

Análise Exploratória da Base de Dados IMDb

Bárbara Boechat¹ Luccas Guidio¹

¹Departamento de Ciência da Computação –
Universidade Federal de São João del Rei / UFSJ

barbs.boechat@gmail.com, guidioo@hotmail.com

1. Introdução

O estudo de redes complexas é algo que vem acontecendo desde a época de 1930, uma rede, corresponde a um grafo, onde existem nós e arestas que os ligam. A diferença entre uma rede simples e uma rede complexa, é o fato de que as redes complexas não seguem um padrão regular. É também um tema interdisciplinar que envolve além da computação, a biologia e a matemática, como explicado por [Metz et al. 2007].

Definindo os nós e arestas de diferentes maneiras, uma rede pode representar características do mundo real como por exemplo clientes de uma biblioteca e os livros que compraram, ou amizades entre pessoas (o princípio das redes sociais), ou até mesmo ideias mais complexas como o cérebro humano, que é o princípio das redes neurais. As representações possíveis para as redes complexas são incontáveis.

Como citado anteriormente, redes complexas não seguem padrões regulares, entretanto, [Metz et al. 2007] explica também, que não há uma definição exata do que é um padrão regular e nem mesmo um consenso sobre o que constitui essas redes. Modelar, estruturar e caracterizar propriedades de uma rede, são os principais objetivos ao estudar uma rede complexa.

Desta forma, este trabalho consiste em estudar os atores do mundo real e os filmes nos quais eles participaram, adotando um caráter exploratório mediante a base de dados IMDb Movies.

2. Conceitos Fundamentais

2.1. Teoria de Grafos

De maneira sucinta, um grafo é um diagrama que apresenta relações entre variáveis, para expressar essas relações e variáveis são utilizadas arestas e vértices. De acordo com [Costa 2011], a Teoria de Grafos é um assunto interdisciplinar, envolvendo diversas áreas, além disso, os autores definem grafo como:

”Um grafo (finito) G é formado por um par $(V(G), A(G))$ onde $V(G)$ é um conjunto finito não vazio e $A(G)$ uma família de pares não ordenados de elementos, não necessariamente distintos, de $V(G)$. Uma família é uma coleção de elementos, os quais podem ser repetidos.”

Ou seja eles são representações de vértices, chamados de $V(G)$ que se relacionam entre si e essas relações são demonstradas por arestas, $A(G)$. E a partir do estudo da Teoria de Grafos é preciso atenção a outros conceitos relevantes derivados deste, como por exemplo, Redes Complexas (seção 2.2).

2.2. Redes Complexas

O estudo de redes complexas também vem sendo facilitado com os avanços tecnológicos do nosso mundo, algo também mencionado por [Metz et al. 2007], quando antigamente eram estudados grafos com no máximo dezenas de arestas, hoje em dia a quantidade de vértices chega à casa dos bilhões. Como mencionado pelos autores de [Metz et al. 2007], o conceito de redes complexas é estudado por diversas áreas como a matemática e física, não apenas pela Ciência da Computação.

Criar uma rede complexa é o mesmo que representar algum aspecto do mundo real em grafos, transformar comunidades em clusters de nós ou relações em arestas, e é algo que vem sendo estudado desde o século XX.

3. Motivação

A realização deste trabalho se motivou no aprofundamento de conhecimentos nas bibliotecas *Pandas* e *Networkx* da linguagem de programação Python. E também na aplicação de conceitos vistos no estudo de Redes Complexas neste trabalho que é a primeira realização de exploração e análise de uma base de dados.

4. Metodologia

4.1. Extração e Tratamento dos dados

A base de dados (dataset) utilizada é fornecidos pelo IMDb, uma base de dados online de informações sobre cinema, música, comerciais de televisão, jogos de computador e entre outras. Contudo, neste trabalho a base escolhida é relativa a 50 mil registros de filmes e pode ser encontrada em [IMDb Movies Extensive Dataset](#).

Uma base de dados pode ser entendida como uma matriz ou tabela, em que os cabeçalhos indicam qual informação contem determinada coluna e suas linhas, neste caso, são pertinentes a cada filme registrado. Para este trabalho nem todas as colunas do dataset eram pertinentes, assim, foi necessário realizar um corte nos dados restando apenas as colunas: *title*, *actors*, *country*, *genre*, *avg_vote* como mostrado na figura 1 uma pequena parte deste subdataset.

	imdb_title_id	title	actors	country	genre	avg_vote
0	tt0000009	Miss Jerry	Blanche Bayliss, William Courtenay, Chauncey D...	USA	Romance	5.9
1	tt0000574	The Story of the Kelly Gang	Elizabeth Tait, John Tait, Norman Campbell, Be...	Australia	Biography, Crime, Drama	6.1
2	tt0001892	Den sorte drom	Asta Nielsen, Valdemar Psilander, Gunnar Helse...	Germany, Denmark	Drama	5.8
3	tt0002101	Cleopatra	Helen Gardner, Pearl Sindelar, Miss Fielding, ...	USA	Drama, History	5.2
4	tt0002130	L'Inferno	Salvatore Papa, Arturo Pirovano, Giuseppe de L...	Italy	Adventure, Drama, Fantasy	7.0

Figura 1. Recorte da Base de Dados

Após a realização do recorte foram removidas todas as linhas que por ventura possuísem algum dado corrompido ou inexistente.

