

# A Social Network Analysis of Successful Movies with Complex Networks

Victória Patrícia Aires  
Instituto de Computação –  
Universidade Federal do Amazonas  
(UFAM)  
Manaus, AM 69077-000  
vpsa@icomput.ufam.edu.br

Fabiola G. Nakamura  
Instituto de Computação –  
Universidade Federal do Amazonas  
(UFAM)  
Manaus, AM 69077-000  
fabiola@icomput.ufam.edu.br

Eduardo F. Nakamura  
Instituto de Computação –  
Universidade Federal do Amazonas  
(UFAM)  
Manaus, AM 69077-000  
nakamura@icomput.ufam.edu.br

## ABSTRACT

Social networks are useful to represent social relationships like friendship and work. In a movie, they can represent the relationship between characters. These networks are called character networks. In this scientific initiation, we are interested in analyse the character networks of five successful movies at box office: Avatar, Titanic, Star Wars: The Force Awakens, The Avengers and Frozen. The aim is to check if there is some kind of pattern in the social networks of these movies. The results show that the character networks have similarities that may be related to their performance at box office. Furthermore, we show that the analysis with centrality measures and communities detection reflects the main characters and the story's main nuclei.

## KEYWORDS

Character networks, social networks, complex networks

## 1 INTRODUÇÃO

Redes sociais modelam interações sociais, como relacionamentos de amizade, trabalho, coautoria e citação em artigos acadêmicos. Dentro deste contexto, existem as redes de personagens: redes sociais que modelam o relacionamento de personagens em um filme ou obra literária. Estas redes, embora sejam pequenas, possuem atributos de redes complexas, como distribuição desigual de grau (poucos personagens muito conectados e muitos personagens periféricos); estrutura de comunidades bem definida; e *small-world*, ou seja, a maior distância entre dois vértices na rede é pequena [7].

Neste projeto de iniciação científica, estamos interessados em redes sociais de personagens em filmes. O objetivo é verificar se filmes de grande sucesso nas bilheterias possuem algum padrão semelhante em suas redes sociais correspondentes, ou seja, se existe uma possível fórmula para que um filme faça sucesso, refletida por meio da rede social do mesmo. Para tanto, foram escolhidos 5 dentre os 10 filmes de maior bilheteria de todos os tempos: Avatar (2009), Titanic (1997), Star Wars: O Despertar da Força (2015), Os Vingadores (2012) e a animação Frozen (2013). Com estas escolhas, esperamos ter uma visão inicial de como são as redes sociais de um filme bem sucedido em termos de bilheteria, possibilitando

posteriormente um estudo mais detalhado com um maior número de filmes.

Este artigo está organizado da seguinte maneira. A Seção 2 apresenta conceitos preliminares. A Seção 3 contém uma visão geral dos trabalhos relacionados. Na Seção 4 é detalhada a metodologia utilizada para o desenvolvimento do trabalho. Na Seção 5 são apresentados os resultados obtidos após a construção e análise das redes de personagens. Por fim, a Seção 6 traz as considerações finais do artigo, seguida das referências.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Redes Complexas

Na natureza e sociedade existe uma variedade de sistemas complexos que podem ser modelados através de redes. Estes fenômenos incluem redes de telecomunicações, relações sociais, a *World Wide Web* e a Internet, entre outros exemplos. Esta modelagem é feita por meio de *redes livres de escala* ou *redes complexas*. Redes complexas possuem como propriedades [15]: (i) distribuição de grau dos nós seguindo a lei de potência (*power-law degree*); (ii) características de mundo pequeno (*small-world*); (iii) alto coeficiente de agrupamento e estrutura de comunidades.

### 2.2 Redes Sociais

Redes sociais representam pessoas ou grupos de pessoas com algum padrão de contato ou interação entre si [14]. Matematicamente, redes sociais podem ser modeladas como grafos  $G(V, E)$ , onde  $V$  corresponde ao conjunto de vértices (indivíduos) e  $E$  corresponde ao conjunto de arestas (interações) do grafo [16]. Desta forma, é possível realizar análises que revelam informações implícitas como o nível de influência de vértices e a padrão de propagação de doenças em uma cidade, por exemplo.

