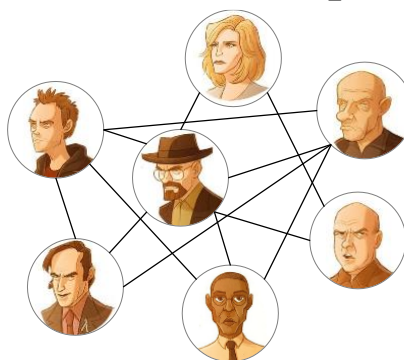


Análise de Redes Complexas



Considerações iniciais

- Códigos copiados da Internet e/ou de ferramentas de Inteligência Artificial Generativa receberão nota 0 (zero);
- Códigos que utilizem bibliotecas externas que fornecem estruturas de grafos ou algoritmos para sua manipulação ou processamento receberão nota 0 (zero);
- Todas as implementações serão avaliadas mediante um teste de autoria. Todos os estudantes do grupo devem apresentar o teste de autoria. A falta de apresentação ou o desconhecimento de partes importantes do código apresentado pelo grupo resultará em nota 0 (zero) ao estudante e poderá prejudicar a nota final de todo o grupo;
- A entrega do código-fonte, link com o vídeo da apresentação e relatório deverá ser feita em um único arquivo **zip**. A falta de qualquer um dos três arquivos resultará em nota 0 (zero) ao grupo;
- Entregas em atraso serão desconsideradas e resultarão em nota 0 (zero).

Visão Geral do Projeto

Neste projeto, você e sua equipe irão:

- 1) Aplicar o código desenvolvido em uma rede complexa de larga escala com o objetivo de extrair informações de um problema real de plataformas de *streaming* de vídeo utilizando a estrutura de grafos e algoritmos adequados;
- 2) Analisar os resultados obtidos, interpretá-los e discuti-los no formato de um relatório técnico de acordo com as atividades solicitadas.

Formato da Entrega

A entrega deve contemplar um arquivo comprimido no formato zip com 3 itens:

- O código-fonte com as implementações solicitadas;
- Link do vídeo de apresentação de até 15 minutos em que se discute a lógica empregada para a implementação de cada uma das funções solicitadas e a realização de testes;
- Um relatório, em formato PDF, com respostas às questões apresentadas neste documento.

Descrição do problema e dos dados

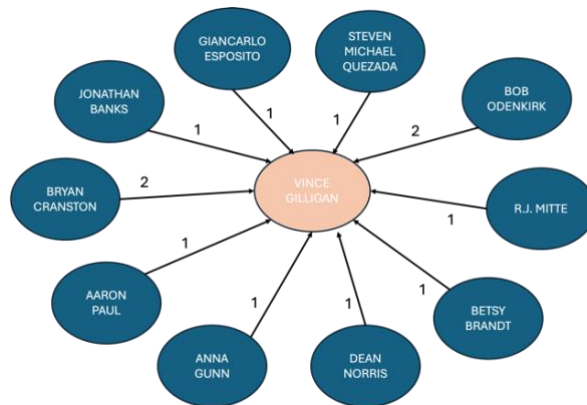
Neste trabalho, você e sua equipe, irão explorar os relacionamentos entre criadores de conteúdo disponível nas principais plataformas de *streaming* (Netflix, Amazon Prime Video e Disney+). A partir de um conjunto de dados¹ com informações sobre 19.621 filmes e séries envolvendo 61.811 atores/atrizes e 10.870 diretores, você deverá construir dois grafos para a condução de análises exploratórias:

1. Um grafo ponderado direcionado que representa as relações entre os atores/atrizes com os diretores das obras, considerando todos os filmes e séries do catálogo;
2. Um grafo ponderado não-direcionado que representa as relações entre os atores/atrizes em uma obra, considerando todos os filmes e séries do catálogo.

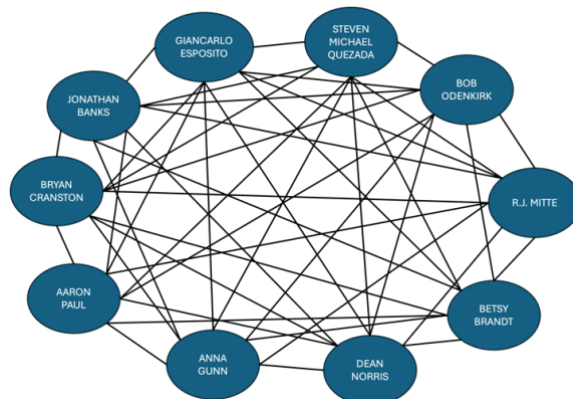
Para ilustrar, considere a entrada abaixo do conjunto de dados com informações da série *Breaking Bad*. As informações sobre o(s) diretor(es) e atores/atrizes são apresentadas nas colunas *director* e *cast*.

index	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
2732	s5941-netflix	TV Show	Breaking Bad	Vince Gilligan, Michelle MacLaren, Adam Bernstein, Colin Bucksey, Michael Slovis, Bryan Cranston, Terry McDonough, Johan Renck, Rian Johnson, Scott Winant, Peter Gould	Bryan Cranston, Aaron Paul, Anna Gunn, Dean Norris, Betsy Brandt, R.J. Mitte, Bob Odenkirk, Steven Michael Quezada, Jonathan Banks, Giancarlo Esposito	United States	2-Aug-13	2013	TV-MA	5 Seasons	Crime TV Shows, TV Dramas, TV Thrillers	A high school chemistry teacher dying of cancer teams with a former student to secure his family's future by manufacturing and selling crystal meth.