4.2. Modelagem dos Dados

Como dito em 4.1 foi realizado o recorte e a eliminação de dados faltantes, neste recorte cada linha representa um filme e na coluna nomeada como *actors* estão os nomes dos atores que atuaram neste filme, então, os nós da rede principal podem ser de dois tipos, nós filmes e nós atores.

A principio foi criado um grafo direcionado em que as relações entre os nós eram dadas a partir de um nó ator em direção a um nó filme, ou seja, como na figura 2 um nó ator aponta para todos os nós filmes em que ele participou.

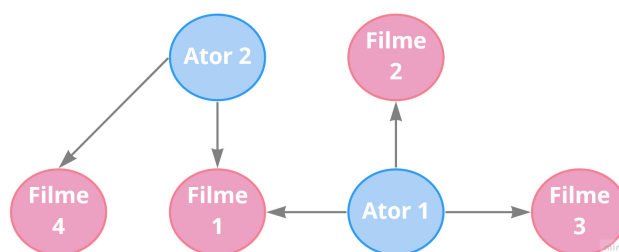


Figura 2. Recorte da Base de Dados

4.3. Redes Ego

Muito utilizadas nas metodologias de análise de redes sociais as redes ego são caracterizadas por seu foco em um nó individual, o ego, o qual é o nó central de análise. Os chamados alteres são seus vizinhos mais próximos e possuem relações diretas com o ego como descrito por [Borgatti 2005].

Além disso, as redes egocêntricas se propagam em níveis (hops), o nível 1, por exemplo, é constituído pelo ego e seus alteres imediatos, o nível 1.5 são as conexões entre os alteres, o nível 2 são os nós imediatos aos alteres. Este conceito é facilmente exemplificado no contexto de redes sociais, em que um usuário possui amigos, seus alteres diretos, por sua vez seus amigos possuem relações entre si e também possuem outros amigos constituindo o nível 2, ou os amigos de seus amigos. Desta maneira, uma rede ego é caracterizada como uma topologia de rede centralizada e hierárquica.

Ao realizar a criação da rede principal, neste trabalho, os subgrafos podem ser observados como redes ego para cada ator, nas quais consta todos os filmes que o nó principal participou que foram mapeados no dataset utilizado. Como mostrado na figura 3 os círculos verde e vermelho destacam um esboço de como se apresentam as redes ego das atrizes Scarlett Johansson e Kate Mara no modelo de rede deste trabalho.

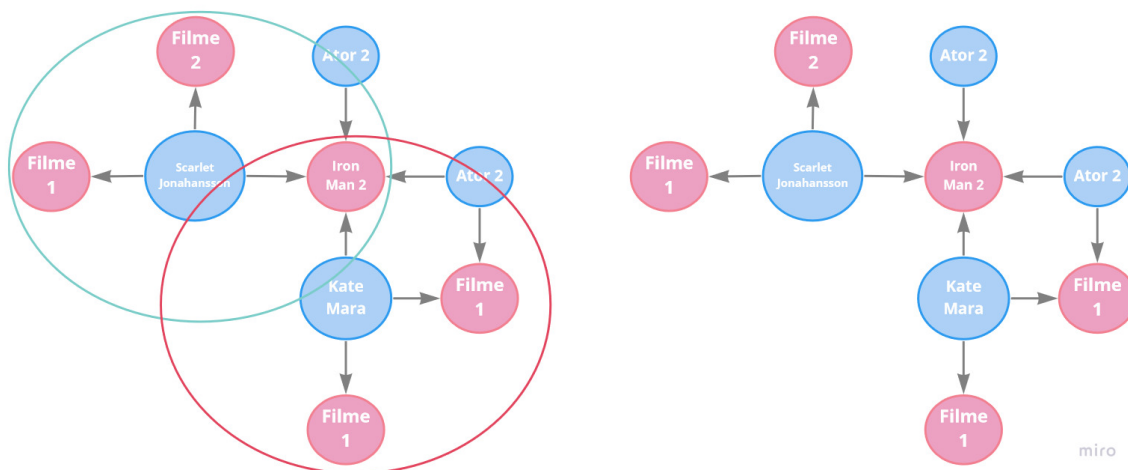


Figura 3. Recorte da Base de Dados

Assim, a partir do modelo de redes ego é possível fazer análises sobre todos os filmes realizados por um ator, seus alteres diretos e também analisar os outros níveis de propagação desta rede a partir de um determinado ator.

