

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Fernando de Moraes Faquim

**Algoritmo Genético para o Ajuste dos
Parâmetros de um Modelo Epidemiológico de
COVID-19 Baseado em Autômatos Celulares**

Uberlândia, Brasil

2026

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Fernando de Moraes Faquim

**Algoritmo Genético para o Ajuste dos Parâmetros de um
Modelo Epidemiológico de COVID-19 Baseado em
Autômatos Celulares**

Trabalho de conclusão de curso apresentado
à Faculdade de Computação da Universidade
Federal de Uberlândia, como parte dos requi-
sitos exigidos para a obtenção título de Ba-
charel em Ciência da Computação.

Orientador: Luiz Gustavo Almeida Martins

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

2026

Fernando de Moraes Faquim

Algoritmo Genético para o Ajuste dos Parâmetros de um Modelo Epidemiológico de COVID-19 Baseado em Autômatos Celulares

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

Trabalho aprovado. Uberlândia, Brasil, 01 de novembro de 2016:

Luiz Gustavo Almeida Martins
Orientador

Professor

Professor

Uberlândia, Brasil
2026

Resumo

A modelagem epidemiológica tem sido amplamente utilizada para compreender a dinâmica de propagação de doenças infecciosas e auxiliar na tomada de decisões de saúde pública. Dentre as diferentes abordagens empregadas na construção dos modelos, os autômatos celulares destacam-se por sua capacidade de representar comportamentos complexos por meio de interações locais entre indivíduos de forma simples e intuitiva. Entretanto, modelos epidemiológicos baseados em autômatos celulares dependem de diversos parâmetros que influenciam seu comportamento. O ajuste manual desses parâmetros pode ser difícil, demorado e exige conhecimento especializado sobre o fenômeno modelado. Nesse contexto, métodos de ajuste automático tornam-se importantes para calibrar o modelo de forma sistemática. Este trabalho emprega um algoritmo genético para o ajuste automático do conjunto de parâmetros utilizado em um modelo epidemiológico SEIR (Suscetível, Exposto, Infectado, Recuperado), baseado em autômato celular, que simula a propagação da COVID-19. O algoritmo genético permite explorar eficientemente o espaço de soluções a fim de encontrar a calibração do modelo capaz de reproduzir o comportamento observado em dados históricos de propagação da doença. Essa abordagem tem como objetivo melhorar a exatidão e a representatividade das simulações epidemiológicas, auxiliando na compreensão dos processos de contágio e oferecendo suporte para a tomada de decisões em questões de saúde pública. Resultados experimentais mostram que o algoritmo genético é capaz de reproduzir satisfatoriamente a dinâmica epidemiológica presente nos dados artificiais usados como referência, embora diferentes combinações de parâmetros possam produzir comportamentos semelhantes. Experimentos adicionais com dados reais de COVID-19 no Brasil, referentes ao ano de 2022, mostram que a abordagem desenvolvida apresenta maior dificuldade em reproduzir integralmente a dinâmica observada, evidenciando limitações estruturais do modelo diante de séries temporais irregulares. De modo geral, os resultados demonstram que o algoritmo genético proposto constitui uma ferramenta promissora para a calibração automática de modelos epidemiológicos baseados em autômatos celulares.

Palavras-chave: Algoritmos Genéticos, Autômatos Celulares, Modelo SEIR, Simulação da Propagação da COVID-19.

Lista de ilustrações

Figura 1 – Principais dimensionalidades de um autômato celular.	16
Figura 2 – Formatos de célula possíveis em um AC.	17
Figura 3 – Vizinhanças mais comuns em um AC bidimensional	18
Figura 4 – Exemplos de tratamento de bordas	19
Figura 5 – Evolução da regra 110 ao longo de 6 iterações	20
Figura 6 – Evolução de um glider ao longo de 5 iterações	21
Figura 7 – Funcionamento típico de um algoritmo genético	22
Figura 8 – Representações comuns de indivíduos em algoritmos genéticos (Fonte: adaptado de (MITCHELL, 1996b))	23
Figura 9 – Crossover de um ponto	25
Figura 10 – Crossover de dois pontos	25
Figura 11 – Crossover uniforme	25
Figura 12 – <i>Bit flip</i>	26
Figura 13 – Perturbação gaussiana	26
Figura 14 – Troca de posição	26
Figura 15 – Exemplo de vizinhança com raio igual à metade do reticulado	31
Figura 16 – Possíveis valores para x em vermelho	35
Figura 17 – Curva de estados da configuração alvo	40
Figura 18 – Evolução temporal dos estados do modelo epidemiológico a partir da configuração de referência.	42
Figura 19 – Evolução da aptidão média do melhor indivíduo ao longo das execuções do AG.	43
Figura 20 – Comparação entre a evolução da doença de acordo com a configuração adotada pelo modelo de simulação	47
Figura 21 – Comparação entre a evolução da doença de acordo com a configuração adotada pelo modelo de simulação para o cenário 1	49
Figura 22 – Comparação entre a evolução da doença de acordo com a configuração adotada pelo modelo de simulação para o cenário 1	50
Figura 23 – Série temporal do número de infectados em 2022, mostrando os dados brutos e a versão suavizada utilizando o filtro de Savitzky-Golay.	52
Figura 24 – Primeira tentativa de ajuste dos dados reais de infectados utilizando o algoritmo genético.	53
Figura 25 – Resultado do ajuste após modificar a condição inicial do modelo, utili- zando um número inicial de células infectadas semelhante ao observado nos dados reais.	54

Figura 26 – Resultado do ajuste considerando apenas a dinâmica dos dados a partir do dia 100.	55
--	----

Lista de tabelas

Tabela 1 – Valores mínimos e máximos para atributos do indivíduo	34
Tabela 2 – Configuração inicial do algoritmo genético	41
Tabela 3 – Melhores parâmetros da configuração inicial	41
Tabela 4 – Resultados dos experimentos com tamanho de população	43
Tabela 5 – Resultados dos experimentos com mutação	44
Tabela 6 – taxa=0,15, $\sigma_0=0,1$	45
Tabela 7 – Resultados dos experimentos com <i>crossover</i>	45
Tabela 8 – Configuração final do algoritmo genético	46
Tabela 9 – Valores encontrados pelo algoritmo para a configuração de parâmetros alternativa 1	48
Tabela 10 – Valores encontrados pelo algoritmo para a configuração de parâmetros alternativa 2	50
Tabela 11 – Comparação dos resultados obtidos nos diferentes cenários. σ Curva representa a média do desvio padrão das curvas de estados por dia e σ RMSE representa o desvio padrão do erro RMSE ao longo das execuções.	51

Lista de abreviaturas e siglas

AG	Algoritmo genético
AC	Autômato celular
EDO	Equação diferencial ordinária
SIR	<i>Susceptible, Infectious, Recovered</i>
SEIR	<i>Susceptible, Exposed, Infectious, Recovered</i>
SEIRD	<i>Susceptible, Exposed, Infectious, Recovered, Dead</i>

Sumário

1	INTRODUÇÃO	10
1.1	Motivação	11
1.2	Objetivos	11
1.3	Estrutura do Texto	12
2	EMBASAMENTO TEÓRICO	13
2.1	Modelos Epidemiológicos	13
2.1.1	Fundamentos dos Modelos Compartmentais	13
2.1.2	Modelo SIR	13
2.1.3	Modelo SEIR	14
2.1.4	Modelo SEIRD	15
2.1.5	Modelos Compartmentais e Autômatos Celulares	15
2.2	Autômatos Celulares	15
2.2.1	Breve Contexto Histórico	15
2.2.2	Espaço Celular	16
2.2.3	Estado e Vizinhança	17
2.2.4	Regras de Transição	19
2.2.5	Funcionamento Básico de um AC	19
2.2.5.1	Regra 110	20
2.2.5.2	<i>Game of Life</i>	20
2.3	Algoritmo Genético	21
2.3.1	Representação do Indivíduo	22
2.3.2	População	23
2.3.3	Avaliação do Indivíduo	23
2.3.4	Seleção	24
2.3.5	Operadores Genéticos	24
2.3.5.1	Cruzamento	24
2.3.5.2	Mutação	25
2.3.6	Reinserção	26
2.3.7	Critério de Parada	27
2.4	Trabalhos Relacionados	27
2.4.1	Autômatos Celulares Epidemiológicos	27
2.4.2	Algoritmos Genéticos no ajuste de parâmetros	28
3	DESENVOLVIMENTO	30
3.1	Reprodução do Modelo de Referência	30

3.1.1	Espaço Celular e Estado das Células	30
3.1.2	Vizinhança	31
3.1.3	Regras de Transição	31
3.2	Algoritmo Genético	32
3.2.1	Representação do Indivíduo	33
3.2.2	Inicialização da População	33
3.2.3	Função de Avaliação	34
3.2.4	Cruzamento	35
3.2.5	Mutação	36
3.2.6	Reinserção	36
3.2.7	Critério de Parada	37
4	EXPERIMENTOS E RESULTADOS	38
4.1	Configuração Experimental	38
4.1.1	Geração dos Dados de Referência Artificiais	38
4.1.2	Avaliação do Algoritmo	40
4.2	Ajuste dos Hiperparâmetros	41
4.2.1	Configuração Inicial do AG	41
4.2.2	Critério de Parada	42
4.2.3	Tamanho de População	43
4.2.4	Mutação	44
4.2.5	Elitismo	44
4.2.6	Cruzamento	45
4.2.7	Função de Aptidão	46
4.3	Análise de Desempenho	46
5	CONCLUSÃO	56
	REFERÊNCIAS	58

1 Introdução

Desde o início do século, observa-se o aumento do uso de tecnologias da informação no controle e previsão de surtos epidêmicos (WALDMAN; SATO, 2016). Durante a pandemia de COVID-19, por exemplo, modelos computacionais tiveram papel crucial na simulação de cenários e avaliação de medidas de distanciamento social (YARSKY, 2021). O desenvolvimento de ferramentas computacionais para modelagem de epidemias pode contribuir significativamente para a saúde pública, fornecendo estimativas sobre a evolução de surtos, identificação de regiões críticas e avaliação do impacto de políticas sanitárias, servindo para nortear as tomadas de decisões e criações de políticas públicas.

A modelagem da propagação de doenças infecciosas tem se mostrado uma ferramenta essencial na compreensão de epidemias e no auxílio à formulação de políticas públicas de controle (WALDMAN; SATO, 2016). Entre as diferentes abordagens propostas na literatura, destacam-se os modelos baseados em autômatos celulares, devido à sua capacidade de representar sistemas complexos por meio de interações locais (WOLFRAM, 1984; SIRAKOULIS; KARAFYLLIDIS; THANAILAKIS, 2000), e as técnicas de otimização inspiradas na evolução biológica, como os algoritmos genéticos, que têm sido empregadas para o ajuste automático de parâmetros em modelos computacionais (YARSKY, 2021; MURILO, 2025).

Diversos trabalhos na literatura exploram a integração entre autômatos celulares e algoritmos evolutivos para simulação e ajuste de modelos complexos. A dissertação de Murilo (2025) propõe o uso dessa técnica para modelar a propagação de incêndios florestais, demonstrando o potencial da técnica em ambientes dinâmicos. Na área da saúde, Fraga (2021) aplica métodos semelhantes para simular a reprodução e dispersão do vetor barbeiro, transmissor da Doença de Chagas. No contexto da COVID-19, Chowdhury, Roychowdhury e Chaudhuri (2022) propõe um modelo baseado em autômatos celulares com vizinhança probabilística para simular a propagação da doença e o estudo de Yarsky (2021) utiliza algoritmos genéticos para ajustar modelos, levando em consideração dados reais de mobilidade e distanciamento social nos Estados Unidos. Esses exemplos evidenciam a viabilidade e aplicabilidade dos métodos bio-inspirados (ACs e AGs) na modelagem da propagação de vetores e doenças.

Nesse sentido, esse trabalho visa desenvolver uma estratégia baseada em algoritmos genéticos para ajustar automaticamente os parâmetros de um modelo de propagação de COVID-19 baseado em autômatos celulares. O modelo será baseado em propostas anteriores de simulação da propagação da COVID-19 (SILVA, 2024). Experimentos foram realizados a fim de avaliar a eficácia do método na adaptação do modelo a dados sintéticos

(artificiais) e reais, buscando refinar sua precisão e utilidade como ferramenta preditiva em contextos epidemiológicos.

1.1 Motivação

A modelagem da propagação de doenças infecciosas é uma importante ferramenta para a compreensão de epidemias e para o apoio à formulação de políticas públicas (WALDMAN; SATO, 2016). Por meio desses modelos, é possível estimar a evolução de surtos, avaliar medidas de controle e identificar cenários de risco, fornecendo informações que auxiliam gestores e profissionais da saúde na tomada de decisões.

Entre as diferentes abordagens de modelagem, os autômatos celulares destacam-se por sua capacidade de representar sistemas complexos a partir de regras locais simples. Nesse tipo de modelo, cada célula interage apenas com sua vizinhança, mas a combinação dessas interações pode gerar comportamentos coletivos complexos. Essas características tornam os autômatos celulares uma alternativa adequada para representar a propagação espacial de doenças infecciosas.

Entretanto, a qualidade das simulações depende diretamente da escolha dos parâmetros que compõem as regras de transição do modelo. A definição desses valores normalmente exige conhecimento especializado e pode demandar um processo extenso de calibração. Em situações envolvendo novas doenças, esse conhecimento pode não estar disponível. Nesse contexto, algoritmos genéticos podem ser empregados para ajustar automaticamente os parâmetros do modelo, buscando reproduzir de forma mais fiel a dinâmica observada em dados epidemiológicos.

1.2 Objetivos

O objetivo principal deste trabalho é desenvolver uma estratégia baseada em algoritmos genéticos para ajustar automaticamente os parâmetros de um modelo epidemiológico baseado em autômatos celulares, de modo que ele seja capaz de reproduzir adequadamente a dinâmica de propagação da COVID-19 observada em dados de referência.

Nesse sentido, ao longo da pesquisa buscou-se atingir os seguintes objetivos específicos:

- Estudar e implementar um modelo epidemiológico baseado em autômatos celulares para simular a propagação da COVID-19.
- Implementar um algoritmo genético para o ajuste automático dos parâmetros do modelo.

- Avaliar a eficácia do algoritmo genético na adaptação do modelo a dados de referência gerados artificialmente.
- Analisar a capacidade do método em reproduzir comportamentos gerados por diferentes configurações de referência.
- Avaliar o desempenho da abordagem utilizando dados reais da propagação da COVID-19.

1.3 Estrutura do Texto

Este trabalho está organizado da seguinte forma: o Capítulo 2 apresenta os fundamentos teóricos relacionados à modelagem epidemiológica, autômatos celulares e algoritmos genéticos. O Capítulo 3 descreve o modelo epidemiológico adotado e a metodologia de experimentos proposta. O Capítulo 4 apresenta os experimentos realizados e a análise dos resultados obtidos. Por fim, o Capítulo 5 apresenta as conclusões e perspectivas para trabalhos futuros.

2 Embasamento Teórico

2.1 Modelos Epidemiológicos

Modelos epidemiológicos são ferramentas matemáticas usadas para prever e analisar a dinâmica de disseminação de doenças em populações ([KERMACK; MCKENDRICK, 1927](#)).

Os modelos compartimentais assumem que a população pode ser dividida em grupos (compartimentos) que representam diferentes estágios da doença. Esta sessão apresenta alguns modelos clássicos aplicados em epidemiologia.