### 2.3 Medidas de Centralidade

Neste trabalho, fazemos uso de medidas de centralidade para a identificação dos personagens mais influentes em cada filme. Tais medidas indicam quais os vértices mais centrais em uma rede complexa. As medidas escolhidas para análise neste trabalho são [13]:

- **Centralidade de grau ponderado (*weighted degree centrality*):** considera que os nós de maior grau são mais centrais. O grau de um vértice é a soma dos pesos das arestas a que ele está conectado. No contexto deste trabalho, considera que os personagens mais importantes são os que possuem mais interações com os demais.

- **Centralidade de intermediação (*betweenness centrality*):** indica o número de vezes que um vértice está no caminho mínimo entre dois outros vértices quaisquer na rede. Desta maneira, um personagem é mais central quando conecta personagens de diversos núcleos diferentes da história, funcionando como ponte entre personagens principais e secundários.
- **A centralidade de PageRank (*PageRank*):** considera o número e a qualidade de conexões de um vértice para determinar a importância do mesmo. Vértices mais importantes são os que recebem mais conexões de outros vértices importantes. No caso dos filmes, os personagens centrais interagem com outros personagens importantes na história.

## 2.4 Comunidades em Redes Complexas

Comunidades são definidas como grupos de vértices cuja densidade das arestas que os interligam é maior do que a densidade das arestas que ligam diferentes grupos [6]. Estes grupos representam indivíduos que interagem direta ou indiretamente, possuindo similaridade ou interesses em comum, por exemplo. O método de *Louvain* [8] é o estado-da-arte em detecção de comunidades. *Louvain* é uma heurística baseada em otimização da modularidade, uma função que mede a qualidade de agrupamentos da rede [5]. Por ser essencialmente linear, permite detecção de comunidades em redes com grande número de nós [1].

## 2.5 Medida de Influência

Para identificar os personagens mais influentes das redes complexas, propomos uma medida de influência  $I(v)$ , que se trata da média geométrica das métricas de centralidade de cada vértice  $v$ :

$$I(v) = \sqrt[3]{GP \times BT \times PR}, \quad (1)$$

onde GP é o grau ponderado, BT o *betweenness* e PR o *PageRank*.

Essa forma de avaliação da influência de personagens nos permite comparar a importância dos personagens com base na média geométrica das métricas, que aponta a tendência independente das grandezas de cada medida. Isso permite uma normalização dos valores das medidas.

## 3 TRABALHOS RELACIONADOS

Redes de personagens, obtidas a partir da extração das redes sociais de uma obra como livros e filmes, vêm sendo analisadas na literatura. Tais trabalhos têm como objetivo principal obter informações sobre uma determinada obra (ou conjunto de obras) a partir da análise das redes sociais [2, 4].

Em Beveridge and Shan [4], foi gerada a rede de personagens do livro *A Tormenta de Espadas (A Storm of Swords)*, terceiro volume da série de livros *As Crônicas de Gelo e Fogo*. Após a geração da rede social, foram aplicadas medidas de centralidade e detecção de comunidades. O objetivo é responder empiricamente quais os protagonistas do livro. Algo semelhante é feito em Aires et al. [2], onde a partir dos scripts (roteiros) e listas de personagens da saga *Senhor dos Anéis*, são criadas as redes de personagens e é feito o uso de medidas de centralidade para enumerar os personagens principais dos filmes e a estrutura de comunidades.

Outros trabalhos têm como objetivo a análise de um grande número de obras em busca de um padrão comum às redes de personagens. Em Gil et al. [9], são desenvolvidos e aplicados métodos para extração de redes sociais de obras de entretenimento e são feitas análises sobre as propriedades das redes resultantes. Os autores utilizam as redes sociais para tentar classificar filmes quanto à sua classificação indicativa e sucesso de público e crítica (através de notas do site *Rotten Tomatoes*<sup>1</sup>). Porém, os classificadores não tiveram bom desempenho em separar os filmes quanto a estas características, o que leva os autores a afirmar que não existe clara relação entre a estrutura do roteiro do filme e seu sucesso.

