Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 11, n. 1, 2025.

Trabalho apresentado no XLIII CNMAC, Centro de Convenções do Armação Resort - Porto de Galinhas - PE, 2024

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Utilizando Modelos SIR e SIRV com Criptografia Totalmente Homomórfica

Aline de L. Z. Lunkes[1]
Unimontes, Montes Claros, MG
Fábio Borges[2]
LNCC, Petrópolis, RJ

Resumo. Neste trabalho, apresentamos um estudo de caso utilizando os modelos Suscetíveis-Infectados-Removidos (SIR) e Suscetíveis-Infectados-Removidos-Vacinados (SIRV), sistemas que modelam a transmissão dinâmica de doenças em populações com três e quatro compartimentos, respectivamente. Para resolver numericamente o sistema de equações diferenciais ordinárias (EDO) acopladas associado a esses modelos, empregamos o método das diferenças finitas. Após o estudo teórico, aplicamos os modelos aos casos dos vírus Influenza A (H1N1), com surto no Brasil em 2009, e COVID-19, emergente em 2019, ambos escolhidos por sua relevância global. Nosso objetivo principal é comparar os resultados obtidos, com e sem o uso de Criptografia Totalmente Homomórfica, em inglês Fully Homomorphic Encryption (FHE), protegendo as informações dos modelos SIR e SIRV e garantindo a integridade dos dados. Os dados são armazenados no Brasil, em um banco de dados do Sistema Único de Saúde (SUS), e devem manter sua autenticidade. Demonstramos que o uso da Criptografia Totalmente Homomórfica preserva a equivalência dos resultados, garantindo a segurança e precisão das informações sem comprometê-las.

Palavras-chave. SIR, SIRV, Criptografia Totalmente Homomórfica, Privacidade, Segurança

1 Introdução

Em 1918, surgiu o vírus Influenza A (H1N1), que ao longo dos anos passou por rearranjos genéticos com outros vírus influenza, tanto humanos (H1N1, H1N2, H3N2) quanto de origem animal. Em 2009, ocorreu a primeira pandemia de gripe do século 21, desencadeada por uma nova cepa do vírus, o Influenza A (H1N1) de origem suína (S-OIV). Acredita-se que a transmissão tenha iniciado no México e nos Estados Unidos, espalhando-se rapidamente pelo mundo, conforme documentado em [4] e [9]. Embora o nome possa sugerir o contrário, o vírus não é transmitido pelo consumo de carne suína, mas sim pelo ar, de forma semelhante à gripe comum, através de gotículas liberadas ao tossir ou espirrar, como descrito em [6] e [10].

Em 2020, a China atraiu a atenção global ao registrar os primeiros casos de COVID-19, conforme citado em [10]. Posteriormente, a propagação da doença e suas altas taxas de contágio afetaram não apenas a Itália, mas também diversos outros países, como destacado em [2]. O vírus SARS-CoV-2, responsável pela COVID-19, tornou-se uma pandemia mundial no início de 2020 e vem sendo gradualmente controlado, em parte devido à evolução de suas variantes. Esse controle tem sido viabilizado pela distribuição global de cinco tipos diferentes de vacinas. Devido à novidade da doença, os sintomas variaram entre os indivíduos no início, conforme discutido em [3]. Inicialmente confundida com a gripe, a COVID-19 rapidamente se tornou foco de pesquisa, com estudos sobre a Influenza A (H1N1) oferecendo valiosas contribuições para o entendimento da nova doença.

[1]alinelunkesazl@gmail.com
[2]borges@lncc.br

A seleção dessas doenças baseia-se em seu impacto significativo no Brasil e em outros países. Assim, o objetivo principal deste trabalho é investigar a propagação dessas enfermidades e aplicar a Criptografia Totalmente Homomórfica para garantir a integridade dos dados, prevenindo qualquer forma de manipulação. Este estudo é dividido entre a formulação matemática já estabelecida na literatura para essas doenças e sua implementação utilizando Criptografia Totalmente Homomórfica.

