

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

Paulo Gabriel Nunes Cançado

Ajuste de Parâmetros do Modelo Epidemiológico SIR

São João del-Rei

2022

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

Paulo Gabriel Nunes Cançado

Ajuste de Parâmetros do Modelo Epidemiológico SIR

Monografia apresentada como requisito da disciplina de Projeto Orientado em Computação II do Curso de Bacharelado em Ciência da Computação da UFSJ.

Orientador: Carolina Ribeiro Xavier

Universidade Federal de São João del-Rei — UFSJ

Bacharelado em Ciência da Computação

São João del-Rei

2022

Paulo Gabriel Nunes Cançado

Ajuste de Parâmetros do Modelo Epidemiológico SIR

Monografia apresentada como requisito da disciplina de Projeto Orientado em Computação II do Curso de Bacharelado em Ciência da Computação da UFSJ.

Trabalho aprovado. São João del-Rei, 4 de janeiro de 2022:

Carolina Ribeiro Xavier
Orientadora

Michelli Marlane Silva Loureiro
Avaliadora

Vinícius da Fonseca Vieira
Avaliador

São João del-Rei
2022

Agradecimentos

Primeiramente, agradeço aos meus pais, Edson e Maria Clara, que sempre me apoiaram e proveram fomento na busca dos meus sonhos. Espero um dia conseguir retribuir tudo o que vocês fizeram por mim.

Agradeço a minha irmã, que sempre me proporcionou amparo e apoio emocional durante toda a minha vida. Você é a minha “irmãe”.

Agradeço a todos os professores que tive durante a minha vida. Guardarei com muito carinho todos os ensinamentos adquiridos com vocês.

Agradeço à minha orientadora, Carolina Ribeiro Xavier, por direcionar o desenvolvimento do meu trabalho de pesquisa.

Por último, gostaria de agradecer aos meus amigos. Guardarei com carinho todos os momentos engraçados que vivemos juntos. Sou muito grato por tê-los conhecido.

*Queremos buscar a verdade,
não importa aonde ela nos leve.
Mas para encontrá-la,
precisaremos tanto de imaginação
quanto de ceticismo.
Não teremos medo de fazer especulações,
mas teremos o cuidado
de distinguir a especulação do fato.
(Carl Sagan)*

Resumo

Devido à pandemia causada pelo coronavírus, o estudo de modelos epidemiológicos tornou-se essencial para a tomada de decisão da população e dos governos. Compreender a dinâmica de propagação do vírus na população possibilita elaborar políticas efetivas de combate à epidemia. Pensando nisso, este trabalho tem como objetivo investigar a propagação do coronavírus na cidade de São João del-Rei. Para isso, foi proposto um algoritmo que utiliza o modelo epidemiológico SIR baseado em redes complexas para simular a propagação do vírus na rede. Os resultados de cada simulação são comparados com dados reais, obtidos no site da Secretaria da Saúde do Estado de Minas Gerais. Por último, os parâmetros β e γ do modelo SIR que representam, respectivamente, a probabilidade de infecção e a probabilidade de recuperação, serão ajustados utilizando um algoritmo genético. Após o término do processo de ajuste, alguns cenários hipotéticos de propagação do coronavírus serão simulados.

Palavras-chaves: SIR, Redes Complexas, Algoritmo Genético, COVID-19.

Abstract

Due to the pandemic caused by the coronavirus, the study of epidemiological models has become essential for decision-making by the population and governments. Understanding the dynamics of virus propagation in the population makes it possible to design effective policies to combat the epidemic. With that in mind, this work aims to investigate the propagation of the coronavirus in the city of São João del-Rei. For this, an algorithm that uses the SIR epidemiological model based on complex networks to simulate the propagation of the virus in the network was proposed. The results of each simulation are compared with real data, obtained from the website of the Secretary of Health of the State of Minas Gerais. Finally, the parameters β and γ of the SIR model, which represent, respectively, the probability of infection and the probability of recovery, will be adjusted using a genetic algorithm. After finishing the adjustment process, some hypothetical coronavirus propagation scenarios will be simulated.

Key-words: SIR, Complex Networks, Genetic Algorithm, COVID-19.

Lista de ilustrações

Figura 1 – Fluxograma usual do algoritmo genético	17
Figura 2 – Série temporal do número de infectados	25
Figura 3 – Série temporal do número de removidos	25
Figura 4 – Simulação utilizando a rede original	26
Figura 5 – Simulação de relaxamento do distanciamento social	26
Figura 6 – Simulação do isolamento horizontal	26
Figura 7 – Propagação do coronavírus em uma população parcialmente imunizada (10%) utilizando a vacinação aleatória	27
Figura 8 – Propagação do coronavírus em uma população parcialmente imunizada (10%) utilizando a vacinação de <i>hubs</i>	27
Figura 9 – Propagação do coronavírus em uma população parcialmente imunizada (30%) utilizando a vacinação aleatória	28
Figura 10 – Propagação do coronavírus em uma população parcialmente imunizada (30%) utilizando a vacinação de <i>hubs</i>	28
Figura 11 – Ajuste do período 01/01/2021 até 01/05/2021	29
Figura 12 – Ajuste do período 18/05/2021 até 15/09/2021	29

Lista de tabelas

Tabela 1 – Dados obtidos dos boletins epidemiológicos	20
Tabela 2 – Dados após as manipulações	21
Tabela 3 – Comparação entre as simulações	24

Lista de abreviaturas e siglas

OMS	Organização Mundial da Saúde
EQM	Erro Quadrático Médio
SIR	Modelo Suscetível/Infectado/Recuperado
AG	Algoritmo Genético
I	Infectantes Ativos
IA	Número de Infectados Acumulados
RE	Número de Recuperados
OA	Número de Óbitos Acumulados
R	Número de Removidos

Lista de símbolos

β	Probabilidade de infecção diária do modelo SIR
γ	Probabilidade de recuperação diária do modelo SIR

Sumário

1	Introdução	12
2	Referencial Teórico	14
2.1	Modelo SIR	14
2.2	SIR Baseado em Redes Complexas	15
2.3	Redes	15
2.3.1	Grafos Aleatórios	15
2.3.2	Erdős-Rényi	15
2.4	Algoritmo Genético	16
2.5	Erro Quadrático Médio	18
2.6	Isolamento Horizontal	18
3	Trabalhos Relacionados	19
4	Metodologia	20
4.1	Dados	20
4.2	Redes	21
4.3	Função Objetivo	21
4.4	Algoritmo Genético	22
5	Resultados	24
5.1	Ajuste	24
5.2	Cenários Simulados	25
5.2.1	Distanciamento Social	25
5.2.2	Vacinação	27
5.3	Limitações	28
6	Considerações Finais	30
6.1	Considerações Gerais	30
6.2	Trabalhos Futuros	30
	Referências	31

1 Introdução

A COVID-19 é uma doença infecciosa causada pelo vírus SARS-CoV-2. Sabe-se que cerca de 80% das infecções são leves ou assintomáticas, 15% são graves e 5% são críticas [1]. As infecções graves e críticas provocam a deficiência de oxigênio no organismo do indivíduo, podendo levar a sua morte caso nenhum tratamento médico seja realizado.