2.1.1 Fundamentos dos Modelos Compartimentais

A modelagem compartimental parte da divisão da população em N grupos, representados por variáveis que mudam em função do tempo. As taxas de transição entre estados são representadas por equações diferenciais, considerando parâmetros como:

- β : taxa de transmissão.
- σ : taxa de progressão de estágio latente para infeccioso.
- γ : taxa de recuperação.
- μ : taxa de mortalidade.

Por motivos de simplicidade, estes modelos assumem uma população homogênea, com todos os indivíduos tendo a mesma probabilidade de contato entre si. Além disso, são ignoradas heterogeneidades espaciais, sociais e demográficas. Dessa forma, torna-se possível descrever essa dinâmica com equações diferenciais ordinárias (EDOs). Extensões destes modelos relaxam essas hipóteses, com abordagens espaciais e estocásticas, entre elas, autômatos celulares epidemiológicos ([SIRAKOULIS; KARAFYLLIDIS; THANAILAKIS, 2000](#)).

2.1.2 Modelo SIR

O modelo SIR, introduzido por [Kermack e McKendrick \(1927\)](#), divide a população em três estados:

- S (Suscetível): indivíduos não infectados e aptos a contrair a doença.

- I (Infectado): indivíduos que se encontram infectados pela doença em estudo.
- R (Recuperado): indivíduos que se recuperaram da doença e, nesse tipo de modelo, se tornam imunes permanentemente.

As equações diferenciais que regem o modelo são:

$$\frac{dS}{dt} = -\beta SI \quad (2.1)$$

$$\frac{dI}{dt} = \beta SI - \gamma I \quad (2.2)$$

$$\frac{dR}{dt} = \gamma I \quad (2.3)$$

Nesse tipo de modelo, a dinâmica é guiada pela competição entre os processos de transmissão (βSI) e recuperação (γI). Portanto, a razão entre a taxa de transmissão e a taxa de recuperação (β/γ) determina se a doença cresce ($\beta/\gamma > 1$) ou se extingue ($\beta/\gamma < 1$).

2.1.3 Modelo SEIR

O modelo SEIR adiciona o estado E (Exposto) ao modelo SIR para contemplar o período latente da doença, ou seja, o intervalo de tempo no qual um indivíduo está infectado, mas não infeccioso (não transmite a doença).

Com isso, o sistema de equações diferenciais é modificado para contemplar o novo estado, como segue:

$$\frac{dS}{dt} = -\beta SI \quad (2.4)$$

$$\frac{dE}{dt} = \beta SI - \sigma E \quad (2.5)$$

$$\frac{dI}{dt} = \sigma E - \gamma I \quad (2.6)$$

$$\frac{dR}{dt} = \gamma I \quad (2.7)$$

O parâmetro σ corresponde ao inverso do período de incubação. A introdução do estado latente provoca um atraso no crescimento do número de infectados ($I(t)$) e maior realismo para doenças com tempo de incubação significativo.

Modelos SEIR foram utilizados durante a pandemia de COVID-19 devido à importância epidemiológica da fase assintomática precoce (KUCHARSKI et al., 2020).

2.1.4 Modelo SEIRD

O modelo SEIRD, baseado no modelo SEIR, busca representar com maior precisão doenças com mortalidade relevante. Essa extensão é feita adicionando o estado D (Morto, do inglês *Dead*), para representar indivíduos mortos após a infecção.

No novo modelo, uma pequena fração de indivíduos infecciosos progridem para o estado de óbito, o que é representado pelas seguintes equações:

$$\frac{dI}{dt} = \sigma E - (\gamma + \mu)I \quad (2.8)$$

$$\frac{dD}{dt} = \mu I \quad (2.9)$$

2.1.5 Modelos Compartmentais e Autômatos Celulares

Esses modelos servem como base conceitual para modelos baseados em autômatos celulares, onde cada célula representa um indivíduo ou região e seu estado corresponde ao compartimento epidemiológico (S, E, I, R, D). As regras de transição são inspiradas nas equações dos modelos compartmentais, podendo ser incorporados parâmetros adicionais para representar heterogeneidade espacial e estocasticidade.

2.2 Autômatos Celulares

Um Autômato Celular (AC) pode ser definido como um sistema dinâmico discreto, estruturado como um reticulado regular de células, cada uma ocupando uma posição fixa no espaço (MITCHELL, 1996a). Cada célula possui um estado, que é atualizado a cada passo temporal, conforme uma função de transição que considera seu próprio estado e os estados de suas células vizinhas. Os ACs destacam-se pela sua capacidade de representar fenômenos complexos a partir de regras locais simples, sendo aplicados em diversas áreas, tais como: física (VICHNIAC, 1984), biologia (ZHAO et al., 1993), ciência da computação (WOLFRAM, 1983), entre outras.

2.2.1 Breve Contexto Histórico

Os autômatos celulares foram introduzidos na década de 1940 por John von Neumann e Stanislaw Ulam no contexto de estudos sobre sistemas auto-replicantes (NEUMANN; BURKS, 1966). Von Neumann procurava um modelo teórico capaz de representar

uma máquina autorreprodutiva universal — um conceito relacionado à sua máquina universal. Ulam sugeriu a utilização de uma grade discreta para modelar essa máquina, onde cada célula poderia assumir um número finito de estados e evoluir de acordo com regras locais.

Décadas mais tarde, o conceito ganhou mais destaque com o desenvolvimento do *Game of Life* por John Horton Conway, em 1970 ([GARDNER, 1970](#)). Esse autômato bidimensional simples, apesar de possuir regras muito básicas, demonstrou padrões extremamente ricos e imprevisíveis, tornando-se um exemplo clássico de emergência e complexidade computacional.

Nas décadas seguintes, Stephen Wolfram popularizou e formalizou a teoria dos autômatos celulares, especialmente a partir de 1983, ao estudar os chamados autômatos celulares elementares — sistemas unidimensionais com apenas dois estados e regras baseadas na vizinhança imediata ([WOLFRAM, 1983](#)). Wolfram propôs uma classificação dos comportamentos de ACs em quatro classes distintas (estáveis, periódicos, caóticos e complexos), e demonstrou que regras simples, como a Regra 30 e a Regra 110, podem produzir padrões extremamente complexos.

2.2.2 Espaço Celular

Um autômato celular é definido a partir de um reticulado, que corresponde a uma estrutura formada por diversas células. Cada célula ocupa uma posição específica nesse reticulado e assume um estado, que pode variar ao longo da evolução do sistema conforme regras estabelecidas.

As células podem ser organizadas em diferentes tipos de reticulados, caracterizados pelo número de dimensões. Assim, é possível construir modelos em uma, duas ou mais dimensões, sendo que os mais usuais são ilustrados na Figura 1.

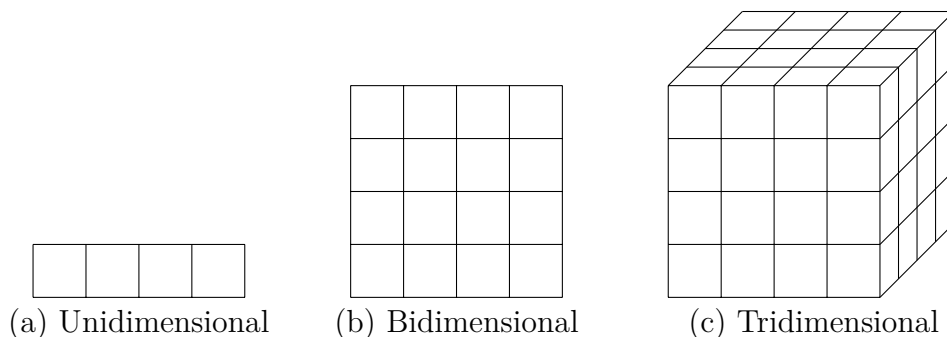


Figura 1 – Principais dimensionalidades de um autômato celular.

As células de um reticulado também podem apresentar diferentes formas geométricas, cuja escolha depende da iteratividade local presente no fenômeno a ser modelado.

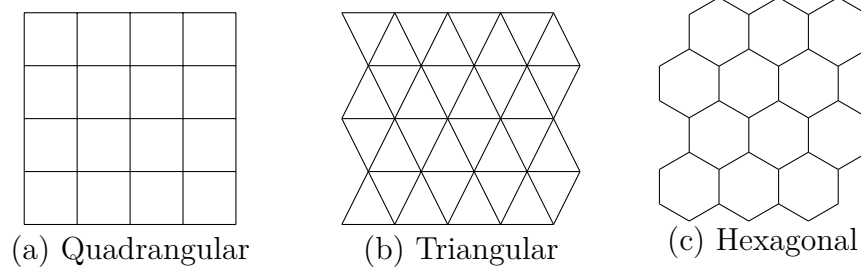


Figura 2 – Formatos de célula possíveis em um AC.

Independentemente da forma adotada, é uma característica fundamental que todas as células de um mesmo reticulado possuam a mesma geometria, garantindo a homogeneidade da estrutura. A Figura 2 ilustra os principais formatos de células adotados em reticulados.

2.2.3 Estado e Vizinhaça

Em um AC típico, o estado de cada célula é escolhido de um conjunto finito, podendo assumir valores binários (0 ou 1), inteiros ou até mesmo reais, dependendo da aplicação. A evolução temporal do sistema ocorre de maneira síncrona, com todas as células atualizando seus estados simultaneamente, com base em uma função de transição. Essa função considera o estado da célula atual e de suas células vizinhas.

A vizinhaça de uma célula é definida pelo conjunto de células próximas segundo uma topologia específica e seu raio de alcance. O raio de uma vizinhaça define até onde, em número de células, a vizinhaça se estende. As vizinhaças mais comuns em um AC bidimensional são:

- **Vizinhaça de von Neumann (raio = 1):** Inclui a própria célula e as quatro vizinhas ortogonais (acima, abaixo, esquerda e direita).
- **Vizinhaça de Moore (raio = 1):** Inclui todas as oito vizinhas ao redor da célula (as quatro ortogonais mais as quatro diagonais).

A Figura 3 ilustra esses dois tipos de vizinhaça para os raios 1 e 2.

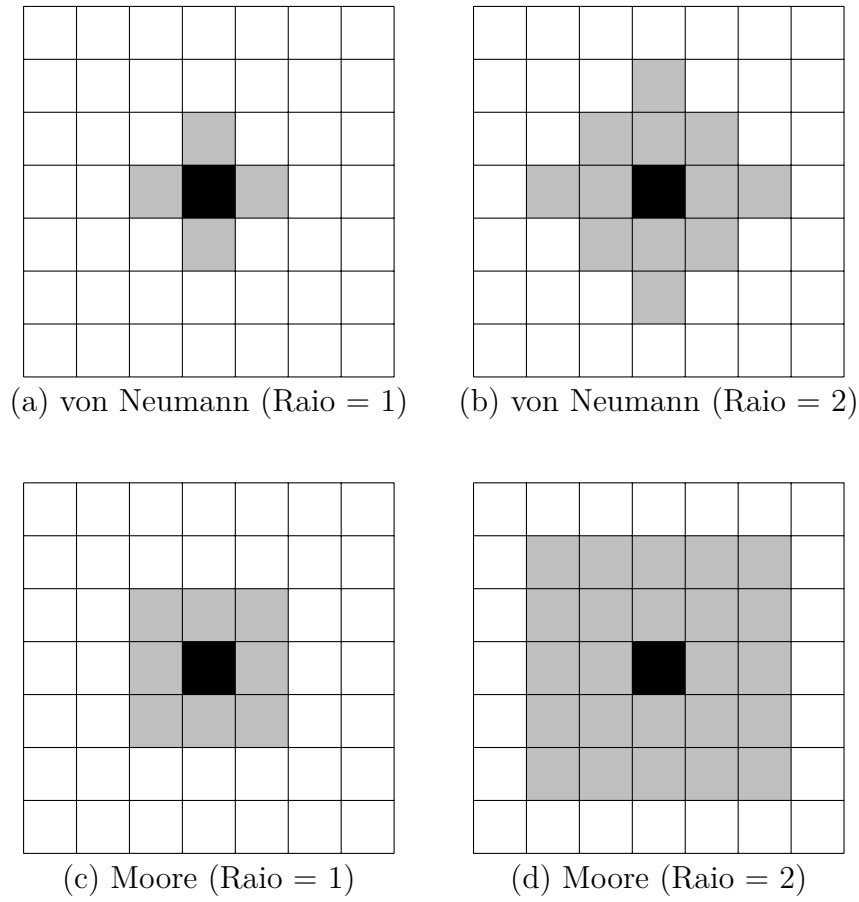


Figura 3 – Vizinhanças mais comuns em um AC bidimensional

Como os ACs geralmente operam em um espaço finito, é necessário adotar estratégias para tratar a vizinhança perto das bordas. As abordagens mais comuns incluem:

- **Periódica:** As extremidades da malha são conectadas umas às outras, criando um efeito de toro.
- **Reflexiva:** As células das bordas refletem os valores de suas vizinhas mais próximas, como se houvesse um espelho.
- **Fixa:** As bordas possuem um valor constante e imutável durante toda a simulação.

A Figura 4 apresenta estas três estratégias. Na abordagem de borda periódica (Figura 4a), assume-se que o reticulado é cíclico, de modo que os valores além do limite são obtidos pela repetição dos elementos da extremidade oposta. Já na borda reflexiva (Figura 4b), os valores externos são definidos pela repetição dos valores das respectivas bordas. Por fim, na borda fixa (Figura 4c), o estado das células vizinhas fora do reticulado assume um valor fixo previamente definido, representado na figura por x, caracterizando uma condição de contorno constante.

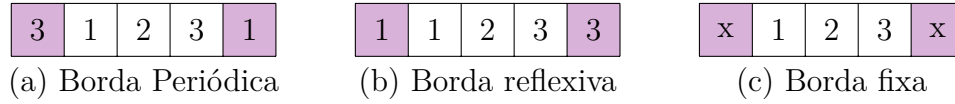


Figura 4 – Exemplos de tratamento de bordas

2.2.4 Regras de Transição

As regras de transição determinam como o estado de uma célula será atualizado com base em sua vizinhança. Essas regras podem ser definidas de forma determinística, sempre resultando no mesmo novo estado para uma dada configuração; ou probabilística, envolvendo características estocásticas que proporcionam uma certa aleatoriedade ao fenômeno modelado.

A formalização matemática de uma regra de transição é uma função:

$$f : S_t^n \rightarrow S_{t+1} \quad (2.10)$$

onde S é o conjunto de estados possíveis e n é o número de células na vizinhança (incluindo a própria célula). O valor de saída é o novo estado da célula.

As regras podem ser definidas por tabelas, expressões lógicas, equações diferenciais discretizadas ou até mesmo outras técnicas inteligentes, como redes neurais, dependendo da complexidade desejada.

Em ACs probabilísticos, uma célula pode assumir diferentes estados mediante uma distribuição de probabilidade para um mesmo estado de vizinhança. Essa característica introduz incerteza na evolução do autômato celular, o que pode resultar em comportamentos complexos e não lineares (GRINSTEIN; JAYAPRAKASH; HE, 1985). Esse tipo de AC têm sido utilizados na modelagem de sistemas estocásticos em áreas, como biologia, física e engenharia (LOZANO, 2017).

2.2.5 Funcionamento Básico de um AC

O funcionamento de um AC pode ser compreendido de forma mais clara por meio de exemplos. Nesta seção, são apresentados dois casos clássicos que ilustram como esses sistemas evoluem ao longo do tempo.