2 Modelagem Matemática

Analisamos o modelo Suscetível-Infectado-Removido (SIR) sob a suposição de que há uma mistura homogênea entre as populações infectadas e suscetíveis e que a população total permanece constante ao longo do tempo. O modelo SIR é descrito por um sistema de três equações diferenciais ordinárias (EDOs) de primeira ordem acopladas, que constituem um sistema dinâmico dessas populações. Esse modelo permite realizar previsões sobre o número de infecções e mortes que podem ocorrer em uma determinada comunidade. No modelo SIR, utilizamos as seguintes populações:

$S(t)$: Um indivíduo suscetível pode ser infectado ou permanecer suscetível;

$I(t)$: Indivíduos infectados são aqueles que já foram infectados pelo vírus e podem transmiti-lo a indivíduos suscetíveis;

$R(t)$: Indivíduos removidos são aqueles que se recuperaram do vírus e são considerados temporariamente imunes, assumimos também que $R(t)$ é formado por aqueles indivíduos que morreram pela doença.

Para obter o sistema de Equações Diferenciais Ordinárias acopladas para o modelo SIR, assumimos que N(t) é constante para qualquer tempo t, e denotamos a população total por $N(t) = S(t) + I(t) + R(t)$. As taxas de mudança dessas populações são governadas pelo seguinte sistema

$$\begin{cases} \frac{dS_t}{dt} = \alpha S(t) + \varepsilon R(t) - \sigma S(t)I(t) - \beta S(t) \\ \frac{dI_t}{dt} = \sigma S(t)I(t) - \gamma I(t) - \mu I(t) \\ \frac{dR_t}{dt} = \gamma I(t) - \varepsilon R(t) - \beta R(t) \\ S(0) = S_0, I(0) = I_0, R(0) = R_0. \end{cases} \tag{1}$$

onde os parâmetros são escolhidos da seguinte forma: α é a taxa de natalidade da população brasileira; β é a taxa de mortalidade da população brasileira nas classes Suscetíveis, Recuperados e Vacinados; μ é a taxa de mortalidade da população brasileira na classe Infectados; σ é a taxa na qual indivíduos da classe Suscetível entram em contato com indivíduos da classe Infectados; γ é a taxa de recuperação da doença na população brasileira; ε corresponde à taxa com que os indivíduos perdem imunidade e passam da classe Recuperados para a classe Suscetível. As condições iniciais são $S(0) = S_0$, $I(0) = I_0$ e $R(0) = R_0$.

Para obter o sistema de equações diferenciais ordinárias acopladas para o modelo SIRV, denotamos a população total por

$$N(t) = S(t) + I(t) + R(t) + V(t),$$

onde $N(t)$ é constante para qualquer valor de t. Além disso, assumimos que a classe $S(t)$ está sujeita a uma vacinação bem-sucedida, ou seja, os indivíduos passam da categoria $S(t)$ para a categoria $V(t)$. Suponhamos inicialmente que os indivíduos vacinados não podem contrair a doença. Portanto, não há impacto nas classes $I(t)$ ou $R(t)$. As taxas de mudança dessas populações são governadas pelo seguinte sistema (consultar os artigos [8], [5], [1] e [7]). Além disso, assumimos

que em um intervalo de tempo Δt, a classe $S(t)$ está sujeita a uma vacinação bem-sucedida, ou seja, as pessoas passam da categoria $S(t)$ para a categoria $V(t)$. Suponhamos inicialmente que os vacinados não podem contrair a doença. Portanto, não há impacto nas classes $I(t)$ ou $R(t)$. Segue o sistema de SIRV,

$$\begin{cases} \frac{dS_t}{dt} = \alpha S(t) + \varepsilon R(t) - \sigma S(t)I(t) - \lambda S(t) - \beta S(t) \\ \frac{dI_t}{dt} = \sigma S(t)I(t) - \gamma I(t) - \mu I(t) \\ \frac{dR_t}{dt} = \gamma I(t) - \lambda R(t) - \varepsilon R(t) - \beta R(t) \\ \frac{dV_t}{dt} = \lambda(S(t)R(t) + R(t)) - \beta V(t) \\ S(0) = S_0, I(0) = I_0, R(0) = R_0, V(0) = V_0, \end{cases} \tag{2}$$

onde p é o parâmetro do coeficiente de vacinação descontínua, e λ é a taxa de vacinação.