O primeiro caso diagnosticado de infecção humana pelo coronavírus ocorreu em dezembro de 2019 na cidade de Wuhan, localizada na China [2]. No dia 11 de março de 2020, a Organização Mundial da Saúde (OMS) elevou o estado de contaminação da COVID-19 à pandemia, indicando que a propagação do vírus estava presente em quase todos os continentes.

De fato, os impactos causados até o momento são assustadores. Estima-se que houveram cerca de 264 milhões de infecções pelo coronavírus no mundo. Infelizmente, cerca de 5,2 milhões de pessoas perderam suas vidas em detrimento desse vírus. O Brasil é o terceiro país com maior número de infecções. Estima-se que ocorreram por volta de 22 milhões de infecções, que ocasionaram o óbito de aproximadamente 615 mil pessoas [3].

Nesse contexto, para entender sobre a propagação de uma doença infecciosa como o COVID-19, é necessário compreender a sua dinâmica de contágio na população alvo, tal tarefa pode ser feita através de um modelo epidemiológico. Assim, o modelo SIR (Suscetível/Infetado/Recuperado) é o mais apropriado para representar essa dinâmica.

A transmissão do coronavírus se dá principalmente por gotículas de saliva e/ou secreções expelidas pelos indivíduos contaminados. Dessa forma, as pessoas que mantêm contato direto com o infectado estão mais propensas a contrair o vírus. Nesse contexto, o modelo SIR baseado em redes complexas [4, 5] poderia ser mais compatível ao problema do que o modelo epidemiológico original, caso a rede de contatos entre as pessoas estivesse disponível. Isso se deve à utilização de uma rede complexa nesse modelo para simular as relações interpessoais dos indivíduos da população, proporcionando simulações de espalhamento mais realistas e introduzindo heterogeneidade no processo de infecção.

Conforme citado anteriormente, cada doença transmissível possui uma dinâmica própria de infecção e de recuperação. Consequentemente, é necessário ajustar individualmente os seus parâmetros probabilísticos, que representam a taxa de infecção e de recuperação dos indivíduos de uma determinada população. Por se tratar de uma busca entre infinitos números entre zero e um, é interessante utilizar alguma meta-heurística. Uma abordagem possível para ajustá-los é por meio de algoritmo genético. Essa técnica de busca permite encontrar soluções aproximadas para o problema, podendo ser aplicada em diferentes contextos.

Portanto, o objetivo deste trabalho é investigar a propagação do coronavírus na cidade de São João del-Rei. Foi proposto um protótipo de algoritmo que efetua simulações da propagação do coronavírus na população com o intuito de encontrar as taxas de infecção e de recuperação para a cidade de São João del-Rei. A propagação do vírus na população foi simulada por meio do modelo epidemiológico SIR, baseado em redes complexas. O resultado de cada simulação é comparado com dados reais, possibilitando mensurar o grau de semelhança da simulação com a realidade. Em suma, a ideia principal é encontrar os valores dos parâmetros β e γ que produzem simulações com a mesma tendência observada nos dados. Em virtude disso, a função do Algoritmo Genético é ajustar o valor desses parâmetros. Após o processo de ajuste, alguns cenários de propagação do vírus são investigados.

O restante do texto está organizado da forma que segue. O Capítulo 2 apresenta o referencial teórico, elucidando alguns conceitos necessários para o entendimento da proposta. No capítulo 3, estão os trabalhos utilizados no embasamento do projeto. Já, no Capítulo 4, são relatados os procedimentos e os métodos adotados no desenvolvimento do projeto. Os resultados obtidos são expostos no Capítulo 5, e as considerações finais serão descritas no Capítulo 6.

2 Referencial Teórico

2.1 Modelo SIR

SIR (Suscetível/Infectado/Recuperado) é um modelo epidemiológico compartimental criado por Kermack e McKendrick [6] em 1927. Nesse modelo, a população é dividida em três compartimentos que representam os possíveis estágios da doença. Matematicamente, o modelo pode ser descrito pelas equações diferenciais 2.1, 2.2 e 2.3,

$$\frac{\partial S}{\partial t} = -\beta S \frac{I}{N}, \quad (2.1)$$

$$\frac{\partial I}{\partial t} = +\beta S \frac{I}{N} - \gamma I \quad (2.2)$$

$$\frac{\partial R}{\partial t} = +\gamma I \quad (2.3)$$

em que S, I e R representam respectivamente a quantidade de indivíduos suscetíveis, infectados e recuperados. Além disso, a população total é representada pela equação 2.4.

$$N = S + I + R \quad (2.4)$$

O parâmetro β representa a taxa probabilística de infecção. Já o parâmetro γ refere-se à taxa de recuperação. Esses dois parâmetros possuem valores diferentes para cada doença, sendo necessário ajustá-los.

Por último, os parâmetros β e γ possibilitam calcular o número de reprodução básica (R_0). Ele é definido como “o número médio de infecções secundárias geradas por um típico indivíduo infectado durante todo o seu período infeccioso quando introduzido em uma população completamente suscetível” [7] e pode ser calculado pela equação 2.5.

$$R_0 = \frac{\beta}{\gamma} \quad (2.5)$$

2.2 SIR Baseado em Redes Complexas

O modelo SIR baseado em redes complexas é uma adaptação do modelo SIR original que tem como objetivo representar a natureza probabilística da propagação da doença na população [8].

A propagação do vírus é simulada em uma rede complexa, que visa representar as interações interpessoais dos indivíduos da população. Isso incorpora heterogeneidade ao sistema, possibilitando que a probabilidade de infecção seja diferente para cada indivíduo.

Nesse modelo, assim como no modelo clássico anteriormente apresentado, existem dois parâmetros probabilísticos. Eles são comumente denominados como β e γ . O parâmetro β é a probabilidade de um nó infectado contaminar outro dado que existe uma aresta entre eles. Já o parâmetro γ é a probabilidade de um nó infectado se recuperar. Como a menor porção de tempo simulada neste trabalho é de um dia, os parâmetros encontrados se referem a probabilidades diárias de infecção e recuperação.

2.3 Redes

“Uma rede é, em sua forma mais simples, uma coleção de pontos unidos em pares por linhas” [9]. Comumente, “os pontos são referidos como vértices ou nós e as linhas são referidas como arestas” [9].

2.3.1 Grafos Aleatórios

Em síntese, “um grafo aleatório é um modelo de rede em que algum conjunto específico de parâmetros toma valores, mas a rede é aleatória em outros aspectos” [9]. Em outras palavras, grafos aleatórios são estruturas que se baseiam principalmente da probabilidade para sua criação.

