base de dados reais e uma situação problema a ser investigada usando elementos da Teoria dos Grafos. *No final deste documento, listo exemplos de bases de dados disponíveis na internet. Outras podem ser sugeridas.*
2. Descrever o contexto que o grupo irá abordar, o qual deve possuir as seguintes características:
  - a. Deve expressar, de forma clara e consistente, elementos e suas relações, tratando-se **exclusivamente** de um caso real. Alguns exemplos:
    - Topologia de uma rede de computadores, onde os elementos seriam os dispositivos conectados e a relação seria a existência de comunicação direta entre eles (elementos de uma rede local não se comunicam diretamente com elementos de outra rede)
    - Para um dado programa, os elementos seriam os comandos atômicos e relação seria de ordem de execução. Quando um comando C1 pode ser executado depois de C2, então teríamos o arco (C1, C2).
  - b. Deve permitir a modelagem de um grafo com, **no mínimo**, 80 vértices e 120 arestas (ou arcos). **ATENÇÃO: o grafo não precisa ser criado neste momento!**
3. A partir da observação dos dados escolhidos, definir 4 perguntas do seu interesse cujas respostas possam ser extraídas a partir da análise dos dados. É importante que as perguntas escolhidas não tenham respostas evidentes. **Sejam criativos!!**

Essas perguntas deverão serem respondidas em etapas posteriores utilizando técnicas de grafos. Além da escolha das perguntas, descrever de forma sucinta seu sentimento sobre a resposta das perguntas apresentadas, sendo apenas hipóteses sua sobre os dados. Exemplos:

  - a. A base de dados escolhida foi a do survey do Stack Overflow:
    - i. Quais linguagens de programação são mais indicadas a se aprender na América Latina? R: Acredito que na América Latina a linguagem mais indicada seja JavaScript.
  - b. A base de dados escolhida foi a do IMDB de filmes:
    - i. Quais os atores mais rentáveis na última década? R: Robert Downey Jr., Tom Cruise e Emma Watson.

### **O que deve ser entregue?**

- Anexar ao classroom um documento PDF contendo a descrição da base de dados escolhida, seu contexto, como será o grafo desses dados (vértices, arcos, etc) e as perguntas a serem respondidas.
- Esse documento será avaliado também em sala de aula a fim de validar se satisfaz os requisitos mínimos.

### **Sugestões de Base de Dados:**

- The Million Playlist Dataset (MPD). Contém 1.000.000 de playlists criadas por usuários do Spotify platform. <https://recsys-challenge.spotify.com/dataset>

- IMDb Datasets. Subconjuntos de dados sobre filmes/series/programas indexados no IMDb  
<https://www.imdb.com/interfaces/>  
<https://datasets.imdbws.com/>

- Stack Overflow Annual Developer Survey  
<https://insights.stackoverflow.com/survey>

- FootballData. Estatísticas sobre futebol e outros esportes:  
<https://github.com/jokecamp/FootballData>  
<https://github.com/openfootball>

- Eclipse failures. Base de dados que relaciona bugs a componentes do Eclipse:  
<https://www.st.cs.uni-saarland.de/softevo/bug-data/eclipse/>

- Amazon Reviews. Base de dados que relaciona reviews de produtos e metadados da Amazon (de Maio de 1996 a Julho de 2014):  
<http://jmcauley.ucsd.edu/data/amazon/>

- Os 7 melhores sites para encontrar datasets para projetos de Data Science  
<https://paulovasconcellos.com.br/os-7-melhores-sites-para-encontrar-datasets-para-projetos-de-data-science-8a53c3b48329>