

Relatório de Modelagem Matemática e Computacional

Pedro Von Zuben¹, Rafael Ribeiro¹, Raphael Mesquita¹

¹Instituto de Computação - CCMN/UFRJ

1. Introdução

A disseminação de notícias falsas, ou *fake news*, é um problema em alta e recente. Este fenômeno traz diversas consequências negativas como prejuízos na economia, segurança, política de um ou mais países, assim como o aumento de conflitos e violências em diversos aspectos da sociedade.

Podemos assumir que a disseminação começa através de pessoas que desejam atingir objetivos de próprio interesse e se propaga através de outras pessoas, elas sabendo ou não da verdade. No entanto, como as pessoas pensam de forma diferente, existem aquelas que se preocupam com as consequências e tendem a buscar a veracidade das notícias e se esforçam para que outros não acreditem nesta.

Portanto, se conseguimos dar esse primeiro passo, que é reconhecer essa disseminação de *fake news* como um problema, então agora precisamos solucioná-lo. Para isso, é necessário que entendamos como funciona esse sistema e como seus fatores se interagem. Pensando nisso, modelamos um sistema dinâmico de disseminação de *fake news* utilizando um sistema de equações diferenciais ordinárias (EDOs). [teoricamente falar sobre equilíbrio, estabilidade, etc do sistema]. Ao final, nós analisamos como a dinâmica do sistema evolui com o tempo, em diversos casos, através da visualização das soluções obtidas com o método numérico de Runge-Kutta de 4ª ordem.

2. Formulação

Antes da formulação da modelagem que propomos, é necessário determinar, ao menos, um cenário específico para tal disseminação. Isso se dá ao fato de que *fake news* disseminam principalmente em canais virtuais, ou *online*, logo são muitos possíveis cenários. Portanto, escolhemos a rede social *WhatsApp* como o meio de comunicação de estudo para este problema.

Ao observar casos de *fake news* sendo espalhadas em grupos de whatsapp genéricos, percebemos as pessoas envolvidas nesse cenário tendem a seguir certos padrões de comportamento. Esses padrões são consequência do estado ou categoria em que a pessoa está. No total, temos quatro categorias de pessoas, com duas dessas podendo ser agrupadas em uma supercategoria chamada de Transmissor:

2.1. Categorias

Pessoa Suscetível: Uma pessoa **Suscetível** é aquela que ainda não sabe da notícia, mas pode ser informada sobre a qualquer momento, através de uma pessoa **Protestante** ou **Infectada**, tornando-se, assim, **Exposta**. Também existe uma chance de tornar-se, diretamente, **Protestante** ou **Infectada**.

Pessoa Exposta: Uma pessoa **Exposta** é aquela que recebeu a notícia de uma pessoa **Transmissora**. Ela pode tornar-se **Infectada** se mantiver contato suficiente com outros **Infectados** ou de forma independente, após certo tempo, após receber a notícia, que chamaremos de *incubação*. Também pode tornar-se **Protestante** se mantiver contato suficiente com outros **Protestantes**.

Pessoa Infectada: Uma pessoa **Infectada** é aquela que escolhe compartilhar a **fake news** propositalmente, de forma que contribua com a transição de uma pessoa **Suscetível** ou **Exposta** para **Infectada**. Pessoas nesta categoria permanecem nesta categoria indefinidamente.

Pessoa Protestante: Uma pessoa **Protestante** é aquela que recebe a notícia mas a verifica e valida, de forma que, se descobrir que esta é falsa, toma medidas necessárias para que uma pessoa **Suscetível** ou **Exposta** também torne-se **Protestante**. Pessoas nesta categoria permanecem nesta categoria indefinidamente.

Essa dinâmica de transições entre categorias do sistema é apresentada na Figura 1 abaixo.