3 Influenza A (H1N1) e COVID-19

Nesta seção, abordamos as doenças Influenza A (H1N1) e COVID-19. Devido à escassez de dados iniciais sobre a COVID-19, muitos pesquisadores utilizaram conceitos estabelecidos para a Influenza A (H1N1) como base para modelar a propagação e o controle da COVID-19, razão pela qual utilizaremos dados reais da COVID-19 neste estudo. Os sintomas da infecção por COVID-19 são semelhantes aos da Influenza A (H1N1), com a adição de ataque cardíaco, distúrbios olfativos e gustativos e obstrução nasal. Além disso, estratégias de contágio, controle e vacinação desenvolvidas para a Influenza A (H1N1) foram adaptadas para enfrentar a COVID-19. Assim, esta revisão abrange tanto a Influenza A (H1N1) quanto a COVID-19, utilizando informações da Influenza A (H1N1) com parâmetros ajustados, provenientes de fontes como a Organização Mundial da Saúde (OMS)[3].

Com base nos resultados apresentados na Seção 2, discutimos a modelagem matemática da dinâmica da Influenza A (H1N1), conforme descrito em [11]. Os estudos [8], [1], [5], e [7] apresentam modelos utilizados para a Influenza A (H1N1) equivalentes aos sistemas descritos nas Equações (1) e (2). O sistema da Equação (1) representa o modelo sem vacinação, enquanto a Equação (2) inclui a vacinação. Além disso, esses modelos descrevem a evolução temporal e as relações entre as quatro populações: S, I, R e V.

Além disso, assume-se que a escala temporal dos modelos SIR e SIRV é curta o suficiente para que os nascimentos e mortes (exceto as mortes causadas pelo vírus) possam ser negligenciados, e que o número de mortes causadas pelo vírus seja pequeno em comparação com a população viva. Com base nessas premissas, os conceitos e as taxas de variação das classes, registradas pelos sistemas de ODEs, constituem os modelos SIR e SIRV utilizados neste estudo, proporcionando uma melhor compreensão de como o vírus se espalha na população ao longo do tempo. Para a análise da transmissão da COVID-19, dispomos dos dados das condições iniciais e dos parâmetros do modelo, conforme apresentados na Tabela 1. Esses dados foram elaborados com base nas informações divulgadas pelo Ministério da Saúde do Brasil[4].

[3] `https://www.gov.br/saude/pt-br`
[4] `https://covid.saude.gov.br/`

Tabela 1: Parâmetros para resolução numérica dos dados publicados pelo Ministério da Saúde do Brasil.

Elementos	Valores
α	17045917×10^{-12}
β	17045917×10^{-12}
μ	41922002×10^{-9}
σ	1130833×10^{-16}
γ	0.004192002
ε	5502466×10^{-12}
λ	1916916×10^{-9}

Os sistemas de EDOs (1) e (2) são discretizados utilizando o método de Diferenças Finitas Regressivas como vemos nas EDOs (3) e (4). A escolha desse método se deve à sua estabilidade, que é garantida por sua capacidade de lidar com equações diferenciais de primeira ordem, assegurando que as soluções numéricas sejam consistentes e convergentes ao longo do tempo. Esse método é particularmente adequado para problemas de evolução temporal, como os modelos SIR e SIRV, uma vez que proporciona uma abordagem robusta para a simulação do comportamento dinâmico das populações ao longo do tempo.

$$\begin{cases} (1-\alpha(\Delta t)+\beta(\Delta t))S_i-(1-\sigma I_{i-1}(\Delta t))S_{i-1})=\varepsilon R_{i-1}(\Delta t) \\ (1+\gamma\Delta t+\mu\Delta t)I_i-(1+\sigma S_{i-1}\Delta t)I_{i-1}=0 \\ (1+\beta(\Delta t))R_i-(1-\varepsilon(\Delta t))R_{i-1}=\gamma I_i(\Delta t). \end{cases} \tag{3}$$

$$\begin{cases} (1-\alpha\Delta t+\lambda\Delta t+\beta(\Delta t)S_i-(1-\sigma I_{i-1}\Delta t)S_{i-1})=\varepsilon R_{i-1}\Delta t \\ (1+\gamma\Delta t+\mu\Delta t)I_i-(1+\sigma S_{i-1}\Delta t)I_{i-1}=0 \\ (1+\beta\Delta t+\lambda\Delta t)R_i-(1-\varepsilon\Delta t)R_{i-1}=\gamma I_i\Delta t \\ (1+\beta\Delta t)V_i-V_{i-1}=\lambda\Delta t(S_i+R_i). \end{cases} \tag{4}$$