Inicialmente, descreve-se a evolução de um autômato celular elementar, por se tratar da forma mais simples de AC: unidimensional, determinístico e com estados binários. Em seguida, é apresentado o funcionamento do *Jogo da Vida*, um autômato celular bidimensional amplamente conhecido.

2.2.5.1 Regra 110

A Regra 110 é um autômato celular unidimensional, onde cada célula tem dois vizinhos (esquerda e direita), e o estado de cada célula é binário (0 ou 1). Essa regra foi demonstrada como sendo Turing completa por Matthew Cook ([COOK, 2004](#)), o que significa que ela pode simular qualquer máquina de Turing, apesar de sua simplicidade.

A regra pode ser representada como a seguinte tabela:

Esquerda	Centro	Direita	Novo estado
1	1	1	0
1	1	0	1
1	0	1	1
1	0	0	0
0	1	1	1
0	1	0	1
0	0	1	1
0	0	0	0

Apesar de parecer simples, a Regra 110 gera padrões altamente complexos e é amplamente estudada em teoria da computação. A Figura 5 ilustra a evolução dessa regra por 6 iterações.

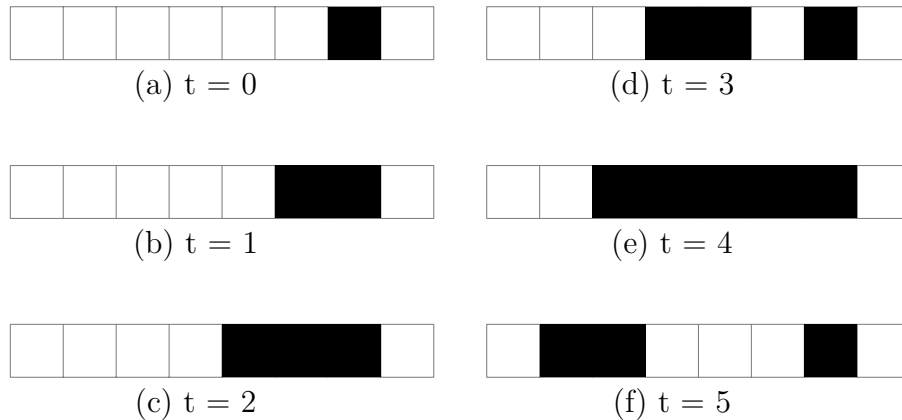


Figura 5 – Evolução da regra 110 ao longo de 6 iterações

2.2.5.2 Game of Life

O Jogo da Vida, criado por John Conway em 1970, é um autômato celular bidimensional onde cada célula pode estar em dois estados: viva (1) ou morta (0). A vizinhança usada é a de Moore (8 vizinhos).

As regras de transição são:

- Uma célula viva com 2 ou 3 vizinhos vivos permanece viva.

- Uma célula viva com menos de 2 ou mais de 3 vizinhos vivos morre.
- Uma célula morta com exatamente 3 vizinhos vivos se torna viva.

Esse sistema é notável por produzir padrões dinâmicos, como *gliders*, que se deslocam na grade de células dentro de ciclos fechados. A Figura 6 ilustra a evolução dessa estrutura ao longo de 5 interações.

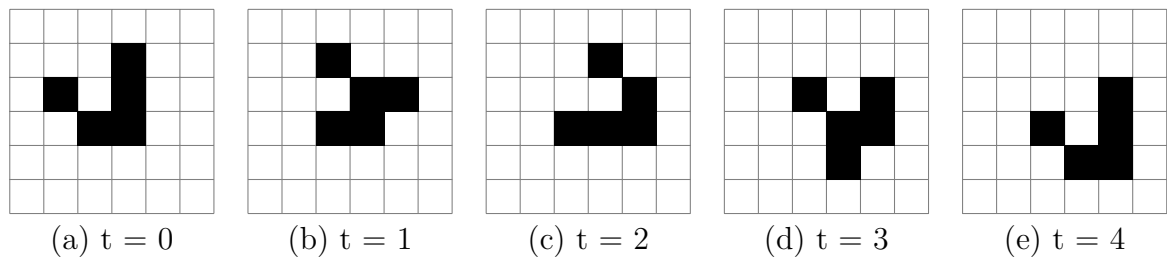


Figura 6 – Evolução de um glider ao longo de 5 iterações

2.3 Algoritmo Genético

Algoritmos Genéticos (AGs) são métodos de otimização e busca inspirados nos mecanismos da evolução biológica natural. Propostos por John Holland na década de 1970 (HOLLAND, 1992), esses algoritmos utilizam conceitos como seleção natural, recombinação genética e mutação para explorar o espaço de soluções de um problema.

AGs já foram empregados para lidar com vários problemas na literatura, como o problema da mochila (CHU; BEASLEY, 1998), problemas de escalonamento de tarefas (CHENG; GEN; TSUJIMURA, 1996) e o problema do caixeiro viajante (POTVIN, 1996). Esses exemplos evidenciam a flexibilidade e a eficiência dos AGs na resolução de problemas de otimização combinatória e em diversos contextos práticos.

A base teórica dos AGs está na Teoria da Evolução de Charles Darwin (DARWIN, 1859). Segundo essa teoria, os organismos que melhor se adaptam ao ambiente têm maior probabilidade de sobreviver e se reproduzir. Esse processo de seleção natural, aliado à recombinação genética e à mutação, leva ao aprimoramento gradual das características das populações ao longo das gerações.

Em AGs, cada solução candidata é representada como um indivíduo, e a busca por soluções é realizada por meio da evolução de uma população, guiada por uma função objetivo. O processo evolutivo inicia-se com a geração de uma população inicial, cuja aptidão é avaliada para verificar o desempenho de cada indivíduo. Caso o critério de parada seja atingido, a execução do algoritmo é encerrada; caso contrário, indivíduos são selecionados para reprodução (*crossover*), seguidos por mutação e reinserção na população, formando assim uma nova geração. A aptidão é então recalculada e o critério de parada é novamente verificado. Esse ciclo se repete até que o critério de término seja alcançado. Cada

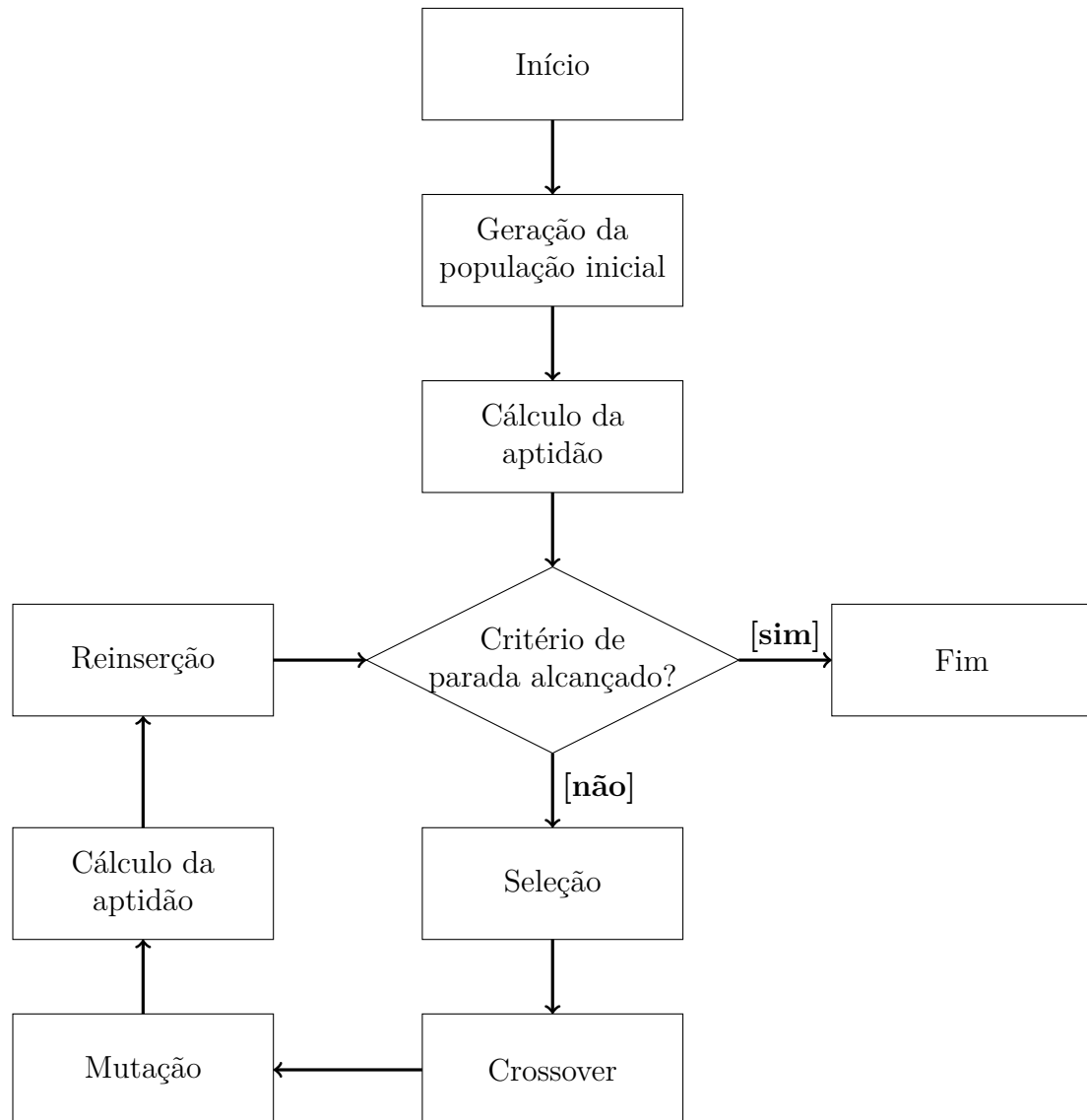


Figura 7 – Funcionamento típico de um algoritmo genético

uma dessas etapas será detalhada posteriormente. O fluxograma da Figura 7 exemplifica o funcionamento básico de um AG.

2.3.1 Representação do Indivíduo

Cada indivíduo em um AG representa uma possível solução para o problema. A forma mais comum de representação é o uso de cadeias binárias (strings de bits), mas outras formas também são utilizadas, como demonstrado na Figura 8.

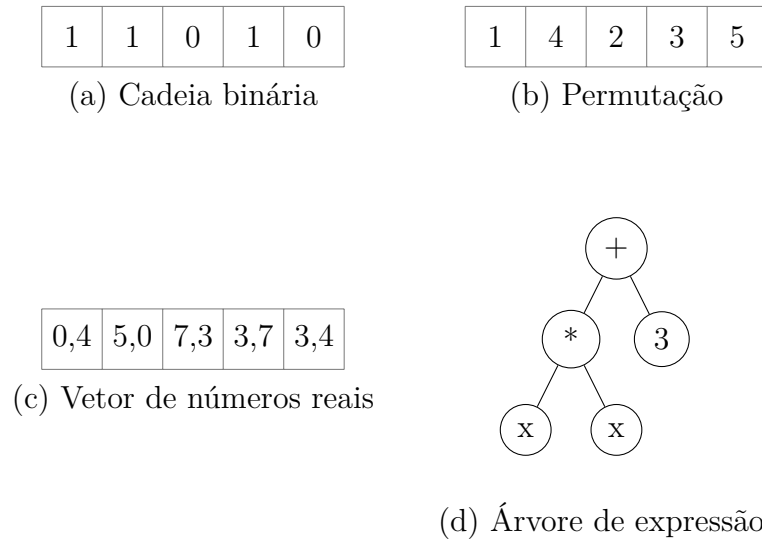


Figura 8 – Representações comuns de indivíduos em algoritmos genéticos (Fonte: adaptado de (MITCHELL, 1996b))

A representação escolhida é determinada pela natureza do problema e influencia diretamente na eficiência e a adequação dos operadores genéticos aplicados.

2.3.2 População

A população é o conjunto de indivíduos (soluções candidatas) que evolui ao longo do tempo. Por meio dela, o processo evolutivo acontece, com indivíduos interagindo, recombina-se e sofrendo mutações. A diversidade da população é um fator essencial para o sucesso do algoritmo, pois evita a convergência prematura para soluções subótimas (MITCHELL, 1996b).

O tamanho da população deve ser escolhido com cuidado. Populações menores tendem a convergir de forma mais rápida, o que pode acelerar a busca, porém reduz a variabilidade genética e eleva o risco de estagnação. Em contrapartida, enquanto populações grandes possibilitam uma ampla exploração do espaço de busca, elas também elevam consideravelmente o custo computacional.

2.3.3 Avaliação do Indivíduo

A função objetivo – também chamada de função de aptidão ou *fitness* – é o mecanismo que avalia a qualidade do indivíduo no contexto do problema abordado (MITCHELL, 1996b). Essa função que guia o processo evolutivo, favorecendo indivíduos com melhor desempenho.

Para problemas de minimização, a função pode ser adaptada para atribuir pontuações maiores a soluções com menor custo. Já em problemas de maximização, a pontuação é diretamente proporcional à qualidade da solução.

A escolha de uma função objetivo adequada é fundamental para o desempenho de algoritmos de otimização, uma vez que funções mal definidas podem direcionar o processo de busca para regiões pouco relevantes do espaço de soluções. Assim, a exploração desse espaço se torna ineficiente, dificultando e retardando a convergência do algoritmo ou, até mesmo, ocasionando sua estagnação em ótimos locais, impedindo a obtenção da solução global desejada.

2.3.4 Seleção

O operador de seleção é responsável por selecionar os indivíduos que participarão do processo de reprodução e ajudarão a formar os indivíduos da próxima geração. Indivíduos mais aptos têm uma chance maior de serem selecionados.

Técnicas comuns de seleção incluem:

- **Roleta (*roulette wheel selection*):** A chance de cada indivíduo ser selecionado é proporcional à sua aptidão relativa em relação à população total.
- **Torneio (*tournament selection*):** Um subconjunto de indivíduos é escolhido aleatoriamente, e o mais apto do grupo é selecionado para reprodução.
- **Ranqueamento (*rank selection*):** Os indivíduos são ordenados por aptidão e a probabilidade de seleção é distribuída com base na posição no ranking, evitando que diferenças extremas de aptidão dominem o processo.

A escolha do método de seleção impacta diretamente a taxa de exploração e a preservação da diversidade populacional.

2.3.5 Operadores Genéticos

Os operadores genéticos são os mecanismos que promovem a evolução da população, criando novos indivíduos a partir de indivíduos existentes. Os dois operadores principais são cruzamento e mutação.

2.3.5.1 Cruzamento

O cruzamento ou *crossover* combina partes de dois indivíduos para gerar novos indivíduos, explorando novas regiões do espaço de busca. Esta operação resulta em dois novos indivíduos, formados a partir da combinação genéticas dos pais previamente selecionados. É o principal mecanismo de exploração do algoritmo, sendo que os tipos mais comuns são:

- **Crossover de um ponto:** Um ponto de corte é escolhido aleatoriamente. O primeiro filho herda os genes anteriores ao ponto de corte do pai 1 e o restante do pai 2, enquanto o segundo filho é formado pela carga genética complementar, conforme ilustrado na Figura 9.