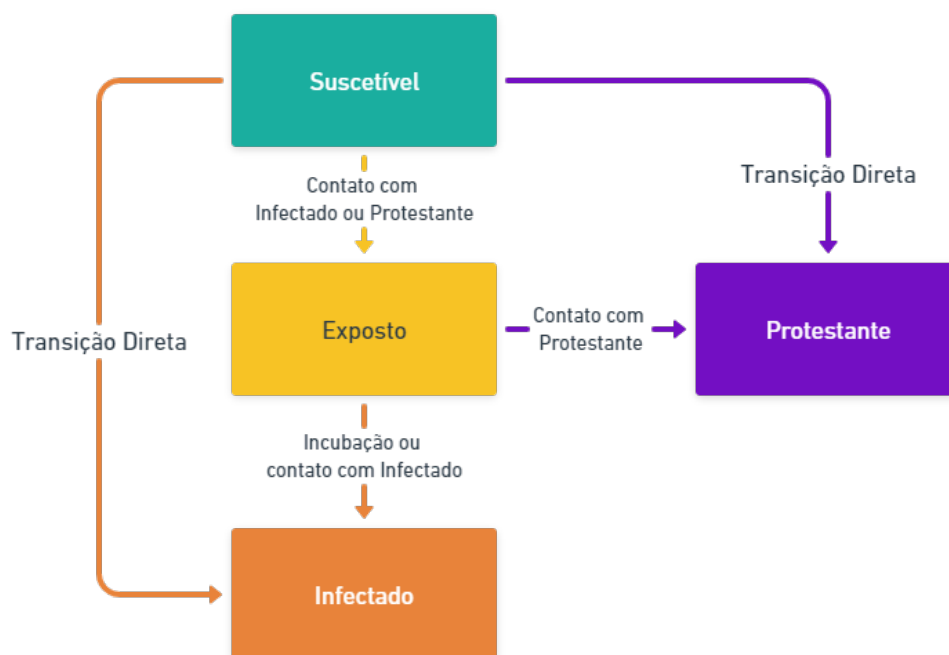


Figura 1. Diagrama da Dinâmica do Sistema

2.2. Variáveis e Parâmetros

Já havendo explicitado a dinâmica do sistema, conseguimos especificar suas as variáveis e introduzir seus parâmetros.

Variáveis:

Todas as variáveis do sistema representam a quantidade de pessoas em cada uma das quatro categorias. No entanto, esta quantidade não é absoluta, ou seja, não conta, literalmente, quantas pessoas pertencem à categoria. Na verdade elas representam as quantidades com relação a um total de pessoas implícito, portanto, uma razão entre as pessoas em certa categoria e todas as pessoas que podem ou não saber da notícia, ou seja, $\frac{\text{pessoas}}{\text{pessoas}} = 1$, logo, são variáveis com grandeza/unidade adimensional. Abaixo, temos a Tabela 1, resumindo as variáveis do sistema:

Tabela 1: Descrição das Variáveis

Variáveis	Descrição	Unidade
S	Pessoas Suscetíveis	Adimensional
E	Pessoas Expostas	Adimensional
I	Pessoas Infectadas	Adimensional
Z	Pessoas Protestantes	Adimensional

Além disso, como cada uma dessas variáveis são uma razão do total, logo

$$S + E + I + Z = 1$$

Parâmetros:

Os parâmetros são valores que determinam o comportamento geral do sistema. Alguns representam a taxa em que duas categorias de pessoas entram em contato e outras determinam a transição destas.

Tabela 2: Descrição dos Parâmetros

Parâmetros	Descrição	Unidade
β	Taxa de contato entre Suscetíveis e Infectados	1/hora
b	Taxa de contato entre Suscetíveis e Protestantes	1/hora
ρ	Taxa de contato entre Expostos e Infectados	1/hora
θ	Taxa de contato entre Expostos e Protestantes	1/hora
ε	Taxa de incubação de Expostos	1/hora
p_1	Probabilidade de transição direta de Suscetíveis para Infectados	Adimensional
l	Probabilidade de transição direta de Suscetíveis para Protestantes	Adimensional

2.3. Sistema e relação entre Variáveis e Parâmetros

Vamos agora apresentar o sistema de equações diferenciais ordinárias proposto por Kumar and Biswas [2019] e descrever como as variáveis e parâmetros se relacionam entre si.

2.3.1. Sistema de EDOs

$$\begin{cases} \frac{dS}{dt} = -IS\beta - SZb & (1) \\ \frac{dE}{dt} = -E\varepsilon - \theta EZ - EI\rho + IS\beta(1 - p_1) + SZb(1 - l) & (2) \\ \frac{dI}{dt} = E\varepsilon + EI\rho + IS\beta p_1 & (3) \\ \frac{dZ}{dt} = \theta EZ + SZbl & (4) \end{cases}$$

2.3.2. Relação entre Variáveis e Parâmetros

No termos do sistema acima podemos perceber algumas relações que possuem padrões em comum.

Em quase todos os termos, há uma relação multiplicativa entre duas variáveis, por exemplo no termo $IS\beta$ as variáveis I e S se multiplicam, representando a quantidade de contato entre pessoas **Infectadas** e **Suscetíveis**, ou seja, IS é uma permutação simples entre pessoas dessas duas categorias. Podemos utilizar a notação S - I para representar esse contato.

Além disso, para todo contato há um parâmetro que o pondera, ou seja, a taxa de contato. Novamente no termo $IS\beta$, o parâmetro é β . Portanto, o termo completo indica, dentre todos os contatos possíveis, quantos de fato ocorrem por hora.

Ainda há outra relação característica no sistema, que podemos observar, por exemplo, no termo $IS\beta p_1$. O termo $IS\beta$ já conhecemos, então sua relação com a probabilidade p_1 significa que, dentre os contatos S - I que de fato ocorrem por hora, quantos **Suscetíveis** transicionam diretamente para **Infectados**. Analogamente, $IS\beta(1 - p_1)$ é quando não ocorre essa transição direta, ou seja, a transição é de **Suscetível** para **Exposto**. Podemos utilizar a notação $S \rightarrow I$ e $S \rightarrow E$ para representar essas transições, respectivamente.

