

# Preditor de Sucesso Cinematográfico

André Albuquerque, Daniel Antunes, Luan Garrido, Juliana Valério

Departamento de Ciência da Computação - Instituto de Matemática  
Universidade Federal do Rio de Janeiro (UFRJ)

andreluiz90@ig.com.br, {danoan2008@gmail.com, lbgarrido89}@gmail.com,

juvianna@dcc.ufrj.br

**Abstract.** *The subjective character in movie analysis makes any prediction an unreliable one, even when made by experts. The propose of this paper is to investigate how human relations contribute to the success of the cultural industry, namely, cinematographic productions. This work was based upon the Susceptible-Infected-Recovered (SIR) epidemiological model and the concept of the influence coefficient and “Peer Pressure” phenomenon.*

**Resumo.** *O caráter subjetivo presente na avaliação de filmes torna qualquer previsão de sucesso incerta, mesmo quando feita por especialistas. A proposta deste trabalho é investigar como as relações humanas contribuem para o sucesso de produtos da indústria cultural, em especial, as produções cinematográficas. O trabalho foi baseado no modelo epidemiológico Suscetíveis-Infetados-Recuperados (SIR) e na definição do conceito de coeficiente de influência e do fenômeno “Pressão por Grupo”.*

## 1. Introdução

O sucesso de um filme, embora muitas vezes possa parecer óbvio, não o é. São muitos os casos em que pequenas produções superam os grandes investimentos de Hollywood [Mlodinow 2009], sugerindo que, embora existam fatores mensuráveis de qualidade, o sucesso não é modelado por uma equação determinística.

A aparente aleatoriedade nos resultados de rentabilidade pode ser entendida analisando-se as características dos produtos da indústria cultural, a qual o cinema integra. Neste tipo de indústria, as interações sociais constituem fator importante para a afirmação de um produto como um sucesso.

O objetivo do trabalho consiste em analisar, a partir de um modelo de equações diferenciais baseado no modelo SIR e por meio de experimentos, como as informações sobre as produções cinematográficas se difundem por meio de interações sociais e, por fim, verificar se os fenômenos mencionados anteriormente contribuem direta ou indiretamente para essa difusão.

Um comportamento típico das relações humanas foi observado em um estudo sobre um mercado virtual de músicas. O fenômeno é denominado “Pressão por Grupo” e consiste no fato de que o grupo com maior número de membros possui, por si só, uma força que atrai novos membros para si [M. J. Salganik and Watts 2006]. No estudo, evidenciou-se que quando uma música obtém uma vantagem inicial (maior

número de downloads ou melhor classificação), os próximos participantes são influenciados a ouvir a música mais baixada e a avaliar de maneira positiva a música com melhor classificação. Resultado semelhante foi observado recentemente no mundo animal[Iain D. Couzin 2011], onde foi constatado que indivíduos tendem a seguir um grupo com maior número de membros.

Verifica-se que, recentemente, as equações diferenciais ordinárias (EDO) são bastante utilizadas para simular o mundo real em situações que se deseja modelar fenômenos de qualquer natureza, inclusive problemas epidemiológicos e difusão de informações[Meeyoung Cha 2011].

## 2. Modelagem

A modelagem SIR foi proposta em 1927 por Kermack e McKendrick[W. O. Kermack 1927] com o objetivo de modelar a propagação de uma epidemia em uma população. O modelo baseia-se na definição de três grupos: Suscetíveis (S), Infectados (I) e Recuperados (R). Dado o fato de o modelo citado servir para uma categoria de transmissão - no seu caso, epidemiológicas -, seu objetivo mantém-se propício quando tratado para difusão de informações. Portanto, a modelagem SIR foi utilizada como base para o modelo deste trabalho, realizando-se algumas modificações. Além disso, o modelo proposto possui as seguintes premissas:

1. Os consumidores, quando desconhecem a existência do filme, não influenciam outras pessoas a assisti-lo ou a não assisti-lo;
2. Os espectadores, uma vez que tenham assistido o filme, e somente o fazem uma vez, formam uma opinião a respeito do produto e passam a influenciar seus amigos de maneira positiva ou negativa. Neste trabalho, espectadores nunca mudam de opinião;
3. Uma pessoa suscetível à informação só receberá influência, positiva ou negativamente, de apenas um indivíduo que tenha assistido o filme;
4. A evolução do sistema se dá em função das relações ocorridas entre indivíduos que já assistiram ao filme e aqueles que não assistiram. Ao final, espera-se ter uma estimativa de quantas pessoas assistiram ao filme e de quantas deixaram de assistir;
5. A população estudada é fixa, representada por  $P$ .