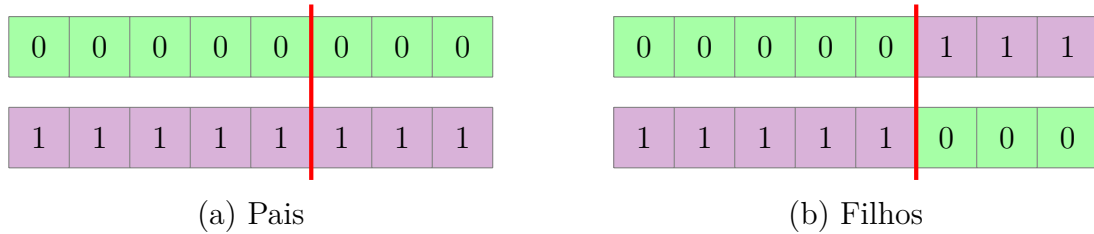


Figura 9 – Crossover de um ponto

- **Crossover de dois pontos:** Dois pontos de corte são escolhidos aleatoriamente. O primeiro filho herda os genes entre os pontos de corte do pai 1 e o restante do pai 2, enquanto o segundo filho é formado pela carga genética complementar, conforme ilustrado na Figura 10.

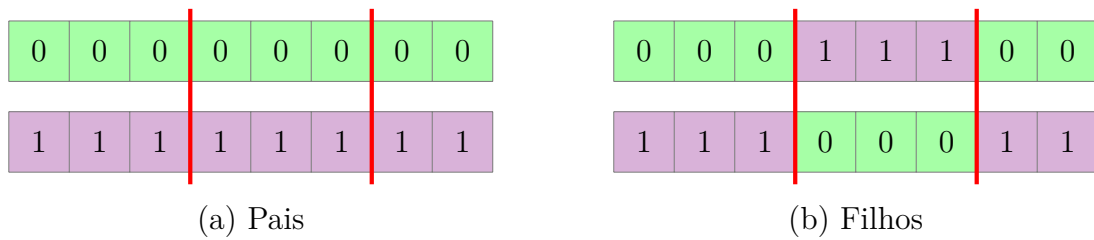


Figura 10 – Crossover de dois pontos

- **Crossover uniforme:** Cada gene do primeiro filho é escolhido aleatoriamente de um dos pais. O outro filho é formado pela carga genética complementar, ou seja, se o primeiro gene do filho 1 é do pai 1, o filho 2 herda o mesmo gene do pai 2. Esse processo se repete para os demais genes, como ilustrado na Figura 11.

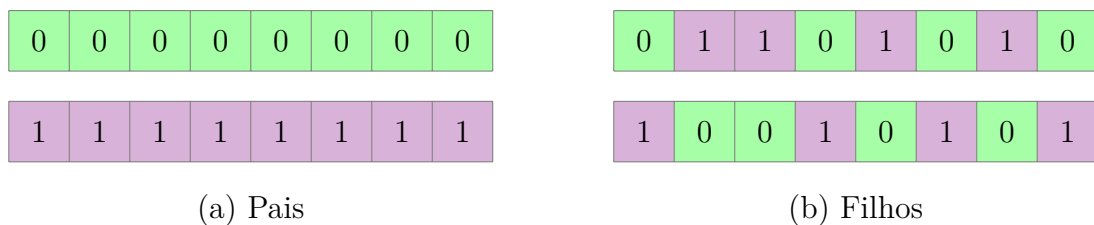


Figura 11 – Crossover uniforme

2.3.5.2 Mutação

A mutação introduz pequenas alterações aleatórias em um indivíduo. Ela tem a função de promover diversidade genética e evitar que a população fique presa em ótimos locais.

Exemplos de mutações:

- **Bit flip:** Esse tipo de mutação é muito usada em representações binárias e consiste na inversão (complemento) de um dos bits do cromossomo. A Figura 12 ilustra a operação de *bit flip*.

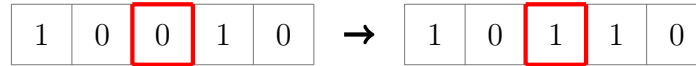


Figura 12 – Bit flip

- **Perturbação gaussiana:** Em representações com valores reais, adiciona-se um ruído aleatório extraído de uma distribuição normal. Esse método é exemplificado na Figura 13.

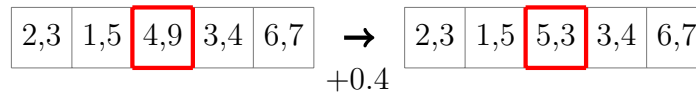


Figura 13 – Perturbação gaussiana

- **Troca de posição:** Em representações por permutação, dois elementos trocam de lugar. A troca de posição é demonstrada na Figura 14.

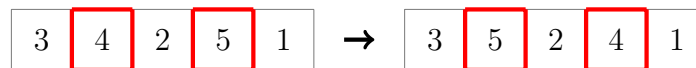


Figura 14 – Troca de posição

2.3.6 Reinscrição

A reinscrição é o processo pelo qual os indivíduos gerados por *crossover* e mutação são incorporados à população, formando a próxima geração em um AG. Essa etapa define quais soluções serão mantidas ou substituídas, influenciando diretamente a qualidade e a diversidade da população ao longo da evolução. O método de reinscrição adotado determina o equilíbrio entre preservar indivíduos de alta aptidão e explorar novas regiões do espaço de busca, sendo, portanto, um componente essencial para a eficiência e robustez do algoritmo (MITCHELL, 1996b).

Principais estratégias de reinscrição:

- **Reinscrição Total:** Todos os indivíduos da geração anterior são substituídos pelos indivíduos gerados. Favorece a exploração, mas permite a perda de boas soluções.
- **Elitismo:** Preserva os melhores indivíduos da geração atual. Garante que as soluções de maior qualidade não sejam perdidas.

- **Reinserção Uniforme:** São aplicadas técnicas de seleção para escolher os indivíduos da próxima geração. Mantém diversidade, mas permite a perda de boas soluções.

2.3.7 Critério de Parada

O término da execução de um algoritmo genético é controlado por critérios de parada que dependem do problema e das restrições computacionais. Os mais comuns incluem:

- **Número máximo de gerações:** O algoritmo é executado por um número fixo de iterações.
- **Convergência da população:** O processo é interrompido quando a variação de aptidão entre os indivíduos cai abaixo de um limite pré-definido.
- **Tempo limite:** A execução é finalizada após um período estabelecido.
- **Aptidão desejada:** A execução para ao encontrar uma solução que satisfaça um critério mínimo de qualidade.

2.4 Trabalhos Relacionados

Esta seção apresenta trabalhos que aplicam autômatos celulares na modelagem de epidemias e o uso de algoritmos genéticos no ajuste de parâmetros, destacando suas contribuições para a simulação e análise de processos dinâmicos.

2.4.1 Autômatos Celulares Epidemiológicos

Em (WHITE; REY; SÁNCHEZ, 2007) é proposto um modelo de autômato celular para simular a propagação de epidemias, fundamentado no clássico modelo SIR (*Susceptible–Infectious–Recovered*). O autômato desenvolvido busca representar a disseminação de uma doença em um espaço finito, incorporando parâmetros como o fator de mobilidade dos indivíduos, a intensidade das conexões entre as células e a densidade demográfica. A simulação apresentada pelos autores evidencia, de forma genérica, como a taxa de contato entre indivíduos e a heterogeneidade populacional influenciam significativamente a dinâmica de evolução de uma epidemia.

No contexto da COVID-19, Chowdhury, Roychowdhury e Chaudhuri (2022) desenvolveram um autômato celular baseado no modelo SEIR (*Susceptible–Exposed–Infectious–Removed*.) que se diferencia por adotar um critério de vizinhança sem barreiras fixas, permitindo que cada célula interaja com qualquer outra célula do reticulado com uma

probabilidade proporcional à $1/d^n$, onde d é a distância euclidiana entre as células e n é um parâmetro ajustável do modelo. Esse parâmetro regula a facilidade de contato entre células distantes, simulando diferentes níveis de confinamento social. A aplicação do modelo aos dados da COVID-19 na Índia revelou que, na segunda onda, a probabilidade de transmissão aumentou e o expoente n diminuiu, indicando menor restrição social e maior taxa de disseminação do vírus. Embora o modelo tenha reproduzido bem os picos das ondas, ele apresentou divergências na fase de declínio dos casos e no platô final, sugerindo que fatores adicionais, como distribuição populacional heterogênea e aspectos socioeconômicos, devem ser incorporados em futuras simulações.

De forma análoga, [Silva \(2024\)](#) apresenta melhorias em relação ao modelo proposto por [Chowdhury, Roychowdhury e Chaudhuri \(2022\)](#), ajustando o raio de transmissão da infecção de maneira a reduzir o custo computacional, sem comprometer a fidelidade do modelo aos dados observados. Também são introduzidos novos parâmetros, como a taxa de vacinação e a possibilidade de reinfecção, e é definida uma borda fixa para o reticulado, contribuindo para uma simulação mais realista da propagação da doença. Este modelo foi escolhido como estudo de caso para o ajuste dos parâmetros com AG, e seu funcionamento será detalhado com maior profundidade no capítulo de desenvolvimento.

2.4.2 Algoritmos Genéticos no ajuste de parâmetros

O uso de algoritmos genéticos para ajustar parâmetros em autômatos celulares é uma técnica consolidada na literatura. [Mitchell, Crutchfield e Das \(1993\)](#) aplicaram essa abordagem a um autômato celular unidimensional, com o objetivo de evoluir regras locais capazes de resolver a tarefa de classificação de densidade. Essa tarefa consiste em, dado um reticulado inicial, formado por células binárias (estados 0 ou 1), evoluir o sistema, após um número finito de transições, a uma configuração homogênea: todas as células no estado 1 se a configuração inicial possui mais células no estado 1, ou todas no estado 0, caso contrário. Esse trabalho evidenciou o potencial desse paradigma para a síntese de sistemas computacionalmente complexos, dispensando a necessidade de engenharia manual das regras locais, o que pode ser desafiador quando se busca projetar comportamentos emergentes específicos e sofisticados.

No contexto de modelos de propagação de incêndios, [Murilo \(2025\)](#) propôs uma abordagem evolutiva para calibrar automaticamente os parâmetros de um modelo probabilístico que simula a propagação do fogo em uma área florestal. Entre as contribuições, destaca-se a introdução de um novo parâmetro a ser ajustado pelo AG: a taxa de amostragem, que determina a quantidade de passos de evolução do autômato que serão executados entre cada avaliação, na qual o reticulado é comparado com uma amostra do conjunto de dados de referência. A inclusão desse parâmetro tornou a abordagem mais versátil e capaz de se adaptar a modelos e entradas distintas e permitiu uma representatividade

mais assertiva dos fenômenos estudados. Além disso, foram avaliadas funções de aptidão alternativas, que consideraram não apenas a presença de fogo, mas também estados de recuperação da vegetação, resultando em maior precisão na convergência do AG.

No âmbito de modelos epidemiológicos, [Fraga \(2021\)](#) aborda o ajuste de parâmetros de um modelo baseado em autômatos celulares que simula a evolução de uma população de insetos vetores da doença de chagas. O estudo explora uma estratégia evolutiva multi-estágio, na qual a evolução do autômato celular é dividido em duas fases, infestação e dispersão, e considera parâmetros demográficos (probabilidades de reprodução, sobrevivência e desenvolvimento) e de movimentação (coeficientes de dispersão e média de insetos que chegam diariamente na aldeia). O método empregado consiste em ajustar alternadamente essas duas categorias de parâmetros. Em cada etapa, o algoritmo genético concentra-se na evolução dos parâmetros de uma categoria, enquanto mantém os demais constantes. Essa abordagem busca equilibrar a adaptação do modelo, evitando ajustes simultâneos que poderiam dificultar a convergência do processo evolutivo. Além disso, foi incorporado um processo de repovoamento, no qual a população era reinicializada entre etapas a partir dos melhores indivíduos da etapa anterior. Tal estratégia possibilitou um ajuste mais refinado dos parâmetros, aproximando o comportamento simulado pelo modelo das dinâmicas observadas nos dados de referência.

No contexto da pandemia de COVID-19, [Yarsky \(2021\)](#) aplicou AGs ao ajuste de parâmetros de um modelo SEIR (*Susceptible–Exposed–Infected–Removed*), utilizando dados empíricos dos Estados Unidos para prever infecções e óbitos. Esse trabalho se destaca por adotar um ajuste incremental, iniciando com a evolução de apenas três parâmetros e, a cada 50 gerações, um novo parâmetro era incorporado, até que o modelo estivesse completo. Essa estratégia tem como objetivo imitar o conceito de "equilíbrio pontuado", proposto por [Eldredge e Gould \(1972\)](#), segundo o qual populações de organismos sexuais permanecem, na maior parte do tempo, em relativa estase evolutiva, experimentando mudanças rápidas e abruptas em períodos pontuais. A estratégia incremental se mostrou útil ao proporcionar picos pontuais de melhora na aptidão da população, favorecendo a exploração e reduzindo a convergência precoce para ótimos locais.

3 Desenvolvimento

Neste capítulo são descritos, de forma detalhada, o modelo de referência adotado (SILVA, 2024) e o algoritmo genético proposto para a calibração de seus parâmetros. Inicialmente, apresenta-se a reprodução do modelo epidemiológico baseado em Autômato Celular (AC), que servirá como base para os experimentos conduzidos ao longo deste trabalho.

3.1 Reprodução do Modelo de Referência

Como etapa inicial, foi realizada a reprodução do modelo proposto por Silva (2024), que por sua vez consiste em uma versão aprimorada do modelo originalmente apresentado por Chowdhury, Roychowdhury e Chaudhuri (2022).

3.1.1 Espaço Celular e Estado das Células

O modelo é estruturado sobre um reticulado bidimensional composto por células quadrangulares e utiliza borda fixa. Essa escolha representa uma modificação em relação ao modelo original, que emprega bordas periódicas. A adoção de bordas fixas tem como objetivo tornar a simulação mais fidedigna a cenários reais de propagação de doenças, nos quais a área analisada é naturalmente delimitada por fronteiras geográficas ou administrativas, como cidades, bairros. Diferentemente da borda periódica, que conecta artificialmente extremidades opostas do domínio, permitindo que indivíduos ou agentes infecciosos "reapareçam" no lado oposto do sistema.

- **Suscetível:** indivíduo ainda não exposto ao vírus;
- **Exposto:** indivíduo que foi exposto ao vírus, mas ainda não o transmite;
- **Infectado:** indivíduo capaz de transmitir o vírus;
- **Recuperado:** indivíduo que superou a infecção;
- **Vacinado:** indivíduo imunizado por vacinação, apresentando resistência à infecção.

Nos experimentos conduzidos neste trabalho foi utilizado um reticulado de dimensão 51×51 , totalizando 2601 células. Essa configuração foi definida de forma empírica, buscando um equilíbrio entre a fidelidade na reprodução dos resultados apresentados por Silva (2024) e o custo computacional das simulações. Além disso, a utilização de uma

dimensão ímpar permite posicionar o indivíduo infectado inicial exatamente no centro do reticulado, garantindo simetria espacial nas condições iniciais do modelo.

3.1.2 Vizinhança

As interações espaciais entre as células são modeladas por meio de uma vizinhança de Moore cujo raio corresponde a 50% da dimensão do reticulado. Assim, em um reticulado 10×10 , o raio da vizinhança é igual a 5, conforme ilustrado na Figura 15.

Essa escolha implica que a influência exercida por células infectadas não se restringe apenas às células imediatamente adjacentes, mas se distribui em camadas concêntricas ao redor da célula central, com intensidade modulada por um fator de decaimento.