4 Análise e Resultados

O método de Criptografia Totalmente Homomórfica é utilizado para preservar a privacidade e a integridade dos dados, mesmo quando processados em sistemas não confiáveis, garantindo que os dados confidenciais ou privados permaneçam protegidos. As operações homomórficas permitem que os dados sejam processados em seu estado criptografado, sem a necessidade de descriptografá-los, o que reduz significativamente o risco de vazamento de informações sensíveis. Dessa forma, a Criptografia Totalmente Homomórfica é uma ferramenta crucial para proteger os dados de saúde dos pacientes. Ressaltamos que utilizamos os resultados numéricos, sem a aplicação da Criptografia Totalmente Homomórfica, apenas como base para comparação com os dados resultantes após a aplicação da criptografia, e destacamos que o uso de dados não criptografados não é adequado para outras aplicações que envolvem dados confidenciais.

Implementamos em linguagem Python os modelos SIR e SIRV para simular a propagação de uma doença em uma população, utilizando os dados reais da COVID-19 no Brasil do período de 2020 à 2023. O modelo calcula a dinâmica das populações de indivíduos suscetíveis, infectados e recuperados, no SIR e vacinados no modelo SIRV. As condições iniciais do sistema e seus parâmetros foram obtidos a partir de dados publicados pelo Ministério da Saúde do Brasil[5], até 15 de fevereiro de 2023. Parâmetros como taxa de transmissão, taxa de recuperação e taxa de vacinação foram

[5] `https://covid.saude.gov.br/`

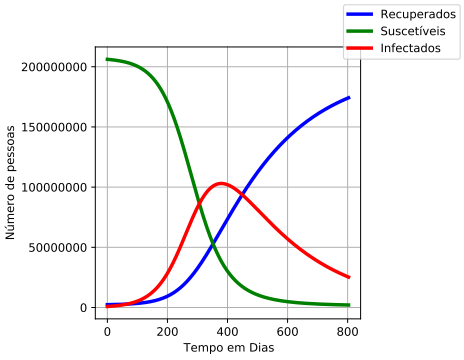
obtidos na Tabela 1. Para garantir a privacidade dos dados, utilizamos a biblioteca Criptografia Totalmente Homomórfica - Microsoft SEAL[6], que realiza operações homomórficas de adição e multiplicação em dados criptografados, sem alterar os resultados, e as chaves pública e secreta são geradas aleatoriamente pela biblioteca SEAL, permitindo o processamento seguro e confidencial dos dados durante as simulações.

Além disso, o esquema escolhido para trabalhar com esta biblioteca foi o CKKS, por oferecer melhor precisão de ponto flutuante. O CKKS suporta aritmética aproximada sobre números reais e complexos. Esse esquema utiliza RLWE internamente, onde Z_q é um quociente polinomial de um anel por $(X^N + 1)$, em potências de 2. Além disso, o CKKS realiza uma etapa adicional para utilizar a Criptografia Totalmente Homomórfica. Essa etapa consiste em transformar um vetor z, que representa seus dados, em um polinômio criptografando o texto simples (o polinômio anterior) e depois descriptografando-o. Ou seja, ele opera do campo dos polinômios não criptografados para o campo dos polinômios criptografados, garantindo a integridade e o sigilo dos dados. Uma vez que a mensagem é criptografada (um par de polinômios), o CKKS fornece diversas operações que podem ser realizadas, como adição, multiplicação e rotação.

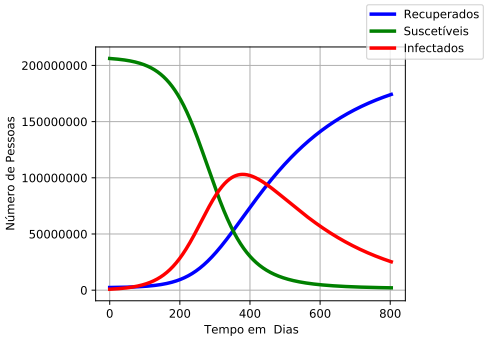
Após uma análise dos artigos [8], [5], [1] e [7], simulamos os modelos SIR e SIRV utilizando um conjunto de EDOs, implementados pelo método de Diferenças Finitas Regressivas. Consideramos 200 iterações (subintervalos) e um período de 800 dias para os dados do Brasil, que descrevem a movimentação de indivíduos entre as populações. A análise foi focada na COVID-19, pois tínhamos acesso aos dados reais e às condições iniciais para o Brasil, conforme apresentados na Tabela 1, são

$$S(0) = 205721209,\ I(0) = 36953492,\ \text{and}\ R(0) = 36083958.$$

Para utilização da Criptografia Totalmente Homomórfica, optou-se por criptografar o cálculo das Equações (3) e (4), já discretizados, utilizando os dados iniciais apresentados na Tabela 1 $S(0) = 206173836,\ I(0) = 834279,\ R(0) = 2384302, V(0) = 38000000)$. O resultado numérico do sistema SIR obtido é apresentado na Figura 1.