4.4. Clusterização dos Subgrafos

Cada rede egocêntrica de um ator foi considerada um subgrafo, então, para criar clusters que reunissem várias redes ego, todos os egos (nós atores) tiveram seus alteres (nós filme) contabilizados baseado nos países em que estes filmes foram realizados como mostrado na figura 4.

	actor	out_degree	Australia	USA	UK	Germany	France	Denmark	Norway	Ukraine	...	Faroe Islands	Guadeloupe	Reunion	Chad	Lesotho	Malawi
14	Frank Mills	8	1	2	5	1	1	0	0	0	...	0	0	0	0	0	0
16	Asta Nielsen	4	0	0	1	2	0	1	1	0	...	0	0	0	0	0	0
50	J.P. McGowan	7	0	7	0	0	0	0	0	0	...	0	0	0	0	0	0
51	Pola Negri	12	0	5	1	5	0	0	0	1	...	0	0	0	0	0	0
52	Emil Jannings	18	0	4	1	13	0	0	0	0	...	0	0	0	0	0	0

Figura 4. Contagem de Filmes de um Ator nos Países

A partir disso foi possível definir os atores que mais fizeram filmes nos países listados pelo IMBd, como a rede possui um número alto de países registrados foram escolhidos arbitrariamente 6 países, cada um formando um cluster com um ator de maior número de filmes feitos naquele país. Os 5 atores com maior destaque em números de filme no país podem ser visualizados a seguir na figura 5.

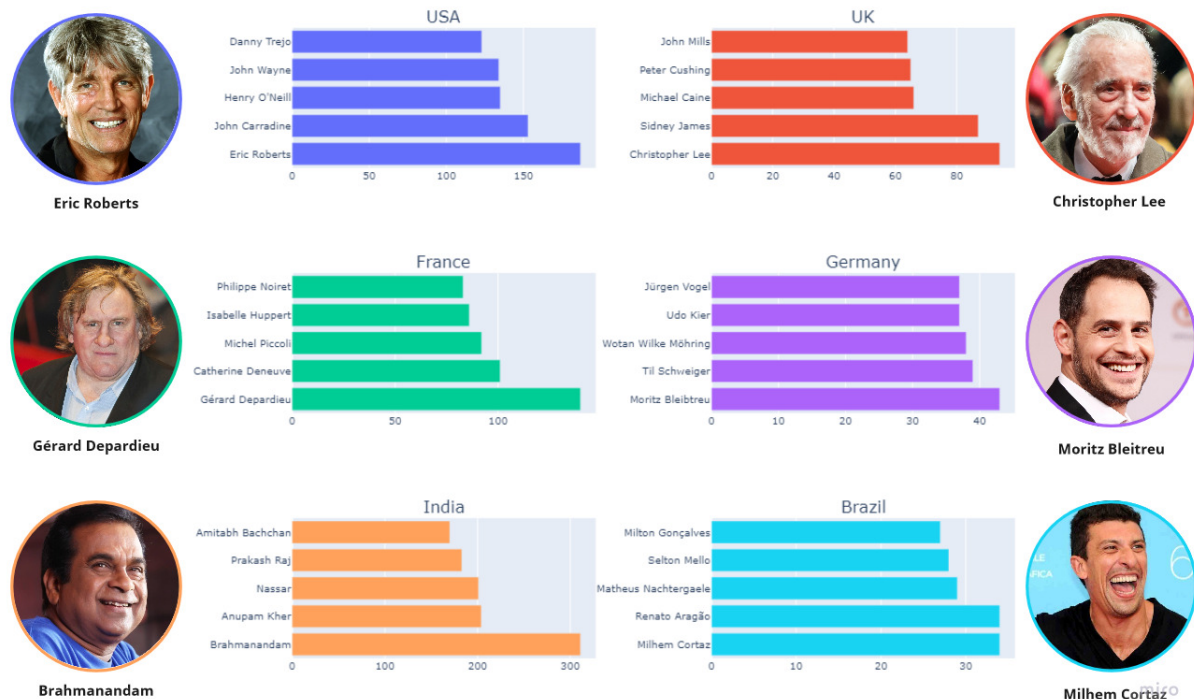


Figura 5. Cluster de Atores por País

4.5. Sistemas de Recomendação

Os sistemas de recomendação buscam oferecer sugestões de itens para um usuário, atualmente, os itens podem ser dos mais variados como músicas, filmes, séries, livros e produtos no geral. Como citado por [Aggarwal 2016] o princípio base das recomendação são as dependências significativas entre a atividade centrada no usuário e no item observado. Por fim, os sistemas de recomendação objetivam ajudar os usuários em suas tomadas de decisão e eles são construídos a partir de duas diferentes abordagens: Baseado em Conteúdo e Baseado em Filtragem Colaborativa (figura 6).