Enfim, temos o termo $E\varepsilon$. O parâmetro ε é a taxa de incubação, ou seja, a magnitude em que pessoas **Expostas** tornam-se **Infectadas** sem precisar de qualquer contato por hora. Portanto, o termo completo indica quantas **Expostas** tornam-se **Infectadas** sem precisar de qualquer contato.

A Figura 2 abaixo demonstra com que magnitude as transições ocorrem, através das taxas de transição. Por exemplo, uma pessoa Suscetível se torna Exposta através dos termos $(1 - p_1)\beta$ e $(1 - l)b$.

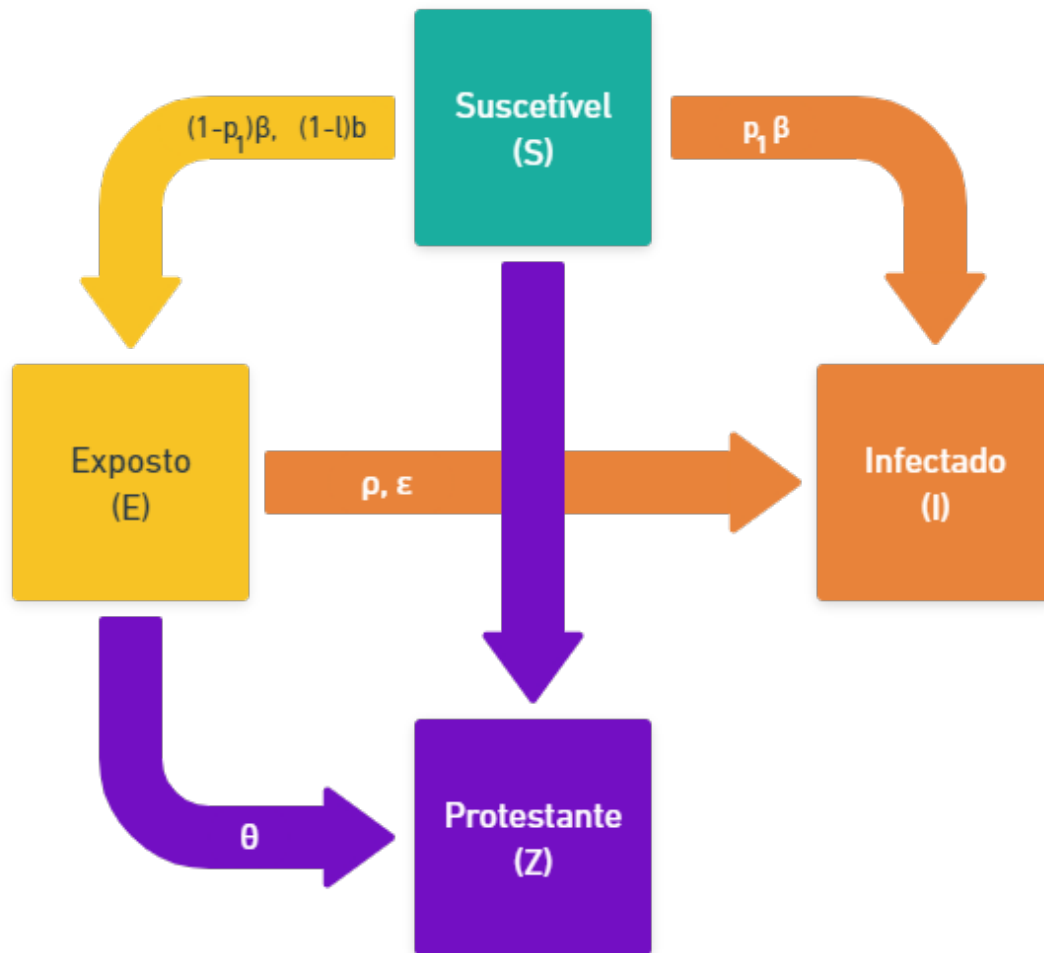


Figura 2. Diagrama da Taxas de Transição

2.3.3. Equações

Já conseguimos identificar a função de cada um dos termos do sistema, agora precisamos entender como eles interagem entre si para formar as equações do sistema. Portanto, analisando equação por equação, deixando as com mais termos por último:

Equação (1):

Como vimos acima, o termo $IS\beta$ é a taxa de transição de $S \rightarrow E$ ou $S \rightarrow I$; Analogamente, o termo ZSb é a taxa de transição de $S \rightarrow E$ ou $S \rightarrow Z$. Portanto, estes dois termos devem diminuir, linearmente, o número de pessoas **Suscetíveis**,

ou seja, basta somá-las e multiplicar o resultado por -1 para obtermos $\frac{dS}{dt}$. Além disso, é sempre verdade que $\frac{dS}{dt} \leq 0$, pois se há, inicialmente, pelo menos algum I ou Z para contato, logo S sempre vai diminuir até chegar a 0.