A principal diferença entre os modelos de grafos aleatórios está nos parâmetros definidos *a priori*. Em certos modelos, por exemplo, define-se a quantidade de vértices e arestas existentes no grafo. Noutros, é estabelecido o número de vértices e a probabilidade de criação de arestas entre cada par de vértices.

2.3.2 Erdős–Rényi

Erdős–Rényi é um modelo proposto por Paul Erdős e Alfréd Rényi [10] para a criação de redes aleatórias. Segundo Satorras *et al* [11], um “grafo $G_p(N)$ é construído a partir de um conjunto de n vértices, em que as $n(n - 1)/2$ possíveis ligações entre esses vértices ocorrem conforme uma probabilidade p ”.

Embora o número de arestas nesse modelo não seja fixo, é possível calcular o valor esperado de arestas no grafo [9]. Assim, essa quantidade de arestas pode ser calculada por meio da equação 2.6,

$$\langle m \rangle = \sum_{m=0}^{\binom{n}{2}} mp(m) = \binom{n}{2} p, \quad (2.6)$$

em que m , n e p correspondem respectivamente ao número de arestas, ao número de vértices e à probabilidade de criação de uma aresta entre dois vértices.

Por esse modelo ser frequentemente utilizado para representar as interações humanas, é relevante ser capaz de determinar a sua distribuição de grau. Ela pode ser calculada pela equação 2.7.

$$P_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} \quad (2.7)$$

Vale-se ressaltar que quando $n \rightarrow \infty$ a distribuição de graus pode ser calculada pela distribuição de Poisson[9], conforme a equação 2.8.

$$P_k = \frac{(n-1)^k}{k!} p^k e^{-c} = \frac{(n-1)^k}{k!} \left(\frac{c}{n-1} \right)^k e^{-c} = e^{-c} \frac{c^k}{k!}, \quad (2.8)$$

2.4 Algoritmo Genético

Os algoritmos evolutivos são heurísticas que se baseiam em mecanismos biológicos para resolução de problemas [12]. Dentro dessa área, destacam-se os algoritmos genéticos, que são técnicas de busca e otimização, baseadas no processo de seleção natural proposto por Charles Darwin e Alfred Wallace [13]. É uma heurística extremamente versátil, possibilitando encontrar boas soluções para problemas complexos.

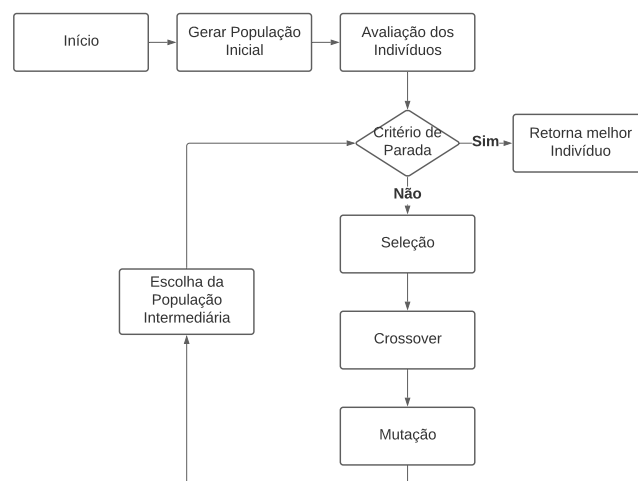
Por se tratar de uma abordagem multidisciplinar, diversos termos pertinentes à biologia foram adaptados. Os termos empregados nos algoritmos genéticos significam:

- **Gene:** Características que compõem um indivíduo. Normalmente, elas assumem valores binários ou inteiros positivos.
- **Cromossomo ou Indivíduo:** Solução candidata formada por um conjunto de genes.
- **População:** Conjunto de soluções candidatas para o problema.

- **Avaliação:** Cada indivíduo da população é ponderado conforme um critério de avaliação (função objetivo). É de responsabilidade da função objetivo mensurar a qualidade da solução em forma de uma pontuação.
- **Seleção dos Pais:** Dois indivíduos da população são selecionados para terem seus genes cruzados. A probabilidade de escolha do indivíduo normalmente é proporcional a sua pontuação (fitness). Algumas das técnicas utilizadas são roleta viciada, torneio ou totalmente aleatório.
- **Crossover ou cruzamento:** Os genes dos dois pais selecionados são combinados gerando uma nova população. Uma das técnicas de *crossover* é o de pontos. Nesta técnica, é sorteado aleatoriamente um ou mais pontos de corte para cada par de pais. Em seguida, os genes dos dois pais são combinados, alternando-se a sequência de genes entre os pontos de corte.
- **Mutação:** Consiste em realizar pequenas mudanças em genes aleatórios com o intuito de explorar novas soluções. Esse processo evita que o algoritmo fique preso em soluções sub-ótimas, ou ótimos locais, que impedem que as melhores soluções sejam encontradas pelo algoritmo.

Na teoria dos algoritmos genéticos parte-se do pressuposto de que os indivíduos mais adaptados tem mais chance de sobreviver. Uma nova população é formada com base nos indivíduos potencialmente mais aptos da população anterior. Consequentemente, as soluções encontradas pelo algoritmo tendem a melhorar a cada nova geração de indivíduos. O fluxo de execução de um algoritmo genético é exemplificado na figura 1.

Figura 1 – Fluxograma usual do algoritmo genético



Fonte: Autor.

2.5 Erro Quadrático Médio

O erro quadrático médio pode ser utilizado para medir a qualidade de uma solução. Ele é definido como sendo a média da diferença entre o valor do estimador e do valor real ao quadrado. Sendo T um estimador qualquer (indivíduos da AG, neste trabalho), e θ o valor real, temos que o erro quadrático médio pode ser calculado pela equação 2.9,

$$EQM(\hat{Y}) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2, \quad (2.9)$$

em que Y são os valores observados e \hat{Y} os valores estimados [14].

2.6 Isolamento Horizontal

O isolamento horizontal é uma medida que visa restringir ao máximo o contato entre as pessoas. “Apesar de causar impactos graves na economia, é essencial para evitar um aumento desenfreado da doença e o colapso no sistema de saúde, o que também causaria danos à economia.” [15]

3 Trabalhos Relacionados

A modelagem de epidemias é uma das áreas que vem se destacando nas últimas décadas devido a sua capacidade de representar a dinâmica de certos eventos que acontecem no mundo real. Ao contrário do que muitos imaginam, a modelagem de epidemias não se restringe a doenças endêmicas. Rumores [16] e “malwares” [17] apresentam uma difusão semelhante à de doenças infectocontagiosas em redes complexas.

Em geral, o modelo SIR é o mais indicado para modelagem de doenças que concedem imunidade permanente aos infectados. Em especial, os modelos SIR baseados em redes complexas levam em consideração as interações dos indivíduos de uma população no processo de propagação de uma doença, possibilitando representar com maior semelhança o comportamento de vírus real e, conseqüentemente, entender melhor o comportamento dessa doença [5].