Baseado em Conteúdo: Tem como foco as propriedades dos itens, a similaridade de algo a ser recomendado será medida a partir do quanto suas propriedades são parecidas com as de um item que o usuário já tenha adquirido ou pesquisado.

Baseado em Filtragem Colaborativa: “A ideia geral por trás da filtragem colaborativa está em extrair grupos de usuários semelhantes, mediante históricos de seus comportamentos, e gerar recomendações para um determinado usuário embasadas nos comportamentos de outros usuários em seu grupo.”[JÚNIOR 2017].

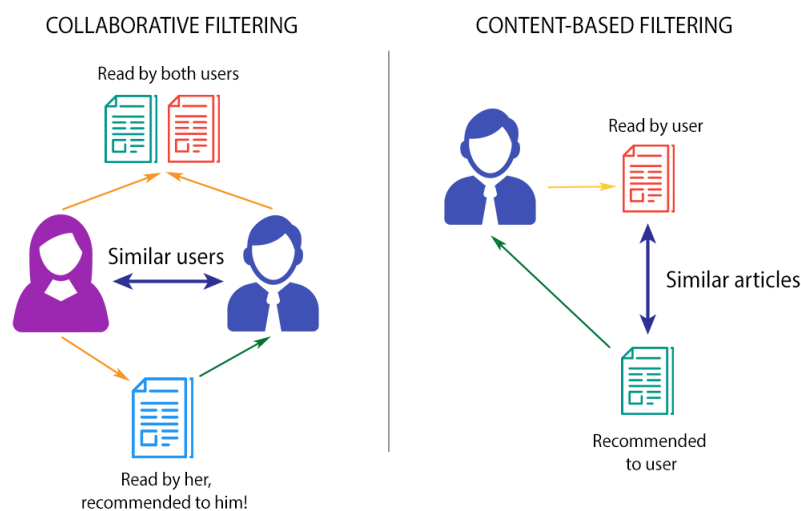


Figura 6. Sistemas de Recomendação por Filtragem e por Conteúdo

4.5.1. Similaridade Cosseno

Tomando como base o modelo de recomendação baseado em conteúdo foi criada uma matriz de similaridade cosseno entre os filmes tomando um único parâmetro de avaliação, os gêneros em que estes filmes foram categorizados. Assim, a fórmula utilizada para o cálculo dessa matriz foi:

$$\cos(a, b) = \frac{\langle a, b \rangle}{||a|| \times ||b||} \quad (1)$$

De acordo as definições trigonométricas acerca do cosseno, quando o ângulo entre dois pontos se aproxima de 0° , o cosseno se aproxima de 1. E quando os vetores são ortogonais, ou seja, o ângulo formado entre eles é de 90° , o cosseno é igual a 0. Então, numa variação de 0 a 1 quanto mais próximo de 1 maior é a similaridade dos vetores e quanto mais próximo de 0 menor é a similaridade. A visualização deste conceito pode ser realizada através da figura 7, a qual mostra a similaridade de gênero entre os filmes de um mesmo ator.

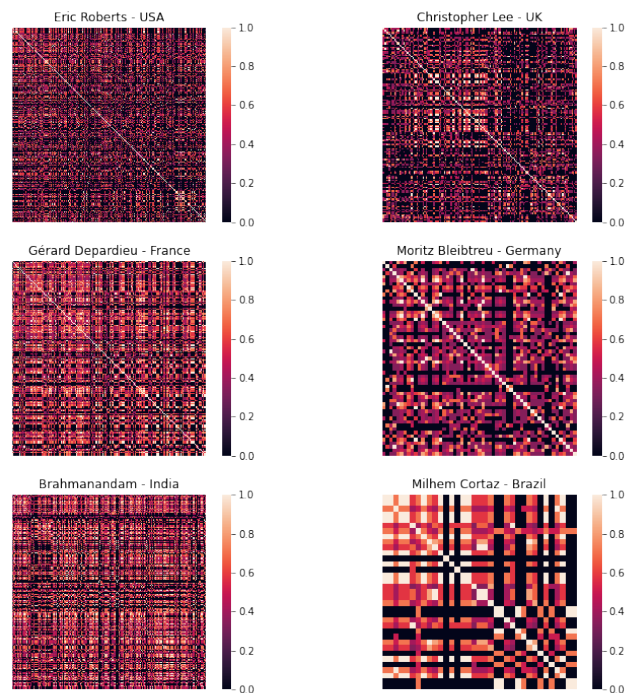


Figura 7. Matrizes de Similaridade dos Filmes de um Ator

Os filmes foram comparados dois a dois para isso eles foram reduzidos a vetores, em que a posição do vetor marcada com 1 indicava que aquele filme era daquele gênero e analogamente a posição marcada com 0 indicava exatamente o contrário. Contudo, a maioria dos pares não possuía a mesma quantidade de itens no vetor binário de gêneros, dessa forma foi necessário normalizar esses vetores de acordo ao maior deles (figura 8).