Diante disso, no presente trabalho buscamos analisar um conjunto de filmes que possuem características em comum: no caso, o sucesso de bilheteria. Desta maneira, pretendemos observar se existe ou não algum tipo de padrão em filmes de alta vendagem. Apresentamos uma análise inicial de cinco filmes, tentando compreender as propriedades das redes sociais dos mesmos para verificarmos se realmente não existe relação entre a estrutura do filme e o sucesso na recepção do público.

## 4 CRIAÇÃO DAS REDES DE PERSONAGENS

Neste trabalho, utilizamos a estratégia descrita em [2], trabalho de mesma autoria do presente artigo, onde é detalhada uma abordagem para extração da rede de personagens de um filme a partir do roteiro (*script* ou *screenplay*) do mesmo. Um roteiro contém a descrição de cada cena, bem como os personagens que aparecem e as falas daqueles que se manifestam nela. Estas informações estão dispostas seguindo um formato semi-estruturado, permitindo o uso de um programa baseado em expressões regulares para extrair os dados necessários para a construção da rede social.

Para auxiliar as informações do roteiro, foram utilizadas também a lista de personagens do filme, obtida a partir do site *IMDb (Internet Movie Database)* [10]. Os roteiros podem ser encontrados no site *IMSDb (Internet Movie Script Database)* [11].

A estratégia funciona da seguinte maneira: são dados como entrada o roteiro do filme e sua lista de personagens. Uma nova cena é indicada no roteiro por uma linha iniciada com INT ou EXT. Em cada nova cena, é verificado se existe a presença de um personagem em cada linha, com o auxílio da lista fornecida. Esta presença se refere tanto a um personagem que fala durante a cena, quanto aos citados nos diálogos ou mencionados na descrição. O personagem é então adicionado a um conjunto que inclui todos os personagens que apareceram ou foram citados naquela cena. Ao encontrar a próxima cena, é adicionada uma aresta entre todos os personagens que estavam no conjunto referente à cena anterior. Caso já exista aresta, seu peso é incrementado.

## 5 ANÁLISE DOS RESULTADOS OBTIDOS

Nesta seção, apresentamos uma análise das redes geradas a partir dos cinco filmes selecionados. Uma vez que o objetivo do trabalho era verificar se filmes de grande sucesso em bilheteria possuem algum tipo de padrão refletido em suas redes sociais, escolhemos cinco dentre os dez filmes de maior bilheteria de todos os tempos. Desta maneira, conseguimos um olhar inicial neste tema, possibilitando algumas conclusões.