Equação (4):

O termo θEZ é a taxa de transição de **Expostos** para **Protestantes** apenas, pois um **Protestante** não transiciona pra nenhuma outra categoria;

O termo $SZbl$ se encaixa em uma das relações descritas na seção acima. Portanto, $SZbl$ é a taxa de transição $S \rightarrow Z$, através do contato $S - Z$.

De forma análoga à (1) estes dois termos devem incrementar o número de pessoas **Protestantes**, portanto basta somá-los para obtermos $\frac{dZ}{dt}$.

Além disso, é sempre verdade que $\frac{dZ}{dt} \geq 0$, pois o **Protestante** não transiciona para nenhuma outra categoria.

Equação (3):

O termo $EI\rho$ é a taxa de transição de **Expostos** para **Infectados** apenas, pois um **Infectado** não transiciona pra nenhuma outra categoria;

O termo $IS\beta p_1$ se encaixa em uma das relações descritas na seção acima. Portanto, $IS\beta p_1$ é a taxa de transição $S \rightarrow I$, através do contato $S - I$.

O termo $E\varepsilon$ é a taxa de transição $E \rightarrow I$, pois uma pessoa Exposta pode transicionar para Infectada através do "tempo de incubação".

De forma análoga à (1) estes dois termos devem incrementar o número de pessoas **Infectadas**, portanto basta somá-los para obtermos $\frac{dI}{dt}$.

Além disso, é sempre verdade que $\frac{dI}{dt} \geq 0$, pois o **Infectado** não transiciona para nenhuma outra categoria.

Equação (2):

Podemos ver as parcelas $E\varepsilon$, θEZ , $EI\rho$ com sinal negativo, pois representam transições de **Expostos** para outra categoria.

Também podemos ver as parcelas positivas $IS\beta(1 - p_1)$ e $SZb(1 - l)$ que são as taxas de não transição direta $S \rightarrow I$, através de $S - I$ e $S \rightarrow Z$, através de $S - Z$, respectivamente. Ou seja, através dos contatos com **Infectados** e **Protestantes**, a taxa de **Suscetíveis** transicionarem para **Expostos**.

Somando todos esses termos, com seus respectivos sinais, obtemos $\frac{dE}{dt}$. Além disso, assim como a quantidade de **Suscetíveis**, os **Expostos** também tendem a 0 se $I(0) > 0$ ou $Z(0) > 0$.

2.3.4. Unidades (ou Grandezas):

As unidades/grandezas se mantêm consistentes por todo o sistema. Todos as parcelas são produtos entre um ou mais termos de grandeza adimensional com um termo com grandeza

1/hora. Logo, como soma de unidades iguais, mantém a unidade, as equações seguem consistentes.

As equações do modelo, ou seja, o sistema, se dá por:

$$\left\{ \begin{array}{l} \frac{dS}{dt} = -IS\beta - SZb \end{array} \right. \quad (1)$$

$$\left\{ \begin{array}{l} \frac{dE}{dt} = -E\varepsilon - \theta EZ - EI\rho + IS\beta(1 - p_1) + SZb(1 - l) \end{array} \right. \quad (2)$$

$$\left\{ \begin{array}{l} \frac{dI}{dt} = E\varepsilon + EI\rho + IS\beta p_1 \end{array} \right. \quad (3)$$

$$\left\{ \begin{array}{l} \frac{dZ}{dt} = \theta EZ + SZbl \end{array} \right. \quad (4)$$

2.4. Exercício

Tentemos encontrar a taxa de transição inicial para cada uma das variáveis.

Vamos considerar valores para as seguintes variáveis e parâmetros:

Tabela 3: Valores de parâmetros para o exercício

Variáveis	Valores	Parâmetros	Valores
$S(0) = S_0$	0,4	β	0,6
$E(0) = E_0$	0,3	b	0,4
$I(0) = I_0$	0,2	ρ	0,1
$Z(0) = Z_0$	0,1	ε	0,03
		θ	0,5
		p_1	0,2
		l	0,4

Se o sistema é

$$\left\{ \begin{array}{l} \frac{dS}{dt} = -I_0 S_0 \beta - S_0 Z_0 b \end{array} \right. \quad (1)$$

$$\left\{ \begin{array}{l} \frac{dE}{dt} = -E_0 \varepsilon - \theta E_0 Z_0 - E_0 I_0 \rho + I_0 S_0 \beta (1 - p_1) + S_0 Z_0 b (1 - l) \end{array} \right. \quad (2)$$

$$\left\{ \begin{array}{l} \frac{dI}{dt} = E_0 \varepsilon + E_0 I_0 \rho + I_0 S_0 \beta p_1 \end{array} \right. \quad (3)$$

$$\left\{ \begin{array}{l} \frac{dZ}{dt} = \theta E_0 Z_0 + S_0 Z_0 b l \end{array} \right. \quad (4)$$