(a) Modelo SIR COVID-19 sem Criptografia Totalmente Homomórfica.

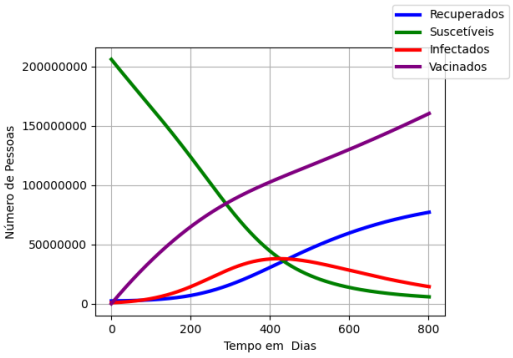


(b) Modelo SIR COVID-19 com Criptografia Totalmente Homomórfica.

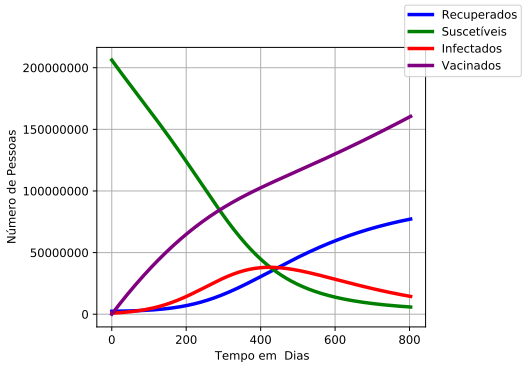
Figura 1: Modelo SIR COVID-19. Fonte:Autores.

Na Figura 2, temos o resultado numérico do sistema SIRV.

[6] `https://github.com/Microsoft/SEAL`



(a) Modelo SIRV COVID-19 sem Criptografia Totalmente Homomórfica.



(b) Modelo SIRV COVID-19 com Criptografia Totalmente Homomórfica.

Figura 2: Modelo SIRV COVID-19. Fonte:Autores.

As operações realizadas em dados criptografados e em texto simples incluem adições, subtrações e multiplicações. Nosso código também rastreia o número de operações de criptografia, descriptografia e operações simples realizadas, com o objetivo de realizar *benchmarking*. Para a operação de divisão, devemos tomar alguns cuidados, pois, embora o esquema CKKS seja capaz de realizar cálculos com números em ponto flutuante e números complexos, a biblioteca TenSEAL não oferece suporte direto a operações de divisão. Para contornar essa limitação, aplicamos a aritmética modular para transformar os parâmetros do Brasil em valores inteiros. Além disso, incorporamos a multiplicação dos parâmetros por 10^{10}, a fim de evitar a perda de informações relacionadas ao cenário pandêmico. Dessa forma, à medida que somamos os resultados de cada expressão nos cálculos de cada classe, é necessário garantir que todos os termos estejam na mesma ordem.

Observamos que, ao criptografar o cálculo de todos os sistemas das Equações (3) e (4), obtemos os mesmos resultados ao decifrá-los, sem perda de informação ou aumento do erro, utilizando uma aproximação de 10^{-14}. O tempo de execução do programa que expressa os sistemas das Equações (3) e (4) de forma homomórfica, com o esquema CKKS, foi de aproximadamente 26 minutos para realizar 100 iterações do modelo, considerando todos os parâmetros e condições iniciais tratados de forma confidencial para os dados do Brasil. Assim, quanto maior o número de iterações, maior será o tempo de execução. No entanto, o tempo gasto é razoável, especialmente considerando o armazenamento e processamento seguro dos dados. Não abordamos a complexidade temporal, pois, para calcular os parâmetros com Δt, o tempo necessário já excede o tempo necessário para realizar 8 iterações do modelo com o esquema CKKS, incluindo as operações de adição e multiplicação homomórficas.