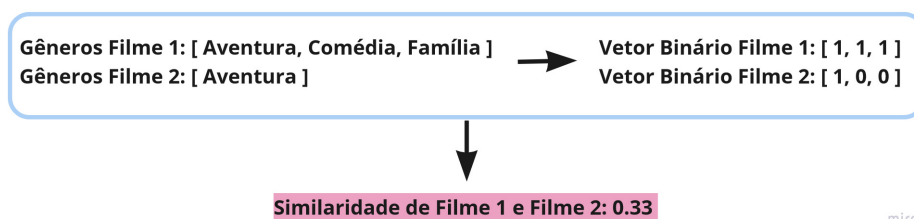


Figura 8. Tratamento dos Vetores de Gênero de Filme

4.6. Heurística Gulosa

As estratégias gulosas ou *greedy strategies* configuram um método guloso de projeto de algoritmos, podendo ser definida como a estratégia que sempre decide pela mesma direção, como um alpinista sempre indo para cima na direção de maior subida com a esperança de chegar ao pico mais alto da montanha, ou melhor dizendo, a estratégia gulosa realiza escolhas locais ótimas em cada fase e espera encontrar uma solução que represente o ótimo global.

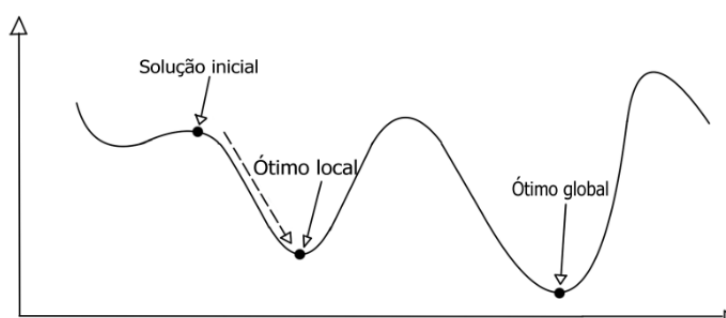


Figura 9. Ótimos Locais e Globais


Um algoritmo guloso escolhe, em cada iteração, o melhor objeto que vê pela frente, as definições do que configura um objeto “apetitoso” ou não são estabelecidas anteriormente, assim, cada escolha realizada passa a fazer parte da solução que o algoritmo está construindo. Dessa maneira, é possível perceber que este tipo de algoritmo não se arrepende e nem volta atrás, ou seja, as escolhas realizadas são definitivas.

5. Resultados

5.1. Listas de Recomendação

A partir de um ator original, encontra-se a matriz de frequência dos gêneros dos filmes deste ator, então são selecionados os dois gêneros mais comuns, então, para recomendar filmes a partir de um segundo ator foi realizada a construção de uma nova matriz de similaridade comparando todos os filmes do segundo ator com os filmes do ator original, assim, é formada uma lista de recomendação de filmes com similaridades correspondentes a 60%, 70% e ou 90%.

Como foram gerados muitos resultados a partir de cada um dos 6 atores anteriormente destacados apenas dois foram selecionados para compor esta parte do trabalho e podem ser verificados nas figuras 10 e 11.