No entanto, pode-se ressaltar que o modelo tradicional SIR não leva em consideração o período de incubação. Esse período influencia diretamente na propagação de uma doença em uma determinada população. Em virtude disso, alguns autores propuseram adaptações ao modelo SIR, embutindo um atraso no processo de propagação [18, 19, 20], possibilitando entender melhor o equilíbrio epidêmico de certas doenças.

É válido ressaltar que o novo coronavírus está fomentando diversas pesquisas na área de modelagem de epidemias. Alguns estudos baseados no modelo SIR em redes complexas foram capazes de fazer previsões assertivas para diferentes cenários de isolamento social, exibindo as possíveis quantidades de infecções, mortes e ocupação de leitos para cada cenário [5]. Outros, demonstraram como a migração pendular corroborou na propagação da COVID-19 na região da Lombardia [21]. Por fim, algumas pesquisas evidenciam a eficiência do isolamento social para contenção do coronavírus [22].

Semelhantemente aos trabalhos descritos acima, este trabalho tem o intuito de ajustar os parâmetros do modelo SIR para o novo coronavírus, utilizando-se o modelo baseado em redes complexas. No entanto, será utilizado algoritmo genético para ajuste desses parâmetros, efetuando simulações e comparando-as com os dados reais.

4 Metodologia

4.1 Dados

Para ajustar os parâmetros β e γ é preciso conhecer a quantidade de indivíduos infectantes (transmitindo a doença) e recuperados existente no intervalo de tempo que se deseja ajustar. Os dados utilizados foram retirados de boletins epidemiológicos disponibilizados no site da Secretaria da Saúde do Estado de Minas Gerais. Nesses boletins, são divulgados os números diários de novas infecções, mortes e a quantidade atual de indivíduos recuperados. A tabela 1 apresenta um exemplo desses dados.

Tabela 1 – Dados obtidos dos boletins epidemiológicos

Data	Confirmados	Óbitos	Recuperados
24/02/2021	8	0	4
25/02/2021	22	0	9
26/02/2021	7	0	13
27/02/2021	12	0	23
28/02/2021	11	3	40

Fonte dos Dados: Secretaria da Saúde do Estado de Minas Gerais.

Após realizar algumas manipulações, pode-se obter a quantidade de indivíduos infectantes ativos e recuperados que existe no período de tempo investigado. Assim, a quantidade de indivíduos infectantes ativos em um determinado período pode ser determinada pela equação 4.1,

$$I_d = IA_d - RE_d - OA_d, \quad (4.1)$$

em que IA é a quantidade total de casos confirmados, RE é a quantidade de pessoas que se recuperaram do coronavírus, OA é o montante de óbitos ocorridos e d é a data.

Além disso, no modelo SIR são considerados como indivíduos recuperados (removidos) todos aqueles que não podem mais contrair a doença, como, por exemplo, os curados e os mortos por ela. Dessa forma, o número de recuperados existentes em um intervalo pode ser obtido por meio da equação 4.2,

$$R_d = RE_d + OA_d, \quad (4.2)$$

em que RE é a quantidade de pessoas que se recuperaram do coronavírus, OA é o montante de óbitos ocorridos e d é a data.

Manipulando os dados da Tabela 1, conforme as equações descritas acima, obtém-se a Tabela 2. Os dados devem estar neste formato para o funcionamento correto do algoritmo.

Tabela 2 – Dados após as manipulações

Data	Infec. Acumul.(IA)	Óbitos Acumul. (OA)	Infec. Ativos (I)	Removidos (R)
24/02/2021	8	0	4	2834
25/02/2021	30	0	21	2842
26/02/2021	37	0	24	2846
27/02/2021	49	0	26	2851
28/02/2021	60	3	17	2855

Fonte: Autor.

4.2 Redes

As redes têm papel fundamental no modelo SIR baseado em redes complexas. São elas as responsáveis por trazer verossimilhança com a realidade, proporcionando heterogeneidade no processo de infecção na simulação. Seu objetivo é tentar representar a dinâmica de interações entre os indivíduos de uma população. Quanto maior for a sua proximidade com a realidade, mais fácil se torna compreender a difusão de uma doença na comunidade analisada. Nesse contexto, esse projeto investigará a propagação do coronavírus no modelo de rede Erdős-Rényi.

Erdős-Rényi é um modelo de rede no qual as interações entre os nós são geradas utilizando aleatoriedade. Para criar esse modelo de rede, é necessário definir a quantidade de nós e a probabilidade de criação de aresta entre dois nós. Como os dados utilizados são referentes aos infectados e recuperados da região de São João del-Rei, o número de nós do grafo gerado deve ser equivalente ao tamanho da população dessa cidade, algo em torno de 90.497 nós. Durante os testes do algoritmo proposto, constatou-se que a probabilidade de criação de aresta deve ser próxima de 0.00007, gerando uma rede na qual o grau médio dos nós é de aproximadamente 6,3 conexões. Valores maiores de probabilidade de criação de aresta formam redes mais densas, que aceleram o processo de difusão da doença na rede. Valores menores produzem redes esparsas, cujas simulações podem terminar precocemente, devido à falta de arestas para propagar a doença.

4.3 Função Objetivo

Em suma, o propósito do algoritmo desenvolvido é encontrar quais valores de β e γ geram simulações de propagação do vírus na rede que seguem a mesma tendência

observada nos dados reais. Ou seja, o algoritmo visa a minimizar o erro entre os resultados da simulação e os dados reais. A função objetivo do algoritmo é definida pela equação 4.3,

$$\min f(s) = \frac{1}{2n} \sum_{i=1}^n (IR_i - IS_i)^2 + \frac{1}{2n} \sum_{i=1}^n (RR_i - RS_i)^2 \quad (4.3)$$

em que IR e RR são, respectivamente, os números de infectados e recuperados observados nos dados. Já as variáveis IS e RS representam, respectivamente, os números de infectados e recuperados ocorridos na simulação. A variável n , na equação, indica a quantidade de dias simulados.

Cada solução candidata simulada recebe uma nota, que será utilizada pelo algoritmo genético no processo de seleção dos indivíduos. Essa nota é dada pela média do EQM do número de infectados e recuperados da simulação em comparação aos dados reais. Quanto menor for o valor dessa nota, maior é a proximidade entre os dados entregues pela simulação e a realidade. A utilização do erro quadrático médio na equação 4.3 tem o intuito de refletir as disparidades nos valores encontrados na simulação. Isso estimula o algoritmo a escolher pelas simulações que seguem a tendência dos dados reais durante todo o período simulado.

4.4 Algoritmo Genético

Algoritmo Genético é uma técnica de busca utilizada para encontrar soluções aproximadas para problemas complexos. Será essa técnica a responsável por encontrar e ajustar os valores β e γ utilizados nas simulações.

Primeiramente, cria-se uma população inicial que será o ponto de partida para o processo de ajuste dos parâmetros. Cada indivíduo dessa população é uma solução candidata. Para o problema em questão, os indivíduos serão compostos por três atributos, sendo eles os parâmetros β , γ e a nota atribuída a simulação. Além disso, esses parâmetros estão na forma de cromossomo binário e são inicializados aleatoriamente. É possível determinar o tamanho do cromossomo binário. No entanto, 10 bits se mostraram suficientes para o problema em foco.