Podemos, primeiro, encontrar alguns termos que se repetem, como os presentes abaixo

$$I_0 S_0 = 0,2 \times 0,4 = 0,08$$

$$S_0 Z_0 = 0,2 \times 0,1 = 0,02$$

$$E_0 Z_0 = 0,3 \times 0,1 = 0,03$$

$$E_0 I_0 = 0,3 \times 0,2 = 0,06$$

O que nos dá

$$\begin{cases} \frac{dS}{dt} = -0,08\beta - 0,02b & (1) \\ \frac{dE}{dt} = -0,3\varepsilon - 0,03\theta - 0,06\rho + 0,08\beta(1 - p_1) + 0,02b(1 - l) & (2) \\ \frac{dI}{dt} = 0,3\varepsilon + 0,06\rho + 0,08\beta p_1 & (3) \\ \frac{dZ}{dt} = 0,03\theta + 0,02bl & (4) \end{cases}$$

Em seguida, substituímos os parâmetros pelos valores no descritos na tabela 3 e ficamos com o sistema da seguinte forma

$$\begin{cases} \frac{dS}{dt} = -0,048 - 0,008 = -0,056 & (1) \\ \frac{dE}{dt} = -0,009 - 0,015 - 0,006 + 0,0384 + 0,0048 = 0,0132 & (2) \\ \frac{dI}{dt} = 0,009 + 0,006 + 0,0096 = 0,0246 & (3) \\ \frac{dZ}{dt} = 0,015 + 0,0032 = 0,0182 & (4) \end{cases}$$

Na (1) obtivemos uma taxa de transição negativa de $-0,056 \times 1/hora$, portanto a quantidade de **Suscetíveis** está diminuindo. Isso faz sentido, pois, pela modelagem, não há formas de uma pessoa tornar-se **Suscetível**.

Na (2) obtivemos uma taxa de transição positiva de $0,0132 \times 1/hora$, portanto a quantidade de **Expostos** está aumentando. Apesar da quantidade de **Expostos**, geralmente, tender a 0 esse crescimento não é estranho, pois enquanto há **Suscetíveis**, ainda há possibilidade de surgirem **Expostos** e a quantidade de **Infectados** e **Protestantes** ainda não é suficientemente grande para que a transição direta ocorra majoritariamente.

Na (3) e (4) obtivemos taxas de transição positivas de $0,0246 \times 1/hora$ e $0,0182 \times 1/hora$, respectivamente, portanto a quantidade de **Infectados** e **Protestantes** está au-

mentando. Isso faz sentido, pois, pela modelagem, essas duas categorias são finais, ou seja, a partir delas não há transição para outras.

2.5. Simulação Numérica

A partir da solução numérica implementada através do método de Runge-Kutta de 4ª ordem, com 500 passos, conseguimos obter facilmente os valores para as variáveis S , E , I , Z ao longo do intervalo de tempo $[t_0, t_f]$, $0 \leq t_0 < t_f$.

Dessa forma conseguimos um modo rápido para analisar a evolução do sistema, plotando tais valores em gráficos. Observemos alguns exemplos:

2.5.1. Exemplo 1:

Neste exemplo, analisamos o sistema com valores já utilizados neste relatório para um exercício. Com um acréscimo do intervalo de tempo em que a simulação irá executar.

Tabela 4: Variáveis e parâmetros para simulação numérica do Exemplo 1

Variáveis	Valores	Parâmetros	Valores
$S(0)$	0,4	β	0,6
$E(0)$	0,3	b	0,4
$I(0)$	0,2	ρ	0,1
$Z(0)$	0,1	ε	0,03
t_0	0	θ	0,5
t_f	25	p_1	0,2
		l	0,4