Porque você assistiu a filmes de Eric Roberts

Categorias Principais do Ator		
Atores	Drama Nome, IMDb Rating	Ação Nome, IMDb Rating
Christopher Lee	I ladro dell'arcobaleno, 5.9 La texana e i fratelli Penitenza, 6.4 Penny and the Pownall Case, 5.8 I Rally piu' pazzo d'Africa, 4.8 Corridors of Blood, 6.5 'Jinnah', 8.0	The Crimson Pirate, 7.3 Star Wars: Episode III – Revenge of the Sith, 7.5 Star Wars: Episode II – Attack of the Clones, 6.5 The Pirates of Blood River, 6.0 Il trono di fuoco, 3.5 Ita privata di Sherlock Holmes, 7.1
Gérard Depardieu	Germinal, 7.1 L'autre Dumas, 6.3 Qualcosa di meraviglioso, 6.8 Il commissario Le Guen e il caso Gassot, 5.4 Cadences obstinées, 4.8 Nemico pubblico n. 1 - L'istinto di morte, 7.5	Lui portava i tacchi a spillo, 7.0 Les compères - Noi siamo tuo padre, 7.0 Temporale Rosy, 5.7 La capra, 7.4 Actors, 6.0
Moritz Bleibtreu	Ich war noch niemals in New York, 5.7 Roads, 6.3 Chiko, 6.8 Liebe deine Nächste!, 5.1 Schutzengel, 5.8 Das kalte Herz, 6.1	Nicht mein Tag, 6.6 Knockin' on Heaven's Door, 8.0 Jerry Cotton, 5.3 Schutzengel, 5.8 La banda Baader Meinhof, 7.4 World War Z, 7.0
Brahmanandam	kkadunnadu, 7.3 Chitemma Mogudu, 4.4 Kamma Rajyam Lo Kadapa Reddilu, 3.5 Greeku Veerudu, 4.6 Alluda Majaaka!, 6.3 Aata, 4.8	Loukyam, 6.4 Hello Brother, 7.4 Anaganaga O Dheerudu, 5.1 Arudu, 3.5 Krishna: The Power of Indrakeeladri, 6.3 Arya 2, 7.4
Milhem Cortaz	Crô: O Filme, 3.3 Sangue Azul, 6.4 Não Por Acaso, 7.0 Canastra Suja, 7.4 Mundo Cão, 6.7 Tungstênio, 6.8	Tropa de Elite - Missão Dada é Missão Cumprida, 8.1 Tropa de Elite 2 - O Inimigo Agora É Outro, 8.0 Boca do Lixo, 5.7 Il più forte del mondo, 7.1

Figura 10. Listas de Recomendação a partir de Eric Roberts



Porque você assistiu a filmes de Milhem Cortaz

Categorias Principais do Ator		
Atores	Drama Nome, IMDb Rating	Comédia Nome, IMDb Rating
Christopher Lee	Rasputin: il monaco folle, 6.3 Il ladro dell'arcobaleno, 5.9 The keeper: il custode, 4.6 Il segreto del narciso d'oro, 5.6 Airport '77', 5.8 Walker Payne, 6.0	The Rosebud Beach Hotel, 4.6 Tradire è un'arte - Boogie Woogie, 5.2 Le incredibili avventure del signor Grand col complesso del miliardo e il pallino della truffa, 6.0 Jocks, 3.8 Jackpot, 2.0 Dark Shadows, 6.2
Gérard Depardieu	Green Card - Matrimonio di convenienza, 6.2 Una pura formalità, 7.7 Due fuggitivi e mezzo, 7.1 A Farewell to Fools, 5.4 '36', 7.0	Buffet freddo, 7.4 Bogus - L'amico immaginario, 5.4 Actors, 6.0 Hello Goodbye, 5.1 Les invincibles, 5.3
Moritz Bleibtreu	Luna Papa, 7.3 The Experiment - Cercasi cavie umane, 7.7 L'letto, 4.8 Gegengerade, 4.1 Jerry Cotton, 5.3 Abgeschnitten, 6.6	Ijay - Il mio amico indiano, 5.3 Lammbock, 7.3 Luna Papa, 7.3 Soul Kitchen, 7.3 Free Rainer, 6.7
Brahmanandam	Kitakitalu, 5.8 Shankardada Zindabad, 5.2 Dongala Mutha, 3.8 Jai Simha, 4.8 Jai Chiranjeeva, 5.8 Aa Okkati Adakku, 7.8	Silly Fellows, 5.9 Seema Tapakai, 6.1 Yamaleela, 7.5 Achari America Yathra, 5.8 Babu Bangaram, 5.3 Race Gurram, 7.2
Eric Roberts	nvito ad uccidere, 6.0 The Beatnicks, 5.2 oments of Clarity, 5.7 L.A. Slasher, 3.3 Cross 3, 2.2	A Talking Cat!?, 2.3 Un viaggio indimenticabile, 4.7 Coca Cola kid, 6.0 50 anni in rosa, 5.1 Rock Slyde, 5.6

miro

Figura 11. Listas de Recomendação a partir de Milhem Cortaz

5.2. Grafo de Conexões

Com o objetivo de analisar a distância, em relação a níveis (hops) entre os atores mais famosos comentados anteriormente, foi criado um grafo de conexões (figura 12) a partir de uma heurística gulosa.