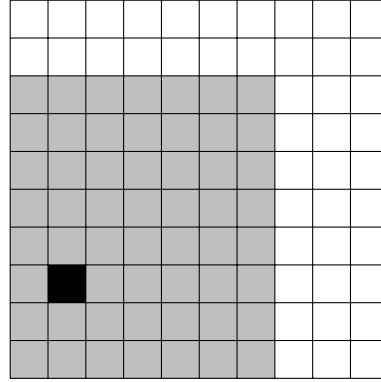


Figura 15 – Exemplo de vizinhança com raio igual à metade do reticulado

3.1.3 Regras de Transição

A dinâmica temporal do modelo é definida por regras de transição que determinam como cada célula evolui entre os diferentes estados ao longo do tempo discreto.

No instante t , uma célula (i, j) no estado Suscetível possui probabilidade de infecção $Q_I(i, j, t)$. Para decidir a transição, gera-se um número aleatório $\mu \in [0, 1]$. Caso $\mu < Q_I(i, j, t)$, a célula transita para o estado Exposto no instante $t + 1$.

Uma vez no estado Exposto, a célula permanece nesse estado por T_i unidades de tempo, representando o período latente da infecção. Após esse intervalo, a célula passa automaticamente para o estado Infectado.

De modo análogo, uma célula Infectada permanece nesse estado por T_r unidades de tempo. Decorrido esse período, ela transita para o estado Recuperado.

O estado Recuperado incorpora a possibilidade de perda gradual da imunidade. Uma célula recuperada pode ser reinfectada com probabilidade $\alpha \cdot Q_I(i, j, t)$, onde $\alpha \in [0, 1]$. Inicialmente, quando a célula entra no estado Recuperado, define-se $\alpha = 0$. A cada T_α unidades de tempo, esse valor é incrementado em $\Delta\alpha$, até atingir o limite máximo 1.

Esse mecanismo modela a redução progressiva da proteção imunológica após a recuperação.

Para células no estado Vacinado, a regra é semelhante à do estado Suscetível. Entretanto, caso ocorra infecção ($\mu < Q_I(i, j, t)$), a célula transita diretamente para o estado Recuperado no instante seguinte, com α inicializado em 0,1. Essa configuração representa a perda progressiva da eficácia da vacina ao longo do tempo.

A probabilidade de infecção $Q_I(i, j, t)$ é dada por:

$$Q_I = q \sum_{l=1}^L p_{int}(l) p_I(l) \quad (3.1)$$

em que

$$p_{int} = \frac{1}{A_n l^n} \quad (3.2)$$

$$A_n = \sum_{l=1}^L \frac{1}{l^n} \quad (3.3)$$

Nessa expressão, q representa a taxa de transmissão da infecção, L corresponde ao número de camadas da vizinhança da célula (i, j) , e n é um parâmetro associado ao nível de distanciamento social. p_{int} é a probabilidade de interação com a camada l , definida pelo valor $\frac{1}{l^n}$ normalizado pelo termo A_n , que representa a soma da chance de interação de todas as camadas da vizinhança. p_{int} é construído de tal forma que, ao aumentar o nível de distanciamento social (n), a probabilidade de interação diminui de forma exponencial conforme a distância da camada em relação à célula central.

O termo $p_I(l)$ indica a proporção de células infectadas na camada l da vizinhança no instante t . Dessa forma, o termo $p_{int}(l) p_I(l)$ representa a probabilidade da célula (i, j) ser infectada por uma célula da camada l .

Adicionalmente, o modelo incorpora uma estratégia de vacinação. A partir do instante t_v , uma fração p_v das células é selecionada aleatoriamente a cada iteração. Caso a célula selecionada esteja no estado Suscetível, ela é convertida para o estado Vacinado no instante seguinte.

3.2 Algoritmo Genético

Nesta seção são descritas as escolhas realizadas na configuração do AG, incluindo a representação dos indivíduos, a estratégia de inicialização da população, a definição da função de avaliação, os operadores genéticos empregados e o critério de parada adotado.

3.2.1 Representação do Indivíduo

A população de um algoritmo genético corresponde a potenciais soluções para o problema analisado, sendo que a forma de representação de seus indivíduos (cromossomos) é definida de acordo com a natureza desse problema. Em nosso caso, cada indivíduo do AG representa uma possível configuração do modelo epidemiológico. Essa configuração é codificada como uma 8-tupla:

$$(q, n, T_i, T_r, \Delta\alpha, T_\alpha, t_v, p_v)$$

em que cada componente corresponde a um parâmetro do modelo:

- q : taxa de infecção;
- n : nível de distanciamento social;
- T_i : duração do período latente da infecção;
- T_r : duração do período agudo da infecção;
- $\Delta\alpha$: incremento aplicado ao valor α de uma célula a cada T_α intervalos de tempo;
- T_α : intervalo entre incrementos sucessivos de α ;
- t_v : tempo de início da vacinação;
- p_v : porcentagem de células vacinadas por instante de tempo.

Os parâmetros q e n assumem valores reais não negativos. Já $\Delta\alpha$ e p_v pertencem ao intervalo $[0, 1]$. Os parâmetros temporais T_i , T_r , T_α e t_v são inteiros não negativos. Dessa forma, o cromossomo combina genes contínuos e discretos, exigindo tratamentos específicos nos operadores genéticos.

3.2.2 Inicialização da População

A população inicial é gerada de forma aleatória, respeitando intervalos pré-estabelecidos considerados plausíveis para cada parâmetro. Essa definição de limites busca restringir o espaço de busca a regiões compatíveis com o comportamento esperado do modelo, sem impor viés excessivo às soluções iniciais.

Os valores mínimos e máximos (inclusivos) adotados são apresentados na Tabela 1.

Atributo	MIN	MAX
q	0,01	1,0
n	0,01	5,0
T_i	1	14
T_r	1	28
$\Delta\alpha$	0,0	1,0
T_α	1	60
T_v	1	100
p_v	0,0	1,0

Tabela 1 – Valores mínimos e máximos para atributos do indivíduo

Cada atributo é amostrado uniformemente dentro de seu respectivo intervalo. Assim, garante-se diversidade inicial suficiente para permitir que o processo evolutivo explore diferentes regiões do espaço de busca.

3.2.3 Função de Avaliação

A função de avaliação (*fitness*) foi definida de modo a quantificar a proximidade entre o comportamento produzido por uma configuração candidata e o comportamento observado em um modelo de referência.

Como o modelo em AC é estocástico, uma única execução não é suficiente para caracterizar adequadamente sua dinâmica. Assim, para avaliar um indivíduo, o modelo é simulado n vezes utilizando a configuração representada pelo cromossomo, ao longo de 250 iterações. Neste trabalho, fixou-se $n = 5$ para todos os experimentos.

O fitness é então calculado como:

$$Fitness = \frac{1}{n} \sum_{t=1}^n \sum_{s \in S} \frac{|X_{s,t} - Y_{s,t}|}{\max\{Y_{s,t}, 1\}} \quad (3.4)$$

onde:

- S é o conjunto dos estados considerados na comparação. Inicialmente, foram utilizados os estados Exposto, Infectado e Vacinado;
- $X_{s,t}$ representa a quantidade de células no estado s no instante t para a solução candidata;
- $Y_{s,t}$ representa a quantidade de células no estado s no instante t extraída da curva de estados alvo.

O uso da normalização por $\max\{Y_{s,t}, 1\}$ evita divisões por zero e reduz a influência desproporcional de instantes em que as quantidades observadas são muito pequenas.

3.2.4 Cruzamento

Para a recombinação genética foi adotado o operador BLX- α (ESHELMAN; SCHAFER, 1993), adequado para genes de natureza contínua. Diferente do *crossover* aritmético, o BLX- α permite explorar regiões além do intervalo delimitado pelos pais, favorecendo a diversidade da população ao longo das gerações.

Dados dois pais p_1 e p_2 , define-se:

$$c_{min} = \min(p_1, p_2) \quad (3.5)$$

$$c_{max} = \max(p_1, p_2) \quad (3.6)$$

$$I = c_{max} - c_{min} \quad (3.7)$$

$$x \sim \mathcal{U}(c_{min} - \alpha I, c_{max} + \alpha I) \quad (3.8)$$

onde $\alpha \in [0, 1]$ controla a expansão do intervalo de amostragem. O gene filho x é então extraído de uma distribuição uniforme no intervalo expandido.

A Figura 16 ilustra esse processo. O intervalo original delimitado pelos pais $[c_{min}, c_{max}]$ é expandido proporcionalmente a αI em ambas as extremidades, permitindo que o valor do filho seja amostrado em uma região maior do espaço de busca.

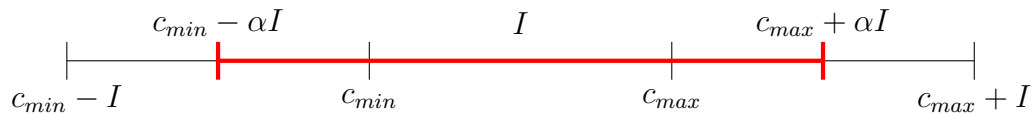


Figura 16 – Possíveis valores para x em vermelho

Exemplo de aplicação. Considere dois valores parentais para determinado gene contínuo: $p_1 = 0,4$ e $p_2 = 0,6$, com $\alpha = 0,5$.

Tem-se:

$$c_{min} = 0,4 \quad c_{max} = 0,6$$

$$I = 0,6 - 0,4 = 0,2$$

O intervalo expandido será:

$$[0,4 - 0,5 \cdot 0,2; 0,6 + 0,5 \cdot 0,2] = [0,3; 0,7]$$

Assim, o gene do filho será um valor amostrado uniformemente no intervalo $[0,3; 0,7]$. Se, por exemplo, o valor sorteado for $x = 0,65$, esse será o valor atribuído ao gene no descendente.

Para genes discretos, o valor gerado é arredondado para o inteiro mais próximo após a aplicação da expressão.

3.2.5 Mutação

A mutação é aplicada de forma independente a cada gene com probabilidade inicial de 0,3. Utiliza-se mutação gaussiana. O desvio padrão da distribuição normal é calculado como:

$$\sigma = (g_{max} - g_{min}) \cdot \sigma_0 \quad (3.9)$$

em que g_{max} e g_{min} representam os limites do gene e σ_0 é um parâmetro do algoritmo.

Para genes discretos, os valores extraídos da distribuição são arredondados para o inteiro mais próximo.

Essa estratégia permite perturbações proporcionais à escala de cada parâmetro, preservando coerência entre diferentes tipos de genes.

Exemplo de aplicação. Considere um gene cujo intervalo permitido é $[0, 10]$ e $\sigma_0 = 0,1$. Nesse caso, o desvio padrão da mutação é calculado como

$$\sigma = (10 - 0) \cdot 0,1 = 1$$

Suponha que o valor atual do gene seja 5. Durante a aplicação da mutação, é sorteado um valor de uma distribuição normal $\mathcal{N}(0, 1)$, por exemplo 0,7. O novo valor do gene passa então a ser

$$5 + 0,7 = 5,7$$

Caso o gene seja discreto, esse valor é arredondado para o inteiro mais próximo, resultando em 6.

3.2.6 Reinscrição

A estratégia de reinscrição combina elitismo e geração de novos indivíduos. A cada geração, 10% da nova população é composta pelos melhores indivíduos da geração anterior.

Os demais indivíduos são gerados por cruzamento e mutação, sendo os pais selecionados por torneio de tamanho 3. Essa escolha busca equilibrar pressão seletiva e manutenção de diversidade populacional.

3.2.7 Critério de Parada

Inicialmente, estabeleceu-se um limite máximo de 200 gerações. Após a análise do comportamento do algoritmo, esse parâmetro poderá ser ajustado.

Além disso, será incorporado um critério adicional baseado em estagnação: caso não haja melhora no fitness do melhor indivíduo por um número pré-definido de gerações consecutivas, a execução será encerrada. De início, esse critério será ignorado a fim de observar o comportamento do algoritmo.

4 Experimentos e Resultados

Neste capítulo são descritos os experimentos conduzidos ao longo desta pesquisa para validar a eficiência e eficácia do algoritmo genético proposto.

Os primeiros experimentos realizados visaram o refinamento dos hiperparâmetros do AG. Diferentemente dos parâmetros do modelo epidemiológico, que representam características do processo de propagação da doença, os hiperparâmetros do AG controlam o comportamento do próprio processo de otimização. Foi adotada uma abordagem experimental, na qual diferentes configurações foram testadas e comparadas sistematicamente. O processo de ajuste foi conduzido de forma incremental: parte-se de uma configuração inicial do algoritmo e, a partir dela, avalia-se o impacto da modificação de determinados hiperparâmetros sobre o desempenho obtido.

Os modelos de simulação foram implementados na linguagem Rust, sem dependência de bibliotecas externas, visando maior controle sobre a execução e otimização de desempenho computacional.

Após a definição da configuração final dos hiperparâmetros, o algoritmo genético será avaliado em cenários adicionais com o objetivo de verificar sua capacidade de generalização. Para isso, o algoritmo foi aplicado inicialmente no ajuste de duas configurações alternativas do modelo de autômato celular, cujos parâmetros serão definidos artificialmente. Esse procedimento permite avaliar se o algoritmo consegue recuperar diferentes conjuntos de parâmetros do modelo a partir de curvas de estados distintas. Posteriormente, o AG foi submetido a dados reais de COVID-19, com o objetivo de verificar sua aplicabilidade em cenários além daqueles controlados pelo próprio modelo de simulação.

4.1 Configuração Experimental

Esta seção descreve a metodologia adotada para conduzir os experimentos de ajuste do algoritmo genético. Inicialmente, um conjunto de dados artificiais foi construído para servir como referência (dados históricos) para o processo de calibração. Em seguida, descreve-se o procedimento de avaliação utilizado para medir o desempenho das diferentes configurações do algoritmo.

4.1.1 Geração dos Dados de Referência Artificiais

Boa parte dos experimentos foram conduzidos considerando um conjunto de dados de referência gerados artificialmente a partir do próprio modelo epidemiológico utilizado pelo AG. Essa estratégia visa criar um ambiente experimental mais bem comportado e

controlado, o qual foi empregado durante as etapas de configuração dos hiperparâmetros do algoritmo genético e sua validação inicial. Nesse processo, os parâmetros do autômato celular que controlam o comportamento do modelo epidemiológico (modelo alvo) foram fixados conforme apresentado em (SILVA, 2024), assumindo os seguintes valores:

- Taxa de infecção (q): 0,2406;
- Nível de distanciamento social (n): 1,3449;
- Duração do período latente da infecção (T_i): 8;
- Duração do período agudo da infecção (T_r): 10;
- Incremento de taxa de reinfecção ($\Delta\alpha$): 0,1;
- Intervalo de incrementos de taxa de reinfecção (T_α): 30;
- Tempo de início da vacinação (t_v): 60;
- Porcentagem de células vacinadas por instante de tempo (p_v): 1%;

Toda simulação é iniciada com uma célula no estado Infectado posicionada no centro do reticulado. Essa escolha busca padronizar as condições iniciais do sistema e facilitar a comparação entre diferentes execuções do modelo.