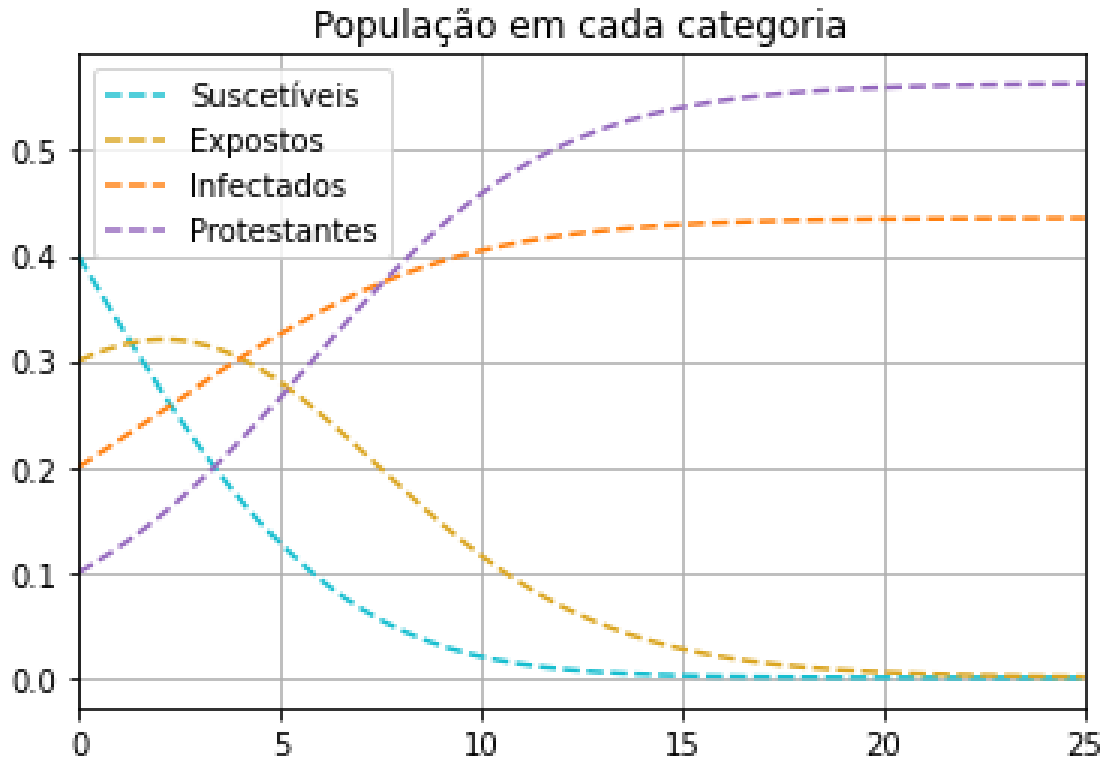


Figura 3. População em cada categoria ao longo do tempo, com os valores informados na Tabela 1

Aqui a atenção deve ser direcionada principalmente aos parâmetros. Enquanto as taxas de contato $\beta = 0,6$ e $b = 0,4$, que definem a taxa de transição $S \rightarrow E$, estão relativamente próximas, a taxa de contato $\rho = 0,1$, que define $E \rightarrow I$, é cinco vezes menor que $\theta = 0,5$. Além disso, a probabilidade $p_1 = 0,2$ da transição $S \rightarrow I$ ocorrer é menor que a probabilidade $l = 0,2$ da transição $S \rightarrow Z$ ocorrer.

Portanto, é de se esperar que a taxa de transição para **Protestante** seja maior que a taxa de transição para **Infectado**. De fato isso ocorre, como podemos observar no gráfico apresentado na Figura 3 acima.

Além disso, no final da simulação temos que $t = t_f = 25$ horas, com as variáveis possuindo os seguintes valores aproximados

Tabela 4: Valores aproximados das variáveis quando $t = t_f$ para o exemplo 1.

Variáveis	Valor
S_{t_f}	0,0000
E_{t_f}	0,0009
I_{t_f}	0,4356
Z_{t_f}	0,5633

Vamos testar a condição $S + E + I + Z = 1$

$$\begin{aligned} S_{t_f} + E_{t_f} + I_{t_f} + Z_{t_f} &= 0,0000 + 0,0009 + 0,4356 + 0,5633 \\ S_{t_f} + E_{t_f} + I_{t_f} + Z_{t_f} &= 0,9998 \approx 1 \end{aligned} \quad (1)$$

Portanto, essa simulação tem o resultado aproximado ao esperado.

2.5.2. Exemplo 2:

Tabela 5: Variáveis e parâmetros para simulação numérica do Exemplo 2

Variáveis	Valores	Parâmetros	Valores
$S(0)$	0,4	β	0,6
$E(0)$	0,2	b	0,6
$I(0)$	0,2	ρ	0,5
$Z(0)$	0,2	ε	0,005
t_0	0	θ	0,5
t_f	25	p_1	0,4
		l	0,4

Neste segundo exemplo mostraremos um caso em que **Infectados** e **Protestantes** se comportam de forma quase idêntica. Assim, escolhemos valores para que $E_0 = Z_0 = 0,4$. Também fizemos o mesmo para os parâmetros, ou seja, as taxas de contato $\beta = 0,6$ e $b = 0,6$, que definem a taxa de transição $S \rightarrow E$, são iguais e as taxas de contato $\rho = 0,5$ e $\theta = 0,5$ também são iguais. Da mesma forma, as probabilidades $p_1 = 0,4$ da transição $S \rightarrow I$ ocorrer e a probabilidade $l = 0,4$ da transição $S \rightarrow Z$ ocorrer são iguais.

Portanto, com todos os parâmetros, com função parecida, idênticos, com apenas ε para trazer alguma diferença, é de se esperar que a taxa de transição para **Protestante** seja bem próxima à taxa de transição para **Infectado**, sendo este último levemente maior por conta do valor baixíssimo de $\varepsilon = 0,005$. De fato isso ocorre, como podemos observar no gráfico apresentado na Figura 4 abaixo.