Os valores de β e γ de cada indivíduo da população inicial são utilizados como probabilidade diária de infecção e de recuperação nas simulações. Para isso, a representação binária desses parâmetros é convertida para decimal conforme a equação 4.4. Em seguida, esse valor em decimal é mapeado para um valor numérico entre 0 e 1, que é efetuado de acordo com a equação 4.5. Em ambas as equações, a variável n refere-se a quantidade de bits do cromossomo. Ademais, a variável x_0 nessa equação é o valor em decimal a ser

convertido.

$$INT(bin) = \sum_{i=0}^{n-1} 2^i * bin[i] \quad (4.4)$$

$$x = \frac{1}{2^n - 1} * x_0 \quad (4.5)$$

São selecionados na rede utilizada dois conjuntos de nós, incumbidos de serem os nós infectados e recuperados iniciais da simulação. Após inicializá-los, efetua-se a simulação e sua nota é calculada conforme a equação 4.3. É válido ressaltar que todas as simulações utilizam o mesmo conjunto inicial de nós infectados e recuperados, possibilitando que haja igualdade nesse processo.

Em cada geração, os indivíduos da população atual são selecionados aos pares para dar origem a uma nova população, filhos da geração atual. A escolha desses pais é feita utilizando a técnica de roleta viciada, em que a probabilidade de escolha do indivíduo é inversamente proporcional a sua nota. Ou seja, quanto menor o EMQ, maior a chance do indivíduo de ser selecionado como pai. De posse desse par de pais, a nova população será concebida combinando os seus cromossomos β e γ . A estratégia de recombinação utilizada foi o *crossover* em um único ponto. Imediatamente, aplica-se o processo de mutação nos cromossomos dos indivíduos gerados, com o intuito de explorar novas soluções do espaço de busca e fugir de ótimos locais. Valores de mutação equivalentes a 5% se mostraram satisfatórios nos testes empíricos. A população atual é substituída por essa nova população e os processos descritos acima serão aplicados novamente.

Desse modo, a ideia geral do AG é gerar novas populações baseadas em indivíduos aptos da população anterior. Esse procedimento acontece até que se tenha alcançado o número de gerações estabelecidas inicialmente. Pode-se ressaltar que o número de indivíduos por população pode ser definido. Assim, o processo de *crossover* tenta proporcionar novos indivíduos entre os bons da geração anterior, tentando realizar melhoria nos mesmos. E a mutação permite que as soluções não fiquem presas em ótimos locais. Portanto, as soluções encontradas para o problema tendem a melhorar ao longo de novas gerações, ajustando os valores de β e γ e, conseqüentemente, minimizando o erro encontrado nas simulações.

5 Resultados

5.1 Ajuste

Por representar a probabilidade de infecção diária, o parâmetro β é extremamente sensível às medidas de combate ao coronavírus, tais como o distanciamento social. Assim, torna-se interessante escolher, como período de ajuste, um intervalo de tempo no qual essas políticas de contenção permaneceram estáveis em seu decorrer. Em virtude disso, foi estabelecido o intervalo de tempo em que a cidade de São João del-Rei esteve na Onda Roxa para realização do ajuste. O algoritmo implementado, os dados utilizados e os resultados obtidos podem ser acessados no GitHub¹.

Foram realizados cinco ajustes para o período estabelecido. Para cada um deles, uma rede Erdős-Rényi de tamanho 90.479 foi criada, utilizando uma probabilidade de criação de aresta entre dois nós equivalente a 0,00007. Além disso, em cada execução foi utilizado um conjunto diferente de nós infectados e recuperados iniciais. Nelas, foram criadas 300 gerações de 20 indivíduos cada, empregando 0.05 como taxa de mutação. O tamanho escolhido para os cromossomos binários foi de dez bits. A tabela 3 exibe os resultados obtidos nesses ajustes.

Tabela 3 – Comparação entre as simulações

Execução	β	γ	EQM Infectados	EQM Recuperados	Tempo de Execução
01	0.01075	0.03030	2473.87	2990.90	6h59 min
02	0.01075	0.03714	925.78	653.58	7h43 min
03	0.01173	0.03910	1148.12	512.36	7h43 min
04	0.01173	0.04007	934.04	461.58	6h52 min
05	0.01173	0.03910	1116.24	515.68	6h57 min

Fonte: Autor.

É possível observar que o processo de ajuste ocorreu muito bem para o período estipulado. É interessante ressaltar que, mesmo utilizando diferentes conjuntos iniciais de nós e redes, os valores encontrados para β e γ são bem próximos nos ajustes. Isso reforça a ideia de que os valores ajustados são, de fato, as probabilidades de infecção e recuperação do período analisado.

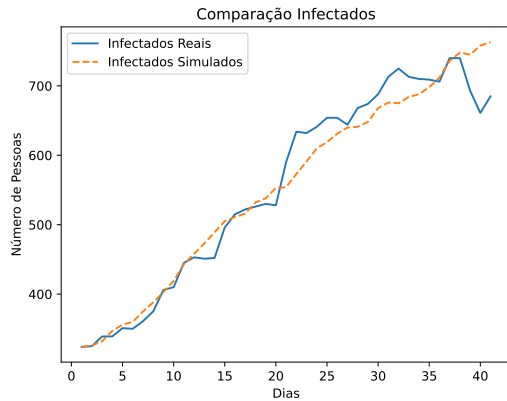
Além disso, a execução 04 obteve a maior performance. Os valores encontrados por ela foram de $\beta = 0.01173$ e $\gamma = 0.04007$. O erro entre as curvas de infectados e recuperados é relativamente baixo, tendo em vista que a métrica utilizada foi o EQM.

¹ <https://github.com/Paulognunes/Ajuste-de-Parametros-do-modelo-SIR.git>

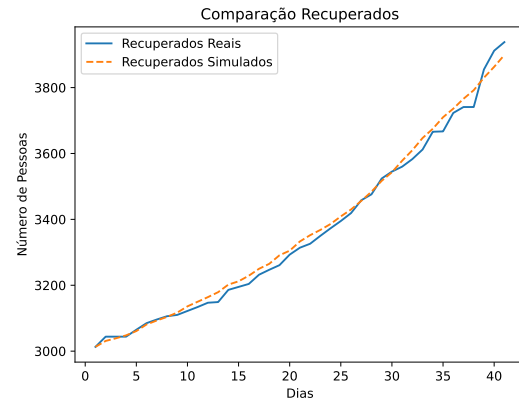
Por fim, as figuras 2 e 3 são séries temporais que comparam os números de infectados e recuperados da simulação em relação aos dados reais.

Figura 2 – Série temporal do número de infectados

Figura 3 – Série temporal do número de removidos



Fonte: Autor.



Fonte: Autor.