Para obter as curvas (evolução temporal) dos estados do modelo SEIR que servirão de referência para o cálculo do *fitness* do algoritmo genético, foi calculada a quantidade média de indivíduos em cada estado ao longo do período simulado, considerando 1000 execuções independentes do autômato celular utilizando a configuração alvo. A utilização de múltiplas execuções é necessária devido à natureza estocástica do modelo, ou seja, diferentes simulações a partir da mesma configuração podem apresentar variações significativas na evolução temporal dos estados epidemiológicos. Ao considerar a média de um grande número de execuções, obtém-se uma representação mais estável do comportamento esperado do sistema.

A Figura 17 apresenta a evolução temporal média de estados epidemiológicos gerada a partir das simulações do modelo usando a configuração alvo. O eixo horizontal representa o tempo de simulação (iterações do modelo), enquanto o eixo vertical indica a quantidade de indivíduos em cada estado. As regiões sombreadas representam o desvio padrão das respectivas curvas médias.

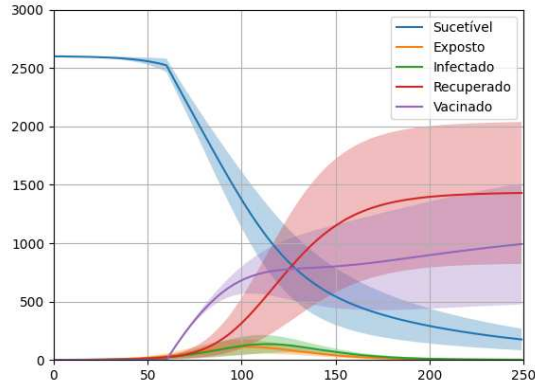


Figura 17 – Curva de estados da configuração alvo

4.1.2 Avaliação do Algoritmo

Para avaliar o impacto dos diferentes parâmetros do algoritmo genético, cada configuração de hiperparâmetros foi executada 15 vezes. Esse procedimento busca reduzir a influência da aleatoriedade inerente ao processo evolutivo.

Ao final dessas execuções, seleciona-se a solução com menor valor de *fitness*. Essa solução é então submetida a uma avaliação mais robusta, a qual é realizada a partir de 1000 execuções independentes do autômato celular utilizando os parâmetros encontrados pelo AG. A partir dessas execuções, calcula-se os valores médios dos estados do sistema, os quais são comparados com os dados de referência, por meio da métrica de erro RMSE (raiz do erro quadrático médio) definida por:

$$RMSE = \sqrt{\frac{1}{T \cdot |S|} \sum_{t=1}^T \sum_{s \in S} (X_{s,t} - Y_{s,t})^2} \quad (4.1)$$

na qual:

- T é a quantidade de iterações do autômato.
- S é o conjunto de possíveis estados da célula.
- $X_{s,t}$ é a quantidade média de células no estado s durante o instante de tempo t do modelo avaliado.
- $Y_{s,t}$ é a quantidade média de células no estado s durante o instante de tempo t do modelo de referência.

A utilização do RMSE permite quantificar a diferença entre as curvas de evolução temporal dos estados epidemiológicos produzidas pelo modelo ajustado e observadas nos dados de referência.

4.2 Ajuste dos Hiperparâmetros

4.2.1 Configuração Inicial do AG

O processo de ajuste dos hiperparâmetros teve início com a definição de uma configuração inicial do algoritmo genético, apresentada na Tabela 2.

Categoria	Parâmetro	Valor
Configuração geral	Número de gerações	200
	Limite de estagnação	-
	Tamanho da população	100
	Elitismo	10%
Operadores	Inicialização	Amostragem uniforme
	Função de fitness	Métricas selecionadas (Expostos, Infectados, Vacinados)
	Seleção	Torneio (tamanho = 3)
	Cruzamento	BLX- α ($\alpha = 0,3$)
	Mutação	Mutação gaussiana (taxa de 0,3 por gene, $\sigma_0 = 0,1$)

Tabela 2 – Configuração inicial do algoritmo genético

Essa configuração inicial serve como ponto de partida para os experimentos subsequentes. A partir dela, diferentes hiperparâmetros são modificados individualmente para avaliar seu impacto sobre o desempenho do algoritmo.

Os resultados obtidos com essa configuração inicial apresentaram um RMSE de 265,0763. A comparação entre os parâmetros encontrados pelo algoritmo e os valores alvo é mostrada na Tabela 3.

Tabela 3 – Melhores parâmetros da configuração inicial

Parâmetro	Encontrado	Alvo
q	0,1933	0,2406
n	1,1311	1,3449
T_i	9	9
T_r	12	10
$\Delta\alpha$	0,2928	0,1
T_α	45	30
T_v	77	60
p_v	0,0195	0,01

A evolução temporal dos estados epidemiológicos obtida com essa solução é apresentada na Figura 18.

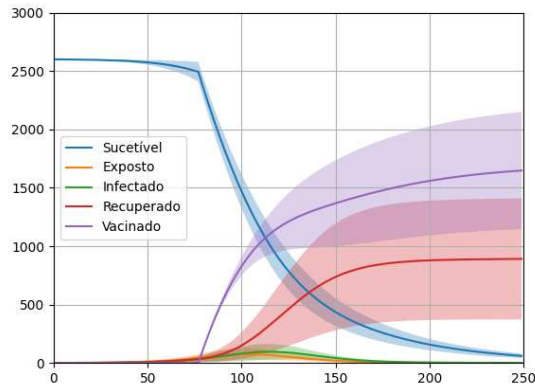


Figura 18 – Evolução temporal dos estados do modelo epidemiológico a partir da configuração de referência.

Embora os valores encontrados estejam relativamente próximos de alguns parâmetros do modelo alvo, observa-se que o erro global ainda é elevado, em especial a relação entre os estados Infectado e Vacinado. A fim de melhorar a eficiência do AG, foi realizado o ajuste individual de cada um dos seus parâmetros, como segue.

4.2.2 Critério de Parada

A primeira configuração foi executada com um limite de 200 gerações, com o objetivo de observar o comportamento do algoritmo ao longo de um número elevado de iterações.

A Figura 19 apresenta a evolução do valor médio do *fitness* do melhor indivíduo ao longo das gerações do AG, considerando as 15 execuções independentes. O eixo horizontal representa o número de gerações, enquanto o eixo vertical indica o valor da função objetivo (aptidão) do melhor indivíduo encontrado em cada geração. A região sombreada representa o desvio padrão da curva.

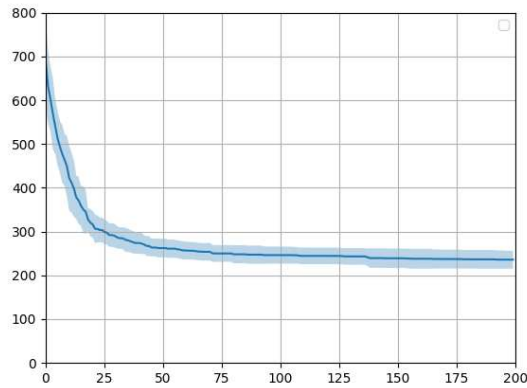


Figura 19 – Evolução da aptidão média do melhor indivíduo ao longo das execuções do AG.

Observa-se que a maior parte da melhoria ocorre nas primeiras gerações. Após aproximadamente 100 gerações, a evolução do fitness torna-se pouco significativa, indicando que o algoritmo tende a estabilizar sua busca.

Com base nessa observação, os experimentos subsequentes passaram a utilizar um limite máximo de 100 gerações. Além disso, foi incorporado um critério adicional de parada baseado em estagnação, no qual, caso o melhor indivíduo não apresente melhoria ao longo de 25 gerações consecutivas, a execução é interrompida antecipadamente.

Essa estratégia permite reduzir o custo computacional das execuções sem comprometer significativamente a qualidade das soluções encontradas.

4.2.3 Tamanho de População

Uma vez definido o critério de parada, investigou-se o impacto do tamanho da população sobre o desempenho do algoritmo. Foram testados três tamanhos populacionais: 100, 130 e 160 indivíduos.

Os resultados obtidos são apresentados na Tabela 4.

Tabela 4 – Resultados dos experimentos com tamanho de população

Tamanho da População	RMSE
100	265,0763
130	145,3052
160	128,6803

Observa-se que o aumento do tamanho da população resulta em uma redução significativa do erro obtido pelo algoritmo. A população de 160 indivíduos apresentou o menor RMSE dentre as configurações testadas.

Entretanto, a melhora no valor do RMSE de 130 indivíduos para 160 não foi tão significativa se comparada ao aumento de 100 para 130 e ao aumento no tempo de execução. Para 130 indivíduos, cada execução do algoritmo demorou, em média, 1,44 horas, enquanto que para 160 indivíduos o tempo médio foi de 1,82 horas. O que caracteriza um aumento de aproximadamente 26,38% no tempo contra uma melhora de 11,44% no valor do RMSE. Diante disso, visando um equilíbrio entre eficácia e eficiência, optou-se por adotar o valor intermediário de 130 indivíduos nos experimentos subsequentes.

4.2.4 Mutação

Após a definição do tamanho da população, investigou-se o impacto do operador de mutação sobre o desempenho do algoritmo.

Foram testadas cinco combinações entre a taxa de mutação aplicada a cada gene e a intensidade do desvio padrão da mutação gaussiana, controlada pelo parâmetro σ_0 .

Os resultados obtidos são apresentados na Tabela 5.

Tabela 5 – Resultados dos experimentos com mutação

Taxa	σ_0	RMSE
0,3	0,1	145,3052
0,15	0,1	76,4000
0,05	0,1	96,3438
0,1	0,2	80,7421
0,1	0,3	77,2096

Observa-se que a configuração originalmente adotada na configuração inicial (taxa de mutação igual a 0,3) apresentou o pior desempenho dentre os valores testados. Esse comportamento sugere que a incidência excessiva de mutações pode estar prejudicando a preservação de soluções promissoras ao longo das gerações.

As configurações com taxas intermediárias apresentaram resultados consideravelmente superiores. Em particular, destacaram-se duas combinações: (taxa = 0,15; σ_0 = 0,1) e (taxa = 0,1; σ_0 = 0,3), ambas produzindo valores de RMSE significativamente menores.

Essas duas configurações foram então selecionadas para os experimentos subsequentes, nos quais se avaliou o impacto de diferentes níveis de elitismo sobre o desempenho do algoritmo.

4.2.5 Elitismo

Foram testadas proporções de elitismo de 5%, 10%, 15% e 20%. Os resultados obtidos para as duas configurações de mutação selecionadas anteriormente são apresentados na Tabela 6.

Tabela 6 – taxa=0,15, $\sigma_0=0,1$

Elitismo	RMSE (taxa=0,15, $\sigma_0=0,1$)	RMSE (taxa=0,1, $\sigma_0=0,3$)
0,1	76,4000	77,2096
0,2	142,1914	119,9328
0,15	115,5039	95,3439
0,05	74,1214	128,5533

Observa-se que, para a configuração com taxa de mutação igual a 0,15 e $\sigma_0 = 0,1$, o elitismo de 5% apresentou o melhor desempenho, resultando no menor valor de RMSE.

Considerando tanto o desempenho obtido quanto a manutenção da diversidade genética, optou-se por adotar o elitismo de 5% e a configuração de mutação (taxa = 0,15; $\sigma_0 = 0,1$) nos experimentos subsequentes.

4.2.6 Cruzamento

O próximo experimento teve como objetivo investigar o impacto do operador de cruzamento sobre o desempenho do algoritmo. Em particular, foi analisado o parâmetro α do operador BLX- α , responsável por controlar a amplitude do intervalo de amostragem utilizado na geração dos descendentes.

Foram testados três valores para o parâmetro α , além de uma configuração alternativa utilizando o operador de *crossover* uniforme.

Os resultados obtidos são apresentados na Tabela 7.

Tabela 7 – Resultados dos experimentos com *crossover*

	RMSE
$\alpha = 0,3$	74,1214
$\alpha = 0,1$	87,7553
$\alpha = 0,5$	70,2997
<i>crossover</i> uniforme	101,81

Observa-se que o aumento do parâmetro α para 0,5 resultou no melhor desempenho dentre as configurações avaliadas. Esse resultado indica que uma maior dispersão na geração dos descendentes contribui positivamente para a exploração do espaço de busca.

Em contraste, o operador de *crossover* uniforme apresentou desempenho inferior ao operador BLX- α , reforçando a adequação deste último para problemas que envolvem parâmetros contínuos.

Dessa forma, adotou-se o operador BLX- α com $\alpha = 0,5$ como configuração padrão do algoritmo.

4.2.7 Função de Aptidão

Por fim, investigou-se o impacto da escolha dos estados considerados na função de fitness. Inicialmente, o cálculo do erro considerava apenas os estados Exposto, Infectado e Vacinado. Entretanto, foi realizado um experimento adicional incluindo também o estado Recuperado na função de avaliação, com o objetivo de verificar se essa informação adicional poderia melhorar a precisão do ajuste.

Os resultados indicaram que a inclusão desse estado resultou em uma degradação do desempenho do algoritmo, produzindo um RMSE de 147,46. Esse comportamento sugere que os estados Exposto, Infectado e Vacinado já fornecem informações suficientes para caracterizar a dinâmica do modelo epidemiológico.

Além disso, observou-se que a quantidade de células no estado Recuperado apresenta elevada variabilidade ao longo das execuções do modelo, o que pode introduzir ruído adicional na função de avaliação. Dessa forma, optou-se por manter a configuração original da função de fitness considerando apenas os estados Exposto, Infectado e Vacinado.

4.3 Análise de Desempenho

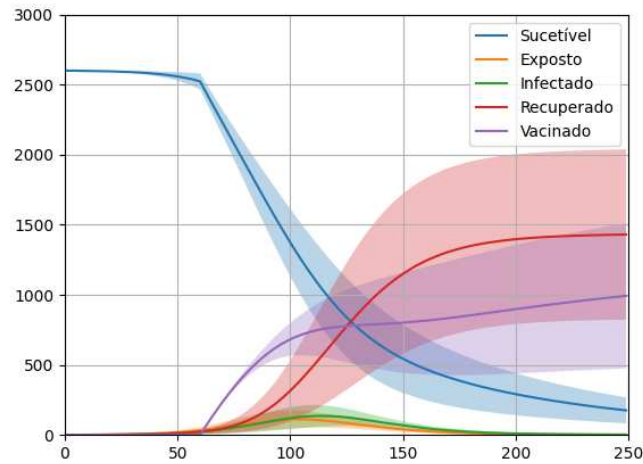
Após a realização dos experimentos descritos na seção anterior, foi definida a configuração final dos hiperparâmetros do algoritmo genético, apresentada na Tabela 8.

Tabela 8 – Configuração final do algoritmo genético

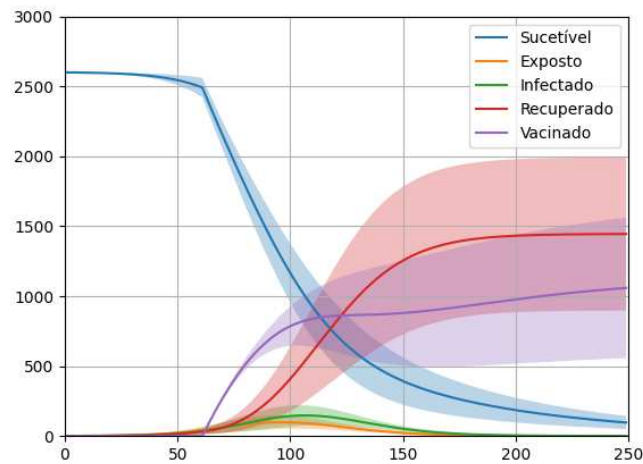
Categoria	Parâmetro	Valor
Configuração geral	Número de gerações	100
	Limite de estagnação	25
	Tamanho da população	130
	Elitismo	5%
Operadores	Inicialização	Amostragem uniforme
	Função de fitness	Métricas selecionadas (Expostos, Infectados, Vacinados)
	Seleção	Torneio (tamanho = 3)
	Cruzamento	BLX- α ($\alpha = 0,5$)
	Mutação	Mutação gaussiana (taxa de 0,15 por gene, $\sigma_0 = 0,1$)

Com essa configuração final, o algoritmo foi capaz de encontrar uma solução com valor de RMSE igual a 70,2997.