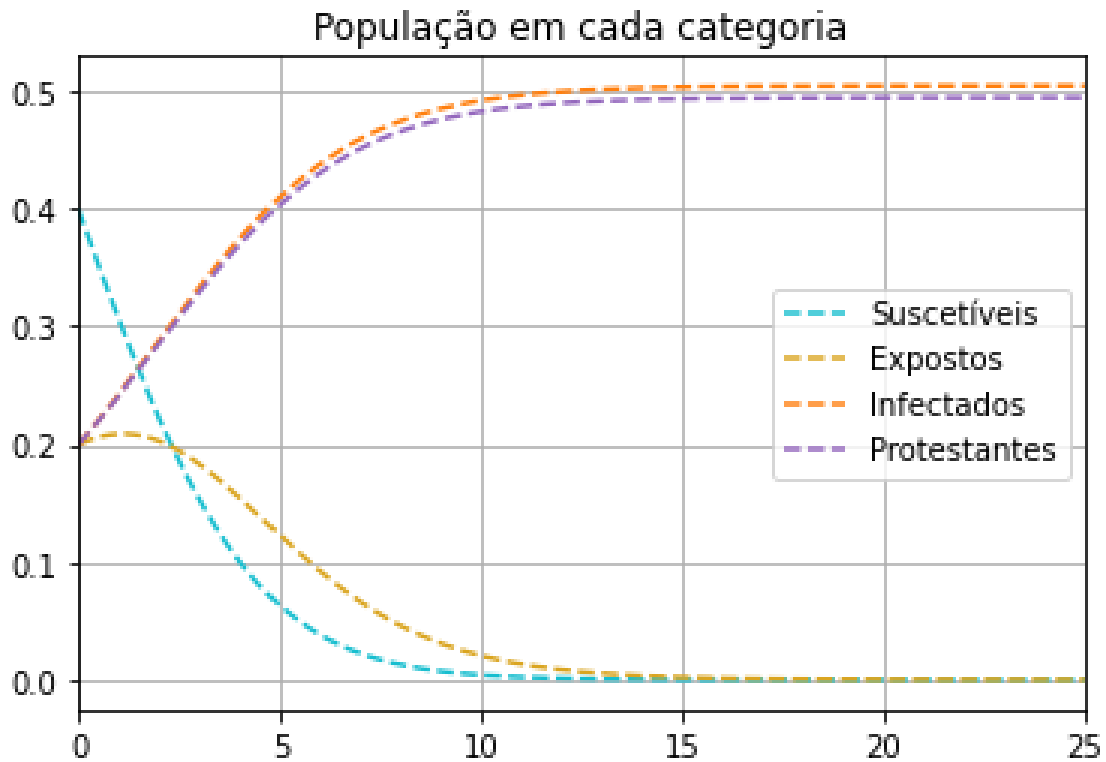


Figura 4. População em cada categoria ao longo do tempo, com os valores informados na Tabela 2

Além disso, no final da simulação temos que $t = t_f = 25 \text{ horas}$, com as variáveis possuindo os seguintes valores aproximados

Tabela 6: Valores aproximados das variáveis quando $t = t_f$ para o exemplo 2.

Variáveis	Valor
S_{t_f}	0,0000
E_{t_f}	0,0000
I_{t_f}	0,5049
Z_{t_f}	0,4950

Testando, novamente, a condição $S + E + I + Z = 1$

$$\begin{aligned}
 S_{t_f} + E_{t_f} + I_{t_f} + Z_{t_f} &= 0,0000 + 0,0000 + 0,5049 + 0,4950 \\
 S_{t_f} + E_{t_f} + I_{t_f} + Z_{t_f} &= 0,9999 \approx 1
 \end{aligned}
 \tag{2}$$

Portanto, essa simulação tem o resultado aproximado ao esperado.

2.5.3. Exemplo 3:

Neste exemplo, vamos analisar a simulação com diferentes valores de parâmetros para **Infectados** e **Protestantes**. Os valores que não forem indicados explicitamente que foram modificados serão definidos pela tabela 7 abaixo.

Tabela 7: Valores de parâmetros base para o exemplo 3

Variáveis	Valores	Parâmetros	Valores
$S(0)$	0,4	β	0,6
$E(0)$	0,2	b	0,4
$I(0)$	0,2	ρ	0,1
$Z(0)$	0,2	ε	0,03
t_0	0	θ	0,5
t_f	25	p_1	0,2
		l	0,4

Tabela 1. Valores de parâmetros para o exemplo 3

A partir das figuras 5 e 6 abaixo, podemos perceber que mesmo alterando as taxas de contato $S-I$ e $S-Z$ a população de **Infectados** e **Protestantes** não se comporta muito diferente ao longo do tempo. Isso se dá pelo fato de que as probabilidades p_1 e l das transição diretas ocorrerem serem pequenas. Indicando o pouco impacto que o primeiro contato com *fake news* tem.