Considere que  $S$  seja o conjunto que define o valor total de pessoas suscetíveis a receber a informação acerca do filme e serem influenciadas a assisti-lo ou não, sendo  $V$  o conjunto de indivíduos que decidiram ver o filme e  $N$  os que decidiram não ver. Podemos escrever, então, que:

$$S(t) + V(t) + N(t) = P \Rightarrow \frac{dS}{dt} + \frac{dV}{dt} + \frac{dN}{dt} = 0. \quad (1)$$

Com isso, é possível observar que a variação dos indivíduos suscetíveis decresce de acordo com a soma das variações dos indivíduos do grupo  $V$  e  $N$ :

$$\frac{dS}{dt} = - \left( \frac{dV}{dt} + \frac{dN}{dt} \right) \quad (2)$$

O grupo  $V$  é dividido em outros dois:  $L$  e  $D$ , que representam, respectivamente, as pessoas que gostaram e as que não gostaram do filme:

$$\frac{dV}{dt} = L(t)s(t)Q(t)I_p + e_1S(t) \quad (3)$$

$$\frac{dN}{dt} = D(t)s(t)Q(t)I_n + e_2S(t) \quad (4)$$

Em poucas palavras, essas equações indicam que as variações de  $V$  e de  $N$  serão dependentes de dois fatores:

1. A quantidade de pessoas resultante do encontro entre os indivíduos que gostaram do filme, no caso de  $V$ , ou os que não gostaram do filme, no caso de  $N$ , e as pessoas suscetíveis restantes;
2. Quantidade de indivíduos suscetíveis influenciadas por um fator externo.

Temos que  $s(t) = \frac{S(t)}{P}$ , representando a proporção de pessoas suscetíveis, ou seja, porcentagem de indivíduos que entrarão em contato com os agentes influenciadores, e  $Q(t)$  representando o número médio de indivíduos diferentes que um membro  $V_i$  de  $V$  tenha tido contato sem que nenhuma outra pessoa  $V_j$  de  $V$  (com  $j \neq i$ ) tenha entrado em contato no mesmo instante. Dessa forma,  $Q(t)$  é limitado superiormente por  $\frac{P}{V(t)}$  e sua expressão é dada por  $Q(t) = \rho_1 \frac{P}{V(t)}$  onde  $\rho_1$  é um número aleatório.

Os coeficientes médios de influência das pessoas são representados por  $I_p$ , o positivo e  $I_n$ , o negativo.

A influência de fatores externos é representada por  $e_{1,2}$ . No presente trabalho, é considerado como influência externa somente o gênero do filme [Nash Information Services 2011]. A expressão de  $e_1$  é dada por  $\rho_2 Ext(gen)$  onde  $Ext$  é uma função que retorna a taxa obtida a partir de estatísticas referentes a preferências de filmes por gênero e  $\rho_2$  um número aleatório. Para o grupo de pessoas que decidiram não ver o filme, utilizamos  $e_2 = (1 - \rho_2)Ext(gen)$ .

Representamos a variação das pessoas que gostaram e que não gostaram de forma proporcional ao grupo de indivíduos que decidiram ver o filme, devido ao fato de somente eles formarem opinião sobre o mesmo:

$$\frac{dL}{dt} = M\left(\lambda, \frac{L}{V}\right) \frac{dV}{dt} \quad (5)$$

$$\frac{dD}{dt} = M\left(1 - \lambda, \frac{D}{V}\right) \frac{dV}{dt} \quad (6)$$

onde  $M(\lambda, rt)$  exerce o papel definido pelo fenômeno “Pressão por Grupo”. Neste trabalho, está sendo proposta uma função exponencial para modelar a função  $M$ , chamada grau de majoritariedade, expressa pela seguinte equação:

$$M(\lambda, rt) = \begin{cases} e^{\lambda x} + (\lambda - 1) & rt \geq 0.5 \\ 1 - M(1 - \lambda, 1 - rt) & rt < 0.5 \end{cases}$$

sendo  $A = \ln(\lambda(\sigma - 1) + 1)$ , com  $\sigma \in [1, \infty)$  e  $x = \frac{(rt-0.5)}{0.5}$ , onde  $\sigma$  representa o poder com que o fator aleatório  $\lambda$  será reforçado por  $M$ . Recomenda-se  $1 \leq \sigma \leq 3$ .

Essa função foi desenvolvida para fortalecer o grupo majoritário e, quando o valor passado a ela ( $\lambda$ ) está em um grupo não majoritário, irá diminuir exponencialmente. Quando está no grupo de maior proporção, aumentamos o seu valor (também de forma exponencial), para que o efeito da “Pressão por Grupo” seja satisfeito.