<sup>1</sup>Rotten Tomatoes – <https://www.rottentomatoes.com>

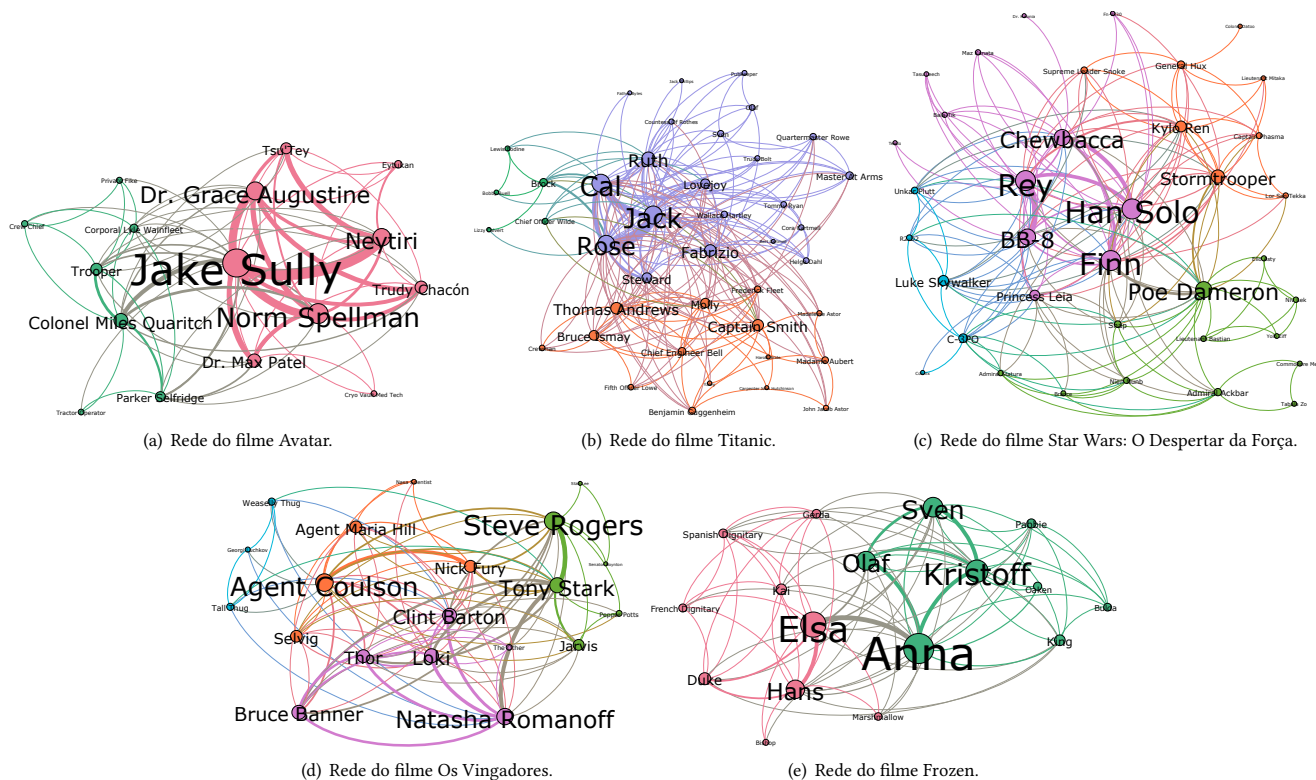


Figura 1: Redes de personagens dos cinco filmes selecionados. O tamanho dos vértices e rótulos reflete o valor de seu PageRank.

Os filmes foram escolhidos com base na lista disponível no site Box Office Mojo [12]. Dentre os dez filmes de maior bilheteria, escolhemos Avatar (2009), Titanic (1997), Star Wars: O Despertar da Força (2015), Os Vingadores (2012) e Frozen (2013). Estes filmes possuem versões de seus roteiros disponíveis no site IMDb [11], permitindo nossa abordagem.

Para realizar os experimentos, foram feitos testes utilizando o software Gephi [3], que calcula as medidas de centralidade e permite a modelagem visual dos grafos que representam as redes. A Tabela 1 apresenta os atributos gerais das redes sociais de personagens de cada filme. Com base nela, podemos observar que os cinco filmes tiveram números similares de vértices (ou seja, personagens), e quatro deles tiveram números de conexões, ou seja, interações entre personagens, semelhantes. O filme Frozen teve um número menor que os demais: isto pode ser explicado pelo fato de ser um filme de animação, de duração mais curta que os demais filmes. Por outro lado, Titanic é o filme mais longo, com 3 horas e 15 minutos de duração, e também o que possui o maior número de arestas em sua rede social.

Também é possível visualizar na tabela que os valores de modularidade dos filmes é bem similar, com pequenas variações. Não são redes onde os grupos são bem separados: esta é uma característica da rede de personagens. Por ser *small-world*, espera-se que esta rede seja altamente conectada, com comunidades misturadas. Ainda assim, é interessante observar que os cinco filmes apresentam estrutura de comunidades parecida, indicada pela modularidade da rede social.

Tabela 1: Atributos gerais das redes de personagens dos filmes selecionados. Star Wars 7 corresponde a Star Wars: O Despertar da Força.

Filme	Nº de vértices	Nº de conexões	Modularidade
Avatar	16	594	0,094
Titanic	43	777	0,094
Star Wars 7	37	754	0,089
Os Vingadores	20	690	0,174
Frozen	17	280	0,177

Ao analisar os resultados individuais de cada filme, vemos que a estrutura de comunidades dos filmes reflete os principais núcleos da história, com pelo menos uma comunidade com o lado do bem e outra, do mal. Esta polaridade é uma característica comum a todos os filmes. Além disso, os personagens principais possuem medidas de influências semelhantes em todos os filmes, indicando um possível padrão nas interações dos protagonistas dos diferentes filmes. Por fim, os personagens mais influentes correspondem aos esperados protagonistas da história e incluem também os principais antagonistas na maioria dos casos. A seguir, é descrita a análise de cada filme individualmente. As redes sociais estão representadas na Figura 1.