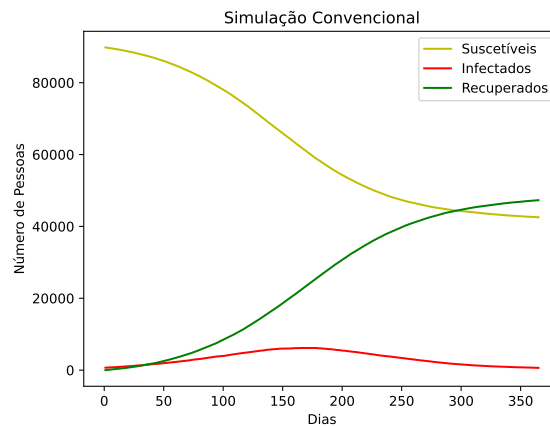
5.2 Cenários Simulados

De posse dos parâmetros ajustados, é possível simular alguns cenários hipotéticos, possibilitando compreender melhor a dinâmica de propagação do vírus mediante suas medidas de contenção. Os cenários simulados utilizam a rede e os parâmetros encontrados pela execução 04, que é o ajuste com a menor média de EQM. Foram inicializados 685 indivíduos infectados nessa rede, que correspondem ao número de infectantes ativos após o fim da Onda Roxa na cidade de São João del-Rei. O número de recuperados iniciais pode variar de acordo com a proposta do cenário.

5.2.1 Distanciamento Social

O distanciamento social refere-se a um conjunto de medidas adotadas com o objetivo de desacelerar a propagação do vírus ao restringir o grau de interação dos indivíduos da população. Uma situação de aumento do distanciamento social, como o isolamento horizontal, pode ser simulado ao remover uma parcela das ligações da rede utilizada. Analogamente, para simular um relaxamento dessa medida de controle, basta acrescentar uma porcentagem de arestas na rede. Assim, esse cenário tem como intuito investigar a efetividade do distanciamento social na contenção inicial da epidemia. Isto é, no período em que toda a população está suscetível ao vírus.

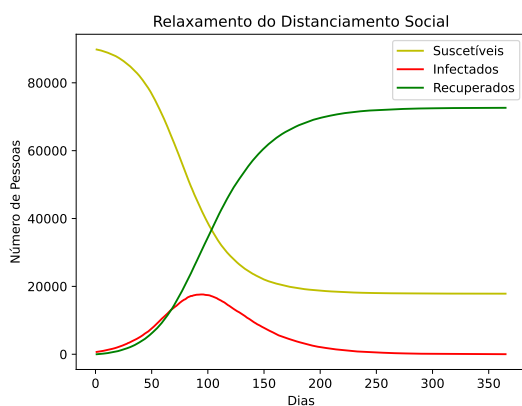
Figura 4 – Simulação utilizando a rede original



Fonte: Autor.

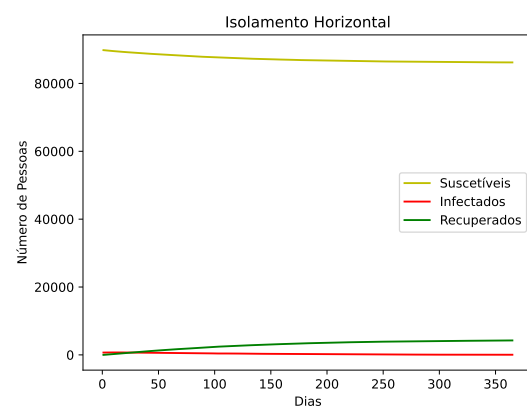
Para isso, três simulações de 365 dias foram efetuadas, iniciando-se no primeiro dia após o término da Onda Roxa. A primeira utiliza a rede original, sem efetuar qualquer alteração. Já, na segunda, houve um acréscimo de 40% de arestas na rede, replicando um relaxamento nas medidas de distanciamento social. Na terceira, 40% das arestas da rede foram removidas, representando um quadro de isolamento horizontal. Todas elas utilizaram o mesmo conjunto inicial de nós infectados. Como o objetivo desse cenário é apurar a propagação do vírus no período inicial da pandemia, nenhum nó recuperado foi inicializado na rede. As figuras 4, 5 e 6 descrevem as séries temporais resultantes das simulações.

Figura 5 – Simulação de relaxamento do distanciamento social



Fonte: Autor.

Figura 6 – Simulação do isolamento horizontal



Fonte: Autor.

Como era de se esperar, os resultados obtidos demonstraram a efetividade do isolamento social para retardar a propagação do coronavírus, aplicando o famoso “achateamento

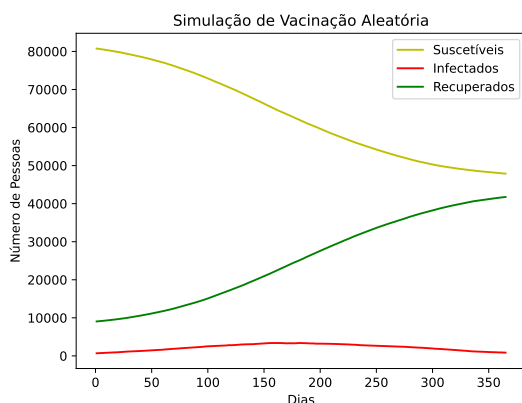
de curva”. Nessas simulações, pôde-se observar uma enorme diferença no número máximo de infectantes ativos e no instante de tempo que isso ocorreu. A primeira simulação apresentou o pico de infectantes no 175º dia, sendo este de 6.199 indivíduos. Já, na segunda, que corresponde ao cenário de relaxamento do distanciamento social, o pico de infectantes ocorreu precocemente no intervalo simulado, ocorrendo no 95º dia e alcançando cerca de 17.621 indivíduos. Por último, o pico de infectantes do cenário de isolamento horizontal ocorreu no 8º dia, atingindo apenas 722 indivíduos.

5.2.2 Vacinação

Sem dúvida, a vacinação é o mecanismo mais eficiente de combate à transmissão do coronavírus na população. Sua função é estimular o corpo a produzir respostas imunológicas, preparando-o para o combate de certas doenças. Vacinar indivíduos de uma população dificulta a propagação do agente infeccioso e, conseqüentemente, acelera o processo de imunidade coletiva.

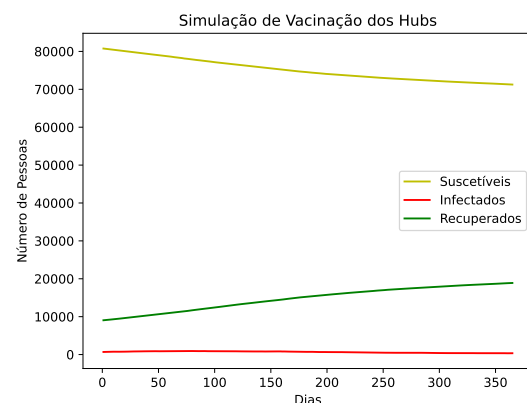
Neste cenário, é examinada a eficácia da vacinação para conter a difusão do coronavírus. Foram efetuadas quatro simulações de 365 dias. Em duas delas é investigado a propagação do vírus em uma população parcialmente imunizada, utilizando-se a estratégia de vacinação aleatória. Nas outras duas, é averiguado a disseminação do vírus em uma população onde os nós com mais interações, também chamados de *hubs*, foram vacinados. Foi suposto que a imunização é uma passagem forçada do estágio $S \rightarrow R$ no modelo SIR, concedendo 100% de imunidade aos indivíduos vacinados. As porcentagens de vacinados na população serão, respectivamente, de 10% para as simulações 1 e 2 e de 30% para as simulações 3 e 4.

