Transmissão de Informação em Redes

Estudante: Pedro Dib, pedroaraujo@discente.ufg.br
Orientador: Valdivino Vargas Júnior, vvjunior@ufg.br
Universidade Federal de Goiás

Agosto, 2021

Resumo

Este texto descreve a modelagem matemática para a seguinte situação hipotética: o compartilhamento de notícias falsas entre indivíduos numa rede. Mais especificamente, esses indivíduos se encontram dispostos numa árvore binária, onde cada um deles têm probabilidades associadas de transmitir a notícia para 0, 1 ou 2 pessoas à sua direita. Foi feita a apresentação do ferramental teórico necessário para construir esse modelo, os resultados esperados de acordo com a teoria, a simulação do modelo na linguagem de programação Python e, por fim, a comparação entre os dados estimados e os valores reais.

Palavras-chave: Probabilidade; Transmissão de Informação; Notícias Falsas; Função Geradora de Probabilidade; Processos de Ramificação; Simulação Computacional; Python.

1 Introdução

A Transmissão de Informação é uma vertente dos chamados Sistemas de Partículas Interagentes, uma área da Probabilidade em grande expansão atualmente. Desde 2020, sua importância ficou ainda mais evidente com o início da pandemia de um novo tipo de coronavírus.

Em meio a essa situação, a internet transborda de informações sobre os mais diferentes tipos de assuntos relacionados à COVID-19, e muitas dessas informações são notícias falsas.

Dito isso, é do nosso interesse saber, por exemplo, a velocidade de disseminação dessas informações falsas na rede e estimar o impacto que elas têm no contexto em que se inserem, a transição de fase, a proporção da população que recebe a informação, entre outras coisas.

Nosso objetivo neste trabalho foi compreender uma dinâmica simples de transmissão de informação numa árvore binária e realizar simulações computacionais para avaliar, em cenários hipotéticos, a disseminação de uma notícia falsa numa rede.

2 Metodologia

2.1 Referencial teórico

2.1.1 Processos de Ramificação

Primeiramente, vamos entender o que são processos de ramificação e vejamos algumas propriedades deles.

Definição 1. *Entendemos por **processo de ramificação** o processo estocástico com as seguintes características:*

- *Tem-se um único indivíduo na geração $n = 0$;*
- *Todo indivíduo vive exatamente 1 geração, produz Y descendentes e então morre;*
- *O número de descendentes, Y, recebe valores $0, 1, 2, ...$ e a probabilidade de produzir k descendentes é $P(Y = k) = p_k$;*
- *Todos os indivíduos se reproduzem independentemente.*
  Indivíduos $1, 2, ..., m$ têm número de descendentes $Y_{n,1}, Y_{n,2}, ..., Y_{n,m}$, onde cada $Y_{n,i}$ tem a mesma distribuição de Y;

Seja Z_n o número de indivíduos na n-ésima geração. O processo de ramificação é o conjunto $\{Z_0, Z_1, Z_2, ...\} = \{Z_n; n \in \mathbb{N}\}$.

Observação 1. *Uma definição equivalente para um processo de ramificação é: O tamanho da n-ésima geração, Z_n, é uma soma aleatoriamente interrompida do número de descendentes $Y_{n,i}$, isto é,*

$$Z_n = \sum_{i=1}^{Z_{n-1}} Y_{n,i}. \tag{1}$$

Esse fato nos permite obter propriedades muito interessantes acerca dos processos de ramificação, como veremos a seguir.

Teorema 2.1 (As iterações de G)**.** *Seja $G(s) = E(s^Y) = \sum_{y=0}^{\infty} s^y P(Y = y)$ a função geradora de probabilidade da distribuição do número de descendentes, Y. Tome $Z_0 = 1$ e Z_n o tamanho da população na geração n. Considere $G_n(s)$ a função geradora de probabilidade de Z_n. Nessas condições,*

$$G_n(s) = G(G(G(...G(s)...))), n \text{ vezes}. \tag{2}$$

$G_n(s)$ é a n-ésima iteração de G.