### 5.1 Rede de personagens do filme Avatar

A rede de personagens de Avatar, mostrada na Figura 1(a), apresenta duas comunidades. A comunidade em rosa reúne o lado dos mocinhos do filme, os cientistas interessados em proteger Pandora, local onde o filme se passa, bem como os principais personagens nativos do lugar. Em verde, temos os

vilões do filme, interessados em destruir o lugar para fazer uso dos recursos naturais. Os cinco personagens com maiores valores de influência foram, em ordem, Jake Sully (1,853), Norm Spellman (0,887), Coronel Miles Quaritch (0,643), Dra. Grace Augustine (0,567) e Neytiri (0,555). Esta classificação incluiu os protagonistas e o principal antagonista do filme.

## 5.2 Rede de personagens do filme Titanic

A rede social do filme Titanic, exibida na Figura 1(b), possui três comunidades. A comunidade em roxo contém o núcleo principal do filme, incluindo o triângulo de protagonistas e personagens secundários. A comunidade em laranja reúne os personagens ligados ao comando do navio e trabalhadores. Por fim, a comunidade em verde apresenta os personagens que estão ouvindo a história contada pela personagem Rose anos após o naufrágio. Os cinco personagens com maiores valores de influência foram, em ordem, Jack (1,361), Cal (1,277), Rose (1,224), Ruth (0,557) e Thomas Andrews (0,479). Estes personagens são o casal protagonista, Jack e Rose, os antagonistas Cal e Ruth e o criador do navio.

## 5.3 Rede de personagens do filme Star Wars: O Despertar da Força

A rede social de Star Wars: O Despertar da Força é exibida na Figura 1(c). Ela apresenta quatro comunidades. Em roxo, os personagens protagonistas e aqueles mais envolvidos nas principais cenas de ação, incluindo Han Solo, Chewbacca, Princesa Leia, Rey, BB-8 e Finn. Em azul, personagens envolvidos com Luke Skywalker. Em verde, os personagens do núcleo da Resistência, incluindo o piloto Poe Dameron. Por fim, em laranja, os antagonistas do filme: o núcleo da Primeira Ordem, incluindo Kylo Ren e Snoke. Os cinco personagens com maiores influências foram Poe (1,049), Finn (0,997), Han Solo (0,853), Rey (0,853) e Chewbacca (0,768). Estes cinco personagens constituem o núcleo de protagonistas do filme.

## 5.4 Rede de personagens do filme Os Vingadores

A rede social de Os Vingadores, presente na Figura 1(d), possui quatro comunidades. Em azul, vilões combatidos pela Viúva Negra no início do filme. Em laranja, os agentes da Shield. Em roxo, o núcleo de Bruce Banner (Hulk), Natasha Romanoff (Viúva Negra), Thor e Loki. Por fim, em verde, o núcleo de Steve Rogers (Capitão América) e Tony Stark (Homem de Ferro). Os cinco personagens com maiores influências foram Capitão América (1,146), Agente Coulson (0,961), Viúva Negra (0,855), Homem de Ferro (0,737) e o vilão Loki (0,577). Estes personagens incluem os principais heróis e o vilão, além do Agente Coulson, um personagem secundário que funciona como motivação para a união dos protagonistas.

## 5.5 Rede de personagens do filme Frozen

Finalmente, a rede social de Frozen (Figura 1(e)) possui duas comunidades. Em rosa, a comunidade de Elsa e figuras do governo presentes no reino de Arendelle. Em verde, a comunidade de Anna, Olaf e Kristoff, personagens que ela encontra na busca pela irmã. Os cinco personagens de maiores influências foram Anna (1,206), Elsa (0,816), Sven (0,688), Kristoff (0,645) e o vilão Hans (0,622). Estes correspondem aos protagonistas do filme e o principal antagonista.