### 3. Implementação

Primeiramente, considere um grupo inicial de pessoas que avaliaram o filme em um pré-lançamento: os críticos. Esse grupo inicial divide-se em outros dois: o das pessoas que gostaram do filme ( $L$ ) e o das que não gostaram ( $D$ ). Para ambos, são calculados os coeficientes de influência  $I_p$  (influência positiva) e  $I_n$  (influência negativa). A fim de calcular o valor de cada uma dessas influências, foram utilizadas as seguintes métricas, que se baseiam em funcionalidades do Facebook:

- Taxa de “curtis” que a pessoa recebe ( $L$ );
- Taxa de “compartilhamentos” ( $S$ );
- Taxa de marcações em fotos ( $P$ );
- Taxa de comentários ( $C$ );
- Taxa de amigos que a pessoa possui ( $F$ ).

Para cada uma dessas métricas, foi estipulado um limitante superior, considerando um valor satisfatório para que elas alcancem a sua maior contribuição para a determinação da influência de cada indivíduo.

É válido ressaltar a existência de inúmeras métricas que poderiam ser usadas, algumas menos, outras mais complexas [Karthik 2010]. Contudo, optou-se por essas para simplificação do modelo, já que não é um objetivo do trabalho definir o conjunto ótimo de métricas que meça a influência da maneira mais apropriada.

A determinação dos coeficientes que serão aplicados a cada métrica foi obtida por meio de uma pesquisa na qual os voluntários deveriam atribuir valores entre 1 e 5, sendo o primeiro o menos importante e o segundo, o oposto. Ao final da pesquisa, com o resultado obtido, foi gerada a seguinte função de influência:

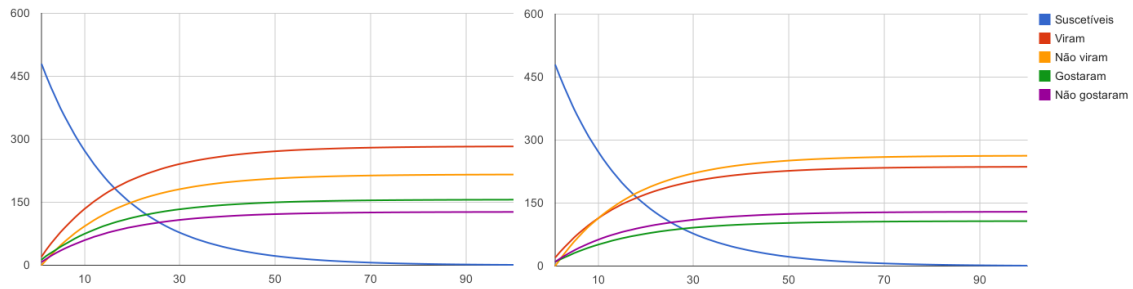
$$Influencia = \frac{(2.6226)L + (3.2930)S + 2P + (2.1698)F + (3.4339)C}{13.5193} \quad (7)$$

Essa função é aplicada para cada um dos críticos iniciais e seu valor é mantido até o final da simulação. A fim de apresentar dados mais consistentes, planeja-se um experimento obtendo essas informações de influência a partir de usuários do Facebook. Entretanto, atualmente, estão sendo gerados valores aleatórios para as métricas.

O fator externo revela-se como parte importante da implementação, na medida em que representa uma grande generalização que pode abranger elementos tais como: gênero do filme, atores do elenco, campanha de marketing, entre outros. A expressão definida na implementação baseia-se somente na popularidade de um gênero. O fator externo também é utilizado como um ponto de partida na dinâmica do modelo em situações onde não se deseja críticos iniciais. Além disso, serve como um perturbador do sistema, contribuindo positiva ou negativamente.