Proposição 1. *Sejam $\{Z_n; n \in \mathbb{N}\}$ um processo de ramificação e Y a distribuição do número de descendentes desse processo. Suponha $E(Y) = \mu$. Assim,*

$$E(Z_n) = \mu^n. \tag{3}$$

Uma das aplicações mais interessantes dos processos de ramificação é o cálculo da probabilidade de eventual extinção. Alguns exemplos: qual a probabilidade de que uma colônia de células cancerígenas se torne extinta antes que ela se alastre pelos tecidos adjacentes? Ou então, qual a probabilidade de uma doença infecciosa morrer antes de alcançar o nível epidêmico? Podemos obter vários resultados acerca da extinção de um processo de ramificação. Veremos alguns agora.

Definição 2. *Dado $\{Z_i; i \in \mathbb{N}\}$ um processo de ramificação, dizemos que a população está* ***extinta na geração*** *n se $Z_n = 0$. Definimos o evento $E_n = \{Z_n = 0\}$. Dessa forma, $\gamma_n = P(E_n)$ é a probabilidade do processo estar extinto na geração n.*

Definição 3 (Extinção final)**.** *A probabilidade do processo estar extinto na geração n, para qualquer valor de n, é chamada* ***probabilidade de extinção final****, ou simplesmente probabilidade de extinção, denotada por γ.*

O Lema a seguir utiliza uma propriedade cuja prova pode ser encontrada em [2].

Lema 1. *Sejam γ e γ_n como definidos acima. Temos que*

$$\gamma = P(\bigcup_{n=0}^{\infty} E_n) = P(\lim_{n\to\infty} E_n) = \lim_{n\to\infty} P(E_n) = \lim_{n\to\infty} \gamma_n. \tag{4}$$

Teorema 2.2. *Seja γ a probabilida de extinção final. Temos que o valor de γ é a menor solução não-negativa da equação $G(s) = s$, onde $G(s)$ é a função geradora de probabilidade da distribuição do número de descendentes, Y.*

Corolário 2.2.1. *Sempre existe solução para $G(s) = s$ em $[0, 1]$. Além disso, $\gamma = \min\left\{1, \bar{s}\right\}$, onde $\bar{s}$ é uma raiz de $G(s) = s$.*

Teorema 2.3 (Condições para a extinção final)**.** *Sejam $\{Z_n; n \in \mathbb{N}\}$ um processo de ramificação e Y sua distribuição do número de descendentes. Suponha $\mu = E(Y)$ a média de Y e γ a probabilidade de extinção final. Nessas condições,*

(i) Se $\mu > 1$, então $\gamma < 1$. A extinção é possível, mas não está garantida;

(ii) Se $\mu < 1$, então $\gamma = 1$. A extinção ocorre com probabilidade 1;

(iii) Se $\mu = 1$ e $P(Y = 1) \neq 1$, então $\gamma = 1$. A extinção também ocorre com probabilidade 1.

Para mais detalhes sobre processos de ramificação, as demonstrações dos teoremas e exemplos, verificar [1]. Precisaremos também de outras ferramentas para trabalhar em nosso modelo. Falaremos agora sobre:

2.1.2 Funções geradoras fracionais lineares

Definição 4. *Uma função geradora de probabilidade é dita ser fracional linear quando tem a forma*

$$G(b,c,s) = 1 - \frac{b}{1-c} + \frac{bs}{1-cs},$$

onde $s, b \in [0,1]$*,* $c \in [0{,}1)$ *e* $b + c \leq 1$*.*

Observação 2. *Equivalentemente, podemos escrever a função geradora de probabilidade funcional linear da seguinte forma:*

$$f(s,\mu,c) = 1 - \mu(1-c) + \frac{\mu(1-c)^2 s}{1-cs}. \tag{5}$$

Proposição 2. *Seja* $G(s)$ *uma função geradora de probabilidade funcional linear e considere funções geradores geradoras* L *e* U *tais que* $L(s) \leq G(s) \leq U(s)$*, com* $s \in [0{,}1]$*. Realizando a composição* n *vezes:*

$$L_n(s) \leq G_n(s) \leq U_n(s).$$

Consequentemente,

$$\sum_{n=0}^{\infty}[1 - U_n(0)] \leq \sum_{n=0}^{\infty}[1 - G_n(0)] \leq \sum_{n=0}^{\infty}[1 - L_n(0)].$$