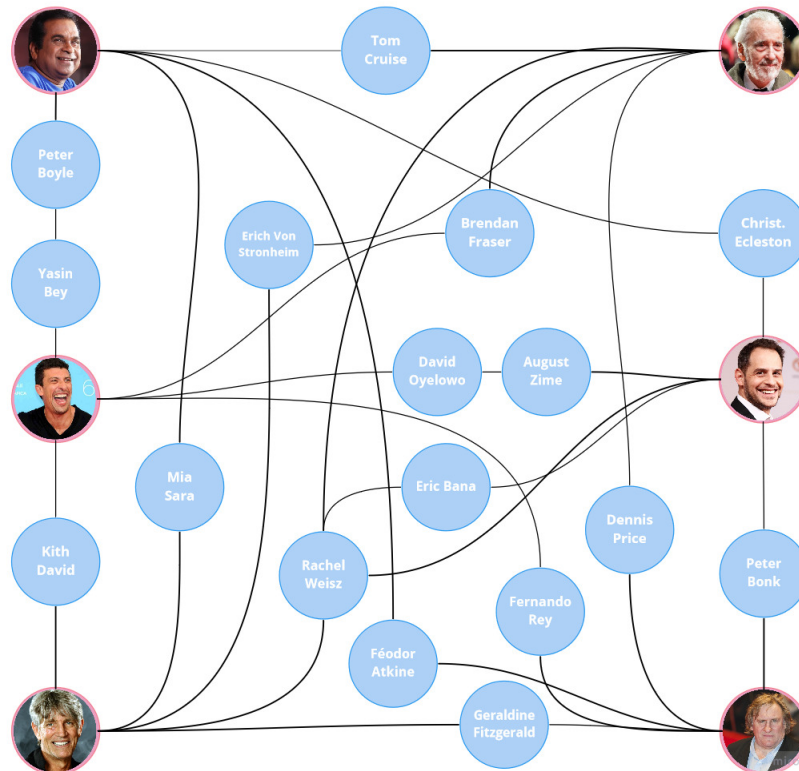


Figura 12. Grafo de Conexões

A heurística gulosa consiste primeiramente em transformar a rede original em um grafo não direcionado, desta forma, a rede complexa utilizada como base se transformaria em um grafo fortemente conectado em que qualquer nó é alcançável a partir de outro. Também é criado um segundo grafo (o grafo de conexões) onde os nós originais são os 6 atores mais ativos dos países:

- Brasil
- Estados Unidos
- Reino Unido
- Índia
- França
- Alemanha

Em seguida 2 nós do grafo de conexão original (6 atores mais ativos) são selecionados para que o menor caminho entre forme o critério inicial do grafo a ser gerado. Para isso foi utilizado o método "shortest_paths" do módulo netowkx da linguagem Python, ele retorna múltiplos caminhos com a mesma distância pois podem existir n maneiras de menor tamanho de alcançar outro nó. Os resultados possuem a seguinte forma (figura 13):

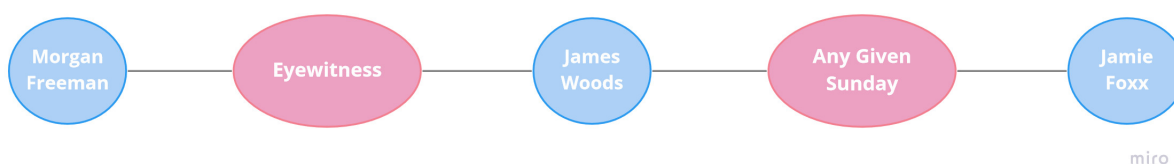


Figura 13. Exemplo de Menor Caminho entre Atores

Em cada caminho, existem "sub caminhos" que são definidos por nó ator 1 (nó azul) \leftrightarrow nó filme (nó rosa) \leftrightarrow nó ator 2, em que o nó filme possui a participação dos atores 1 e 2, logo eles estão diretamente conectados. Por conta disso, foi realizada uma limpeza no vetor de resultados, no qual todos os filmes foram retirados por serem irrelevantes, já que não é pertinente a este grafo saber como os atores são conectados, mas apenas a medida binária da conexão, ou seja, se eles são conectados ou não.

Com os vetores são fatorados, cada um deles passa por uma verificação que busca saber se algum de seus elementos já se encontra no grafo, essa verificação é o critério para tentar criar um grafo com a menor quantidade de nós novos possível e caso não exista nenhum resultado com nós pertencentes ao grafo atual, é escolhido um caminho aleatoriamente, então, todos os atores pertencentes são adicionados ao grafo de conexões.

Caso existam múltiplos caminhos com elementos pertencentes ao grafo que está sendo criado, escolhe-se aquele que possui o maior número de elementos que se encaixam nesta verificação. Assim, todos os caminhos entre uma dupla de atores são analisados, o mesmo é feito para outra dupla até que todos os atores originais estejam conectados entre si.