A Figura 20 apresenta a comparação entre a curva de evolução dos estados epidemiológicos do modelo de referência e a curva produzida pelo modelo calibrado pelo algoritmo genético.



(a) Modelo de referência



(b) Modelo configurado pelo AG

Figura 20 – Comparação entre a evolução da doença de acordo com a configuração adotada pelo modelo de simulação

Observa-se que as curvas apresentam comportamento bastante semelhante, indicando que o algoritmo foi capaz de reproduzir de forma satisfatória a dinâmica do modelo de referência.

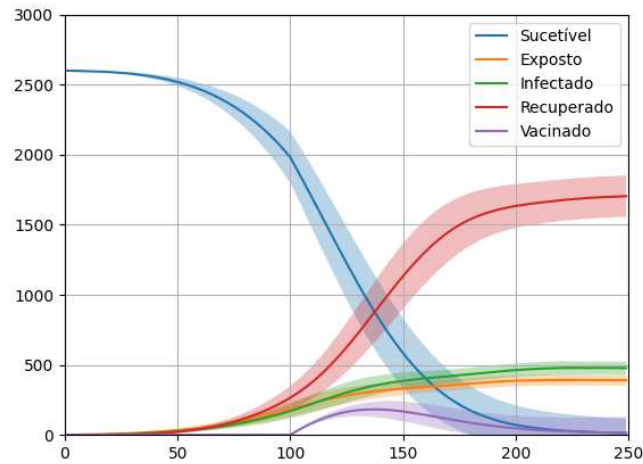
Para validar a generalização do AG configurado, o algoritmo foi aplicado a dois cenários com parâmetros de AC distintos (cenários 1 e 2), os quais foram gerados artificialmente alterando os valores dos parâmetros usados no modelo de AC.

No cenário 1, o algoritmo genético foi capaz de encontrar parâmetros próximos aos valores alvo para a maioria das variáveis (Tabela 9), indicando boa capacidade de ajuste. Entretanto, observou-se uma discrepância significativa nos parâmetros Δ_α e T_α , relacionados ao controle da taxa de reinfeção.

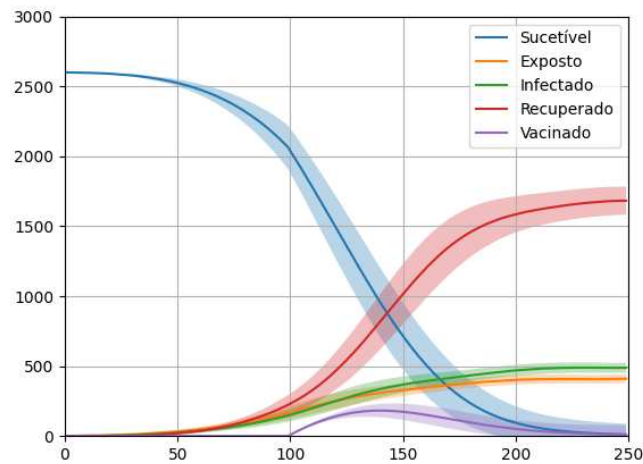
Apesar dessas diferenças, o gráfico de estados da solução encontrada apresenta comportamento semelhante ao da configuração alvo (Figura 21), indicando que o modelo conseguiu reproduzir adequadamente a dinâmica geral do sistema. Esse resultado sugere que diferentes combinações de parâmetros podem gerar curvas de evolução similares.

Tabela 9 – Valores encontrados pelo algoritmo para a configuração de parâmetros alternativa 1

Parâmetro	Encontrado	Alvo
q	0,5378	0,5
n	3,0274	3,0
T_i	10	8
T_r	12	10
$\Delta\alpha$	0,7653	0,2
T_α	27	10
T_v	100	100
p_v	0,004	0,005



(a) Modelo de referência



(b) Modelo configurado pelo AG

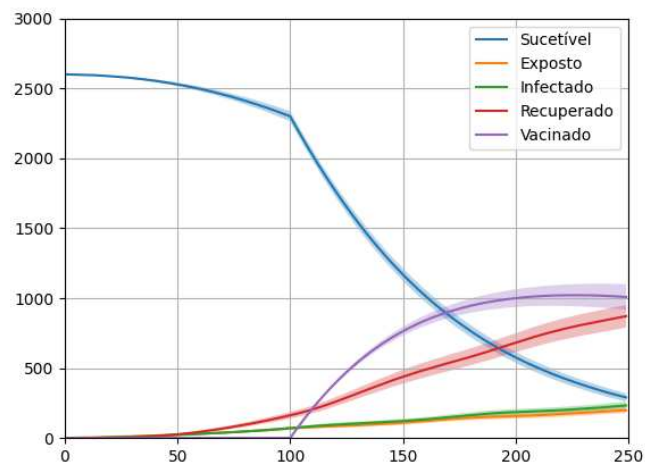
Figura 21 – Comparação entre a evolução da doença de acordo com a configuração adotada pelo modelo de simulação para o cenário 1

No cenário 2, o algoritmo genético apresentou um desempenho ainda mais consistente, encontrando valores próximos aos parâmetros alvo em praticamente todas as variáveis (Tabela 10). Embora ainda exista alguma diferença nos parâmetros Δ_α e T_α , os demais valores mostraram boa convergência.

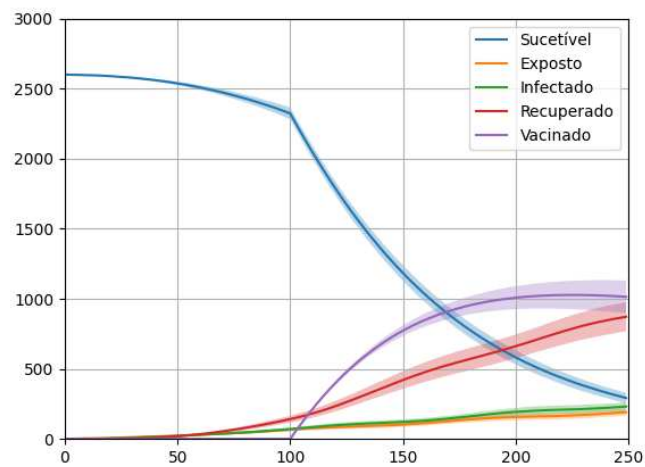
O gráfico de estados obtido demonstra alta similaridade com a configuração alvo (Figura 22), indicando uma reprodução fiel do comportamento esperado do sistema.

Tabela 10 – Valores encontrados pelo algoritmo para a configuração de parâmetros alternativa 2

Parâmetro	Encontrado	Alvo
q	0,6805	0,7
n	4,7540	5,0
T_i	9	8
T_r	12	10
$\Delta\alpha$	0,7319	0,5
T_α	27	20
T_v	100	100
p_v	0,0099	0,01



(a) Modelo de referência



(b) Modelo configurado pelo AG

Figura 22 – Comparação entre a evolução da doença de acordo com a configuração adotada pelo modelo de simulação para o cenário 1

Foi possível observar uma correlação entre a estabilidade da configuração alvo com a eficácia do algoritmo. A Tabela 11 mostra que, conforme o desvio padrão médio observado na curva gerada pela configuração alvo diminui, o erro obtido pelas configurações encontradas pelo algoritmo se torna menor. Isso se explica pelo fato de que uma simulação com valores mais estáveis favorece a amostragem limitada da função de *fitness*, permitindo que ela se torne mais representativa.

Tabela 11 – Comparação dos resultados obtidos nos diferentes cenários. σ Curva representa a média do desvio padrão das curvas de estados por dia e σ RMSE representa o desvio padrão do erro RMSE ao longo das execuções.

Configuração	σ Curva	RMSE Médio	Melhor RMSE	σ RMSE
Modelo de Referência	763,6100	136,5905	70,2997	37,1902
Cenário Alternativo 1	357,0114	89,3378	40,1809	37,9218
Cenário Alternativo 2	120,0980	36,7973	9,6273	17,8743

Além disso, o desempenho do algoritmo também foi testado no ajuste de dados reais de COVID-19. Para isso, foram utilizados dados referentes ao número de casos confirmados no Brasil, obtidos por meio da plataforma Brasil.IO. Essa plataforma reúne dados oficiais divulgados pelo Ministério da Saúde e pelas secretarias estaduais de saúde, disponibilizando registros atualizados sobre a evolução da pandemia em nível municipal ([Brasil.IO, 2025](#)).

Para este teste, foram consideradas exclusivamente as informações relativas ao número de pessoas infectadas, agrupadas por data. Isso se deve ao fato de o conjunto de dados não oferecer informações referentes a outros estados epidemiológicos contemplados pelo modelo, restando apenas a quantidade de casos por dia como estimativa do número de infectados. A partir desse conjunto de dados, foram extraídos os registros correspondentes ao ano de 2022, permitindo a análise da evolução diária dos casos confirmados ao longo desse período.

Entretanto, os dados brutos de número de casos são ruidosos e sujeitos a variações constantes, enquanto o AC foi projetado para produzir curvas com variações mais suaves. Dessa forma, antes de serem utilizados pelo algoritmo, os dados foram suavizados por meio de um filtro de Savitzky-Golay ([SAVITZKY; GOLAY, 1964](#)). As curvas de infectados, tanto a original quanto a suavizada, são apresentadas na Figura 23.

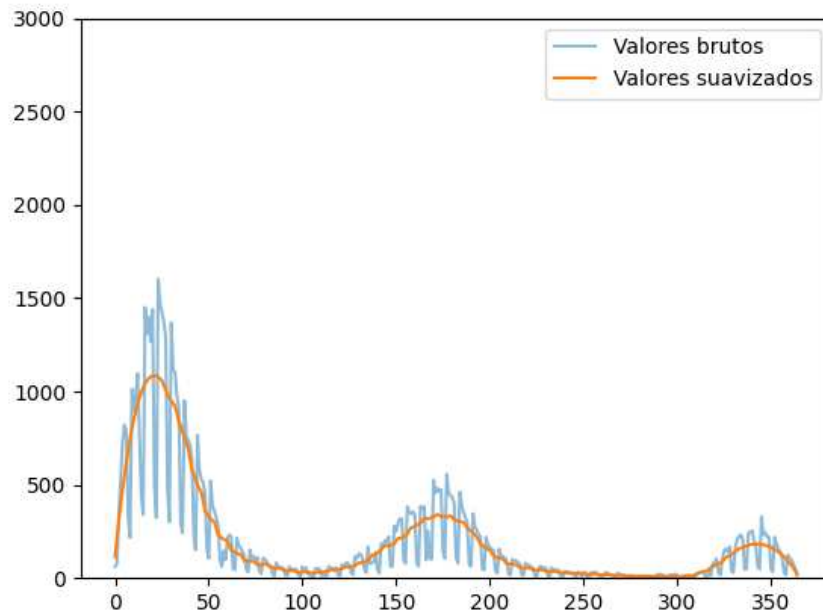
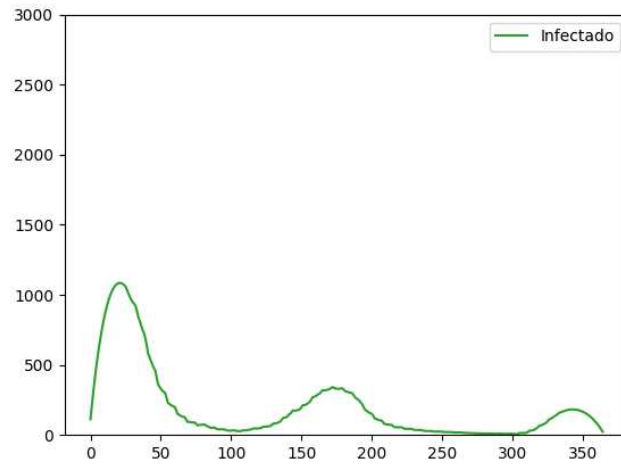


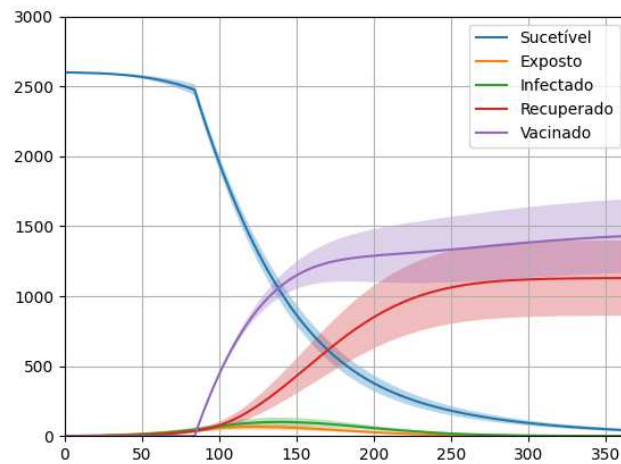
Figura 23 – Série temporal do número de infectados em 2022, mostrando os dados brutos e a versão suavizada utilizando o filtro de Savitzky-Golay.

Para englobar toda a duração da dinâmica dos dados, durante a etapa de avaliação do algoritmo (*fitness*), foram realizadas simulações de 365 iterações, ao invés das 250 iterações utilizadas nos experimentos anteriores. Além disso, devido à limitação dos dados de referência, o *fitness* foi configurado para avaliar apenas as diferenças no número de células no estado **Infectado**.

A primeira tentativa de ajustar esses dados não foi satisfatória. Nenhum dos picos foi reproduzido e o número de infectados permaneceu baixo ao longo da simulação, apresentando apenas um pequeno aumento em torno do dia 150 (Figura 24).



(a) Dados de referência



(b) Modelo configurado pelo AG

Figura 24 – Primeira tentativa de ajuste dos dados reais de infectados utilizando o algoritmo genético.

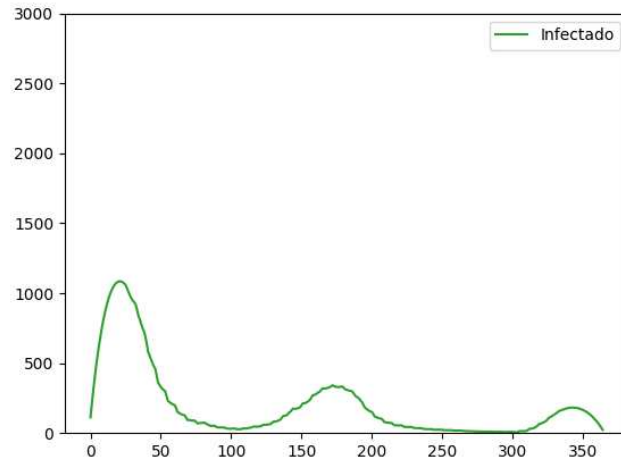
Para os experimentos seguintes, foram avaliadas duas hipóteses.