Isto é, se a função geradora $G(s)$ *admite limitantes superior e inferior* $U(s)$ *e* $L(s)$*, respectivamente, então, ao realizarmos a composição* n *vezes, a desigualdade é mantida.*

Teorema 2.4. *Seja* $G(s)$ *uma função geradora tal que* $G'(1) = \mu$ *e* $f(s,\mu,c)$ *uma família de funções geradoras fracionais lineares do tipo 5. Suponha que*

$$G'(1)G'(s)(1-s)^2 - (1-G(s))^2 < 0, \text{ para todo } s \in [0{,}1).$$

Nessas condições,

(i) $f(s,\mu,c_u)$ *é a melhor função geradora fracional linear de limite superior para* $G(s)$*, onde* $c_u = \frac{G'(1)+G(0)-1}{G'(1)}$*;*

(ii) $f(s,\mu,c_l)$ *é a melhor função geradora fracional linear de limite inferior para* $G(s)$*, onde* $c_l = \frac{G''(1)}{2G'(1)+G''(1)}$*, com* $G''(1) < \infty$*.*

Observação 3. *Matematicamente, escrevemos*

$$f(s,m,c_l) \leq G(s) \leq f(s,m,c_u),$$

onde $G'(1) = \mu = f'(1,\mu,c_l) = f'(1,\mu,c_u)$*.*

Corolário 2.4.1. *A Proposição 2 é válida quando temos que*

$$G'(1)G''(s) - G'(s)G''(1) \geq 0, \forall s \in [0,1].$$

Corolário 2.4.2. *Nas condições da Proposição 2,*

$$f_n(s, \mu, c_l) \leq G_n(s) \leq f_n(s, \mu, c_u).$$

Mais informações acerca de funções geradoras fracionais lineares, bem como outros exemplos de utilização delas podem ser encontradas em [3].

2.2 O modelo

No presente trabalho, temos o seguinte processo de transmissão. Numa árvore binária, cada indivíduo, ao receber a notícia falsa, tenta repassá-la para seus vizinho à direita. Inicialmente, apenas 1 indivíduo começa com a informação. Essa é a geração 0 do processo. A Figura 1 ilustra como seria o processo.

Cada linha vertical de círculos representa uma geração do processo, e o número de círculos coloridos em cada geração representa o tamanho Z daquela geração. Isto é, na figura, temos: $Z_0 = 1$, $Z_1 = 2$, $Z_2 = 4$, $Z_3 = 6$, e assim por diante. Nesse processo, consideramos a seguinte lei de probabilidade:

$$P_k = P(X = k) = \begin{cases} p, & \text{se } k = 2 \\ q, & \text{se } k = 1 \\ r = 1 - p - q, & \text{se } k = 0 \end{cases}$$

onde P_k é a probabilidade de um indivíduo transmitir a informação para k indivíduos à direita.

A partir disso, foram testados, no programa que pode ser acessado através deste link, diferentes valores para p e q e os seus alcances médios, para esses respectivos valores, de onde podemos inferir o tempo de extinção do processo.

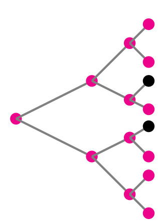

Figura 1: Árvore binária.

O programa foi implementado a partir do seguinte algoritmo:

Algorithm 1 Obtenção do alcance médio do processo

```
Parâmetros: n, p e q
DEFINA n, p e q
r ← 1 − p − q

para j variando de 1 até n faça                          :
    contador ← 0
    y ← uniforme(0;1)
    se 0 ≤ y < r então                                   :
        l ← 0
    fim se
    se r ≤ y < r + q então                               :
        l ← 1
    fim se
    se y ≥ r + q então                                   :
        l ← 2
    fim se

    se l > 0 então                                       :
        contador ← +1
    fim se

    enquanto l > 0 faça                                  :
        y ←l uniforme(0;1)
        aux ← l
        l ← 0

        para i de 1 até aux faça                         :
            se r ≤ y(i) < r + q então                    :
                l ← l + 1
            fim se
            se r + q ≤ y(i) < 1 então                    :
                l ← l + 2
            fim se
        fim para
        se l > 0 então                                   :
            contador ← contador +1
        fim se
    fim enquanto
    x(j) ← contador
fim para
```

media $\leftarrow \frac{\sum_{j=1}^{n} x(j)}{n}$

3 Resultados obtidos

3.1 Resultados teóricos

Partindo de teoremas amplamente conhecidos sobre teoria de Processos de Ramificação, alguns resultados teóricos sobre o modelo em questão foram obtidos para servirem como referencial para avaliarmos se a simulação representa bem, ou não, o nosso modelo.