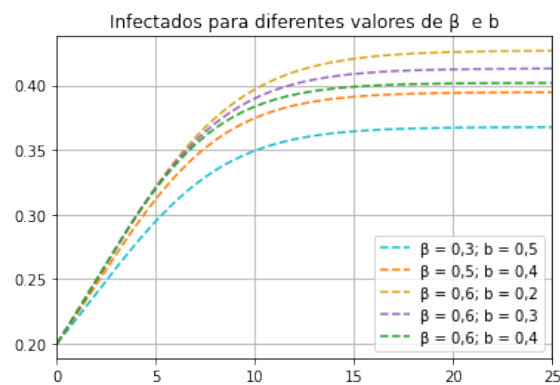


Figura 5. Infectados para diferentes valores das taxas de contato $S-I$ e $S-Z$, respectivamente.

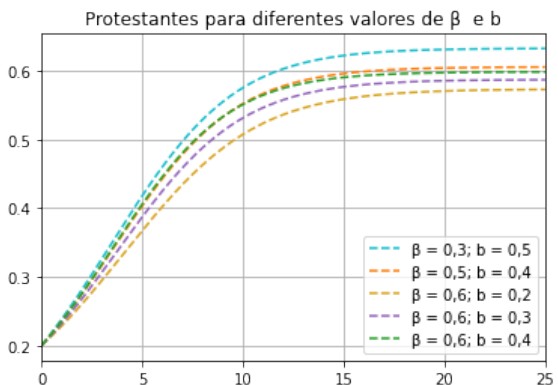


Figura 6. Protestantes para diferentes valores das taxas de contato $S-I$ e $S-Z$, respectivamente.

No entanto, nas figuras 7 e 8 temos uma situação diferente. Ao alterar as taxas de contato $E-I$ e $E-Z$ a população de **Infectados** e **Protestantes** variou notavelmente seu

comportamento. A justificativa é análoga ao das figuras 5 e 6. Isso indica que o maior impacto que uma **fake news** possui é quando uma pessoa que já tem noção da notícia e entra em contato com um **Infectado** ou **Protestante**.

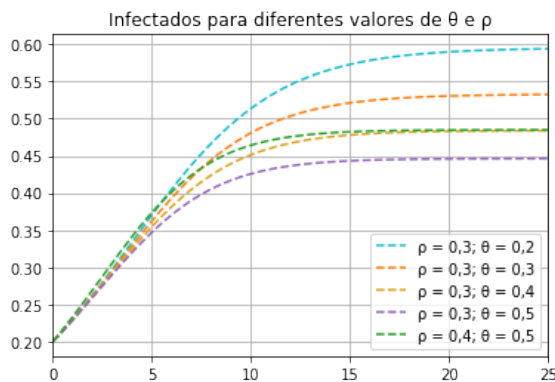


Figura 7. Infectados para diferentes valores das taxas de contato $E-I$ e $E-Z$, respectivamente.

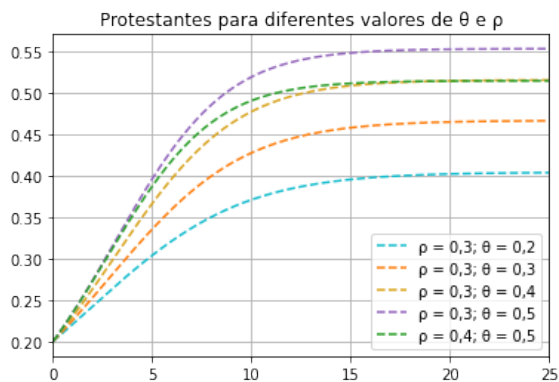


Figura 8. Protestantes para diferentes valores das taxas de contato $E-I$ e $E-Z$, respectivamente.

2.6. Conclusão

A partir das figuras e análises acima conseguimos, em certo nível, descrever e definir uma solução: queremos que a população de **Protestantes** seja significativamente maior que a população de **Infectados**.

Considerando um mundo ideal em que as probabilidades de transição direta são sempre relativamente baixas, então nossa principal preocupação seria com os parâmetros ρ , ε e θ , que vimos o impacto que têm através das figuras 7 e 8.

Pensando no cenário que estabelecemos, uma forma de manipular esses parâmetros em nosso favor seria identificar notícias falsas e redirecionar os possíveis **Expostos** a um contato **Protestante**. Assim, parte da taxa de contato $E-I$ se transformaria em $E-Z$, reproduzindo o comportamento favorável da figura 8.

No entanto, como a população de **Infectados** não diminui, ainda vai existir parte da população nessa categoria, no entanto se eles não conseguem mais entrar em contato inocentes ou vulneráveis, não resta mais nada para eles fazerem com a notícia falsa.

3. Próximos trabalhos

4. Relatório de trabalho em equipe

Referências

A. Kumar and M. H. Biswas. Modeling the dynamics of spreading rumors and fake news through online and social media., 2019.