Na construção do primeiro grafo (**direcionado**), é necessário estabelecer conexões ponderadas de acordo com a quantidade de colaborações partindo de cada ator/atriz até o nome de cada diretor. Na figura abaixo, o processo é ilustrado considerando os 10 atores e 1 dos diretores. No entanto, o processo deve ser realizado para cada um dos diretores listados (Michelle MacLaren, Adam Bernstein, etc).



Na construção do segundo grafo (**não-direcionado**), é necessário estabelecer conexões ponderadas entre todos os atores/atrizes que participaram de um mesmo filme/série, conforme ilustrado a seguir. Em ambos os grafos, o peso das conexões é equivalente a quantidade de colaborações em diferentes obras. Por simplicidade, o grafo abaixo não contém pesos, mas você deve considerá-los.



¹ Disponível em:

https://drive.google.com/file/d/1UmuuWDb9qN5B6dwarX_aqJdWH2EwEJR/view?usp=drive_link

Atividades de implementação para a análise das redes complexas (6 pontos)

- 1) (1 ponto) Construção dos dois grafos solicitados (direcionado e não-direcionado) utilizando lista de adjacências. Durante o processo de construção, todos os nomes devem ser padronizados em letras maiúsculas e sem espaços em branco no início e no final da string. Entradas do conjunto de dados onde o nome do diretor e/ou nome do elenco estão vazias, devem ser ignoradas. Após a construção de cada grafo, retorne a quantidade de vértices e arestas.
- 2) (1 ponto) Função para a identificação e contagem de componentes. Para o grafo direcionado, a função deve contar a quantidade de componentes fortemente conexas. Para o grafo não-direcionado, a função deve retornar a quantidade de componentes conexas.
- 3) (1 ponto) Função que recebe como entrada um vértice X (por exemplo, BOB ODENKIRK) e retorna a Árvore Geradora Mínima da componente que contém o vértice X, bem como o custo total da árvore (i.e., a soma dos pesos das arestas da árvore). Essa função deve ser executada somente no grafo não-direcionado.
- 4) (1 ponto) Função que calcula a Centralidade de Grau (*Degree Centrality*) de um vértice, retornando um valor entre 0 e 1.
- 5) (1 ponto) Função que calcula a Centralidade de Intermediação (*Betweenness Centrality*) de um vértice, retornando um valor entre 0 e 1.
- 6) (1 ponto) Função que calcula a Centralidade de Proximidade (*Closeness Centrality*) de um vértice, retornando um valor entre 0 e 1.

Relatório com a análise das redes (4 pontos)

Elabore um **relatório analítico** no formato PDF que responda às questões listadas a seguir. **Justifique cada resposta por meio de texto explicativo e figuras** (como histogramas gerados a partir da extração das informações ou outros gráficos que julgar relevante) que comprovem e ilustrem suas conclusões.

- 1) (0.5 ponto) Qual é a distribuição de graus de ambos os grafos? Essa distribuição apresenta características típicas de rede complexas?
- 2) (0.5 ponto) Quantas componentes conexas (grafo não-direcionado) e fortemente conexas (grafo direcionado) existem? Qual a distribuição de ordem dessas componentes (número de vértices)? O que essa distribuição indica sobre as características do problema?
- 3) (0.5 ponto) Para o grafo direcionado, quais são os 10 diretores mais influentes perante a métrica de centralidade de grau? O que essa métrica representa nesse contexto?
- 4) (0.5 ponto) Para o grafo direcionado, quais são os 10 diretores mais influentes perante a métrica de centralidade de intermediação? O que essa métrica representa nesse contexto?
- 5) (0.5 ponto) Para o grafo direcionado, quais são os 10 diretores mais influentes perante a métrica de centralidade de proximidade? O que essa métrica representa nesse contexto?
- 6) (0.5 ponto) Para o grafo não-direcionado, quais são os 10 atores/atrizes mais influentes perante a métrica de centralidade de grau? O que essa métrica representa nesse contexto?
- 7) (0.5 ponto) Para o grafo não-direcionado, quais são os 10 atores/atrizes mais influentes perante a métrica de centralidade de intermediação? O que essa métrica representa nesse contexto?
- 8) (0.5 ponto) Para o grafo não-direcionado, quais são os 10 atores/atrizes mais influentes perante a métrica de centralidade de proximidade? O que essa métrica representa nesse contexto?