A primeira hipótese considera que o modelo é incapaz de reproduzir o primeiro pico devido às condições iniciais da simulação. Em todas as execuções do AC, a simulação é inicializada com apenas uma célula infectada no centro do reticulado. Entretanto, nos dados de referência, o primeiro dia já apresenta aproximadamente 112 infectados. Assim, o crescimento abrupto observado nos dados não pode ser reproduzido devido à discrepância nas condições iniciais, restando ao algoritmo ajustar apenas a dinâmica posterior, caracterizada por uma quantidade menor de infectados.

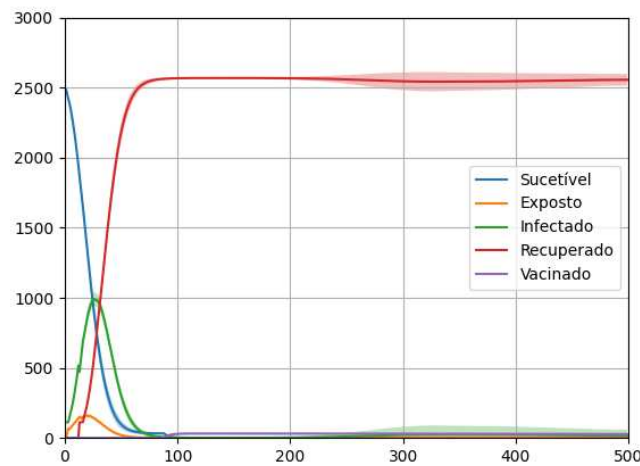
Nesse experimento, o reticulado foi inicializado com uma quantidade de células infectadas semelhante à observada nos dados de referência, distribuídas em posições ale-

atórias.

Ao adotar essa estratégia de inicialização, os resultados foram mais promissores (Figura 25). O primeiro pico pôde ser reproduzido de forma satisfatória, embora com uma altura ligeiramente inferior à observada nos dados de referência. No entanto, a dinâmica posterior ainda não pôde ser adequadamente capturada pelo modelo.



(a) Dados de referência



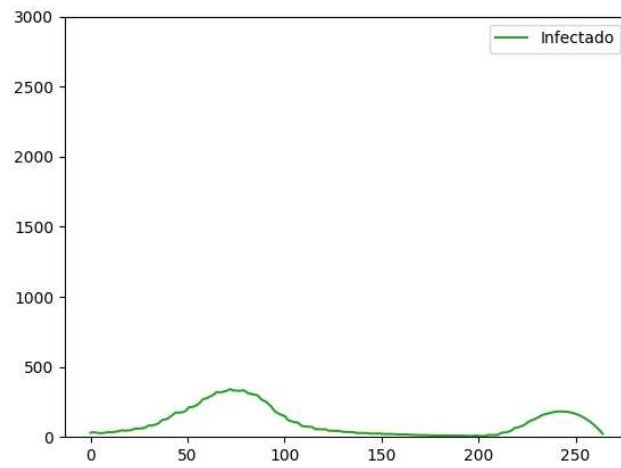
(b) Modelo configurado pelo AG

Figura 25 – Resultado do ajuste após modificar a condição inicial do modelo, utilizando um número inicial de células infectadas semelhante ao observado nos dados reais.

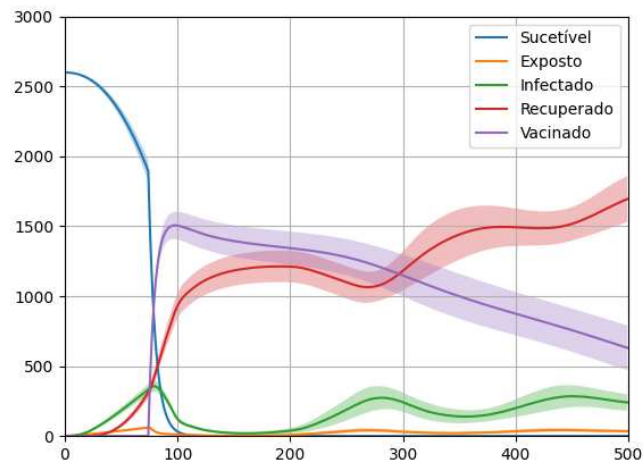
A segunda hipótese considera a possível limitação do modelo em reproduzir dinâmicas muito irregulares. Para avaliar essa possibilidade, em vez de ajustar toda a série temporal, foi utilizado um recorte dos dados a partir do dia 100, buscando selecionar

um trecho com comportamento mais regular e, potencialmente, mais reprodutível pelo modelo.

Os dois picos presentes no recorte selecionado puderam ser observados nos resultados (Figura 26). No entanto, o primeiro pico apresentou um crescimento mais lento do que o observado nos dados de referência. O segundo pico foi reproduzido com um pequeno atraso e, ao final da simulação, o número de infectados permaneceu acima dos valores registrados na série original.



(a) Dados de referência



(b) Modelo configurado pelo AG

Figura 26 – Resultado do ajuste considerando apenas a dinâmica dos dados a partir do dia 100.

5 Conclusão

Este trabalho propôs a utilização de um algoritmo genético para o ajuste automático dos parâmetros de um modelo epidemiológico de COVID-19 baseado em autômatos celulares.

Os resultados experimentais obtidos indicam que o algoritmo é capaz de reproduzir satisfatoriamente as curvas de evolução dos estados epidemiológicos geradas artificialmente pelo próprio modelo. Esse resultado sugere que, quando os dados de referência pertencem ao mesmo domínio do modelo utilizado na calibração, o processo evolutivo consegue identificar configurações de parâmetros que reproduzem adequadamente a dinâmica do sistema.

Entretanto, observou-se que os parâmetros encontrados pelo algoritmo nem sempre coincidem exatamente com os valores utilizados para gerar os dados de referência. Esse comportamento indica que diferentes combinações de parâmetros podem resultar em dinâmicas epidemiológicas semelhantes, produzindo curvas de estados bastante próximas entre si. Dessa forma, mesmo quando os parâmetros originais não são recuperados com precisão, o algoritmo ainda é capaz de encontrar configurações que reproduzem o comportamento global do modelo.

Outro aspecto observado durante os experimentos está relacionado à estabilidade das curvas utilizadas como referência para o cálculo do *fitness*. Configurações do modelo que geram dinâmicas altamente instáveis tendem a introduzir maior variabilidade nas simulações, o que pode gerar ruído adicional na função de avaliação e dificultar o processo de calibração. Em contrapartida, curvas mais estáveis mostraram-se mais facilmente reproduzíveis pelo algoritmo genético, uma vez que fornecem um sinal mais consistente para orientar o processo de busca no espaço de parâmetros.

O algoritmo também foi aplicado ao ajuste de dados reais de COVID-19, utilizando a série temporal de casos confirmados no Brasil durante o ano de 2022. Diferentemente dos experimentos realizados com dados gerados artificialmente, nesse cenário o modelo possui acesso apenas ao número de infectados ao longo do tempo, o que limita o processo de calibração à comparação entre essa única variável observável. Os resultados obtidos indicam que o modelo apresenta maior dificuldade em reproduzir integralmente a dinâmica dos dados reais.

Os experimentos sugerem que essa dificuldade é resultado de limitações do próprio modelo. Uma vez que ele opera utilizando uma configuração de regras fixa, torna-se difícil reproduzir comportamentos muito irregulares, que possivelmente exigiriam um modelo mais expressivo. Uma possibilidade seria utilizar um modelo cuja configuração possa variar

ao longo do tempo, permitindo representar dinâmicas associadas a mudanças de políticas públicas, como medidas de distanciamento social e campanhas de vacinação, ou mesmo à evolução do próprio vírus.

De maneira geral, os resultados indicam que o algoritmo genético proposto constitui uma ferramenta promissora para a calibração automática de modelos baseados em autómatos celulares. No entanto, sua eficácia depende da qualidade e da compatibilidade entre os dados utilizados como referência e as hipóteses estruturais do modelo adotado.

Referências

- Brasil.IO. **COVID-19 Dataset - Casos confirmados e óbitos por município no Brasil**. 2025. <<https://brasil.io/dataset/covid19/files/>>. Acessado em: 6 jun. 2025. Citado na página 51.
- CHENG, R.; GEN, M.; TSUJIMURA, Y. A tutorial survey of job-shop scheduling problems using genetic algorithms—i. representation. **Computers & Industrial Engineering**, v. 30, n. 4, 1996. ISSN 0360-8352. Citado na página 21.
- CHOWDHURY, S.; ROYCHOWDHURY, S.; CHAUDHURI, I. Cellular automata in the light of covid-19. **The European Physical Journal Special Topics**, v. 231, n. 18–20, p. 3619–3628, 2022. Disponível em: <<https://link.springer.com/article/10.1140/epjs/s11734-022-00619-1>>. Citado 4 vezes nas páginas 10, 27, 28 e 30.
- CHU, P. C.; BEASLEY, J. E. A genetic algorithm for the multidimensional knapsack problem. **Journal of Heuristics**, v. 4, n. 1, 1998. ISSN 1381-1231. Citado na página 21.
- COOK, M. Universality in elementary cellular automata. **Complex Systems**, v. 15, n. 1, 2004. ISSN 0891-2513. Citado na página 20.
- DARWIN, C. **On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life**. London: John Murray, 1859. Disponível em: <<https://www.gutenberg.org/ebooks/1228>>. Citado na página 21.
- ELDREDGE, N.; GOULD, S. J. Punctuated equilibria: An alternative to phyletic gradualism. In: SCHOPF, T. J. M. (Ed.). **Models in Paleobiology**. San Francisco: Freeman, Cooper and Company, 1972. p. 82–115. Citado na página 29.
- ESHELMAN, L. J.; SCHAFFER, J. D. Real-coded genetic algorithms and interval-schemata. In: WHITLEY, L. D. (Ed.). **Foundations of Genetic Algorithms 2**. [S.l.]: Morgan Kaufmann, 1993. Citado na página 35.
- FRAGA, L. M. **Estratégias Evolutivas para o Ajuste de Parâmetros de um Modelo Epidemiológico Baseado em Autômatos Celulares Probabilísticos**. Dissertação (Mestrado) — Universidade Federal de Uberlândia, 2021. Citado 2 vezes nas páginas 10 e 29.
- GARDNER, M. Mathematical games: The fantastic combinations of john conway’s new solitaire game “life”. **Scientific American**, v. 223, n. 4, 1970. ISSN 0036-8733. Citado na página 16.
- GRINSTEIN, G.; JAYAPRAKASH, C.; HE, Y. Statistical mechanics of probabilistic cellular automata. **Physical Review Letters**, v. 55, n. 23, 1985. ISSN 0031-9007. Citado na página 19.
- HOLLAND, J. H. Genetic algorithms. **Scientific American**, v. 267, n. 1, 1992. ISSN 0036-8733. Citado na página 21.

KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. **Proceedings of the Royal Society A**, Royal Society, 1927. Citado na página 13.

KUCHARSKI, A. J.; RUSSELL, T. W.; DIAMOND, C.; LIU, Y.; EDMUNDS, J.; FUNK, S. et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. **The Lancet Infectious Diseases**, Elsevier, v. 20, n. 5, 2020. Citado na página 15.

LOZANO, K. K. C. **Autômatos Celulares Probabilísticos com Aplicações a Sistemas Biológicos**. Tese (Tese de Doutorado) — Universidade Federal do Rio de Janeiro, 2017. Citado na página 19.

MITCHELL, M. Computation in cellular automata: A selected review. In: GRAMß, T.; BORNHOLDT, S.; GROß, M.; MITCHELL, M.; PELLIZZARI, T. (Ed.). **Nonstandard Computation**. [S.l.]: VCH Verlagsgesellschaft, 1996. ISBN 978-3527293550. Citado na página 15.

_____. **An Introduction to Genetic Algorithms**. Cambridge, MA: MIT Press, 1996. ISBN 978-0-262-13316-6. Citado 3 vezes nas páginas 4, 23 e 26.

MITCHELL, M.; CRUTCHFIELD, J. P.; DAS, R. The evolution of emergent computation. **Proceedings of the National Academy of Sciences**, National Academy of Sciences, v. 90, n. 23, p. 10742–10746, 1993. Citado na página 28.

MURILO, L. V. **Ajuste evolutivo de um modelo baseado em autômatos celulares para representar a dinâmica de propagação de incêndios florestais**. Dissertação (Dissertação (Mestrado em Ciência da Computação)) — Universidade Federal de Uberlândia, Uberlândia, Brasil, 2025. Defendida em 31 de janeiro de 2025. Disponível em: <<https://repositorio.ufu.br/handle/123456789/44848>>. Citado 2 vezes nas páginas 10 e 28.

NEUMANN, J. von; BURKS, A. W. **Theory of Self-Reproducing Automata**. Urbana, IL: University of Illinois Press, 1966. ISBN 978-0252727339. Citado na página 15.

POTVIN, J.-Y. Genetic algorithms for the traveling salesman problem. **Annals of Operations Research**, v. 63, 1996. ISSN 0254-5330. Citado na página 21.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, American Chemical Society, v. 36, n. 8, 1964. Citado na página 51.

SILVA, G. D. C. **Modelo de propagação da COVID-19 baseado em Autômatos Celulares Probabilísticos**. 2024. Trabalho de Conclusão de Curso (Graduação em Sistemas de Informação), Universidade Federal de Uberlândia, Uberlândia. 43 f.; defesa em 30 de outubro de 2024. Disponível em: <<https://repositorio.ufu.br/handle/123456789/44057>>. Citado 4 vezes nas páginas 10, 28, 30 e 39.

SIRAKOULIS, G. C.; KARAFYLLIDIS, I.; THANAILAKIS, A. A cellular automaton model for the effects of population movement and vaccination on epidemic propagation. **Ecological Modelling**, Elsevier, v. 133, n. 3, 2000. Citado 2 vezes nas páginas 10 e 13.

VICHNIAC, G. Y. Simulating physics with cellular automata. **Physica D: Nonlinear Phenomena**, Elsevier, v. 10, n. 1-2, 1984. ISSN 0167-2789. Citado na página 15.

WALDMAN, E. A.; SATO, A. P. S. Trajetória das doenças infecciosas no brasil nos últimos 50 anos: um contínuo desafio. **Revista de Saúde Pública**, Faculdade de Saúde Pública, USP, v. 50, 2016. Disponível em: <<https://www.scielo.br/j/rsp/a/>>. Citado 2 vezes nas páginas 10 e 11.

WHITE, S. H.; REY, A. M. del; SÁNCHEZ, G. R. Modeling epidemics using cellular automata. **Applied Mathematics and Computation**, v. 186, n. 1, p. 193–202, 2007. Citado na página 27.

WOLFRAM, S. Statistical mechanics of cellular automata. **Reviews of Modern Physics**, v. 55, n. 3, 1983. ISSN 0034-6861. Citado 2 vezes nas páginas 15 e 16.

_____. Cellular automata as simple self-organizing systems. **Physics D: Nonlinear Phenomena**, Elsevier, v. 10, n. 1-2, 1984. Citado na página 10.

YARSKY, P. Using a genetic algorithm to fit parameters of a covid-19 seir model for us states. **Infectious Disease Modelling**, Elsevier, v. 6, p. 51–62, 2021. Disponível em: <<https://doi.org/10.1016/j.idm.2021.01.006>>. Citado 2 vezes nas páginas 10 e 29.

ZHAO, L.; BRESLIN, S. E.; HEARN, A. C.; MURRAY, J. D. Cellular automata approaches to biological modeling. **Journal of Theoretical Biology**, v. 160, n. 1, 1993. ISSN 0022-5193. Citado na página 15.