5.2.1. Análises do Grafo

Como mencionado em 5.2 o grafo gerado pela heurística gulosa com o objetivo de calcular e analisar distância entre os atores definidos previamente é um grafo fortemente conectado, uma vez que a maneira com que foi construído e a forma que a heurística funciona garante que qualquer nó poderá ser atingido por qualquer outro nó no grafo.

Com o grafo construído, uma nova análise foi realizada, ela consiste em calcular um *rating* médio entre todos os atores intermediários para dar uma "nota" geral ao grafo. A nota em questão foi obtida da seguinte maneira (figura 14):

```
('Tom Cruise', 6.897368421052632)
('Peter Boyle', 5.941304347826086)
('Yasiin Bey', 6.14375)
('Keith David', 5.610891089108909)
('Mia Sara', 5.423529411764705)
('Erich von Stroheim', 6.464516129032259)
('Rachel Weisz', 6.456756756756755)
('Féodor Atkine', 6.10677966101695)
('Eric Bana', 6.791304347826087)
('David Oyelowo', 6.211538461538461)
('Geraldine Fitzgerald', 6.496551724137931)
('August Zirner', 6.219999999999998)
('Fernando Rey', 6.177631578947366)
('Brendan Fraser', 5.784615384615385)
('Dennis Price', 5.77236842105263)
('Peter Bonke', 5.905882352941177)
('Christopher Eccleston', 6.504761904761905)
Média do grafo: 6.17
```

Figura 14. Recorte da Base de Dados

Foram construídos novos grafos, mais especificamente, um para cada nó intermediário, ou seja, para cada ator no grafo de conexões que não é um dos 6 atores originais. Os grafos construídos são grafos egocêntricos onde todas as arestas saem do ator em questão, e os nós de entrada destas arestas são os filmes nos quais o ator do grafo egocêntrico participou.

Com os grafos egocêntricos criados, existe uma facilidade em calcular o *rating* de cada um dos filmes participados por cada ator e fazer uma média destes *ratings*. Após o cálculo da média individual de cada ator, é realizada uma nova conta para obter a média das médias obtidas, esta média de médias é a nota geral para o grafo de conexões.

Os resultados obtidos através dos cálculos citados anteriormente são

mostrados na figura 14, é possível visualizar a nota média de cada ator individualmente, e também a média geral do grafo como um todo. Lembrando que os 6 atores originais não estão sendo considerados para estes cálculos uma vez que são constantes em qualquer resultado obtido e portanto irrelevantes na hora de analisar diferentes resultados.

Apesar da conta ter sido feita apenas para o grafo de exemplo obtido e apresentado neste artigo, a função funciona para qualquer grafo de conexões obtidos, é válido este comentário final, pois, os grafos obtidos durante cada execução podem ser diferentes, uma vez que existe um grau de aleatoriedade na construção desses grafos, algo intencional que visava gerar mais exemplos de resultados para serem analisados posteriormente.

Os trabalhos posteriores que foram considerados envolvem alterar a heurística que gera o grafo de conexões de forma que além de buscar o menor grafo possível, busque encontrar o maior *rating* médio, levando em consideração os filmes que conetam cada ator no grafo gerado. Além de criar uma nova heurística, algo que pode ser considerado futuramente quando o tempo não for um empecilho, é tentar criar um grafo de conexões com diferentes atores iniciais e com maiores restrições para os atores intermediários, como por exemplo procurar atores intermediários de um único país.

6. Conclusão

Neste trabalho foi possível realizar análises sobre redes ego dos atores dos filmes, clusterizá-los em diferentes países, gerar listas de recomendação a partir dos filmes feitos por um ator e até mesmo produzir uma heurística gulosa para gerar novos grafos a partir das conexões entre atores base. Então, defronte os resultados apresentados pode-se dizer que foram satisfatórios pois este trabalho tem caráter meramente exploratório da base de dados, de conceitos de Redes Complexas e de aprendizado de algumas bibliotecas da linguagem Python, dessa maneira, as dificuldades encontradas se deram por conta da curva de aprendizado sobre como aplicar e reunir todos os conceitos e habilidades citadas anteriormente na análise da base de dados escolhida.

Referências

- [Aggarwal 2016] Aggarwal, C. C. (2016). *Recommender Systems: The Textbook*. Springer Publishing Company, Incorporated, 1st edition.
- [Borgatti 2005] Borgatti, S. (2005). Ego networks.boston.
- [Costa 2011] Costa, P. P. d. (2011). Teoria dos grafos e suas aplicações.
- [JÚNIOR 2017] JÚNIOR, F. D. H. C. (2017). Avaliação de técnicas de filtragem colaborativa para sistemas de recomendação.
- [Metz et al. 2007] Metz, J., Calvo, R., Seno, E. R., Romero, R. A. F., Liang, Z., et al. (2007). Redes complexas: conceitos e aplicações.