Figura 7 – Propagação do coronavírus em uma população parcialmente imunizada (10%) utilizando a vacinação aleatória



Fonte: Autor.

Figura 8 – Propagação do coronavírus em uma população parcialmente imunizada (10%) utilizando a vacinação de *hubs*

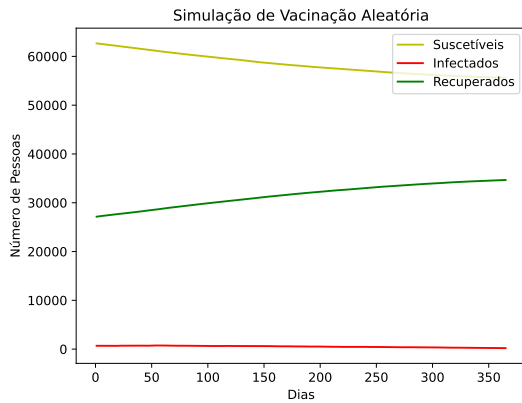


Fonte: Autor.

Comparando-se as figuras 7 e 8 com a figura 4, que é um cenário onde não há

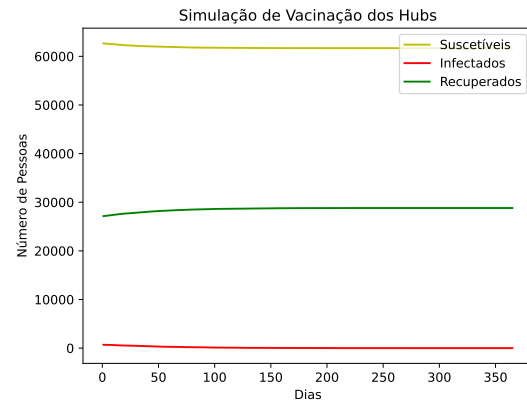
nenhum indivíduo vacinado, constata-se que a vacinação de 10% da população foi capaz de reduzir o avanço do vírus. Neste cenário de vacinação aleatória, houve um pico de 3.427 infectantes ativos, ocorrendo no 161º dia da simulação. Já a estratégia de vacinação dos *hubs* apresentou um resultado mais promissor. O pico de infectantes foi de apenas 961 indivíduos, que aconteceu no 73º dia dessa simulação.

Figura 9 – Propagação do coronavírus em uma população parcialmente imunizada (30%) utilizando a vacinação aleatória



Fonte: Autor.

Figura 10 – Propagação do coronavírus em uma população parcialmente imunizada (30%) utilizando a vacinação de *hubs*



Fonte: Autor.

Conforme pode ser observado nas figuras 9 e 10, a vacinação de 30% da população conteve drasticamente a propagação do vírus. Neste cenário de vacinação aleatória, o pico de infecções foi de 743 indivíduos, sucedendo no 73º dia de simulação. No cenário de vacinação dos *hubs*, houve um pico de 691, decorrendo no 3º dia da simulação. Além disso, a partir do 263º dia, o número de infectados ativos caiu para zero, mantendo-se assim até o término da simulação deste cenário.

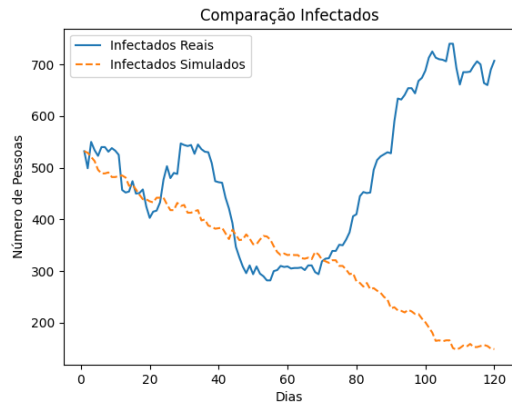
É interessante salientar que, embora a estratégia de vacinação dos *hubs* seja mais eficiente que a vacinação aleatória, ela é pouco factível no mundo real porque seria necessário conhecer toda a estrutura da rede para aplicá-la [11].

5.3 Limitações

Durante o processo de teste do algoritmo implementado, pode-se constatar que ele apresenta dificuldade de ajustar grandes lapsos de tempo, geralmente superiores a 120 dias. Além disso, períodos nos quais ocorrem inversões no sentido da curva de infectados se mostraram inviáveis para o processo de ajuste com este modelo. Ambas as limitações ocorrem porque, provavelmente, o valor real de β modificou-se ao longo do período observado. Assim, o algoritmo não consegue encontrar um valor de β que generalize a taxa de

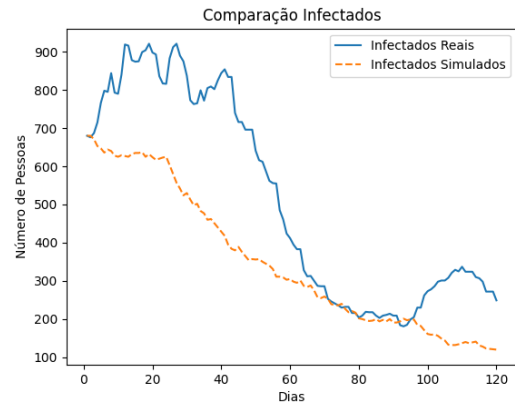
infecção de todo o período. As figuras 11 e 12 ilustram o resultado do ajuste de β para os cenários descritos acima.

Figura 11 – Ajuste do período 01/01/2021 até 01/05/2021



Fonte: Autor.

Figura 12 – Ajuste do período 18/05/2021 até 15/09/2021



Fonte: Autor.

Desse modo, foi observado que o tamanho da rede influencia no tempo de execução do algoritmo. Quanto maior a quantidade de nós da rede, maior será o tempo gasto para realizar cada simulação. Isso explica o alto tempo de execução de cada simulação da tabela 3, já que as redes utilizadas possuem 90.479 nós, o que corresponde ao tamanho da população de São João del-Rei.

6 Considerações Finais

6.1 Considerações Gerais

O objetivo deste trabalho foi investigar a propagação do coronavírus na cidade de São João del-Rei. Para isso, foi desenvolvido um algoritmo que utiliza o modelo SIR baseado em redes complexas para simular a propagação do vírus na população. Cada simulação efetuada tem os seus resultados confrontados com os dados de boletins epidemiológicos da cidade, possibilitando mensurar a sua distância da realidade. Por último, os parâmetros β e γ do modelo SIR foram ajustados por meio de um algoritmo genético.