Lema 2. *Sejam X a distribuição do número de pessoas que recebem a informação de um indivíduo à sua esquerda e p e q como definidos anteriormente. Temos que*

$$\mu = E(X) = 2p + q \tag{6}$$

e também

$$G(s) = E(s^X) = ps^2 + qs + 1 - p - q \tag{7}$$

onde μ é a média de X e $G(s)$ sua função geradora de probabilidade.

Demonstração.

Das definições,

$$E(X) = \sum_{x=0}^{2} xP(X = x) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) = q + 2p$$

E também

$$\begin{aligned} G(s) = E(s^X) &= \sum_{x=0}^{2} s^x P(X = x) \\ &= s^0 \cdot P(X = 0) + s^1 \cdot P(X = 1) + s^2 \cdot P(X = 2) \\ &= (1 - p - q) + qs + ps^2. \end{aligned}$$

□

Proposição 3 (Transição de fase)**.** *Seja V o evento onde o processo sobrevive. Assim, $P(V) > 0$ se, e somente se, $2p + q > 1$.*

Demonstração.

Um resultado essencial da teoria de processos de ramificação é que um processo sobrevive com probabilidade positiva se, e somente se, a média μ do número de sobreviventes é maior que 1.
Sendo assim, do Lema 1, temos que $\mu = 2p + q \Rightarrow P(V) > 0 \Leftrightarrow 2p + q > 1$. □

Teorema 3.1. *Sendo V o evento onde o processo sobrevive, temos que*

$$P(V) = \max\left\{0, \frac{2p+q-1}{p}\right\}. \tag{8}$$

Demonstração.

Primeiramente, calculemos a probabilidade de extinção γ. Ela é a menor solução não-negativa da equação $G(s) = s$, isto é, da equação:

$$ps^2 + (q-1)s + 1 - p - q = 0.$$

Como já sabemos que $s = 1$ sempre é solução para $G(s) = s$, podemos dividir o polinômio acima por $(s-1)$ para obtermos a outra raiz.
Procedendo assim, obtemos a seguinte fatoração:

$$ps^2 + (q-1)s + 1 - p - q = (s-1)(ps + p + q - 1) = 0 \Rightarrow \gamma = \min\left\{1, \frac{r}{p}\right\}.$$

Desse modo, temos que a probabilidade de sobrevivência é

$$P(V) = 1 - \gamma = 1 - \min\left\{1, \frac{r}{p}\right\} = \max\left\{0, 1 - \frac{r}{p}\right\} = \max\left\{0, \frac{2p+q-1}{p}\right\},$$

como queríamos demonstrar. $\square$

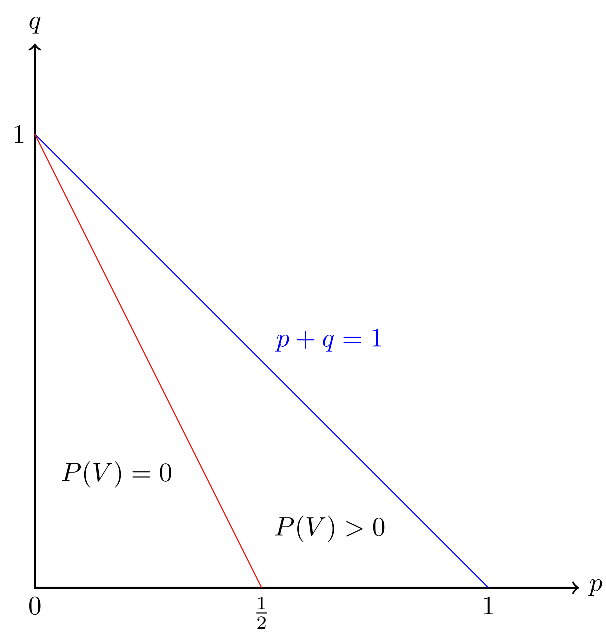

Figura 2: Gráfico da transição de fase.