5 Considerações Finais

Neste estudo, investigamos as soluções numéricas para modelos de propagação de doenças, como os modelos SIR e SIRV, utilizando abordagens com e sem Criptografia Totalmente Homomórfica, permitindo uma análise comparativa detalhada. Nosso principal objetivo foi comparar os dados das classes compartimentais após a execução dos programas, destacando a equivalência obtida através do esquema CKKS na biblioteca SEAL. Optamos por criptografar tanto o cálculo do sistema de EDOs quanto a discretização por Diferenças Finitas Regressivas, com o intuito de minimizar perdas de informação e erros. Esta abordagem analítica demonstra a eficácia da criptografia na

preservação da integridade dos dados, apesar das limitações da operação de divisão homomórfica. Os resultados numéricos sem Criptografia Totalmente Homomórfica foram utilizados como base de referência, não devendo ser aplicados em dados confidenciais. No contexto da vacinação, nosso foco foi avaliar a eficácia global da vacinação, sem distinção entre os laboratórios de vacinas, levando em conta apenas o número de doses administradas e sua eficácia.

Agradecimentos

Agradeço a agência de fomento CNPq e a FAPEMIG pelo auxílio financeiro.

Referências

[1] A. Algarni, A. B. Hamed, M. Hamdi, H. Elmannai e S. Meshoul. "Mathematical COVID-19 model with vaccination: a case study in Saudi Arabia". Em: **Peer J Comput Sci** 8.1 (2022), pp. 2–20. DOI: 10.7717/peerj-cs.959.

[2] I. Cooper, A. Mondal e C. Antonopoulos. "A SIR model assumption for the spread of COVID-19 in different communities". Em: **Chaos Solitons Fractals** 139:110057 (out. de 2020). DOI: 10.1016/j.chaos.2020.110057.

[3] S. Fulchand. "Covid-19 and cardiovascular disease". Em: **BMJ** 369 (2020). DOI: 10.1136/bmj.m1997. eprint: https://www.bmj.com/content/369/bmj.m1997.full.pdf. URL: https://www.bmj.com/content/369/bmj.m1997.

[4] M. Girard e et al. "The 2009 A (H1N1) influenza virus pandemic: A review". Em: **Vaccine** 28.31 (2010), pp. 4895–4902. ISSN: 0264-410X. DOI: 10.1016/j.vaccine.2010.05.031. URL: https://www.sciencedirect.com/science/article/pii/S0264410X1000719X.

[5] J. Jonnalagadda. "Epidemic Analysis and Mathematical Modelling of H1N1 (A) with Vaccination". Em: **Nonautonomous Dynamical Systems** 9.1 (2022), pp. 1–10. DOI: 10.1515/msds-2020-0143. URL: https://doi.org/10.1515/msds-2020-0143.

[6] Y. Kim, A. Barber e S. Lee. "Modeling influenza transmission dynamics with media coverage data of the 2009 H1N1 outbreak in Korea". Em: **PLOS ONE** 15.6 (jun. de 2020), pp. 1–21. DOI: 10.1371/journal.pone.0232580. URL: https://doi.org/10.1371/journal.pone.0232580.

[7] I. Moneim. "Modeling and simulation of the spread of H1N1 flu with periodic vaccination". Em: **International Journal of Biomathematics** 09 (2016), p. 1650003. DOI: 10.1142/S1793524516500030. URL: https://api.semanticscholar.org/CorpusID:123722768.

[8] A. Mutalik. "Models to predict H1N1 outbreaks: a literature review". Em: **International Journal Of Community Medicine And Public Health** 4.9 (ago. de 2017), pp. 3068–3075. DOI: 10.18203/2394-6040.ijcmph20173814. URL: https://www.ijcmph.com/index.php/ijcmph/article/view/1751.

[9] G. Neumann, T. Noda e Y. Kawaoka. "Emergence and pandemic potential of swine-origin H1N1 influenza virus". Em: **Nature** 459 (2019), pp. 931–939. DOI: 10.1038/nature08157.

[10] R. Verity e et al. "Estimates of the severity of coronavirus disease 2019: a model-based analysis". Em: **The Lancet Infectious Diseases** 20.6 (2020), pp. 669–677. ISSN: 1473-3099. DOI: 10.1016/S1473-3099(20)30243-7. URL: https://www.sciencedirect.com/science/article/pii/S1473309920302437.

[11] X. Zhou e Z. Guo. "Analysis of an influenza A (H1N1) epidemic model with vaccination". Em: **Arab. J. Math. 1** (2012), pp. 267–282. DOI: 10.1007/s40065-012-0013-6.