## 6 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho de iniciação científica, apresentamos uma análise das redes sociais de personagens de cinco filmes com algumas das maiores bilheteria de todos os tempos. Através do uso de medidas de centralidade e detecção de comunidades, foi possível observar que as redes de personagens obtidas refletem claramente o enredo das obras. Os personagens mais centrais correspondem aos protagonistas e antagonistas de cada um dos filmes, bem

como as comunidades representam núcleos bem definidos do enredo de cada filme. A partir desta análise, observamos que as redes possuem semelhanças entre si: os valores de modularidade, número de vértices e número de conexões são próximos. Isto indica que a estrutura de comunidades da rede é semelhante, embora o número de comunidades seja variável. Todas as redes, apesar de serem bem conectadas, possuem agrupamentos bem definidos e relacionados aos respectivos filmes.

A abordagem escolhida permitiu observar um indicio de que um filme de sucesso nas bilheteria possui um determinado número de vértices (personagens envolvidos) e uma estrutura de comunidades bem definida (ou seja, núcleos principais envolvendo os lados do bem e do mal). Estes valores podem variar de acordo com o tipo de filme: uma aventura pertencente a uma saga, um romance, um filme de fantasia e animação.

Apesar de positivos, estes resultados constituem apenas uma análise inicial. Desta maneira, para validar o estudo pretendemos prosseguir coletando um grande número de filmes a partir do site IMDb. A partir disto, é possível uma análise mais detalhada por gênero de filme, desempenho nas bilheteria e classificação pela crítica especializada. Esperamos assim compreender que características possuem filmes de sucesso de público e crítica, em diferentes gêneros, e observar se estes atributos refletem-se nas redes sociais de personagens extraídas dos filmes.

## REFERÊNCIAS

- [1] Victoria P. Aires and Fabíola G. Nakamura. 2017. Detecção de Comunidades em Redes Sociais: Relacionando o Método de Louvain a Medidas de Centralidade. In *Proc. of the 36th Concurso de Trabalhos de Iniciação Científica*.
- [2] Victoria P. Aires, Paulo R. Oliveira, and Fabíola G. Nakamura. 2017. Construção e Análise das Redes Sociais de Personagens dos Filmes da Franquia O Senhor dos Anéis. In *Proc. of the 6th Brazilian Workshop on Social Network Analysis and Mining*.
- [3] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: an open source software for exploring and manipulating networks.. In *Proc. of the 3th Int'l AAAI Conf. On Web and Social Media*. 361–362.
- [4] Andrew Beveridge and Jie Shan. 2016. Network of thrones. *Math Horizons* 23 (2016), 18–22.
- [5] Vincent Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast Unfolding of Communities in Large Networks. *Journal of statistical mechanics: theory and experiment* 2008 (2008).
- [6] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D Hwang. 2006. Complex Networks: Structure and Dynamics. *Physics reports* 424 (2006), 175–308.
- [7] Anthony Bonato, David R. D'Angelo, Ethan R. Elenberg, David F. Gleich, and Yangyang Hou. 2016. Mining and modeling character networks. In *Proc. of the 13th Int'l Conf. on Algorithms and Models for the Web Graph*. 100–114.
- [8] Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, and Alessandro Provetti. 2011. Generalized Louvain Method for Community Detection in Large Networks. In *Proc. of the 11th Int'l Conf. on Intelligent Systems Design and Applications*. 88–93.
- [9] Sebastian Gil, Laney Kuenzel, and Suen Caroline. 2011. Extraction and analysis of character interaction networks from plays and movies. *Retrieved June 15* (2011), 2016.
- [10] IMDb. 2017. Internet Movie Database (IMDb). Disponível em: <http://www.imdb.com/>. (2017). Acesso em: 02/2017.
- [11] IMDb. 2017. Internet Movie Script Database (IMSDb). Disponível em: <http://www.imsdb.com/>. (2017). Acesso em: 02/2017.
- [12] Box Office Mojo. 2017. All Time Worldwide Box Office Grosses. <http://www.boxofficemojo.com/alltime/world/>. (2017). Acesso em: 08/2017.
- [13] Nicola Perra and Santo Fortunato. 2008. Spectral centrality measures in complex networks. *Physical Review E* 78 (2008).
- [14] John Scott. 2012. *Social network analysis*. Sage.
- [15] Hua-Wei Shen. 2013. *Community structure of complex networks*. Springer Science & Business Media.
- [16] Stanley Wasserman and Katherine Faust. 1994. *Social network analysis: Methods and applications*. Vol. 8. Cambridge university press.