O algoritmo implementado se mostrou satisfatório para a realização do processo de ajuste. O erro encontrado na maioria das simulações é relativamente pequeno, tendo em vista que a métrica utilizada para mensurar foi do EQM. Além disso, após a realização do ajuste, foi possível simular alguns cenários hipotéticos, possibilitando investigar a efetividade das medidas de combate ao coronavírus. Os resultados desses cenários confirmam o que diversos cientistas alertam: a vacinação e o distanciamento social são ferramentas extremamente efetivas para o controle da propagação do vírus na população.

Deve-se ressaltar que algumas limitações foram constatadas no algoritmo. Ele apresenta dificuldade de ajustar longos intervalos de tempo. Além do mais, a metodologia proposta neste trabalho não é capaz de ajustar períodos nos quais há uma inversão no sentido da curva de infectantes ativos nos dados. Isso é totalmente compreensível, já que o parâmetro β pode mudar ao longo do tempo.

6.2 Trabalhos Futuros

Para trabalhos futuros, pretendem-se propor algumas adaptações no algoritmo implementado, possibilitando o ajuste dos parâmetros para quaisquer períodos, inclusive aqueles nos quais ocorrem a inversão do sentido da curva de infectantes. Assim, seria extremamente interessante desenvolver uma rede complexa utilizando os dados municipais de São João del-Rei, semelhantemente ao trabalho realizado pelo José Mauro [23]. Isso aumentaria a autenticidade do processo de simulação, permitindo que o algoritmo seja capaz de realizar predições.

Referências

- 1 ORGANIZATION, W. H. *Coronavirus disease 2019 (COVID-19): situation report, 46*. [S.l.], 2020-03-06. 9 p. p.
- 2 ORGANIZATION, W. H. *Coronavirus disease 2019 (COVID-19): situation report, 94*. [S.l.], 2020-04-23. 12 p. p.
- 3 CORONAVIRUS Cases - Worldometer. 2021. Disponível em: <[https://www-worldometers.info/coronavirus](https://www.worldometers.info/coronavirus)>. Acesso em: 2021-12-02.
- 4 NEWMAN, M. E. Spread of epidemic disease on networks. *Physical review E*, APS, v. 66, n. 1, p. 016128, 2002.
- 5 SCABINI, L. F.; RIBAS, L. C.; NEIVA, M. B.; JUNIOR, A. G.; FARFÁN, A. J.; BRUNO, O. M. Social interaction layers in complex networks for the dynamical epidemic modeling of covid-19 in brazil. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 564, p. 125498, 2021.
- 6 KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, The Royal Society London, v. 115, n. 772, p. 700-721, 1927.
- 7 WANG, Y.; CAO, J.; ALOFI, A.; ABDULLAH, A.-M.; ELAIW, A. Revisiting node-based sir models in complex networks with degree correlations. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 437, p. 75-88, 2015.
- 8 JÚNIOR, R. A.; VIEIRA, V. F.; XAVIER, C. R. Estudo de critérios de distribuição de vacinas com doses limitadas-uma abordagem baseada em redes complexas. In: SBC. *Anais do X Brazilian Workshop on Social Network Analysis and Mining*. [S.l.], 2021. p. 103-114.
- 9 NEWMAN, M. E. J. *Networks : an introduction*. Oxford; New York: Oxford University Press, 2010. ISBN 9780199206650 0199206651. Disponível em: <[https://www.amazon-de/Networks-Introduction-Mark-Newman/dp/019920665](https://www.amazon.de/Networks-Introduction-Mark-Newman/dp/019920665)>.
- 10 ERDOS, P.; RÉNYI, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, Citeseer, v. 5, n. 1, p. 17-60, 1960.
- 11 PASTOR-SATORRAS, R.; CASTELLANO, C.; MIEGHEM, P. V.; VESPIGNANI, A. Epidemic processes in complex networks. *Reviews of modern physics*, APS, v. 87, n. 3, p. 925, 2015.
- 12 MITCHELL, M. *An introduction to genetic algorithms*. [S.l.]: MIT press, 1998.
- 13 DARWIN, C.; WALLACE, A. R.; LYELL, S. C.; HOOKER, J. D. On the tendency of species to form varieties: and on the perpetuation of varieties and species by natural means of selection. In: LINNEAN SOCIETY OF LONDON. [S.l.], 1858.

- 14 MORETTIN, P. A.; BUSSAB, W. O. *Estatística básica*. [S.l.]: Saraiva Educação SA, 2017.
- 15 COVID-19: Quais as diferenças entre isolamento vertical, horizontal e lockdown? 2020. Disponível em: <<https://portal.fiocruz.br/pergunta/covid-19-quais-diferencas-entre-isolamento-vertical-horizontal-e-lockdow>>. Acesso em: 2021-12-01.
- 16 QIAN, Z.; TANG, S.; ZHANG, X.; ZHENG, Z. The independent spreaders involved sir rumor model in complex networks. *Physica A: Statistical Mechanics and Its Applications*, Elsevier, v. 429, p. 95–102, 2015.
- 17 KLEINBERG, J. The wireless epidemic. *Nature*, Nature Publishing Group, v. 449, n. 7160, p. 287–288, 2007.
- 18 HUANG, C.; CAO, J.; WEN, F.; YANG, X. Stability analysis of sir model with distributed delay on complex networks. *PloS one*, Public Library of Science San Francisco, CA USA, v. 11, n. 8, p. e0158813, 2016.
- 19 XIA, C.; WANG, L.; SUN, S.; WANG, J. An sir model with infection delay and propagation vector in complex networks. *Nonlinear Dynamics*, Springer, v. 69, n. 3, p. 927–934, 2012.
- 20 ARQUAM, M.; SINGH, A.; SHARMA, R. Modelling and analysis of delayed sir model on complex network. In: SPRINGER. *International Conference on Complex Networks and their Applications*. [S.l.], 2018. p. 418–430.
- 21 LOMBARDI, A.; AMOROSO, N.; MONACO, A.; TANGARO, S.; BELLOTTI, R. Complex network modelling of origin–destination commuting flows for the covid-19 epidemic spread analysis in italian lombardy region. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 11, n. 10, p. 4381, 2021.
- 22 LYMPEROPOULOS, I. N. # stayhome to contain covid-19: Neuro-sir–neurodynamical epidemic modeling of infection patterns in social networks. *Expert Systems with Applications*, Elsevier, v. 165, p. 113970, 2021.
- 23 RIBEIRO, J. M.; ALENCAR, R.; MARTINS, G.; XAVIER, C.; EVSUKOFF, A.; VIEIRA, V. Caracterização da relação entre redes sociais e mobilidade de indivíduos em contextos urbanos. In: SBC. *Anais do X Brazilian Workshop on Social Network Analysis and Mining*. [S.l.], 2021. p. 32–43.