Teorema 3.2. *Seja C o número de indivíduos que recebem a informação ao longo do processo. Suponha que:*

(i) $2p + q < 1$. *Então,* $E(C) = \frac{1}{1-2p-q}$.

(ii) $2p + q > 1$. *Então,* $E(C|V^{\complement}) = \frac{1}{2p+q-1}$.

Demonstração.

No caso de $\mu = 2p + q < 1$,

$$E(C) = E\Big(\sum_{n=0}^{\infty} Z_n\Big) = \sum_{n=0}^{\infty} E(Z_n) = \sum_{n=0}^{\infty} \mu^n = \frac{1}{1-\mu}.$$

Para $\mu = 2p + q > 1$, temos que:

$$\begin{aligned} E(C|V^{\complement}) = E\Big(\sum_{n=0}^{\infty} Z_n|V^{\complement}\Big) &= \sum_{n=0}^{\infty} E(Z_n|V^{\complement}) \\ &= \frac{1}{1-G'(\gamma)} = \frac{1}{1-(2p(\frac{r}{p})+q)} = \frac{1}{2p+q-1}. \end{aligned}$$

□

Teorema 3.3 (Extinção)**.** *Seja $A = \max\{n > 0; Z_n > 0\}$ o alcance da mensagem. Tem-se que*

$$\frac{c_1(1-\mu^{n+1})}{c_1-\mu^{n+1}} \leq P(A \leq n) \leq \frac{c_2(1-\mu^{n+1})}{c_2-\mu^{n+1}}, \tag{9}$$

onde $c_1 = 1 + \frac{(2p+q)(1-2p-q)}{p}$ *e* $c_2 = \frac{(1-p-q)(2p+q)}{p}$ *e* $\mu = 2p + q$.
Além disso,

$$(c_2 - 1)\sum_{n=0}^{\infty} \frac{\mu^{n+1}}{c_2-\mu^{n+1}} \leq E(A) \leq (c_1-1)\sum_{n=0}^{\infty}\frac{\mu^{n+1}}{c_1-\mu^{n+1}}. \tag{10}$$

Demonstração.

Considere $T = \min\{n \geq 0; Z_n = 0\}$ o tempo de extinção do processo.
A ideia aqui é utilizar a Proposição 2 e o Teorema 2.4. Isto é, queremos encontrar funções geradoras $L_n(s)$ e $U_n(s)$ tais que

$$L_n(0) \leq P(T \leq n) \leq U_n(0).$$

Como $T = A + 1$, temos que

$$L_{n+1}(0) \leq P(A \leq n) \leq U_{n+1}(0).$$

Perceba que

$$E(A) = \sum_{n=0}^{\infty} P(A > n) = \sum_{n=0}^{\infty} [1 - P(A \leq n)]$$

e que

$$1 - U_{n+1}(0) \leq 1 - P(A \leq n) \leq 1 - L_{n+1}(0),$$

o que implica

$$\sum_{n=0}^{\infty} [1 - U_{n+1}(0)] \leq \sum_{n=0}^{\infty} [1 - P(A \leq n)] \leq \sum_{n=0}^{\infty} [1 - L_{n+1}(0)]. \tag{11}$$

Ou seja,

$$\sum_{n=0}^{\infty} [1 - U_{n+1}(0)] \leq \sum_{n=0}^{\infty} E(A) \leq \sum_{n=0}^{\infty} [1 - L_{n+1}(0)]. \tag{12}$$

Encontremos, então, $L_n(0)$ e $U_n(0)$. Primeiramente, devemos verificar se $V(s) = G'(1)G''(s) - G'(s)G''(1) \geq 0$. Ora, como $G(s) = ps^2 + qs + 1 - p - q$, temos que:

- $G'(s) = 2ps + q$,
- $G''(s) = 2p$,
- $G(0) = 1 - p - q$,
- $G'(1) = 2p + q = \mu$,
- $G''(1) = 2p$,

de modo que $V(s) = (2p + q)p - (2ps + q)2p \geq 0 \Leftrightarrow 2p + q \geq 2ps + q$, o que, de fato, acontece para todo $s \in [0, 1]$.
Encontremos, agora, nossas constantes c_1 e c_2. A primeira tem a forma

$$c_1 = \frac{2G'(1) + G''(1) - 2[G'(1)]^2}{G''(1)} = 1 + \frac{(2p + q)(1 - 2p - q)}{p}.$$

Já a segunda,

$$c_2 = \frac{G(0)G'(1)}{G'(1) + G(0) - 1} = \frac{(1 - p - q)(2p + q)}{p}.$$

Assim, $L_n(0)$ e $U_n(0)$ serão:

$$L_n(0) = \frac{c_1(1 - \mu^n)}{c_1 - \mu^n} = \frac{\left[1 + \frac{(2p+q)(1-2p-q)}{p}\right][1 - (2p + q)^n]}{1 + \frac{(2p+q)(1-2p-q)}{p} - (2p + q)^n},$$

$$U_n(0) = \frac{c_2(1-\mu^n)}{c_2-\mu^n} = \frac{\left[\frac{(1-p-q)(2p+q)}{p}\right][1-(2p+q)^n]}{\frac{(1-p-q)(2p+q)}{p} - (2p+q)^n}.$$

Obtidas as funções geradoras em função de n, teremos que

$$\frac{\left[1+\frac{(2p+q)(1-2p-q)}{p}\right][1-(2p+q)^n]}{1+\frac{(2p+q)(1-2p-q)}{p} - (2p+q)^n} \leq P(A \leq n)$$
$$\leq \frac{\left[\frac{(1-p-q)(2p+q)}{p}\right][1-(2p+q)^n]}{\frac{(1-p-q)(2p+q)}{p} - (2p+q)^n}.$$

Substituindo os valores de $L_{n+1}(0), U_{n+1}(0)$ e $P(A \leq n)$ na equação (11) nos dá o resultado desejado. $\square$

3.2 Resultados práticos

Ao realizar a simulação para $n = 100.000$ repetições, para diferentes valores de p e q, obtemos os valores médios $\tilde{m}$ do alcance do processo e o comparamos com a média μ do número de descendentes do processo, as constantes c_1 e c_2 e o intervalo teórico m estabelecidos pelo Teorema 3.3. Os resultados são apresentados na tabela abaixo.

p	q	μ	c_1	c_2	m	$\tilde{m}$
0.30	0.30	0.90	1.3000	1.2000	[2.95, 3.71]	3.28
0.20	0.50	0.90	1.4500	1.3500	[4.01, 4.52]	4.22
0.27	0.45	0.99	1.0366	1.2666	[9.21, 11.70]	10.95
0.25	0.48	0.98	1.0784	1.0584	[7.90, 9.69]	9.03
0.18	0.61	0.97	1.1616	1.1316	[8.81, 9.98]	9.39

Figura 3: Valores experimentais para o alcance médio do processo.

4 Conclusão

Diante do exposto, conclui-se que os resultados presentes na literatura são suficientes para encontrar limitantes para a distribuição e a média do alcance do processo de ramificação avaliado. A proposta do trabalho foi obter, através de simulação computacional, uma estimativa pontual para o valor real da média desse alcance. Essa estimativa se encontra bem centralizada em relação aos limitantes teóricos propostos, de modo que o objetivo foi alcançado.

Referências

[1] Course Notes, STATS 325 : Stochastic Processes, Department of Statistics, University of Auckland.

[2] ROSS, S. Probabilidade: Um Curso Moderno com Aplicações, 8ª Edição Bookman.

[3] SANTOS, G. C. **Função Geradora de Probabilidade e Processos de Ramificação**. Trabalho de Conclusão de Curso (Bacharelado em Estatística) - Instituto de Matemática e Estatística, Universidade Federal de Goiás. Goiânia, p. 44. 2021.

5 Certificados

Não foi possível comparecer às atividades do Programa Diálogos em Pesquisa e Inovação pois a maioria das palestras ocorreram em horários que chocavam com as disciplinas obrigatórias cursadas no período do programa (14:00 e 16:00). Quanto às palestras que aconteceram em outros horários, não pude comparecer pois estava ocupado com as atividades assíncronas das matérias obrigatórias.