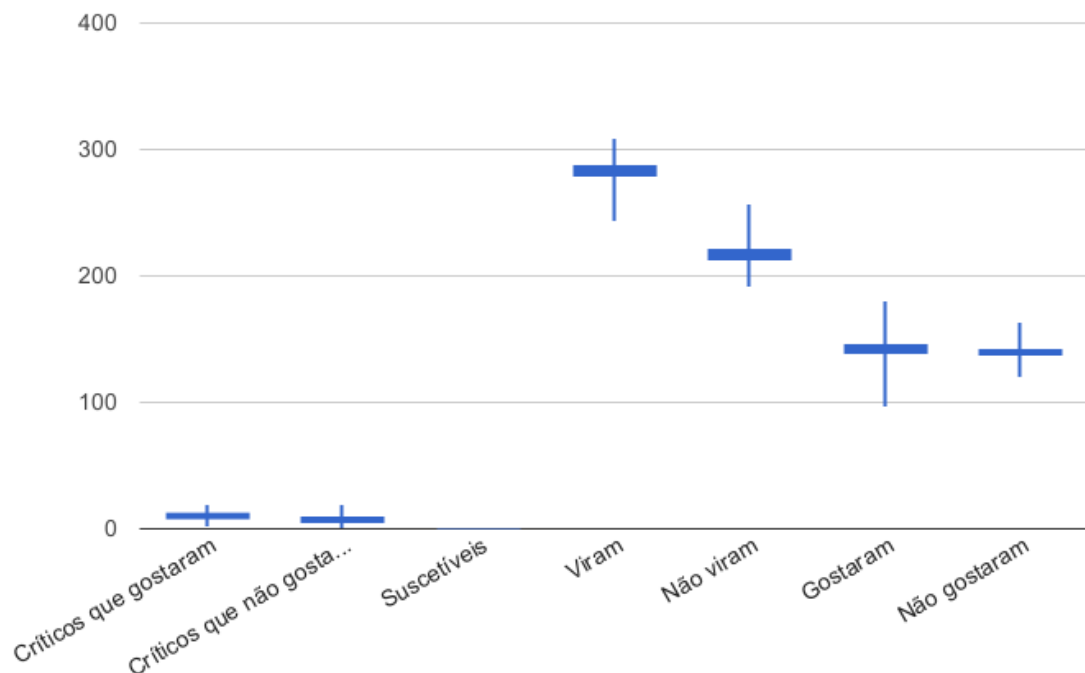
## 4. Resultados

As simulações abaixo foram iniciadas com os mesmos valores, indicando uma tendência positiva maior do que a negativa. Entretanto, mesmo tendo uma tendência positiva, pode acontecer de a negativa ganhar, o que de fato era o esperado devido aos fatores aleatórios presentes no programa.



**Figure 1. Pessoas que gostaram inicialmente = 11, Pessoas que não gostaram inicialmente = 9, Influência positiva = 0.7, Influência negativa = 0.5, População = 500**

Utilizando intervalos de confiança para as mesmas condições iniciais acima, rodamos o algoritmo um número suficiente de vezes para realizar uma aproximação pela distribuição normal, obtendo os intervalos cujas respostas irão se encontrar em pelo menos 95% das vezes.



**Figure 2. Intervalos de Confiança utilizando as mesmas condições iniciais anteriores**

## 5. Trabalhos Futuros

A modelagem proposta apresenta as atuais características a fim de simplificar o problema e tentar modelar a aceitabilidade de uma produção cinematográfica dentro das premissas

apresentadas. Entretanto, é possível aumentar suas funcionalidades para agregar outras situações:

1. Os coeficientes de influência médio são calculados apenas uma vez. As interações sociais seriam mais fielmente representadas se tratadas individualmente, com coeficientes de influência únicos para cada indivíduo. Devido ao custo dessa implementação, foi adotada, a princípio, a abordagem de coeficiente médio, onde este valor é constante. Entretanto, é possível que uma função de influência que varie com o tempo apresente um comportamento mais realístico;
2. O fator externo está implementada de forma simplista, levando em conta apenas o gênero do filme, como já fora citado. Uma forma de tornar mais verossímil, seria adicionar outras categorias como diretor, elenco, investimento em marketing, etc;
3. Apenas um grupo fechado com quantidade fixa de pessoas é considerado. Portanto, um trabalho futuro seria estender o modelo para aceitar comportamentos dinâmicos (e.g., crescimento) na população.

Atualmente, uma aplicação web está sendo desenvolvida utilizando a implementação feita em cima do modelo proposto e com utilização de dados oriundos de perfis do Facebook para obtenção das influências e críticos iniciais. Planeja-se tornar essa ferramenta para uso comercial.

## References

- Iain D. Couzin, Christos C. Ioannou, e. a. (2011). *Uninformed Individuals Promote Democratic Consensus in Animal Groups*. Number 6062. 334 edition.
- Karthik, N. (2010). *How to Measure Facebook Influence*, <http://www.dailybloggr.com/2010/10/how-to-measure-your-facebook-influence/>.
- M. J. Salganik, P. S. D. and Watts, D. J. (2006). *Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market*. Number 5762. 311 edition.
- Meeyoung Cha, Fabricio Benevenuto, Y.-Y. A. K. P. G. (2011). *Delayed Information Cascades in Flickr: Measurement, Analysis, and Modeling*.
- Mlodinow, L. (2009). *Peering Through the Eyepiece of Randomness*. Knopf Doubleday Publishing Group, 1st edition.
- Nash Information Services, L. (2011). *US Movie Market Summary 1995 to 2011*, <http://www.the-numbers.com/market>.
- W. O. Kermack, A. G. M. (1927). *A Contribution to the Mathematical Theory of Epidemics*.